IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 November 2023, accepted 9 December 2023, date of publication 18 December 2023,
date of current version 21 December 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3343780*

RESEARCH ARTICLE

# YOLO Network Optimization With a Single Circular Bounding Box for Detecting Defective Cigarettes

**HEE-MUN PARK**[1] **AND JIN-HYUN PARK**[2]
[1]Department of Computer and Mechatronics Engineering, Gyeongsang National University, Jinju 52725, Republic of Korea
[2]School of Mechatronics Engineering, Gyeongsang National University, Jinju 52725, Republic of Korea

Corresponding author: Jin-Hyun Park (uabut@gnu.ac.kr)

**ABSTRACT** The manufacturing industry utilizes computing technology, robot technology, artificial intelligence, and IoT to improve production processes and quality. In particular, object detection technology is used in various industrial fields, and object detection methods based on deep learning are attracting attention. The tobacco processing industry requires automated production facilities, and quality control for defects in product appearance is essential. Mainly because tobacco products are sold at high prices, poor appearance is a significant issue in terms of consumer complaints and processing costs. Therefore, accurate cigarette detection is essential. We propose a modified network structure based on the YOLOv4-Tiny network, and use it to build a network optimized for cigarette detection. The modified network uses a single circular bounding box for learning and fast detection. It utilizes visual techniques, such as gradient-weighted Class Activation Mapping (Grad-CAM) to analyze the degree of activation of the network to construct an optimal network. This reduces the size of the network and increases processing speed, while maintaining detection accuracy. This paper is expected to play an important role in quality control and efficient production in the manufacturing industry.

**INDEX TERMS** Circular bounding box, detecting defective cigarettes, tobacco processing, you only look once (YOLO).

## I. INTRODUCTION

The modern manufacturing industry is making efforts to produce high-quality products with high efficiency while minimizing human intervention in the manufacturing process due to technological advancements in various fields, such as computing technology, robot technology, artificial intelligence, and the Internet of Things (IoT) [1], [2]. Defects in product appearance significantly impact product quality for consumers, so many methods include inspection processes. While the process was initially carried out by utilizing the visual ability of skilled workers, it was not easy to inspect all produced products. In addition, it is inefficient, as there is a great deal of variation depending on the worker's ability. Since then, with the development of advanced computer vision-related technologies, such as digital cameras, product exteriors can be visually inspected in even faster production processes, and have been applied to many processes [3], [4]. These machine vision systems use image processing technologies in computer vision to distinguish features that only defective products demonstrate, such as shape, color, and brightness, distinguishing between standard and defective products in product exterior images obtained from cameras. Among them, object detection technology is advantageous compared to other technologies in applications such as multi-object detection, as it can distinguish the location and the type of object in the image. It has been widely studied throughout academia and industry due to its diverse application potential. Object detection methods incorporating deep learning technology, which has excellent generalization and high classification performance while enabling real-time processing, are also widely used [5], [6], [7].

In particular, among manufacturing industries, the tobacco processing industry is an equipment industry that requires boundaries of scale and automated production facilities

The associate editor coordinating the review of this manuscript and approving it for publication was Guangcun Shan.

for many processes, from processing raw materials to final packaging. Despite controversy over the dangers of cigarettes, demand is high due to steady growth due to population growth in developing countries, and national tax policies are being used as a primary source of revenue for national finances, leading to a significant increase in the tobacco processing industry. Due to the high price of these tobacco products, there are many consumer complaints about defective product appearance, and high processing costs are incurred in compensating for defective products, making them more sensitive to quality control. In addition, production costs that are adjusted by external factors make it challenging to maintain limited consumer prices, so optimized automation equipment is essential to improve production capacity per unit time, and reduce losses through accurate defect detection. Most of the cigarettes currently sold on the market are cigarette-type, in which well-dried tobacco leaves are cut into small pieces, rolled in paper, combined with a filter, and shaped into sticks. They are generally packaged in multi-packs of 20 cigarettes, and distributed in packs. When the consumer opens the pack cover and tries to remove the cigarette, if the quantity of cigarettes is insufficient, or the cigarette is aligned in the wrong direction, this can cause great dissatisfaction from the consumer. These defects can occur when the produced cigarettes are turned over while being transferred to the conveyor, or are packaged in an unstable alignment during packaging. If such a defect is discovered during the packaging process, all cigarettes used in packaging must be discarded, so for efficient production, accurate detection is essential. Early inspection methods included a sensor that directly touches a cigarette with a rod of varying length, and a photo sensor, which uses the characteristics of light reflection depending on the presence or absence of a cigarette. In addition, image processing is utilized by applying a machine vision system. Still, continuous performance detection is difficult due to the influence of the surrounding environment, such as dust and lighting. Additionally, a speed of up to 600 packs per min is required in automated cigarette manufacturing devices for defect detection. This requires high-speed image processing at 100 ms per pack. Recently, there have been cases of using a deep learning-based object detection system with superior classification accuracy. Still, there is a lack of related research, so research is needed for accurate cigarette detection [8], [9], [10], [11], [12].

Deep learning-based object detection technology uses deep learning to distinguish objects of interest in an image from the background, proposes an area by regressing the object location on the image to the bounding box coordinates, and classifies the type of bounding box to determine the final object. In the early days, many studies consisted of separate neural networks for region proposal, and neural networks for classification. As more and more studies aimed to improve accuracy, fast performance, and learning efficiency, it gradually developed into a single-step detection method. The Region-based Convolutional Neural Network (R-CNN) [13], an early representative detection method, is a selective search algorithm for network input images. The R-CNN groups areas with a high probability of objects, extracts feature through CNN, and detects objects using Support Vector Machine (SVM). However, the CNN, SVM, and regression learning steps are all separated in a structure necessary for learning, which takes too much time. Fast R-CNN [14] further simplifies the learning step by outputting the softmax classifier loss and bounding box regression loss simultaneously, reducing the time required for learning. Still, it is inefficient, because a separate network for the candidate region proposal exists. Faster R-CNN [15] applies the Region Proposal Network (RPN) to directly propose candidate regions, which speeds up the overall processing time, making it usable. However, it is insufficient for real-time object detection. You Only Look Once (YOLO) [16], which has dramatically improved this processing time problem, is a representative first-stage detection method. There is no separate network for extracting the region of interest, the overall configuration of a single network is relatively simple, and it solves object recognition problems. This is a structure that approaches the task as a single regression problem. Since learning was performed on all input images, it showed high detection accuracy even for new images that were not used for learning, but had the disadvantage of low accuracy for small-sized objects. Afterward, YOLOv2 [17] was proposed to increase detection accuracy further, while YOLOv3 [18] has improved overall detection accuracy compared to YOLOv2. However, detection takes more processing time, due to a more profound and complex network configuration. Since then, detection performance has been improved in many studies designed with various methods and structures. There are differences in detection performance depending on the environment or field to which it is applied [19], [20], [21], [22]. A general deep learning-based object detection method detects objects using several predetermined bounding boxes to increase the probability of accurate object estimation. The greater the number of predetermined bounding boxes, the greater the amount of calculation. In factory automation, many types of objects are recognized, while there are many cases where only simple, standardized objects are recognized. Recognition of these standardized objects does not require recognizing many types of objects, and requires a network optimized for those to be identified. Recently, when detecting objects of the same size, such as cigarette detection, some studies reduce the computation time by using only a single bounding box [23], and instead of the commonly applied rectangular bounding box, a circular bounding box is applied to detect the object. One study also proposed a network with less surrounding background [24]. Therefore, when only specific, standardized objects need to be detected, such as cigarette detection, a single circular bounding box that fits the object shape is more effective. Additionally, since the size of the object to be detected does not change significantly, a layer for recognizing the same objects with different object ratios can be removed from the network.

