Research article

Enhanced SSD framework for detecting defects in cigarette appearance using variational Bayesian inference under limited sample conditions

Shichao Wu[1,†], Xianzhou Lv[2,†], Yingbo Liu[1,*], Ming Jiang[2], Xingxu Li[1], Dan Jiang[1], Jing Yu[1], Yunyu Gong[1] and Rong Jiang[3]

[1] School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming 650221, China

[2] Hongyun Honghe Tobacco (Group) Co., Ltd. Huize Cigarette Factory, Qujing 654200, China

[3] Yunnan Key Laboratory of Service Computing, Kunming 650221, China

[†] These authors contributed equally to this work.

* **Correspondence:** Email: liuyb@ynufe.edu.cn.

Abstract: In high-speed cigarette manufacturing industries, occasional minor cosmetic cigarette defects and a scarcity of samples significantly hinder the rapid and accurate detection of defects. To tackle this challenge, we propose an enhanced single-shot multibox detector (SSD) model that uses variational Bayesian inference for improved detection of tiny defects given sporadic occurrences and limited samples. The enhanced SSD model incorporates a bounded intersection over union (BIoU) loss function to reduce sensitivity to minor deviations and uses exponential linear unit (ELU) and leaky rectified linear unit (ReLU) activation functions to mitigate vanishing gradients and neuron death in deep neural networks. Empirical results show that the enhanced SSD300 and SSD512 models increase the model's detection accuracy mean average precision (mAP) by up to 1.2% for small defects. Ablation studies further reveal that the model's mAP increases by 1.5%, which reduces the computational requirements by 5.92 GFLOPs. The model also shows improved inference in scenarios with limited samples, thus highlighting its effectiveness and applicability in high-speed, precision-oriented cigarette manufacturing industries.

Keywords: SSD; cigarette appearance defects; variational Bayesian inference; tiny target detection

1. Introduction

For cigarettes, appearance defects are critical indicators of manufacturing quality and significantly influence brand reputation and market sales. The high-speed production rates of cigarette

manufacturing machines exacerbate the challenge of detecting small-scale defects. These machines typically operate at speeds ranging from 7000 to 12,000 cigarettes per min [1], with some achieving speeds of up to 20,000 per min [2, 3]. Adherence to stringent quality standards thus necessitates a delicate balance between detection accuracy and operational efficiency.

Traditional defect detection methods, such as area-ratio-based segmentation, have difficulty precisely identifying tiny defects due to low sensitivity and inadequate segmentation quality assessment [4, 5]. In contrast, modern deep learning approaches employing convolutional neural networks (CNNs), such as YOLO [6] and SSD [7], are significantly faster and more accurate. However, these methods require extensive training data and complex supervision, potentially prohibiting their use in high-speed cigarette manufacturing machines.

Furthermore, while current deep learning algorithms based on CNNs or Transformer [8] architectures deliver impressive detection results, they often inadequately address uncertainty in data-sparse regions, potentially leading to overconfident decision-making [9]. The sporadic nature of certain defects poses significant challenges for collecting sufficient training samples, at times providing little to no data for developing a robust model. This irregular occurrence of defects complicates the task of assembling a comprehensive dataset and frequently results in underrepresentation or complete absence of specific defect types in the training process [10]. In response to these issues, this study introduces an enhanced SSD model that uses variational Bayesian inference and incorporates advanced loss and activation functions. This methodology proficiently manages weight uncertainty in scenarios with sparse data and limited sample size. The refined model balances accuracy and performance, effectively detecting minor defects at the operational speeds of cigarette machines. Validated in a high-speed production environment and tested using an industrial camera with a resolution of 1280×280, the proposed model confirms its practical efficacy in real-world scenarios, demonstrating its capability to overcome the challenges of high-speed detection and limited sample availability.

