Article

Design and Implementation of Defect Detection System Based on YOLOv5-CBAM for Lead Tabs in Secondary Battery Manufacturing

Jisang Mun [1,2], Jinyoub Kim [1], Yeji Do [1], Hayul Kim [2], Chegyu Lee [1] and Jongpil Jeong [1,*]

[1] Department of Smart Factory Convergence, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Republic of Korea; qpepfgh1@g.skku.edu (J.M.); kjy333k@naver.com (J.K.); hyginoyj@hygino.co.kr (Y.D.); leechgyu@skku.edu (C.L.)

[2] Hygino AI Research Laboratory, 25, Simin-daero 248 Beon-gil, Dongan-gu, Anyang 14067, Republic of Korea; hygino@hygino.co.kr

[*] Correspondence: jpjeong@skku.edu; Tel.: +82-10-9700-6284 or +82-31-299-4267

Abstract: According to QYResearch, a global market research firm, the global market size of secondary batteries is growing at an average annual rate of 8.1%, but fires and casualties continue to occur due to the lack of quality and reliability of secondary batteries. Therefore, improving the quality of secondary batteries is a major factor in determining a company's competitive advantage. In particular, lead taps, which electrically connect the negative and positive electrodes of secondary batteries, are a key factor in determining the stability of the battery. Currently, the quality inspection of secondary battery lead tab manufacturers mostly consists of visual inspection after vision inspection with a rule-based algorithm, which has limitations on the types of defects that can be detected, and the inspection time is increasing due to overlapping inspections, which is directly related to productivity. Therefore, this study aims to automate the quality inspection of lead tabs of secondary batteries by applying deep-learning-based algorithms to improve inspection accuracy, improve reliability, and improve productivity. We selected the YOLOv5 model, which, among deep-learning algorithms, has a benefit for object detection, and used the YOLOv5_CBAM model, which replaces the bottleneck part in the C3 layer of YOLOv5 with the Convolutional Block Attention Module (CBAM) based on the attention mechanism, to improve the accuracy and speed of the model. As a result of applying the YOLOv5_CBAM model, we found that the parameter was reduced by more than 50% and the performance was improved by 2%. In addition, image processing was applied to help segment the defective area to apply the SPEC value for each defective object after detection.

Keywords: automatic defect detection; YOLOv5; deep learning; object detection; lead tap; computer vision; CBAM

1. Introduction

The global market research firm QYResearch forecasts that the global market for lithium-ion battery lead taps will grow at an average annual rate of 8.1% from USD 75.6 billion in 2022 to 1.33 billion by 2029 [1]. In addition, the growth of the EV market is expected to accelerate from 2023 due to increasing EV purchase subsidies under the Inflation Reduction Act (IRA) [2]. Most industrialized countries are targeting a 100% reduction in carbon dioxide emissions from the automotive sector by 2035 [3], and the global demand for EVs is surging. According to industry researchers, e.g., the global automotive industry market research firm LMC Automotive [4–6], the global average BEV + PHEV sales penetration rate is expected to exceed 10% in 2022, with China and Europe leading regionally with 27% and 15%, respectively. In contrast, the USA is a bit slower at 5%; however, the Biden administration's green policies are in full swing, therefore opening vast opportunities for domestic battery companies with favorable relationships with them.

Currently, most quality inspections of secondary battery lead tab manufacturers involve visual inspections using a rule-based algorithm followed by microscopic inspections. The subsequent microscopic inspection is time-consuming and represents a bottleneck in the inspection process. Recently, electric vehicles catching fire has emerged as an ongoing problem. The ignition of lithium-ion batteries is such that the entire battery must be immersed in water to be incinerated, and battery fires can result in fatal accidents. Please note that approximately 70% of the world's total supply of secondary battery lead taps are used in electric vehicles, and this demand is expected to increase, therefore raising the bar for lead tap quality. There are approximately 58 types of defects in lead taps with small (approximately 1×1 mm) defects in each type. Thus, if the product is repeatedly inspected visually by workers, worker fatigue accumulates, which can negatively impact quality control. If only nonvisual inspections are performed, the effectiveness of the inspection is limited; thus, visual and microscopic inspections by skilled workers are necessary.

Thus, the goal of this study was to realize reliable automatic quality inspection of secondary battery lead taps. There are three primary types of defects in lead tabs, i.e., those in the material production process, those in the material transportation process, and those in the production process. To organize the lead tap quality inspection framework, we must consider inspection speed and accuracy. The lead tap production facility is composed of a single line from material input to completion of inspection; thus, a problem occurring in any process will result in a bottleneck that prevents the entire process from proceeding. If defective products are produced and delivered due to a lack of reliability in the lead tap quality inspection processes, battery fires may occur, which can cause human casualties. The inspection speed affects productivity, and poor inspection accuracy can cause major accidents. Thus, it is necessary to select an algorithm in consideration of inspection speed and inspection accuracy. In this study, we suggest a quality-checking technique based on the YOLOv5 model, which is actively used in the object detection field. In our investigation, 4K resolution images were used; however, typically, the defects to be detected are as small as 10×10 pixels, which is very fine. Comparing the YOLO method to other algorithms, its detection speed is great; however, it is limited in terms of detecting small objects. Thus, we employed the YOLOv5x model, which is the deepest layer of the YOLOv5 model, to check the object detection speed and accuracy. This model demonstrated an improvement in detection accuracy compared to the existing rule-based method, and we confirmed that it improved the detection speed. In addition, to improve detection speed and accuracy we apply the CBAM, which is based on the attention mechanism.

The paper's contributions can be outlined as follows:

- Secondary Battery Lead Tab Quality Inspection Automation: Secondary battery lead tab quality inspection automation using AI can give companies an edge in competitiveness and increase productivity by reducing worker fatigue and improving inspection speed.
- YOLOv5_CBAM: The CBAM based on the Attention mechanism is applied to the Bottleneck part of YOLOv5 to reduce the amount of computation and improve the accuracy.
- Accuracy: Instead of simply adding layers to improve accuracy, we improved accuracy using an algorithm based on an attention mechanism that remembers important information and suppresses unnecessary information.

In this paper, an improved YOLOv5_CBAM model is proposed to detect small defects in secondary battery lead taps. We demonstrate that the detection speed and accuracy are improved compared to existing methods. The proposed object detection algorithm utilizes an attention mechanism; thus, the layer is not deeply organized, and therefore information loss is avoided. As a result, detection speed and accuracy are improved compared to existing methods. Thus, the defect rate of secondary battery lead taps is reduced, productivity is improved, and companies can gain a competitive advantage.

The paper in question has the following format. An overview of related literature is given in Section 2. The suggested flawed object identification method is presented in Section 3, covering the process of gathering data, preprocessing those data, and using the upgraded YOLOv5_CBAM model to achieve quick inference speed. In Section 4, we summarize and discuss the results of the experiment. In Section 5, which also examines the study's limitations and potential future research areas, the work is completed.