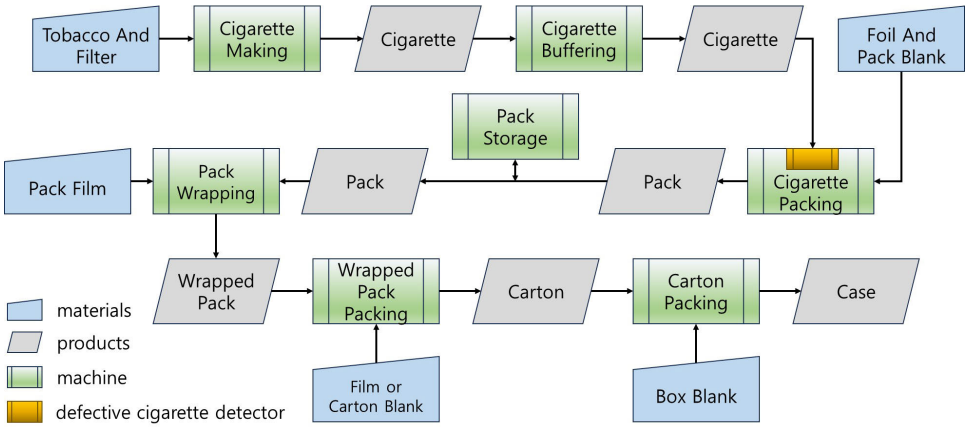



**FIGURE 1.** The Cigarette Manufacturing Process.

This study proposes a modified network structure optimized for cigarette detection based on the YOLOv4-Tiny network, which is widely used for object detection in images. The modified network achieves fast learning and performance speed using a single circular bounding box. It recognizes objects by visually analyzing the activation degree of the network, such as gradient-weighted Class Activation Mapping (grad-CAM), for faster network computation. We want to construct an optimal network by cutting out networks that do not affect the main network. The proposed network shows detection accuracy equivalent to the YOLOv4-Tiny network while reducing the size of the network to about 12%, requiring less memory to run the network, and enabling processing that is about twice as fast. Therefore, it was confirmed that the proposed networks are entirely usable. In addition, we wish to introduce the process of effectively increasing a small number of learning data using rotation transformation and image synthesis for efficient network learning.

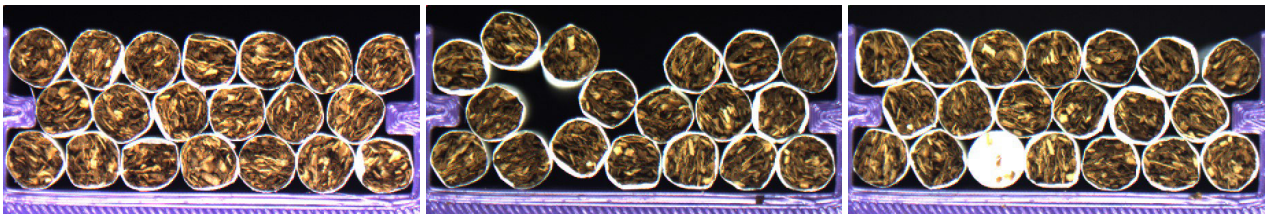
## II. RELATED WORK

### A. CIGARETTES MANUFACTURING PROCESS

Cigarettes are made by cutting well-dried tobacco leaves into thin pieces, grinding them into powder, rolling them in paper to form a stick, and combining them with a filter that blocks smoke appropriately [8], [25]. Fig. 1 shows the general manufacturing process for producing, packaging, and shipping these cigarettes. In the Cigarette Making Machine, tobacco powder, a raw material, is wrapped in paper, made into a cigarette-shaped bar, cut into appropriate sizes, combined with an input filter, and then combined to produce cigarettes. Only good quality cigarettes that have passed the Cigarette Inspection System (CIS) are transferred to the conveyor, and stored sequentially in the Cigarette Buffering Machine. The Cigarette Buffer Machine has a First In, First Out (FIFO) structure, in which the first stored cigarettes are sequentially transferred to the Cigarette Packing Machine by conveyor. During this process, the glue used to make cigarettes in the Cigarette Making Machine is sufficiently dried while passing through the Cigarette Buffering Machine. The Cigarette Packing Machine receives a Pack Blank for cigarette packaging, and shapes it into a pack. Cigarettes transferred from the Cigarette Buffering Machine are collected in a certain quantity, and bundled into a 2- or 3-layer structure on paper or aluminum foil, and one pack is packaged by assembling it with a pre-folded Pack Blank. In the next pack assembly process, normal packs that have passed various quality inspection devices are delivered to the Pack Wrapping Machine to wrap the film to maintain moisture. Afterward, the pack with completed film packaging is packaged into a carton by the Pack Packing Machine, which receives paper or film, and repackages the pack to bundle it in a certain quantity. The cartons made this way are transferred to the Carton Packing Machine by conveyor, packed into box blanks, and then transported to the shipping location as a case on a conveyor belt. In this study, we aim to design a defective cigarette detector with an optimized YOLO network in cigarettes transported to the Cigarette Packing Machine.

### B. CIGARETTE DETECTION METHOD

In manufacturing cigarettes, the process in which cigarettes are missing or aligned in the reverse direction mainly occurs at the packaging stage with the Pack Blank. Reverse alignment may arise if the direction is changed during high-speed production in the Cigarette Making Machine, or if the alignment of cigarettes is disturbed and the direction is wrong, due to various reasons during the conveyor transfer process connecting the Buffering Machine and Packing Machine. In addition, the cigarettes transferred to the Cigarette Packing Machine are arranged in a multi-stage format according to the quantity to be included in the pack, and the cigarettes placed in multi-stages are pushed into the foil, and wrapped. During this process, problems may arise if the quantity required for cigarette packaging is insufficient. Fig. 2 shows a turret created to imitate the operating situation in a Cigarette Packing Machine, and three situations in which cigarettes are aligned. The number of cigarettes that must be contained in a pack is 20, and they must be packaged in three



(a) (b) (c)

**FIGURE 2.** Examples of cigarette packs before packaging. (a) Normal; (b) Missing; (c) Inverse.

layers in a 7-6-7 structure. Fig. 2a shows a case where the cigarettes are usually well aligned, while Fig. 2b shows a case where cigarettes are missing, and Fig. 2c shows a case where a cigarette is aligned in the reverse direction.

The initial method to filter out these defects utilized capacitive or optical sensors in pockets connected to a rotating turret or conveyor belt to inspect the presence or absence of cigarettes or normal alignment [26], [27], [28]. This method makes it difficult to accurately match the operating points of the sensors and instruments used in the inspection method; durability issues due to long-term use and detection errors due to surrounding foreign substances are frequent, making management difficult. Since then, with the development of control technology and camera technology, machine vision systems with excellent performance have been utilized, and the device configuration for inspection has been simplified, resulting in a relatively wide inspection range and improved accuracy and usability. Park et al. [29] separated colors from RGB images captured by a camera to inspect the quality of packaged cigarettes, and applied the K-mean clustering technique to classify them by individual cigarette type. Sarkar et al. [30] used Vision Builder, a machine vision-based automatic inspection program available on the LabVIEW® platform of Next Instruments (NI), and Vision Assistant, an auxiliary program, to inspect the number of cigarettes inside the pack and obtain satisfactory results. Qu et al. [31] reduced the detection error by reprocessing 3D images for differences in brightness depending on height to accurately measure the number of filters in the Filter Tray supplied as raw materials during the cigarette manufacturing process. Wang et al. [32] designed a geometric model based on a normalized intensity distribution histogram and showed robust performance, despite changes in lighting brightness. These methods also require resetting detection parameters by professional personnel to obtain optimal detection performance under changes in the surrounding environment, and changes in production brands. These factors can affect production efficiency by reacting sensitively to small environmental changes. Recently, a study by Park et al. [23] proposed three improved network structures based on YOLOv4-Tiny for fast cigarette detection as a deep learning-based object detection method that shows robust performance even in changes in the surrounding environment, and excellent performance. The proposed method can process operations about two times faster and with a network size of 10 %, compared to YOLOv4-Tiny. Nevertheless, there are limitations to its actual use in high-speed automation equipment in low-level embedded environments. Therefore, an optimal network design applicable to high-speed automated equipment for detecting cigarette defects is needed. A typical object detector must classify objects of different sizes and many types. The convolution layer is set to be very complex and large to extract all the features of various objects from the network. As the network is connected to lower layers, the resolution is reduced, and the number of channels is increased through down-sampling. This reduces network memory and changes the output layer that is classified depending on the object's size. Generally, the largest object is estimated in the final output layer, and the output of the upper layer estimates the smallest object. Therefore, when the object size is constant, multiple output layers are unnecessary. An object detection system suitable for an automated cigarette defect detection system requires small network memory and fast execution speed.