2. Related works

Small-target detection is a significant challenge that currently stands at the forefront of computer vision research [11]. Innovative algorithms primarily based on CNNs and Transformer technologies have been developed, with notable examples including the YOLO [6] and SSD [7] series, demonstrating broad applicability across various industries.

Significant progress has been made in the detection of appearance defects in cigarettes. Yang and Meng [12] refined the YOLOv5 algorithm to obtain more accurate aerial-image target detection by integrating the convolutional block attention (CBAM) module and modifying the spatial pyramid pooling (SPP) with atrous spatial pyramid pooling (ASPP), alongside a new detection head in feature pyramid network (FPN), effectively enhancing the detection of small aerial targets. Likewise, Li et al. [13] developed a faster, lightweight, region-based CNN algorithm optimized for detecting cigarette capsule defects by replacing the VGG16 network with MobileNet v1, achieving efficient real-time detection.

Furthermore, various improvements have been introduced to enhance defect detection in cigarette production. For instance, Kim et al. [14] introduced an improved YOLO v5 model for detecting small hazardous targets in industrial settings. Diers and Pigorsch [15] reviewed unsupervised learning-

based methods for industrial defect detection, highlighting cost-saving techniques that do not require predefined defect types or manual labeling. Qu et al. [10] and Wang [16] notably improved the SSD and YOLO v5 models for detecting cigarette defects, achieving high accuracy and addressing problems in traditional methods. Further contributions by Yuan et al. [17], Liu et al. [18], and Li et al. [19] have significantly improved the accuracy and efficiency of detecting defects in cigarette appearance by employing various innovative techniques. Liu et al. [18] introduced the C-CentreNet method to address issues in traditional cigarette appearance defect detection. They incorporated a convolutional attention mechanism and deformation convolution to improve the accuracy of defect detection. Yuan et al. [20] enhanced YOLO v4 to detect defects in cigarette appearance by incorporating a channel attention mechanism, a k-means++ algorithm, and α-CIoU loss. The experimental results indicate an improved overall performance. Ma et al. [21] proposed the CJS-YOLO v5n model based on YOLO v5n for detecting defects in cigarette appearance. The experimental results indicate superior performance in defect detection. Yuan et al. [17] proposed a classification method based on ResNeSt to address cigarette appearance defects in tobacco production, resulting in enhanced classification accuracy.

Recent studies by Liu et al. [22], Feng et al. [23], Yuan et al. [20], Qu [24], Liu [25], Peng [26], Ma et al. [21], Peng et al. [27], and Li et al. [28] have further extended the capabilities of deep learning models in this area. These studies introduced enhancements such as the local characteristic similarity metric, channel attention mechanisms, and various algorithmic improvements to the YOLO and SSD models, resulting in significantly improved detection of defects in cigarette appearance. Despite the wide applicability of these methods, their suitability for the stringent demands of defect detection in the cigarette industry requires further investigation and exploration.

In the realm of SSD and Bayesian techniques, Liu et al. [7] initially developed an SSD network with a multiscale architecture, enhancing its performance on various datasets such as the Pascal VOC [29], COCO [30], and ILSVRC [31]. Zhang et al. [32] introduced additional improvements, such as a residual structure and convolutional attention module, coupled with additional fusion upsampling, significantly boosting the model's performance, particularly for small targets. Leng et al. [33] proposed an energy-saving and -securing data algorithm that merges feature maps from different layers for better feature fusion, yielding superior results on datasets such as Pascal VOC. In Bayesian networks, Graves [34] used stochastic variational inference to simplify weight pruning, whereas Blundell et al. [35] fine-tuned weights with their Bayesian backprop algorithm, delivering results on the MNIST dataset on par with Dropout. Shridhar et al. [9] introduced a Bayesian convolutional neural network with variational inference, effectively addressing the challenge of uncertainty representation in sparse data scenarios. This study applies variational Bayesian inference to enhance the convolutional layer of the VGG16 network within the SSD framework. The goal is to address the challenges of limited uncertainty expression in network weights and reduce overconfidence in decision-making under sparse data conditions, thereby improving the accuracy of the SSD model for target detection.