2. Related Work

2.1. Yolov5

Traditional methods for image segmentation have attained a considerable level of maturity; nevertheless, they necessitate the extraction of features for each defect, which is labor-intensive and reduces efficiency. Many target detection algorithms currently exist [7,8], and beginning in 2012, deep-learning methods have been developed and proposed in many studies. YOLO is a type of deep-learning network [9] proposed in 2016 [10]. The YOLO network exhibits high efficiency and good generalizability when detecting small targets. YOLOv5, recognized as one of the most extensively employed detection networks, finds application across a range of industries and use cases. These include production processes [11], autonomous driving [12], monitoring and safety [13], surface defect detection [14,15], as well as target detection [16] in various industries and applications. The YOLOv5 network can realize high object detection accuracy and good inference speed. The YOLOv5 object detection algorithm outperforms various other algorithms, e.g., the Fast R-CNN algorithm. The YOLOv5 model is divided into YOLOv5n (nano), YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), and YOLOv5x (xlarge), according to the depth of the network, where a deeper network yields higher accuracy at the expense of increasing computational time.

2.2. Attention Mechanism

Much research [17–19] has revealed that human attention plays an important role in how we perceive. The fact that human vision does not attempt to process the entire scene at once is one of the most important characteristics of human vision. Instead, humans use a series of partial glimpses and selective focus on salient parts to better capture visual information [20]. For example, the center of the human eye has a higher resolution than the periphery [21]. To process visual information efficiently and adaptively, the human visual system processes space and focuses on salient regions in an iterative manner [20].

2.2.1. Cross-Modal Attention

In multimodal settings, attention mechanisms are a common technique when processing needs to be conditional on other modalities. One well-known task that is useful in such cases is visual question answering (VQA), which involves predicting answers to questions about counting, object location, properties, inferences, and more when solving problems between natural language and images. VQA can be considered a dynamically changing set of tasks that require processing a given image based on a given question. Attention mechanisms smoothly select aspects from the image features that are relevant to the task (i.e., the question). As proposed by Yang et al. [22], a given question serves as a query to generate attention maps for image features and retrieve features that are relevant to the question. The final determination is then categorized using the accumulated image features. Another approach is to use bidirectional inference to generate attention maps for both text and images, as proposed by Hyeonseob Nam et al. in [23]. Attention maps are a useful tool for conditionally solving tasks in this body of literature. but are acquired in stages for specialized tasks.

2.2.2. Self-Attention

To train attention generation and feature extraction together end-to-end using DNNs, a variety of ways to incorporate attention have been developed. There have been several attempts to use attention to effectively apply it to common classification tasks [24,25]. For example, Wang et al. proposed a residual attention network using an hourglass module to generate three-dimensional (3D) attention maps for intermediate features [25]. Due to the created attention maps, it should be noted that this architecture is resilient to noisy labels; however, it involves high computational/parameter overhead due to the large number of steps involved in the 3D map generation process. Hu et al. proposed a compact squeeze-and-excitation module [24] to effectively utilize the relationship between channels. Although it is not stated specifically in their paper, this method can be thought attention mechanism used on the channel axis. However, we are missing the spatial axis, which is an important factor in inferring the correct attention map.

2.2.3. Adaptive Modules

In several earlier experiments, adaptable modules that dynamically alter the output based on the input were employed. For example, the dynamic filter network [26] generates convolutional features according to the input features for flexibility. In addition, the spatial transformer network [27] utilizes input features to generate the hyperparameters of an affine transformation adaptively to ensure that the target region feature map is well-aligned, which can be seen as paying careful attention to the feature map. In the Deformable Convolutional Network [28], only pertinent features are pooled for convolution since the pooling offset is dynamically produced from the input features. In the same vein as these methods, the Block Attention Module (BAM) is also a self-contained adaptive module that dynamically suppresses or emphasizes the feature map through an attention mechanism.

2.3. CBAM (Convolutional Block Attention Module)

By adding more layers, which raises the network's complexity, the performance of a network can be enhanced. This allows neural networks to approximate higher-dimensional functions, and VGGNet [29], ResNet [30], and AlexNet [31] reflect this concept, with VGGNet having twice as many layers as AlexNet. ResNet also has 22 times more layers than VGGNet. WideResNet [32] and PyramidNet [33] show that, with more channels and convolution, you can obtain higher performance than simply increasing the number of layers.

Bottleneck Attention Module (BAM) is a module designed for simple and easy integration into common DNN and CNN architectures in the direction of investigating the influence of attention in the structure of DNN and improving the representational power of the network very effectively, as opposed to the conventional approach of adding numerous layers and complicated stacking to enhance network performance [34]. BAM infers a 3D attention map $F \in R^{C \times H \times W}$ for a given input feature map $M(F) \in R^{C \times H \times W}$, and the refined feature map F′ is calculated as follows.

$$F' = F + F \otimes M(F) \qquad (1)$$

To encourage gradient flow, we use a residual learning technique with an attention mechanism. Here, $\otimes$ denotes elementwise multiplication. Design an efficient and robust module by computing channel attention $M_c(F) \in R^C$ and spatial attention $M_s(F) \in R^{H \times W}$ in two separate branches and computing attention map $M(F)$ as follows.

$$M(F) = \sigma(M_c(F) + M_s(F)) \qquad (2)$$

The sigmoid function, σ, is used here. Before being added, both branch outputs are scaled by $R^{C \times H \times W}$.

The BAM was designed to replace all bottlenecks by focusing on the attention structure of DNNs rather than simply stacking layers to improve network performance. Also, BAM was designed to facilitate simple integration into CNNs. Here, given a feature map F, the BAM branches into the channel and spatial attention to compute and apply the attention map. The CBAM adopts a sequential application method, as depicted in Figure 1, to apply operations progressively and reduce the number of operations while improving performance [35]. As a result, CBAM produces a feature map that emphasizes important information and suppresses unnecessary noise.
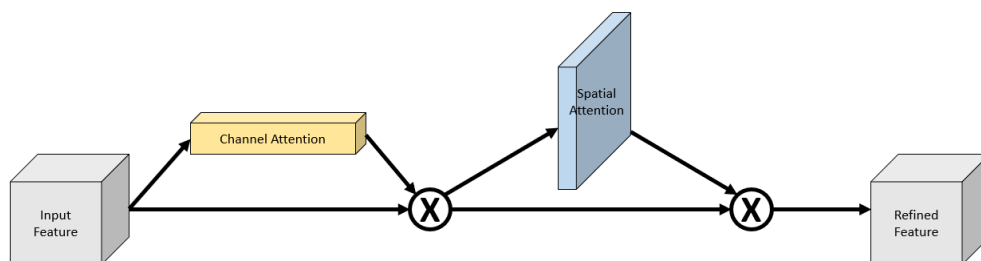
Figure 1. Overview of CBAM.

Earlier, we introduced BAM, a module designed to replace all the bottlenecks, focusing on the attention structure of DNNs rather than simply stacking layers to improve the performance of the network, and designed for simple integration into CNNs. The idea is that given a feature map F, it branches into two separate attention (channel, spatial) to compute and apply the attention map. Although BAM had a structure where channel attention and spatial attention were branched from the feature map and combined, CBAM is a module that adopts a sequential application method as shown in Figure 1 to reduce operations and improve performance by applying them sequentially as shown in the following figure [35]. As a result, it produces a feature map that emphasizes important information and suppresses unnecessary noise.