### C. OBJECT DETECTION IN THE YOLO NETWORK WITH A SINGLE CIRCULAR BOUNDING BOX

The YOLO object detection method has a relatively simple structure, because the input and output are composed of one network, and directly learn the entire loss function of the network. This is a representative one-step detection method that simultaneously processes object detection and classification, and is characterized by fast object detection. As the input image passes through the network, the candidate bounding box coordinates that estimate the object location and type and the probability value for the object type are output as network output. This method selects the final bounding box using the Non-Maximum Suppression (NMS) method, which among the multiple estimated candidate bounding boxes determines the one with the highest probability value. Objects in an image vary in size and number, and two objects of different sizes may simultaneously be in the same location. Objects may overlap in the image, reducing the detection probability. To solve this problem, recent deep learning-based object detection methods increase the likelihood of accurate object estimation by determining the number of specific boxes by searching or directly setting the average object size in the learning data. However, if too many predetermined numbers of specific boxes are designed, the amount of calculation increases, while the detection performance is low for new objects that have not been learned [33]. Therefore, when the object's size is almost constant, such as the cigarette detection

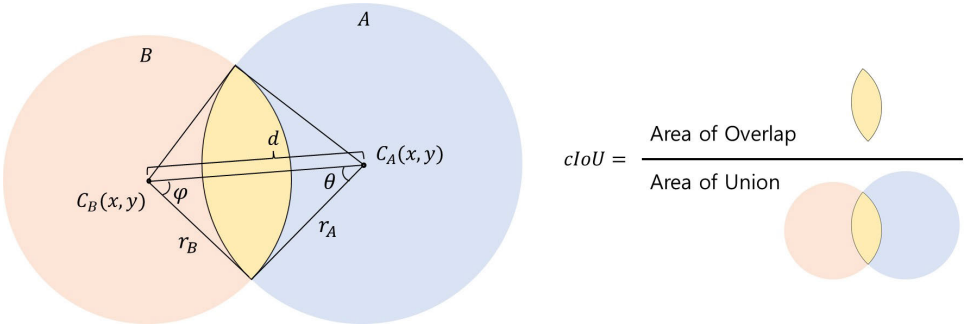



**FIGURE 3.** Circle Intersection of Union [34].

problem we wish to address in this study, applying it as just one specific box can reduce the amount of calculation, and fast computation can be expected. Additionally, no prior learning data analysis work is required to determine a particular box, and the preparation process for network learning is simple and effective. In addition, the rectangular bounding box used to designate the object area can express the position and size, so it can be fully utilized even when the object size is diverse, and the arithmetic operation is simple, so it is used in many object-detection models. However, if the object's shape is complex, or the object is rotated more than a certain angle, a lot of surrounding background other than the object is included within the designated area. In methods such as deep learning, which finds and learns the characteristics of objects on their own, the background surrounding the object can also learn and affect detection performance, so better performance can be expected by applying a bounding box type that can specify only the object [34], [35]. Therefore, since the tip of a cigarette is shaped like a circle, using a circular bounding box suitable for this is expected to increase the learning process's efficiency, and improve detection performance.

Fig. 3 shows the process of calculating the degree to which two circular bounding boxes overlap. The center points of the two circular bounding boxes are $C_A$ and $C_B$, respectively, while the radii are $r_A$ and $r_B$, respectively, with $r_A$ being larger. The distance between the center points of the two circular bounding boxes is $d = \sqrt{(C_Bx - C_Ax)^2 + (C_By - C_Ay)^2}$. For two circular bounding boxes to overlap, the $|r_A - r_B| \leq d \leq |r_A + r_B|$ condition must be satisfied. If satisfied, the degree of overlap between the two circular bounding boxes can be defined as shown in (1):

$$cIoU = \frac{Area(C_A \cap C_B)}{Area(C_A \cup C_B)} \tag{1}$$

The area where the two circles overlap is defined as (2), and all areas where the two circles overlap can be obtained using (3):

$$Area(C_A \cap C_B) = \theta r_A^2 + \varphi r_B^2 - \frac{1}{2} r_A^2 sin\ 2\theta - \frac{1}{2} r_B^2 sin\ 2\varphi \tag{2}$$

$$Area(C_A \cup C_B) = \pi r_A^2 + \pi r_B^2 - Area(C_A \cap C_B) \tag{3}$$

where, $\theta = cos^{-1}\ \frac{r_A^2+d^2-r_B^2}{2r_Ad}$, and $\varphi = cos^{-1}\ \frac{r_B^2+d^2-r_A^2}{2r_Bd}$.

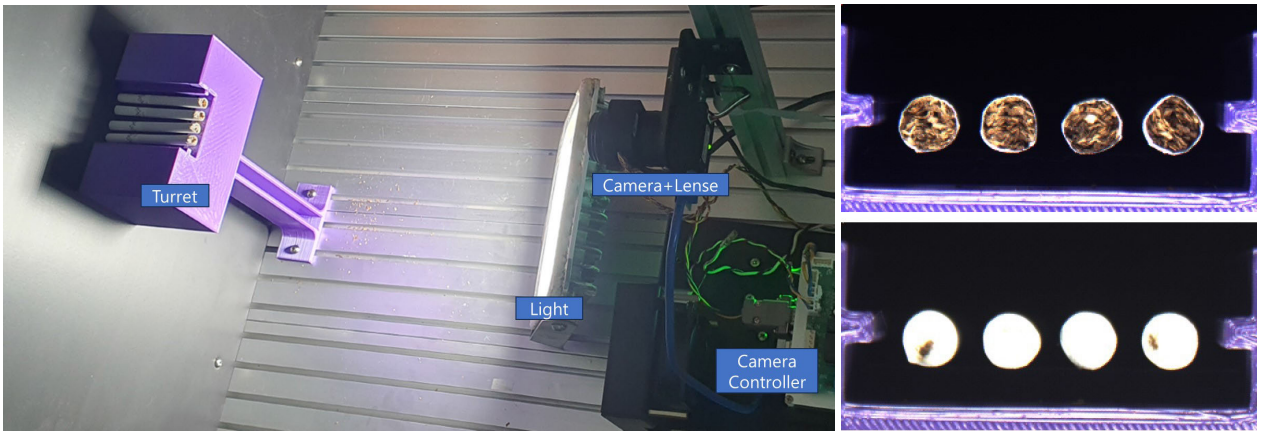
## III. THE PROPOSED METHOD

### A. TRAINING IMAGE COLLECTION AND IMAGE LABELING

Training data for cigarette detection is not publicly available, and related research is lacking. Due to security issues, filming data collection in manufacturing plants is also difficult. The turret of the cigarette manufacturing machine was made using a 3D printer, and a filming system consisting of the manufactured turret, a camera, and a controller for filming and storage was configured on a test bed. Cigarette images were collected to simulate defects that could occur during the manufacturing process. The cigarette used in this study is KT&G's SEASON® [36] product, which is readily available at retail stores in Korea. The form of the cigarette is cylindrical, being about 8 mm in diameter and 84 mm in length, and is sold in packs. One pack contains 20 cigarettes, and measures 56 mm × 88 mm × 22 mm of width × height × depth. Fig. 4 shows the environment of the personally constructed test bed, and the images obtained by alternately placing normal-aligned and reverse-aligned cigarettes on the Turret. The resolution of the acquired images was 608 × 304 pxl; 85 normal-aligned cigarette images were taken under various environmental conditions, and 39 reverse-aligned cigarette images were acquired. The objects in the images number 340 normal-aligned cigarettes, and 156 reverse-aligned cigarettes.

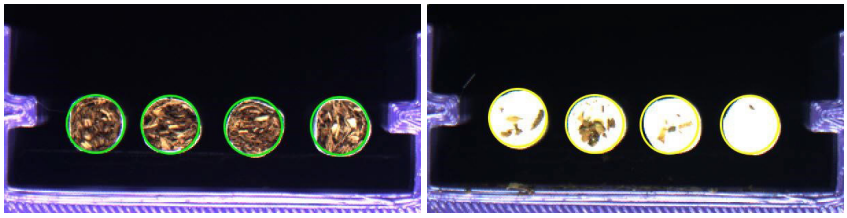
This study aims to detect cases where the number of cigarettes packed in a pack is missing, or individual cigarettes are packaged in the wrong direction. The objects to be classified are in two states. Therefore, the objects in the image were labeled by dividing them into two types, and labeled using a single circular bounding box to match the characteristics of a cigarette with a circular shape of a specific size. Equation (4) gives the definition of the circular bounding box for labeling. Table 1 summarizes the labeling, color, and number of objects, while Fig. 5 shows the labeling results for cigarettes in the image expressed as a circular bounding box.

$$cBBox_{ij} = (Bx_{ij}, By_{ij}, Br_{ij}) \tag{4}$$




(a) (b)

**FIGURE 4.** Image acquisition system and the acquired images. (a) Testbed for image acquisition; (b) Acquired images (normal and inverse aligned).



(a) (b)

**FIGURE 5.** Examples of data labeling.(a)Normal labelling; (b)Inverse labelling.

**TABLE 1.** Summary of image acquisition and data labeling.

| Label | Label Index | Color | Object count | Cigarette alignment |
|---|---|---|---|---|
| Inverse | 0 | Yellow | 156 | Inverse |
| Normal | 1 | Green | 340 | Normal |

where, $i$ is the circle Bounding Box ($cBBox$) number, $j$ is the label index, ($Bx_{ij}$, $By_{ij}$) is the center coordinate of the $cBBox$, and $Br_{ij}$ is the radius of the $cBBox$.

### B. DATA AUGMENTATION

Cigarettes are made by wrapping tobacco powder in paper, and cutting each cigarette to a specific size. The shape of the cut surface at the end of the cigarette appears in countless forms depending on the type of tobacco powder, the direction in which it is collected, and the amount collected, so it has numerous characteristics. Additionally, the number of image data collected for learning is small, so expansion of the learning data is necessary for effective network learning. Therefore, in this study, to effectively increase learning data, labeled raw data is used to cut out the cigarette portion of the image into a circular shape to create each cigarette image. Then, the individual cigarette images are rotated at a certain angle to increase the number of individual images. In addition, the number of learning data was expanded by combining individual cigarette images pre-stretched in a randomly created circular bounding box so that they did not overlap with the background-free image to the size of the raw image. Fig. 6 shows the process of increasing raw data. The rotation of individual cigarette images was set to 355° at 5° intervals. All individual cigarette images created by rotation were made into a list, and the order of the list was randomly shuffled to create the final five new lists. In addition, a single learning datum was created by combining a randomly generated circular bounding box with an overlap of less than 0.1 in an image without a background and individual cigarette images sequentially selected from five pre-created lists. This method was repeated to ensure that all images in the five lists were included, creating the final learning data set. Additionally, the brightness was randomly adjusted when selecting an image from the list to reduce the influence of the same image due to repeated use of individual cigarette images. Through this process, 12,200 items of learning data were finally created, and the total number of objects in the learning data was 56,143 for reverse-aligned cigarettes, and 122,364 for normal-aligned cigarettes.

### C. ANALYSIS THROUGH NETWORK LAYER VISUALIZATION

The capacity of a deep learning network is determined by the number of parameters that can be learned. In general, the network is connected to lower layers by down-sampling and increasing the number of channels, and depending on the shape of the bounding box, the type of object, and the number of specific bounding boxes, there may be slight differences in the final output layer. However, in the case of the YOLOv4-Tiny model, the CBL module consisting of the Convolution Layer, Batch normalization Layer, and LeakyReLU activation function is repeatedly appropriately

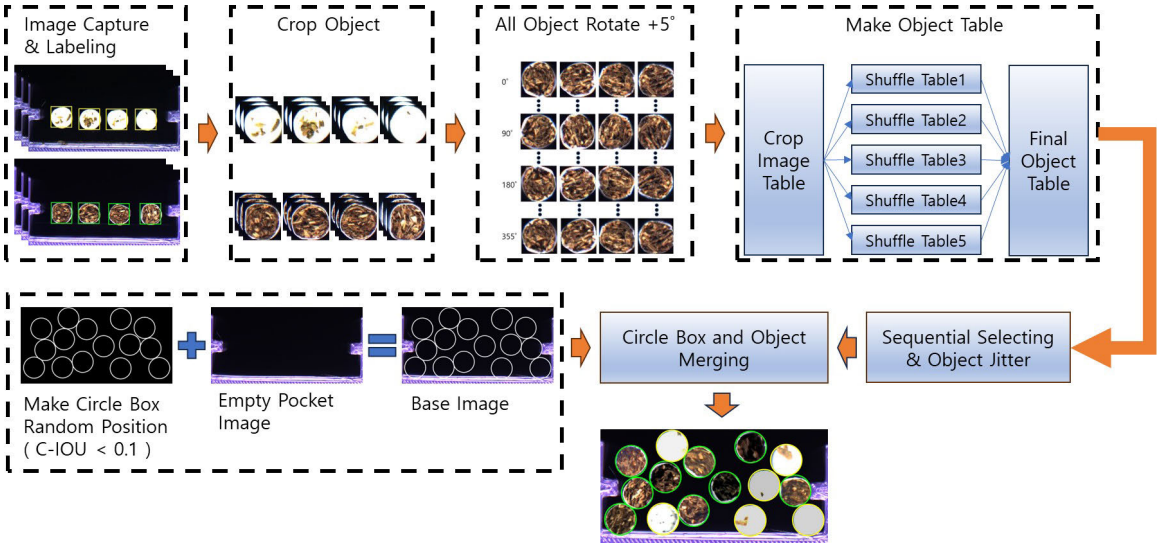



**FIGURE 6.** Data augmentation.

placed and connected, and the total number of learnable parameters in the entire network is about 5.8 Mb. Fig. 7 shows the network structure of the YOLOv4-Tiny model for cigarette detection with a circular bounding box, and the parameters of each layer are based on the size of the prepared input image.

The number of learning parameters is generally determined by the kernel size and channel depth of the convolution layer, and as the number increases, detection accuracy increases. However, as the number of parameters increases, the space occupied in memory increases, while the resulting calculation time also increases, which may affect the performance in areas where fast processing is required. When fast computational processing is necessary for an embedded environment to be applied to manufacturing machines, such as a cigarette detection system, a network configuration that allows rapid processing with a small memory footprint is required. Park's method [23], proposed by modifying the YOLOv4-Tiny model for fast cigarette detection, used only a single bounding box due to the constant size of the cigarette, reducing the learning and computational processing speed. However, a layer designed to improve detection performance for object size changes may not be necessary.

To design a network optimized for cigarettes, checking how each network element judges the cigarette object would be more efficient. Existing research includes methods such as Grad-CAM [37], which adds the connection strength from the network output to the backward-connected layer, and allows visual confirmation by projecting it onto a test image. However, Grad-CAM is a visualization technique that is designed to solve classification or regression problems of CNN networks. It is not easy to apply to structures that predict objects using bounding boxes, as in this study. Therefore, in this study, we focused on YOLO-Head, the final output stage where images are input into the network, to check the network activity level visually. The structure of YOLO-Head contains bounding box information for each channel. When input to the network, an image passes through all layers. This indicates the degree of prediction of the object at the final output stage, depending on each layer's activation degree. The YOLO-Head consists of predicted values for the location, reliability, and type of bounding box, and a channel connects each element. Therefore, the area affecting object detection can be partially confirmed by checking the value of a specific channel of YOLO-Head. In selecting an object among many object candidates, critical values are the reliability and the predicted value of the object type. Therefore, one map is created by multiplying the predicted value and reliability for each object type among YOLO-Head channels by the element, and the size of this map is upscaled using bicubic interpolation to equal the input image size. When projected onto an image, the degree of network activity for object detection can be visually confirmed, as with the input image. Fig. 8 shows the activation level of YOLO-Head1 and YOLO-Head2 according to the normal and reverse alignment states of cigarettes in the YOLOv4-Tiny model. The lower activation value of the map is blue, the intermediate value is green, and the maximum value is red. Fig. 8a is an image of a cigarette aligned in the reverse direction at the upper left edge of the test image. Fig. 8b and Fig. 8c show the normal state and reverse aligned cigarette activation level in YOLO-Head1. Both pictures represent the object to be detected. Fig. 8d and Fig. 8e show the activation level in YOLO-Head2, there is no red area, and it has no effect on object detection. Therefore, in this study, we wish to propose an optimization structure by removing such unnecessary networks.

### D. PROPOSED NETWORK STRUCTURE AND VISUALIZATION ANALYSIS

The embedded environment for detecting defective cigarettes in cigarette manufacturing machines requires fast computational processing and a network configuration with a small