A two-dimensional (2D) spatial attention map $M_s \in R^{1 \times H \times W}$ and a one-dimensional (1D) channel attention map $M_c \in R^{C \times 1 \times 1}$ are sequentially inferred by CBAM from an intermediate feature map $F \in R^{C \times H \times W}$ as input (Figure 1). The complete attention process is summarized as follows.

$$F' = M_c(F) \otimes F, \tag{3}$$

$$F'' = M_s(F') \otimes F' \tag{4}$$

When multiplying, the attention values are broadcast (i.e., copied) appropriately. Channel attention values are broadcast along with spatial attention and vice versa, and the final refined output is F''.

2.3.1. Channel Attention Module

The channel attention module, created by leveraging the relationships between channels in a feature in CBAM, focuses on 'what' is meaningful in the image, and each channel in the feature map is considered a feature detector. To compute the channel attention efficiently, average pooling is a common method to aggregate the spatial attention of the input feature map. We argue that max-pooling in CBAM can obtain other important information about features to detect a more accurate channel attention map. Thus, as shown in Figure 2, CBAM demonstrates the effectiveness of using average and max-pooling.
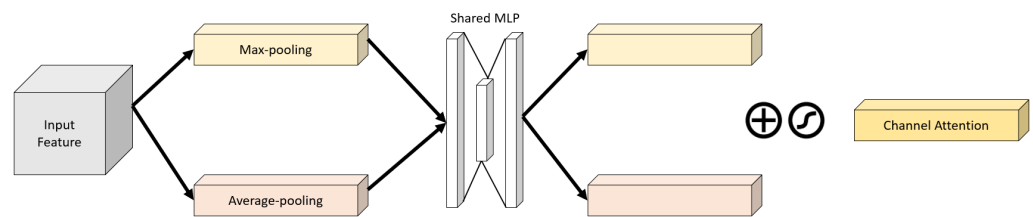
Figure 2. Channel attention module.

2.3.2. Spatial Attention Module

Making use of the spatial correlations between the features, the spatial attention map is produced. Unlike channel attention maps, spatial attention maps supplement channel attention maps by concentrating on "where" information is in the image. As seen in Figure 3, to calculate the spatial attention map, along the channel axis, maximum and average pooling are initially applied, and these operations are then concatenated to produce effective feature descriptors. It has been demonstrated that using pooling activities along the channel axis can effectively highlight information regions. We also apply convolution to the concatenated feature descriptors.
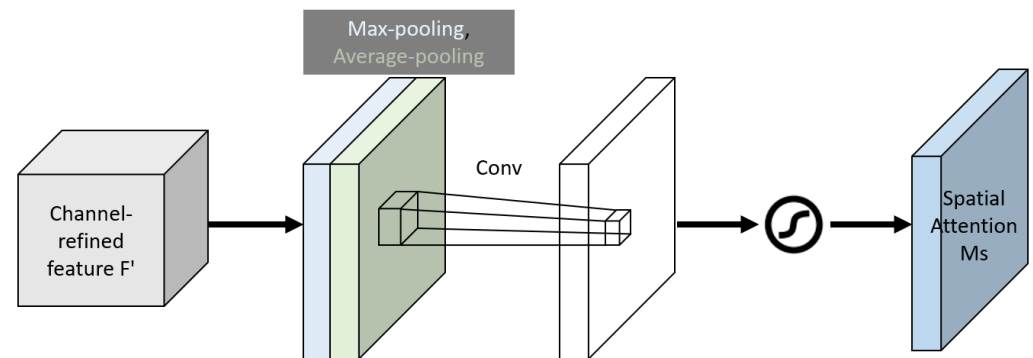
Figure 3. Spatial attention module.

3. Yolov5 _CBAM-Based Inspection

3.1. System Architecture

Figure 4 provides a schematic representation of the system's architecture. The accuracy of neural networks increases with their depth in the area of object detection; however, computational costs increase with increasing depth. Thus, as we construct deeper neural networks to improve accuracy, high-performance hardware is required. Most object detection problems require increasingly lightweight and high-performance solutions. Thus, the problem should be solved such that performance is improved without constructing deep neural networks. For this purpose, various studies are underway, and we have confirmed that the proposed CBAM algorithm contributes to performance and accuracy improvements. YOLOv5's network structure can be categorized as having a neck, a backbone, and output components, and then it goes through Non-Max-Suppression (NMS) to infer the closest bounding box. Unlike previous versions, YOLOv5 adopts a CSPNet-based backbone to improve performance and accuracy. Here, we want to replace bottlenecks with CBAM in all C3 layers (Figure 4) of the existing CSPNet-based backbone to reduce computational costs and improve accuracy.
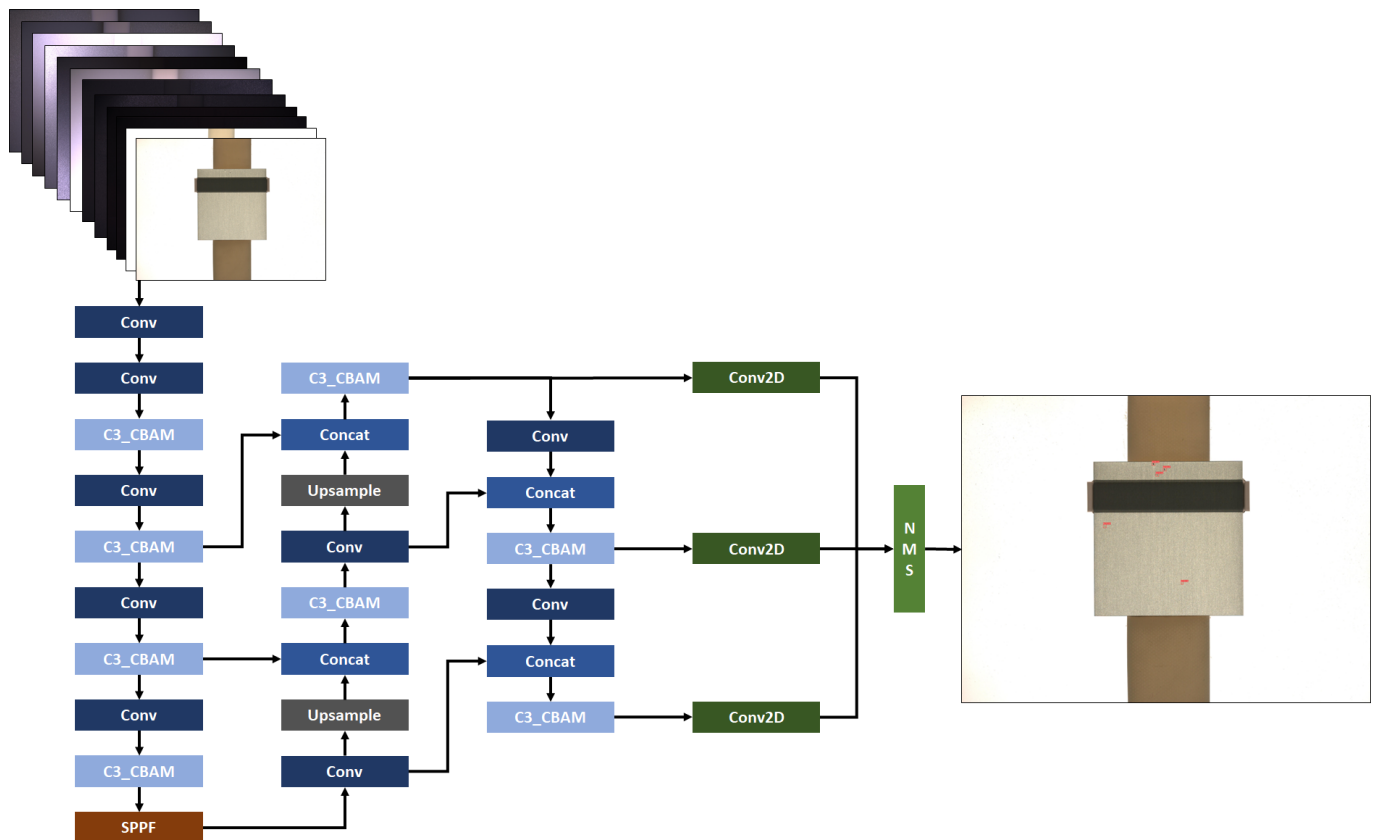
Figure 4. Architecture of the proposed system.