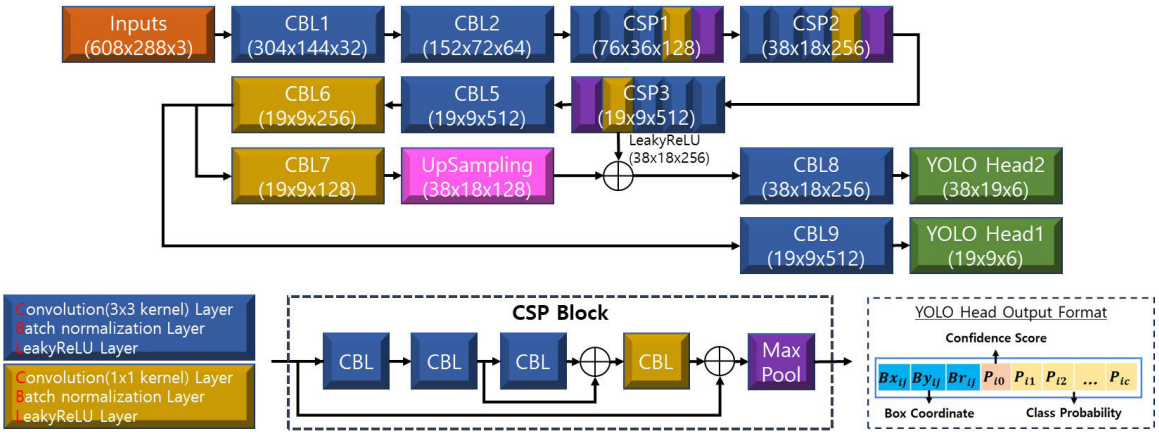



**FIGURE 7.** YOLOv4-Tiny network structure.

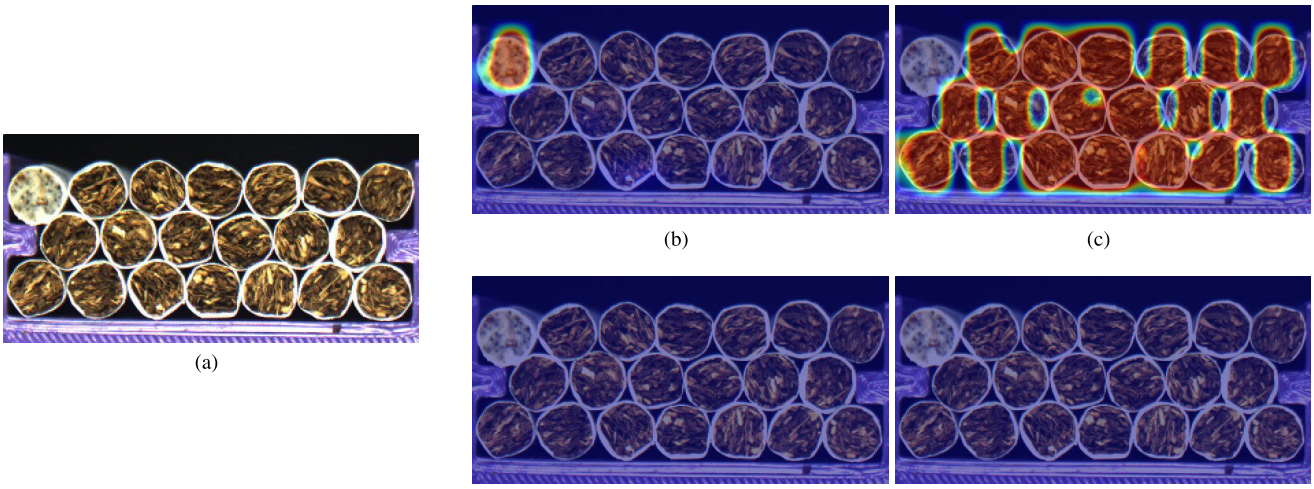



**FIGURE 8.** Test image and Activation map of the YOLOv4-Tiny network. (a) Test image; (b) Activation map in YOLO Head1 for Inverse-aligned object; (c) Activation map in YOLO Head1 for Normal-aligned object; (d) Activation map in YOLO Head2 for Inverse-aligned object; (e) Activation map in YOLO Head2 for Normal-aligned object.

memory footprint. The method of Park et al. [23] learns using only a single bounding box due to the constant size of the cigarette. Then, the YOLOv4-Tiny network's repeatedly placed Convolution Layer+Batch normalization Layer+LeakyReLU (CBL) module and Cross Stage Partial Connections (CSP) module are simply configured to reduce memory size and operation processing speed. Here, it is highly advantageous to check how network elements judge cigarette objects to design a network structure that is optimized for cigarettes. Suppose the activation level of objects in the network output is predicted, and unnecessary network elements are removed. In that case, a network can be designed that requires less memory and can perform fast calculations.

Fig. 9 shows an activation map using the network with the best performance among the networks proposed by Park et al. [23] The results in Fig. 9 show that the degree of activation for the two classes cannot be confirmed in the Head2 of the networks proposed by Park et al. [23]. Therefore, in this study, we intend to design a new network by removing the network associated with Head2.

Table 2 summarizes the number of learnable parameters of the YOLOv4-Tiny, Park's, and the proposed networks. The order of the Module column in Table 2 indicates the hierarchical connection of the proposed network. If the value of the cell parameter in the row exists, it is included in the network structure. Cells marked with '-' are not included in the network structure. The Proposed-Net1 reduces the overall number of parameters by reducing the capacity of the CBL5 module with the most significant number of parameters in the YOLOv4-Tiny structure. By repeating the CSP module of the upper layer in YOLOv4-Tiny, the increased number of channels is removed by removing the CSP1 and CSP2 modules. The number of parameters has been reduced by replacing the CBL3 module and CBL4 module, so that the number of channels of the CBL5 module is adjusted. CBL7, up-sampling, and CBL8 associated with network Head2 were removed. The final number of learning parameters is reduced by more than 5.4 Mega bytes compared to YOLOv4-Tiny, showing a clear difference in the number of parameters. In addition, the Proposed-Net2 and the Proposed-Net3 were configured similarly by setting the

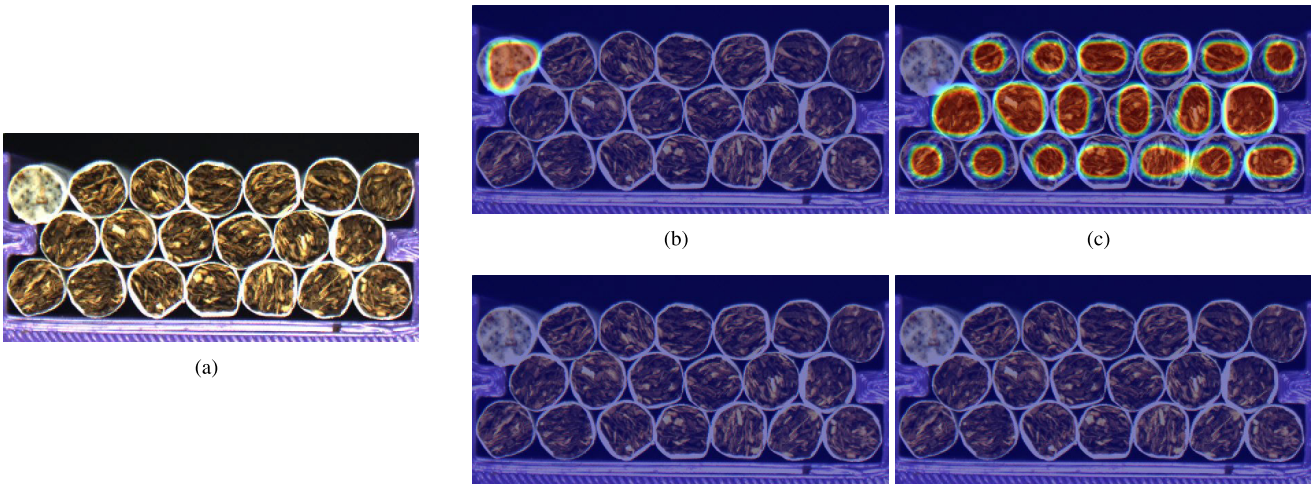



**FIGURE 9.** Test image and Activation map of the Park's network. (a) Test image; (b) Activation map in Head1 for Inverse-aligned object; (c) Activation map in Head1 for Normal-aligned object; (d) Activation map in Head2 for Inverse-aligned object; (e) Activation map in Head2 for Normal-aligned object.

**TABLE 2.** Comparison of the number of learning parameters of the networks.

| Module | YOLOv4-Tiny | The Proposed Networks | | |
|---|---|---|---|---|
| | | Net1 | Net2 | Net3 |
| CBL1 | 960 | 960 | 480 | 480 |
| CBL2 | 18,624 | 18,624 | 4,704 | - |
| CBL3 | - | 37,056 | - | - |
| MaxPool1 | 0 | 0 | 0 | 0 |
| CSP1 | 59,968 | - | - | - |
| CBL4 | - | 37,056 | - | - |
| CSP2 | 238,720 | - | - | - |
| MaxPool2 | 0 | - | - | - |
| CSP3 | 952,576 | 59,968 | 41,536 | 10,528 |
| MaxPool3 | 0 | 0 | 0 | 0 |
| CBL5 | 2,360,832 | 147,840 | 147,840 | 37,056 |
| CBL6 | 131,840 | 16,768 | 16,768 | 4,288 |
| CBL7 | 33,152 | - | - | - |
| UpSamp. | 0 | - | - | - |
| CBL8 | 885,504 | - | - | - |
| Head1 | 1,542 | 774 | 774 | 390 |
| CBL9 | 1,181,184 | 47,840 | 47,840 | 37,056 |
| Head2 | 3,078 | - | - | - |
| Total | 5,867,980 | 466,886 | 359,942 | 89,798 |

network input image differently, reducing the number of parameters by removing the CBL module, and adjusting the number of channel parameters to match the size of the final output.

Fig. 10 shows the proposed network structure as a block diagram. Compared to YOLOv4-Tiny, the network is very simply configured. The CBL module expressed colors differently depending on the filter size of the convolution layer. The blue series uses a $3 \times 3$ filter, while the yellow series uses a $1 \times 1$ filter. Additionally, if the Maxpool layer is included in the CBL module, it is indicated separately.

Fig. 11 shows the activation map and detection results after inputting a test image into the proposed networks. The proposed networks using a single circular bounding box distinguishes and analyzes objects well in the visualization analysis even though the YOLO-Head2 network has been removed. Figs. 11a, 11b, and 11c are the activation maps for the reverse and normal-aligned object and detection results in the Proposed-Net1, respectively. Figs. 11d, 11e, and 11f are the results of the Proposed-Net2 and Figs. 11g, 11h, and 11i are the results of the Proposed-Net3. Fig. 11g shows that although a weak predicted value appears in the middle of the Inverse-aligned cigarette activation map image, it is well detected in the results. This results from a more vital prediction value for the Normal-aligned cigarette in selecting a candidate bounding box, as shown in Fig. 11h.

## IV. EXPERIMENT RESULTS AND EVALUATION

### A. EXPERIMENT SETTINGS

To compare the performance of the proposed networks with the YOLOv4-Tiny and Park's [23] networks, the same parameters and learning algorithms required for network learning are used. The optimization algorithm required for learning used Stochastic Gradient Descent with Momentum (SGDM), the initial learning rate was set to 0.0001, the mini-batch size was 32, and the maximum number of iterations was set to 100. Learning was conducted by randomly shuffling the data order for all data. In this study, since the cigarette size is constant and only one specific bounding box is used, algorithms such as K-means clustering, which is used to obtain a particular predefined box from the entire data, were not used. The size of the bounding box is directly specified to be the same size as the raw data object image ($84 \times 84$ pxl). Then, the raw data image is scaled to match the size of the input layer of each network, and the bounding box data is also adjusted and applied to the same ratio. For fast processing during the network learning process, a method is used that does not compare the predicted bounding box with the raw data if the reliability of the predicted bounding box is less than 0.5. The degree of overlap between the bounding box of the raw data and the predicted bounding box is 0.5. The network loss function was defined by adding the error for object classification, the

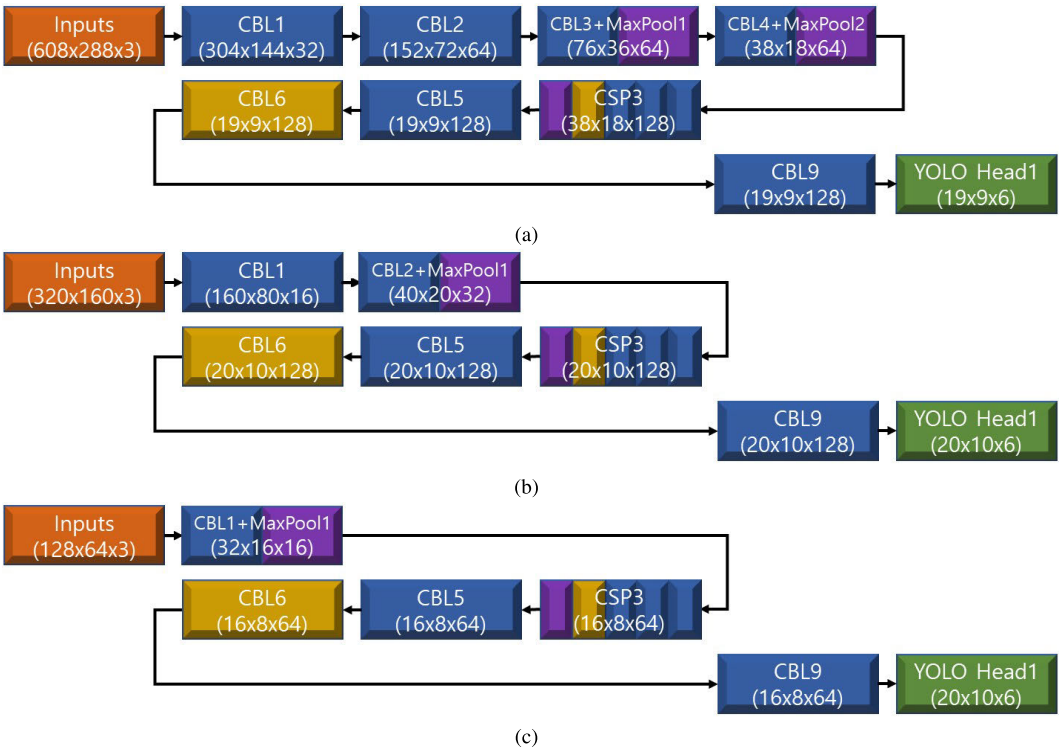



**FIGURE 10.** The proposed networks. (a) The Proposed-Net1; (b) The Proposed-Net2; (c) The Proposed-Net3.

error for object reliability, and the error for the bounding box location and size, as shown in (5). The binary cross-entropy function was applied for object classification error and object reliability error, and the least squares error function was used for bounding box error. The weights for each error are defined as $a$, $b$, and $c$, and were all used as 1 in this study.

$$TotalLoss = a \times clssloss + b \times objloss + c \times boxloss \quad (5)$$

where, $clsloss$ is the object classification loss, $objloss$ is the object confidence loss, and $boxloss$ is the box loss of $cBBox$. The $boxloss$ also includes the angular error of the $cBBox$, while $a$, $b$, and $c$ are the weight factors.

The main specifications of the hardware used in the experiments of this study are central processing unit Intel i9-9900K, system memory RAM 32 GB, and graphics card NIVIDA RTX2090. Additionally, MathWorks' MATLAB® was used as a program for network learning and evaluation, and Deep Learning Toolbox [38] was installed and used to employ the YOLOv4-Tiny model provided by the program.

### B. NETWORK LEARNING AND PERFORMANCE EVALUATION METHOD

All learning-related parameters and learning methods of the networks were implemented under the same conditions, and a small number of image data were effectively augmented to learn with 12,200 images. Table 3 summarizes the learning time and final error for all networks:

Table 3 shows that the proposed network learns faster than YOLOv4-Tiny and Park's networks under the same learning conditions. As for the overall error, the larger the network size, the smaller the final error. In the case of Net3, the simplest one proposed in this study, the overall error is 0.1245, smaller than that of Park's H-Net3. Therefore, all networks have been sufficiently trained. In this study, to evaluate the performance of the proposed network, we defined and used (6), which is commonly used in object detection methods. True Positive (TP) is defined as the case where the degree of overlap between the final predicted classification degree of the object detector and the Ground Truth (GT) labeled in the image is more than 0.5, and vice versa, it is defined as False Positive (FP). In addition, if a bounding box exists in GT but does not exist in the object detector output, it was defined as False Negative (FN), and the mean Average Precision (mAP) of the object detector was calculated as shown in (7) [23].

$$AP = \sum_{i=1}^{N} precision(i)\Delta Recall(i) \quad (6)$$

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k \quad (7)$$

where, $N$ is the total number of test images, $Precision = TP/(TP + FP)$ is the precision value for each object class, $Recall = TP/(TP + FN)$ is the object recall value, and $\Delta Recall(i)$ is the change in recall value between $i$ and $i-1$ images; and where $n$ is the total number of object classes, $k$ represents the object class, and $AP_k$ is the average precision of each object.

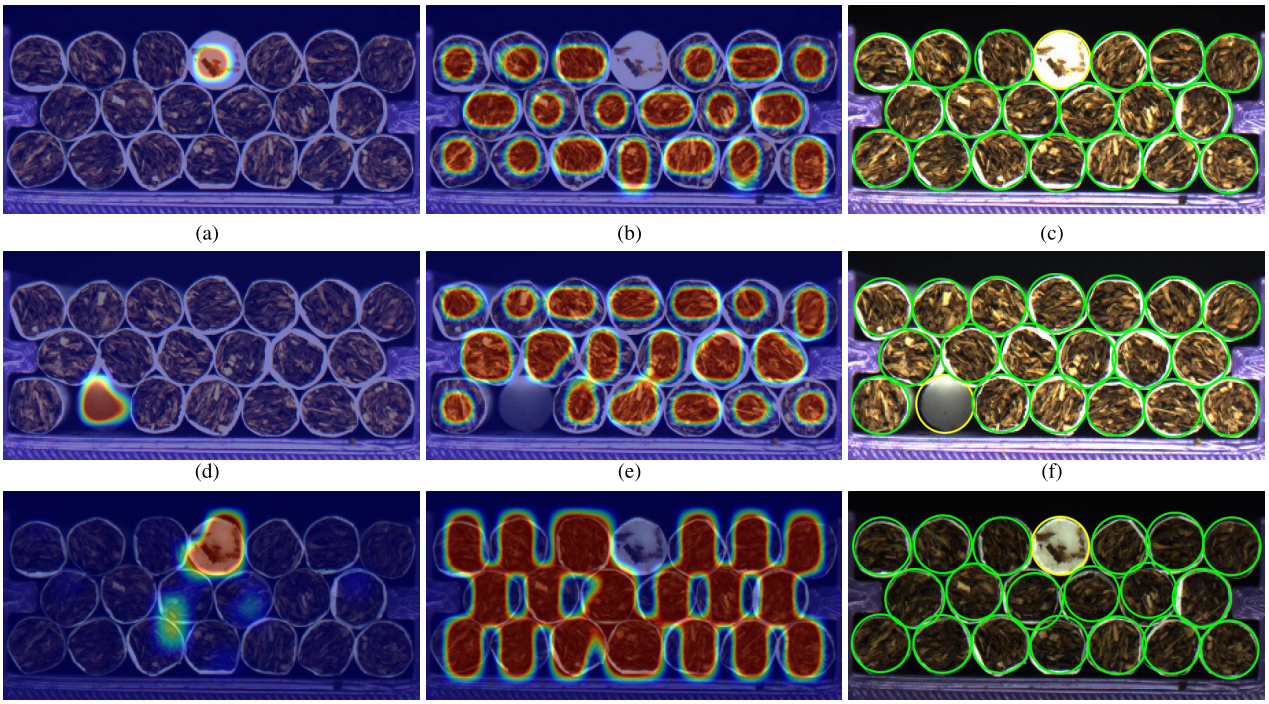



**FIGURE 11.** Activation map and results for the proposed networks. (a), (d), and (g) are the activation maps for the reverse-aligned object and (b), (c), and (h) are the activation maps for the normal-aligned object in the proposed networks. And (c), (f), and (i) are the detection results of the proposed networks.