3.2. C3 _CBAM

Figure 5 depicts the C3_CBAM module, which enhances concentration by utilizing the attention mechanism. The pre-existing C3 module mitigates computational expenses by forcibly reducing computation volume through a bottleneck after convolution, followed by its reassembly. This reduces the amount of computation sufficiently; however, information may be lost in this process. Thus, we apply CBAM based on the attention mechanism rather than the bottleneck to reduce computational costs and improve concentration by utilizing the relationship between channel and spatial.
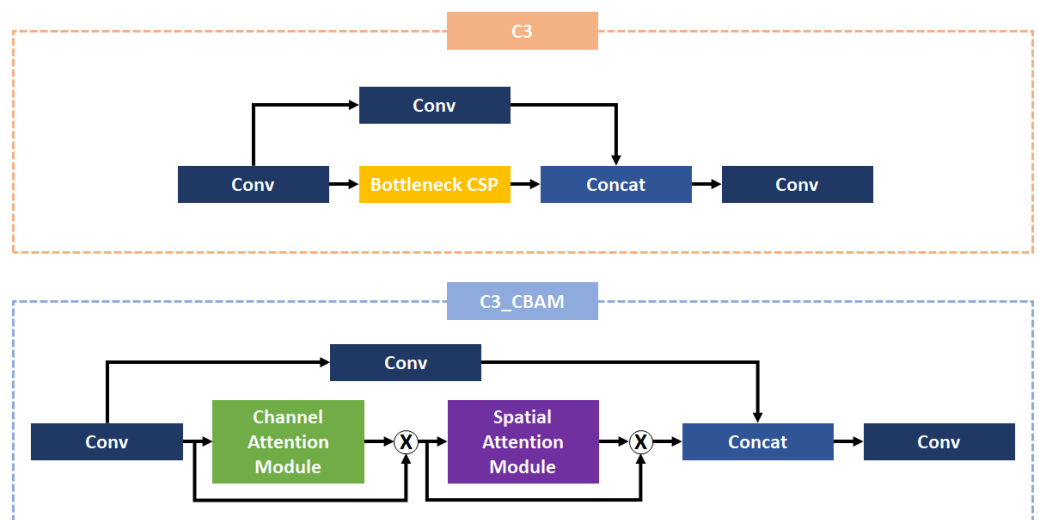

Figure 5. C3 and C3_CBAM.

3.3. Data Postprocessing

Lead tabs have specifications for each type of defect. The specification measures the area, width, and diameter of an identified defective area, and if it does not meet the standard, it is classified as a defective product. Here, it is necessary to segment the defective area within the bounding box of the inference result. Image processing techniques are applied to cut the area of the bounding box and further emphasize and segment the boundaries of the defective area, which helps segment the precise area. As shown in Figure 6, the bounding box area is inferred, the area is cut out, and then image processing is applied to segment the area by applying a threshold.
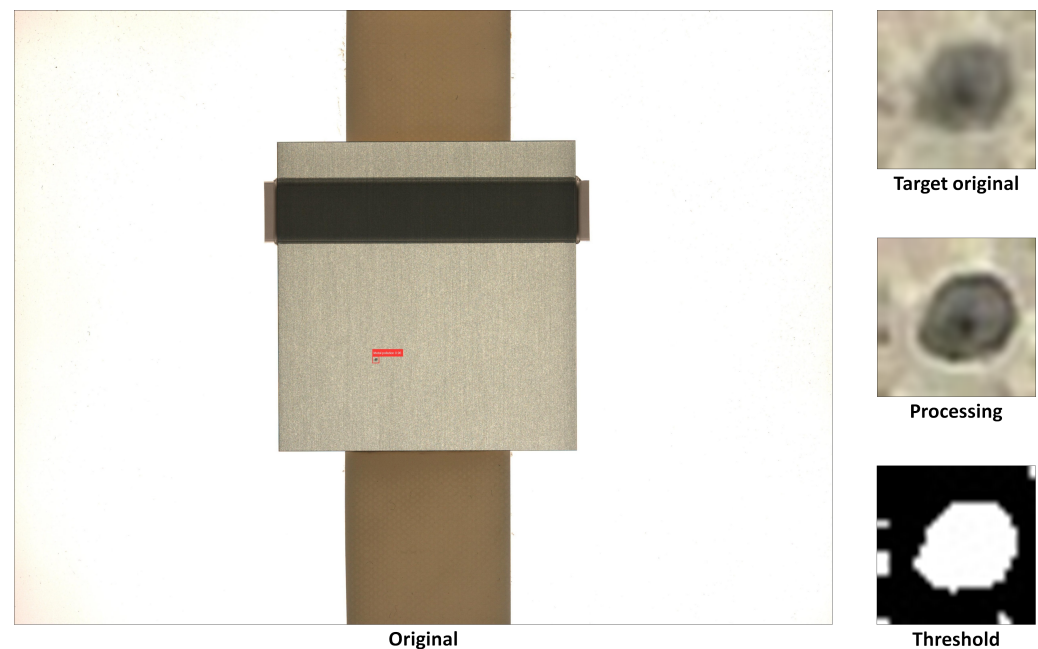
Figure 6. Postprocessing.

In this study, we applied a postprocessing technique to clearly show the defective area. Typically, the defective area exhibits an increase or decrease in the pixel value that differs from the surrounding background. Thus, we applied and experimented with various filters by applying a convolution filter to the image, and we investigated segmenting the defective area. Investigating the use of image processing techniques confirmed that the target is most clearly identified when applied in the order of midianBlur $\rightarrow$ sharpening filter $3 \times 3 \rightarrow$ average blurring 2×2.

First, the graininess caused by the material of the surrounding metal appears as noise in the image; thus, we smooth it out. Various filters help remove noise, e.g., Midian blur, Gaussian blur, and bilateral filters. To identify a feature, the target should be distinct compared to the surrounding background. Thus, we also apply a sharpening filter to sharpen the boundaries between pixels. The sharpening filter applies a 2D filter to the target image and convolves it. The sharpening filter comprises a 3×3 kernel and maximizes the difference with the surrounding pixels to increase the contrast at the boundaries. Figure 7 shows the expression for a sharpening filter with a 3×3 kernel of strength 9 and the result of applying this filter to enlarge the defect object in the image.
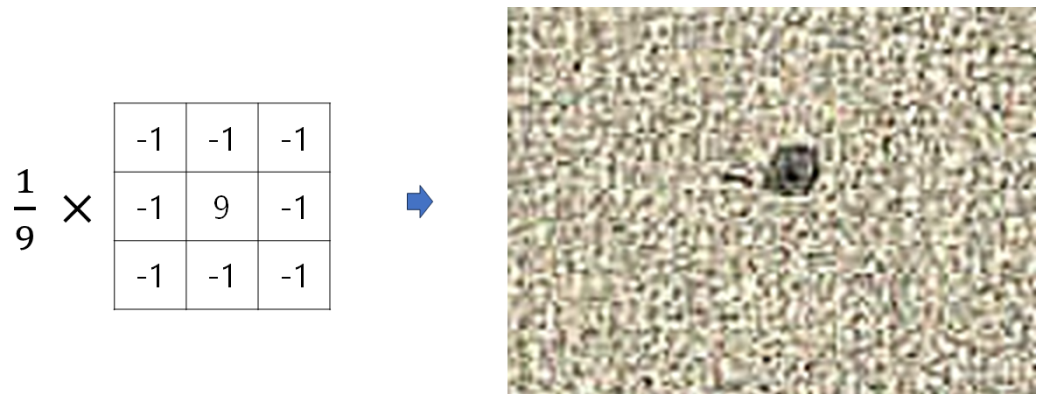
Figure 7. Sharpening filter with a 3 × 3 kernel of strength 9.

Figure 8 shows a set of results obtained after applying various noise removal filters, sharpening the boundaries, and applying average blurring to acquire a more natural appearance. The original image in this example is the one with the black border (bottom right), while the one with the green border is the one that has the most recognizable qualities. The filters identified at the bottom of each result are written in order of application. From the top: bilateral, gaussianBlur, midianBlur, and no filters to remove noise. From the left: 3 × 3 average blur, 2 × 2 average blur, and average blur are applied after applying a noise filter. The results show that the surrounding noise is emphasized in the case of processing without removing noise, and the results of processing with noise and applying a filter show the sharpest features, as indicated by the image highlighted by the green border.
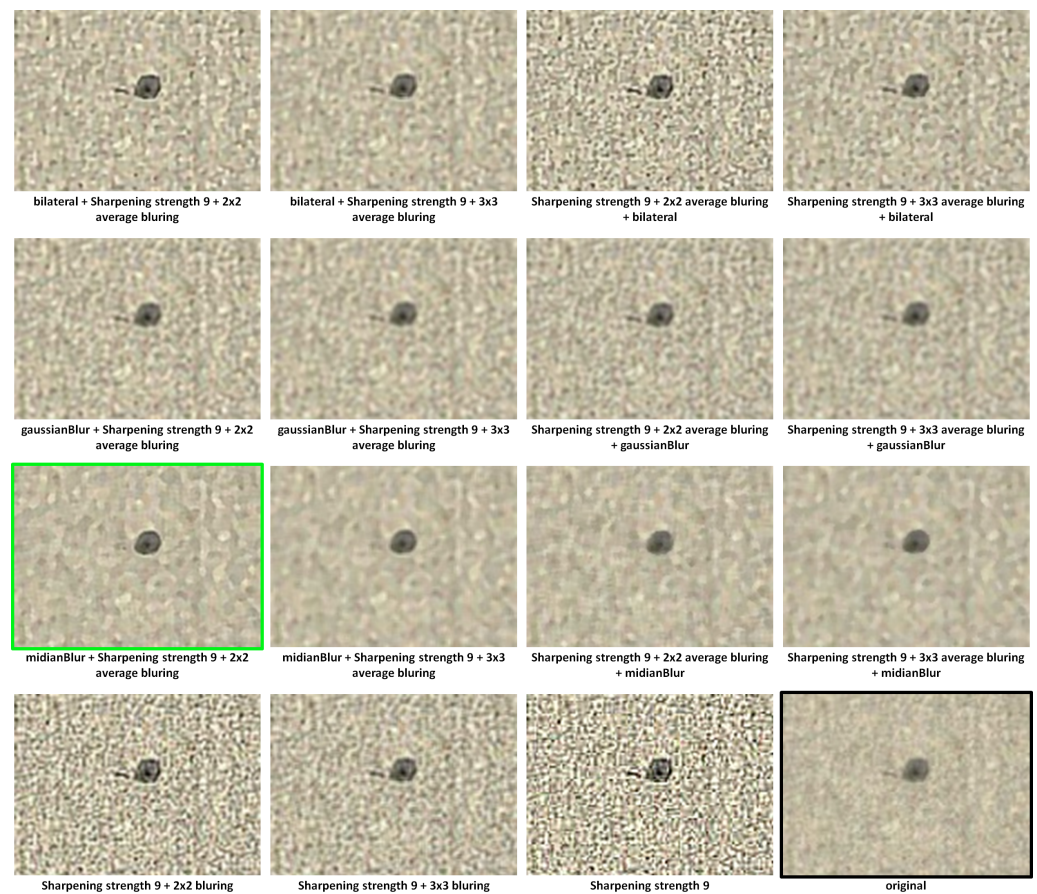

Figure 8. Comparison of filtering results.

4. Performance Analysis

4.1. Experimental Environments

In this study, a manufacturer of lead taps for secondary batteries in Seongnam City, Korea used a camera to acquire images by illuminating the top and bottom of the taps from 12 different directions. The lead taps were illuminated from 12 directions because defective objects appear differently depending on the illumination conditions. One set of 12 cut images was acquired for each lead tab, as shown in Figure 9. These 12 images were compressed into a single image format (.vid4). The compressed images needed to be decompressed separately. After decompression, 4096 × 3072 lead tab images were collected.
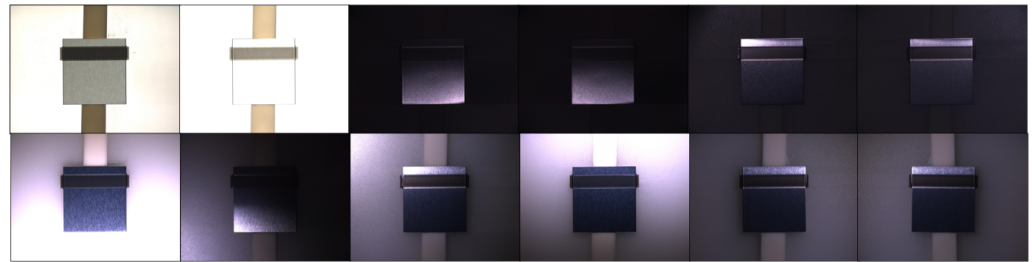

Figure 9. Lead tab image acquired under 12 illumination direction conditions.

Experiments were conducted using the acquired images and the PyTorch framework. Here, the hardware comprised an Intel i3-13900KF CPU, 64 GB memory, and an NVIDIA GeForce RTX 4090 graphics processing unit.

4.2. Experimental Datasets

The experimental dataset consists of two materials and seven types of defects, as shown in Table 1. Please note that there are several defects other than those considered in this study; however, we did not consider these defects because their impact on the system is small compared to the seven defects covered in our experiments. We collected a total of 2500 images (1250 images for each material). In this case, 2000 of the total photos (80%) were chosen for training and 20% (500 images) were used for the performance evaluation. The original images of lead tab defects (4096 × 3072 pixels) were resized to 1280 × 1280 pixels for the experiment.