**TABLE 3.** Comparison of the learning time of the networks.

| Module | YOLOv4-Tiny | Park's Networks [23] | | | The Proposed Networks | | |
|---|---|---|---|---|---|---|---|
| | | H-Net1 | H-Net2 | H-Net3 | Net1 | Net2 | Net3 |
| Training Time | 4h 27m | 3h 20m | 2h 57m | 2h 23m | 2h 15m | 1h 31m | 1h 15m |
| Total Loss | 0.028 | 0.0438 | 0.0748 | 0.1527 | 0.0426 | 0.0733 | 0.1245 |

### C. PERFORMANCE EVALUATION OF TEST IMAGES

In this study, cigarettes not used for learning were used as test images for performance evaluation. One hundred ninety-three test images were taken in cases where 20 cigarettes were usually placed in the turret, where several cigarettes were missing, or were aligned in the inverse direction. In addition, images were taken under various brightness conditions and environments to evaluate robust performance. To assess the performance of the proposed networks, the performance of the YOLOv4-Tiny network and the proposed networks by Park et al. [23] were compared and analyzed, and Table 4 shows the performance results:

All networks proposed in this study showed excellent performance with an *mAP* of 0.98 or higher, and the network performance speed was faster than other models. Even in the case of Proposed-Net1, which showed the slowest computation speed, it showed the same *mAP* performance as the YOLOv4-Tiny network, even though it showed a faster computation speed than the lightest H-Net3 proposed by Park et al. [23], and the proposed network was sufficient. It can be confirmed that it is effective. In particular, the Proposed-Net3 detects objects in images the fastest, with an average processing time of 0.0237 seconds. It has an excellent *mAP* of 0.98 compared to the YOLOv4-Tiny network, even though its memory size is only 1/65th that of the YOLOv4-Tiny network. In general, automated machines used in the cigarette manufacturing process require a speed of up to 600 pack per min, which is 100 ms per pack. The proposed models have a much smaller number of parameters, and the processing time for detection is fast, so it can be seen that they are all satisfactory in terms of processing time for cigarette detection. The mean $cIoU$ value compared to test data GT to measure the accuracy of object detection location shows about 5 % lower performance than the methods proposed by Park et al. [23]. However, in a system for cigarette detection, this is an acceptable error.

Fig. 12 shows an example in which the cigarette detection results were inaccurate. Even though Proposed-Net1 is a much smaller network than YOLOv4-Tiny, there are no incorrect classification results. Fig. 12c is a test image in which two cigarettes are missing. Proposed-Net2 recognized these test images as 18 normal-aligned cigarettes and one reverse-aligned cigarette. This is a case of detecting an object not in the test image. Because the space created by missing cigarettes was strongly illuminated, the activation characteristics for Inverse-aligned were strong, as shown in Fig. 12a, resulting in incorrect detection of other objects. Figs. 12d to 12f show the result of incorrect object detection

**TABLE 4. Performance evaluation by networks.**

| Network | Image Size (pxl) | Net Size | Run Time (s) | *mAP* | mean *cIoU* |
|---|---|---|---|---|---|
| YOLOv4-Tiny | 608 × 288 | 5.8M | 0.057 | 1 | 0.9622 |
| H-Net1 [23] | 608 × 288 | 817.1k | 0.041 | 0.9998 | 0.9576 |
| H-Net2 [23] | 320 × 160 | 415.3k | 0.035 | 0.9989 | 0.9427 |
| H-Net3 [23] | 128 × 64 | 181.6k | 0.032 | 0.9061 | 0.9175 |
| The Proposed-Net1 | 608 × 288 | 466.8k | 0.0314 | 1 | 0.9191 |
| The Proposed-Net2 | 320 × 160 | 359.9k | 0.03 | 0.9987 | 0.8991 |
| The Proposed-Net3 | 128 × 64 | 89.7k | 0.0237 | 0.9861 | 0.8774 |

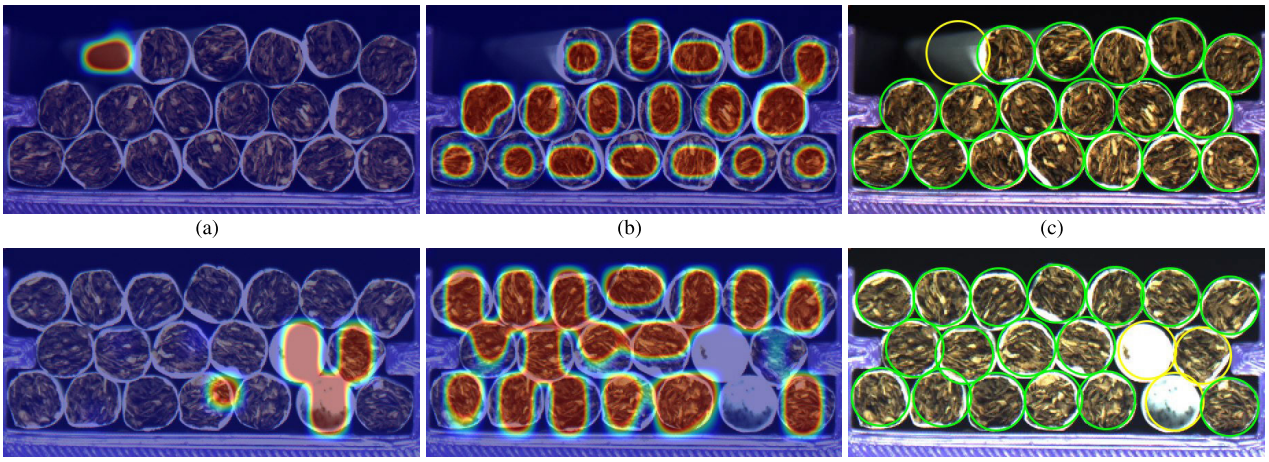



**FIGURE 12. Activation map and incorrect detected results for the proposed networks. (a), (b), and (c) are the activation maps for the reverse and normal-aligned object and incorrect detected results in the Proposed-Net2, respectively. (d), (e), and (f) are the results of the Proposed-Net3.**

in Proposed-Net3, where one Normal-aligned cigarette was incorrectly recognized as one Inverse-aligned cigarette. In Figs. 12d & e, the Inverse-aligned features appear strongly in the right part of the activation map. The test image is very different from the training data in its overall brightness and color. Therefore, if the network is trained with more diverse learning images, this problem will be solved. In the actual cigarette manufacturing process, if there is even one cigarette aligned in the reverse direction, it must be discarded, and it is infrequent that the background brightness around the Turret changes significantly. Therefore, if an appropriate surrounding environment is set, it will not be a big problem if applied to the actual process.

## V. CONCLUSION

In this study, we proposed a deep learning-based optimized cigarette detection method to achieve high quality while increasing productivity by reducing the number of discarded cigarettes through fast and accurate defect detection in the cigarette manufacturing process that requires large-scale production equipment. We propose an optimized network structure based on YOLOv4-Tiny with a single circular bounding box to quickly and accurately detect packaging defects in cigarette production, such as omission, or the reverse alignment of single cigarettes. Network optimization selected the optimal network structure through visualization analysis, and confirmed excellent performance regarding the network size, number of learnable parameters, network processing time, and detection accuracy. In particular, while Proposed-Net1 shows detection accuracy equivalent to the YOLOv4-Tiny network, the network size is as small as 12 % that of YOLOv4-Tiny, so the memory usage required to run the network is less, and processing is possibly about twice as fast. Other proposed networks can also be fully utilized. In addition, as a learning data augmentation method, the learning data is augmented by cutting and separating the original image into an object-based image, and compositing it with a bounding box generated in a randomly generated table without a background, to generate learning data efficiently.