For each type of defect, there will be defects of inconsistent sizes and features. In addition, there are cases where different types of defects have similar appearance, as shown in Figure 10. Thus, it is necessary to classify the defects to help the learning model assess them accurately. In particular, there are cases where there is little distinction between defect types; thus, it was necessary to collect balanced data to avoid class imbalance problems. As a result, classification becomes even more important. For the classification of defects, it is necessary to diversify the background of the objects in the images to facilitate effective learning.
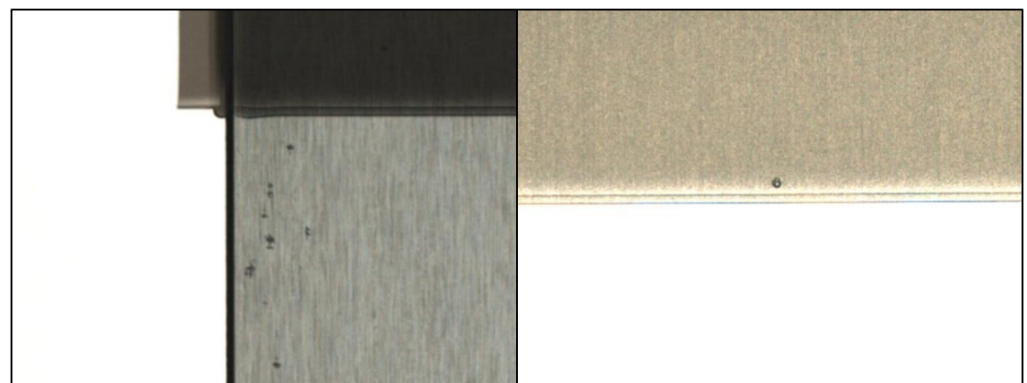
Figure 10. Defects: (**left**) pollution and (**right**) surface bubble.

The test dataset is set up as depicted in Table 2. The same number of images were used for each material, with 700 images for Faultless and 100 images for each type of defect, making a total of 2800 images for the test dataset.

Table 1. Training datasets.

Image Size	Material Type	Training Images	Defect Type	Training Objects
1280 × 1280	Al	1050	Metal pollution	513
			Surface bubble	270
			Ripped off	538
			Film alien substance	377
			Metal alien substance	192
			Teflon	150
			Jinjeop	479
	Ni	1050	Metal pollution	629
			Surface bubble	661
			Ripped off	873
			Film alien substance	495
			Metal alien substance	762
			Teflon	150
			Jinjeop	445

Table 2. Test datasets.

Defect Type	Material Type	
	Al	Ni
Faultless	700	700
Metal pollution	100	100
Surface bubble	100	100
Ripped off	100	100
Film alien substance	100	100
Metal alien substance	100	100
Teflon	100	100
Jinjeop	100	100
Total		2800

4.3. Evaluation Index

We evaluated the classification performance using a Confusion Matrix, which represents the difference between the inference result of the learning model and the actual inspection result. TP data were predicted to be good and were actually good; FN data were predicted to be bad but were actually good; FP data were predicted to be good but were actually bad; and TN data were predicted to be bad and were actually bad. Precision represents the percentage of cases that the model classified as correct that were actually correct. Recall is the percentage of answers that the model correctly identified as correct.

In this experiment, we use the F1 score as a performance metric, which evaluates the model's performance as the harmonic average of recall and accuracy. GPU FLoating-point Operations Per second (GFLOPs) is the usage of the GPU used for training, which means the amount of floating-point operations per second.

$$Presicion = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1\ score = \frac{2 * precision * Recall}{Precision + Recall} \tag{7}$$

4.4. Results

This experiment was conducted by resizing the 4096 × 3072 image to 1280 × 960, and the experiment was executed using a configuration of 1000 epochs and batch sizes of 12. The result of the experiment using Figure 11 displays the Confusion Matrix, and the actual value is a pure image with no defective objects, and the predicted value is judged as a case where one defective object is found. The test result focuses on determining whether the lead tab is a normal product or a non-normal product, and does not judge how many defective objects exist. The model-specific experimental results are presented in Table 3, and all models are defective, but no case is classified as normal.

Predicted / Actual	Positive	Negative
Positive	1400 (TP)	38 (FN)
Negative	0 (FP)	1362 (TN)

Figure 11. Confusion Matrix.

Table 3. An evaluation of Precision, Recall, and F1 Score Comparison.

| Model	Precision	Recall	F1-Score
YOLOv5	1.0	0.93	0.96
YOLOv5_CBAM_Backbone	1.0	0.78	0.87
YOLOv5_CBAM_Neck	1.0	0.94	0.97
YOLOv5_CBAM_All	1.0	0.97	0.98

The inspection result focuses on determining whether the lead tab is a normal product or a defective product; it does not determine how many defective objects exist. We conducted an experiment to apply the CBAM to the C3 layers in the backbone and neck, respectively, and to apply CBAM to all C3 layers. The outcomes of the experiments are presented in Table 4. The performance of the YOLOv5_CBAM model has demonstrated improvement., except for the case of applying CBAM to only the backbone. The results

demonstrate that the best performance was obtained when the CBAM was applied to all C3 layers throughout the network.

Table 4. Experimental results for each model.

| Model	Parameters	GFLOPs	F1-Score
YOLOv5	86.6 M	205.8	0.96
YOLOv5_CBAM_Backbone	55.9 M	114.4	0.87
YOLOv5_CBAM_Neck	61.3 M	153.8	0.97
YOLOv5_CBAM_All	30.6 M	62.3	0.98

As shown in Figure 12, the YOLOv5_CBAM model did not exhibit over-detection because it focuses on important parts of the target image and does not detect unnecessary parts, which is the intended purpose of the proposed system. In addition, the YOLOv5 model detected objects outside the lead tap inspection region of interest. Figure 13 shows the inference results obtained using the YOLOv5_CBAM model.
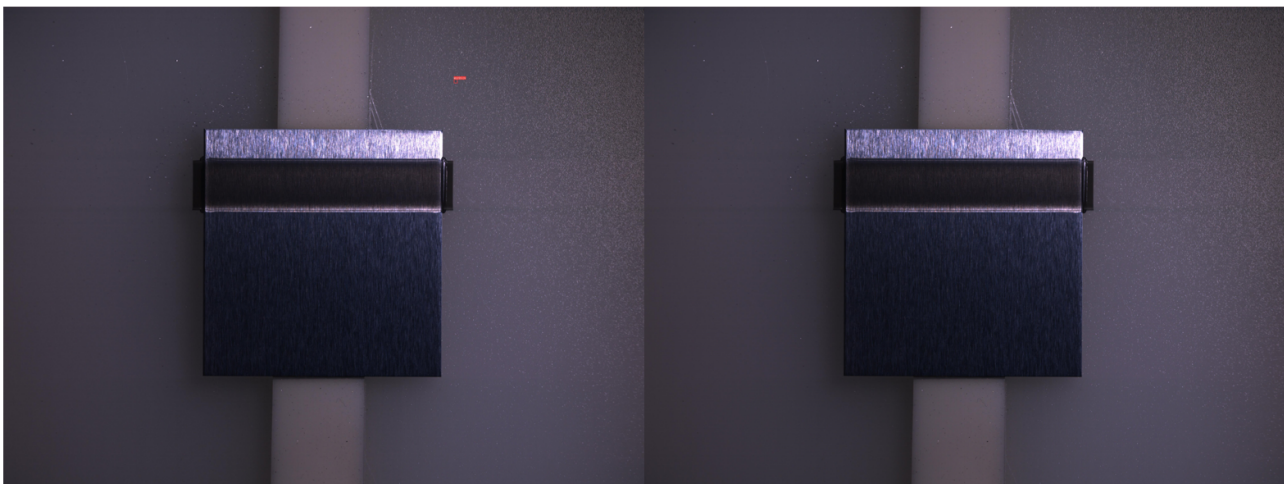
Figure 12. Left YOLOv5, right YOLOv5_CBAM.

4.5. Discussion

Although the YOLOv5_CBAM algorithm has been used to focus on important parts and not detect unnecessary parts, there is still a problem that some images are detected outside the detection area. The U-Net algorithm, which is widely used in the medical imaging field, also had a problem of segmenting non-organ parts, and to solve this problem, the Classification-Guided Module (CGM) was introduced in U-Net3+ to prevent over-segmentation by designing it to predict first. In the field of inspecting lead tabs like this, it seems that there is a way to solve the problem by adding a classification module like CGM to set the area of the lead tab, or to set the region of interest (ROI) through postprocessing.
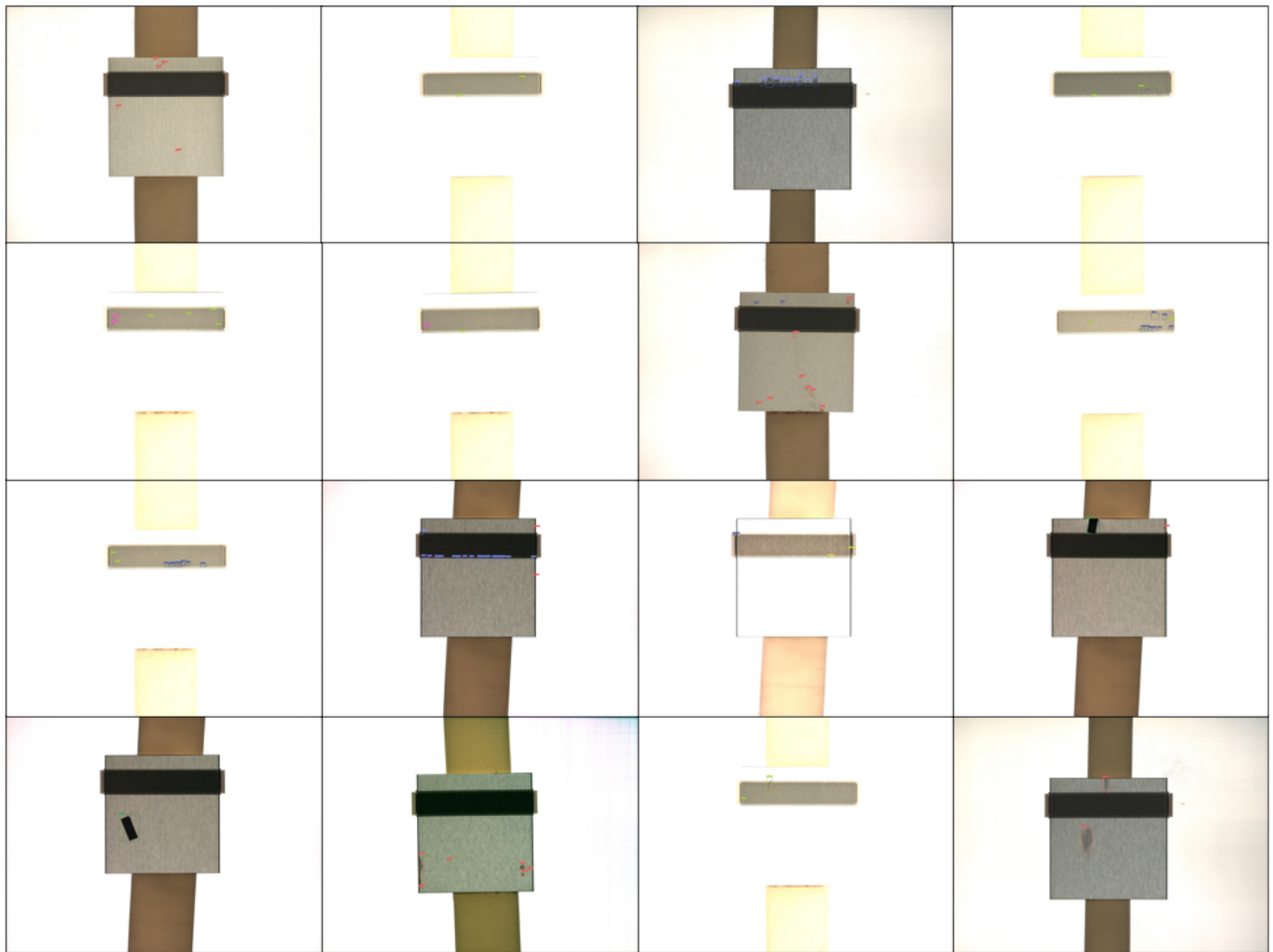
Figure 13. Detection results using the YOLOv5_CBAM model.

5. Conclusions

In this study, we proposed the YOLOv5_CBAM algorithm to detect defects in images of secondary battery lead taps. Our investigation has established that the suggested algorithm can enhance the performance of defect detection, which is expected to contribute to the production of high-quality lead taps and improve competitiveness in the global lead tap market. The proposed YOLOv5_CBAM algorithm, which is an improvement over the existing YOLOv5 algorithm, replaces the existing bottleneck layer with the CBAM, therefore reducing the number of parameters and GFLOPs by more than 50%, and performance was improved by 2%. We found that performance varied depending on whether the CBAM was applied to all C3 layers in the network or only the C3 layers in the backbone or neck components of the network.

In forthcoming work, our intentions include expanding our defect detection capabilities to encompass additional defect types, including those that are difficult to detect in 2D images, e.g., folds (metal slightly folded in half), stamps (metal or film stamped with something), and various other defects. Lead tabs are a key component of secondary batteries, which are the first step toward building a globally ecofriendly system. In this study, CBAM was employed as an attention mechanism to focus on only necessary areas; however, some parts are not detailed sufficiently to detect abnormalities using only 2D images. We note that the CBAM is a bottleneck layer that reduces computational costs; however, information loss may occur when using the CBAM. Henceforth, our future endeavors involve devising

a methodology for identifying smaller and more varied features with the aim of further enhancing the efficacy of the proposed algorithm.

In future research, we plan to detect more defects from the 7 defects we detected, as well as defects that are difficult to detect in 2D images, such as folds (metal slightly folded in half), stamps (metal or film stamped with something), and other defects. Lead tabs are one of the key components of secondary batteries, which are the first step toward building a globally ecofriendly system. In this study, we used CBAM as an attention mechanism to help us focus on only the necessary areas, but some parts are not detailed enough to detect abnormalities using only 2D images. It is true that CBAM is also a bottleneck layer that reduces computation, but there is still a concern about information loss. In the future, we will work on capturing smaller and more diverse features.