In actual cigarette manufacturing plants, there may be more defect situations, compared to the defects set in this study. Therefore, if it is possible to secure defects that may occur in the actual manufacturing process, quality standards for them, and corresponding image data, it is expected that the optimized method for cigarette detection proposed in this study can be utilized by applying it to actual high-speed automated equipment.

## REFERENCES

[1] F. Psarommatis, J. Sousa, J. P. Mendonça, and D. Kiritsis, "Zero-defect manufacturing the approach for higher manufacturing sustainability in the era of Industry 4.0: A position paper," *Int. J. Prod. Res.*, vol. 60, no. 1, pp. 73–91, Oct. 2021.

[2] A. Saihi, M. Awad, and M. Ben-Daya, "Quality 4.0: Leveraging Industry 4.0 technologies to improve quality management practices—A systematic review," *Int. J. Quality Rel. Manage.*, vol. 40, no. 2, pp. 628–678, Jan. 2023.

[3] M. Reyna, G. Delgado, A. Akundi, S. Luna, and E. Chumacero, "Product digital quality inspection using machine vision systems—A categorical review," in *Proc. 17th Annu. Syst. Syst. Eng. Conf. (SOSE)*, Rochester, NY, USA, Jun. 2022, pp. 37–42.

[4] H. Golnabi and A. Asadpour, "Design and application of industrial machine vision systems," *Robot. Comput.-Integr. Manuf.*, vol. 23, no. 6, pp. 630–637, Dec. 2007.
[5] J. Wang, H. Dai, T. Chen, H. Liu, X. Zhang, Q. Zhong, and R. Lu, "Toward surface defect detection in electronics manufacturing by an accurate and lightweight YOLO-style object detector," *Sci. Rep.*, vol. 13, no. 1, p. 7062, May 2023.
[6] J. Bao, J. Jing, and Y. Xie, "A defect detection system of glass tube yarn based on machine vision," *J. Ind. Textiles*, vol. 53, Jan. 2023, Art. no. 152808372311528.
[7] T.-T.-H. Vu, D.-L. Pham, and T.-W. Chang, "A YOLO-based real-time packaging defect detection system," *Proc. Comput. Sci.*, vol. 217, pp. 886–894, Jan. 2023.
[8] H. K. Kim, "Global tobacco industry," *J. Korean Soc. Tobacco Sci.*, vol. 36, no. 1, pp. 34–47, Jan. 2014.
[9] F. Sheng, S. Song, and S. Xia, "A real-time cigarettes counting and loose ends detection algorithm," in *Proc. IEEE Int. Conf. Online Anal. Comput. Sci. (ICOACS)*, Chongqing, China, May 2016, pp. 11–15.
[10] W. Ying, X. Hu, and L. Zhao, "Detection of cigarette missing in packing based on deep convolutional neural network," in *Proc. IEEE 3rd Inf. Technol. Mechatronics Eng. Conf. (ITOEC)*, Chongqing, China, Oct. 2017, pp. 1252–1256.
[11] D. Harjadi, D. Fatmasari, and A. Hidayat, "Consumer identification in cigarette industry: Brand authenticity, brand identification, brand experience, brand loyalty and brand love," *Uncertain Supply Chain Manage.*, vol. 11, no. 2, pp. 481–488, 2023.
[12] K. Karanam, "Analysis of cigarette market," *SSRN J.*, Jun. 2022. Accessed: Oct. 20, 2023. [Online]. Available: https://ssrn.com/abstract=4347623
[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
[14] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1440–1448.
[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.
[17] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6517–6525.
[18] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[19] K. K. Song, M. Zhao, X. Liao, X. Tian, Y. Zhu, J. Xiao, and C. Peng, "An improved bearing defect detection algorithm based on Yolo," in *Proc. Int. Symp. Control Eng. Robot. (ISCER)*, Changsha, China, Feb. 2022, pp. 184–187.
[20] R. Huang, J. Gu, X. Sun, Y. Hou, and S. Uddin, "A rapid recognition method for electronic components based on the improved YOLO-V3 network," *Electronics*, vol. 8, no. 8, p. 825, Jul. 2019.
[21] H. Jung and J. Rhee, "Application of YOLO and ResNet in heat staking process inspection," *Sustainability*, vol. 14, no. 23, p. 15892, Nov. 2022.
[22] K. Zhao, Y. Wang, Y. Zuo, and C. Zhang, "Palletizing robot positioning bolt detection based on improved YOLO-V3," *J. Intell. Robotic Syst.*, vol. 104, no. 3, p. 41, Feb. 2022.
[23] H.-M. Park, H.-S. Jun, K.-B. Hwang, and J.-H. Park, "YOLO-based cigarette detection system using a single circular bounding box," *J. Korea Inst. Inf. Commun. Eng.*, vol. 27, no. 8, pp. 913–925, Aug. 2023.
[24] H.-M. Park and J.-H. Park, "YOLO network with a circular bounding box to classify the flowering degree of chrysanthemum," *AgriEngineering*, vol. 5, no. 3, pp. 1530–1543, Aug. 2023.
[25] O. Geiss and D. Kotzias. *Tobacco, Cigarettes and Cigarette Smoke*. Accessed: Oct. 21, 2023. [Online]. Available: https://core.ac.uk/download/pdf/38616313.pdf
[26] H. Focke and H. Mutschall, "Optical sensor for monitoring cigarette groups," U.S. Patent 4 678 901 A, Jul. 7, 1987. [Online]. Available: https://patents.google.com/patent/US4678901A/en
[27] M. Spatafora, A. Casagrande, and S. Morelli, "Device for transferring and inspecting groups of cigarettes," U.S. Patent 7 779 846 B2, Aug. 24, 2010. [Online]. Available: https://patents.google.com/patent/US7779846B2/en
[28] G. Pezzi, "Cigarette ends firmness detector," U.S. Patent 3 930 406 A, Jan. 6, 1976. [Online]. Available: https://patents.google.com/patent/US3930406A
[29] M. Park, J. S. Jin, S. L. Au, and S. Luo, "Pattern recognition from segmented images in automated inspection systems," in *Proc. Int. Symp. Ubiquitous Multimedia Comput.*, Hobart, TAS, Australia, Oct. 2008, pp. 87–92.
[30] A. Sarkar, T. Dutta, and B. K. Roy, "Counting of cigarettes in cigarette packets using LabVIEW," in *Proc. Int. Conf. Commun. Signal Process.*, Melmaruvathur, India, Apr. 2014, pp. 1610–1614.
[31] H. Qu, P. Zhang, K. Zhang, and J. Wu, "Research on cigarette filter rod counting system based on machine vision," in *Proc. Int. Conf. Life Syst. Modeling Simulation (LSMS), Int. Conf. Intell. Comput. Sustain. Energy Environ. (ICSEE)*, Nanjing, China, Sep. 2017, pp. 513–523.
[32] H. Wang, D. K. Han, Q. Shi, and H. Ko, "A novel probabilistic appearance model for cigarette detection under illumination change," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Auckland, New Zealand, Jan. 2019, pp. 1–2.
[33] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 9756–9765.
[34] G. Liu, J. C. Nouaze, P. L. Touko Mbouembe, and J. H. Kim, "YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3," *Sensors*, vol. 20, no. 7, p. 2145, Apr. 2020.
[35] H. Yang, R. Deng, Y. Lu, Z. Zhu, Y. Chen, J. T. Roland, L. Lu, B. A. Landman, A. B. Fogo, and Y. Huo, "CircleNet: Anchor-free glomerulus detection with circle representation," in *Proc. 23rd Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Lima, Peru, Oct. 2020, pp. 35–44.
[36] KT&G. *Brand Detail View: Seasons*. Accessed: Oct. 10, 2023. [Online]. Available: https://en.ktng.com/brandProductView?ntNo=15&cmsCd=CM0039&brandCd=A210108
[37] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 618–626.
[38] The MathWorks Inc. (2022). *Deep Learning Toolbox Version: 14.5 (R2022b)*. Accessed: Oct. 10, 2023. [Online]. Available: https://www.mathworks.com

**HEE-MUN PARK** received the B.S. and M.S. degrees in mechatronics engineering from the Gyeongnam National University of Science and Technology, Jinju, South Korea, in 2006 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Department of Mechatronics Engineering, Gyeongsang National University. He is also the Technical Operator of British American Tobacco Korea Ltd. His research interests include power electronics, microprocess, intelligent control, and deep learning.

**JIN-HYUN PARK** received the B.S., M.S., and Ph.D. degrees in electrical engineering from Pusan National University, Pusan, South Korea, in 1992, 1994, and 1997, respectively. He is currently a Professor with the Department of Mechatronics Engineering, Gyeongsang National University. His research interests include robotics, intelligent control, and deep learning.