Author Contributions: Conceptualization, J.M. and J.J.; methodology, J.M., J.K., C.L. and J.J.; software, J.M. and Y.D.; validation, Y.D., C.L. and J.J.; formal analysis, J.M. and Y.D.; investigation, J.M., H.K. and C.L.; resources, J.K. and Y.D.; data curation, J.M.; writing—original draft preparation, J.M.; writing—review and editing, J.M. and J.J.; supervision, J.J.; project administration, J.M.; All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: GitHub (https://github.com/qpepfgh1/YOLOv5_CBAM) (accessed on 4 September 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. QYResearch KOREA. Lithium-Ion Battery Lead Tabs Market Report 2023. Revised. Available online: https://qyresearch.co.kr/post-one/%eb%a6%ac%ed%8a%ac%ec%9d%b4%ec%98%a8%eb%b0%b0%ed%84%b0%eb%a6ac-%eb%a6%ac%eb%93%9c%ed%83%ad-lead-tabs-%ec%8b%9c%ec%9e%a5%eb%b3%b4ea%b3%a0%ec%84%9c-2023%eb%85%84-%ea%b0%9c%ec%a0%95%ed%8c%90 (accessed on 16 March 2023).
2. U.S. Department of the Treasury. Treasury Releases Proposed Guidance on New Clean Vehicle Credit to Lower Costs for Consumers, Build U.S. Industrial Base, Strengthen Supply Chains. Available online: https://home.treasury.gov/news/press-releases/jy1379 (accessed on 31 March 2023).
3. Council of the EU. First 'Fit for 55' Proposal Agreed: The EU Strengthens Targets for CO2 Emissions for New Cars and Vans. Available online: https://www.consilium.europa.eu/en/press/press-releases/2022/10/27/first-fit-for-55-proposal-agreed-the-eu-strengthens-targets-for-co2-emissions-for-new-cars-and-vans/ (accessed on 27 October 2022).
4. LMC Automotive. The Batteries Fuelling Global Light Vehicle Electrification. 5. Available online: https://www.thebatteryshow.com/content/dam/Informa/amg/novi/2022/docs/10_15%20-%20Riddell.pdf (accessed on 21 August 2023).
5. Autoview. By 2022, 1 in 10 New Cars Worldwide Will Be Electric Vehicles... Ranked 2nd in Exports to China. Available online: http://www.autoview.co.kr/content/article.asp?num_code=78987&news_section=world_news&pageshow=1&page=1&newchk=news (accessed on 17 January 2023).
6. The Guru. 'Milestone' of 10% Global Share of EVs in 2022...7.8 Million Units Sold. Available online: https://www.theguru.co.kr/news/article_print.html?no=48371 (accessed on 18 January 2023).
7. Zaidi, S.S.A.; Ansari, M.S.; Aslam, A.; Kanwal, N.; Asghar, M.; Lee, B. A survey of modern deep learning based object detection models. *Digit. Signal Process.* **2022**, *126*, 103514. [CrossRef]
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
9. Diwan, T.; Anirudh, G.; Tembhurne, J.V. Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Multimed. Tools Appl.* **2022**, *82*, 9243–9275. [CrossRef] [PubMed]
10. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
11. Wang, Z.; Jin, L.; Wang, S.; Xu, H. Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. *Postharvest Biol. Technol.* **2022**, *185*, 111808. [CrossRef]
12. Zhang, Y.; Guo, Z.; Wu, J.; Tian, Y.; Tang, H.; Guo, X. Real-Time Vehicle Detection Based on Improved YOLO v5. *Sustainability* **2022**, *14*, 12274. [CrossRef]
13. Li, Z.; Xie, W.; Zhang, L.; Lu, S.; Xie, L.; Su, H.; Du, W.; Hou, W. Toward Efficient Safety Helmet Detection Based on YoloV5 with Hierarchical Positive Sample Selection and Box Density Filtering. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–14. [CrossRef]

14. Wang, L.; Liu, X.; Ma, J.; Su, W.; Li, H. Real-Time Steel Surface Defect Detection with Improved Multi-Scale YOLO-v5. 2023. Available online: https://www.mdpi.com/2227-9717/11/5/1357 (accessed on 25 April 2023).

15. Liu, W.; Xiao, Y.; Zheng, A.; Zheng, Z.; Liu, X.; Zhang, Z.; Li, C. Research on Fault Diagnosis of Steel Surface Based on Improved YOLOV5. 2022. Available online: https://www.mdpi.com/2227-9717/10/11/2274 (accessed on 31 October 2022).

16. Cao, Z.; Fang, L.; Li, Z.; Li, J. Lightweight Target Detection for Coal and Gangue Based on Improved Yolov5s. 2023. Available online: https://www.mdpi.com/2227-9717/11/4/1268 (accessed on 18 April 2023).

17. Corbetta, M.; Shulman, G.L. Control of goal-directed and stimulusdriven attention in the brain. *Nat. Rev. Neurosci.* **2002**, *3*, 201–215. [CrossRef] [PubMed]

18. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **1998**, *20*, 1254–1259. [CrossRef]

19. Rensink, R.A. The dynamic representation of scenes. *Vis. Cogn.* **2000**, *7*, 17–42. [CrossRef]

20. Larochelle, H.; Hinton, G.E. Learning to combine foveal glimpses with a third-order Boltzmann machine. *Adv. Neural Inf. Process. Syst.* **2010**. Available online: https://papers.nips.cc/paper_files/paper/2010/hash/677e09724f0e2df9b6c000b75b5da10d-Abstract.html (accessed on 18 April 2023).

21. Hirsch, J.; Curcio, C.A. The spatial resolution capacity of human foveal retina. *Vis. Res.* **1989**, *2*, 1095–1101. [CrossRef] [PubMed]

22. Yang, Z.; He, X.; Gao, J.; Deng, L.; Smola, A. Stacked attention networks for image question answering. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

23. Nam, H.; Ha, J.-W.; Kim, J. Dual attention networks for multimodal reasoning and matching. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2156–2164.

24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. *arXiv* **2017**, arXiv:1709.01507

25. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. *arXiv* **2017**, arXiv:1704.06904.

26. Jia, X.; De Brabandere, B.; Tuytelaars, T.; Gool, L.V. Dynamic filter networks. *Adv. Neural Inf. Process. Syst.* **2016**. [CrossRef]

27. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* **2015**. [CrossRef]

28. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. *CoRR* **2017**, *1*, 3. [CrossRef]

29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for largescale image recognition. *arXiv* **2014**, arXiv:1409.1556.

30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 December 2012.

32. Zagoruyko, S.; Komodakis, N. Wide residual networks. *arXiv* **2016**, arXiv:1605.07146.

33. Han, D.; Kim, J.; Kim, J. Deep pyramidal residual networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6307–6315.

34. Park, J.; Woo, S.; Lee, J.-Y. In So Kweon, BAM: Bottleneck Attention Module. 514. 2018. Available online: https://arxiv.org/abs/1807.06514v2 (accessed on 17 July 2018).

35. Park, J.; Woo, S.; Lee, J.-Y. In So Kweon, CBAM: Convolutional Block Attention Module. 2018. Available online: https://arxiv.org/abs/1807.06521 (accessed on 17 July 2018).