Although these methods excel in accuracy, they often neglect the practical constraints of high-speed cigarette production, particularly the expression of uncertainty in network weights in data-sparse environments, which can lead to overly confident decisions. The present work bears certain similarities to the efforts of Li [13] and Qu [10], whose work we endeavor to extend in the following respects:

- **Balanced approach.** In parallel with existing research, we emphasize a balanced approach that strives for both detection accuracy and real-world performance while recognizing the complex

nature of practical applications.

- **Application of variational inference.** Based on similar aspirations to enhance the SSD network's weight estimates, this study integrates variational inference techniques using probability distributions to judiciously mitigate overconfident predictions.
- **Optimizing for small-scale targets.** The present research, in conjunction with previous target detection studies, specifically concentrates on targets with a relative scale ≤ 1 %, a niche area that is crucial in the field.
- **Addressing small-sample challenges.** Echoing the concerns of limited data in model training, the proposed model functions well under small-sample conditions and bolsters accuracy in these challenging scenarios.

3. Motivation

3.1. Small-scale targets

Quantifying the concept of "small-scale targets" (namely tiny targets) presents a unique challenge in the field of computer vision, particularly when focusing on minute defects in cigarettes. To address this issue, our research leverages computer vision methods to define absolute and relative scales for detecting these defects. The definition of the absolute scale T_A varies between studies and datasets [36]. For instance, Zhu et al. [37] considered objects covering 20% of a traffic sign image as small targets, whereas Torralba et al. [38] used a pixel count threshold. The present study adopts the T_A definition of Torralba et al., which, based on a specific pixel resolution threshold, is pertinent to the COCO dataset [30]. Concerning the relative scale T_R, we follow the approach of Chen et al. [39], which defines small targets based on the target-to-image area ratio. Specifically, we consider the median ratios of the target's border area to the image area, falling within $[0.08\%, 0.58\%]$.

Therefore, by integrating prior analyses with real-world challenges in actual cigarette production, this study establishes the following criteria for classifying defects in cigarette appearance:

- **Absolute Scale T_A.** The pixel resolution for specific defects such as punctures and stains falls within 15×15, making up about 59% of the defect samples. This measurement significantly undershoots the COCO standard. Being less than half of MS COCO's definition, we categorize it as a "tiny target."
- **Relative Scale T_R.** The entire image has a pixel resolution of $1,200 \times 280$. The ratio of the border area for target defects (such as punctures and stains) to the total image area varies from 0.054% to 0.075%, which is less than the 0.08% threshold suggested by Chen et al. [39] and far less than the 20% criteria set by Zhu et al. [37]. As a result, these defects are clearly identified as "tiny targets."

3.2. Cigarette defect types

Before delving into the detection of appearance defects in cigarettes, we must classify cigarette types. Although previous studies provided a rather simplistic view of this categorization, Peng et al. [27] laid the groundwork by proposing a more detailed and comprehensive classification. Building upon this and in alignment with practical requirements, this study selects eight of the most common defect types for scale identification (Table 1). Despite these advances, a significant challenge persists

Table 1. Cigarette defect types.

No.	Code	Type	Tiny target
1	FCP01	Puncture anomaly	✓
2	FJM01	Foam anomaly	✓
3	FZZ01	Fold anomaly	✗
4	FZD01	Masking anomaly	✗
5	FJZ01	Splicing paper abnormality	✗
6	FBK01	Abnormal explosion of mouth	✗
7	FZC01	Normal cigarette paper	✗
8	FZC02	Normal cigarette splicing paper	✗

in detecting specific defects such as stains and punctures because of their rarity or subtlety. In the following sections, while addressing the performance and accuracy of defect-type detection, we also consider "tiny defects," which reflects our commitment to thoroughly identifying these elusive defects. Among these defects, puncture defects (FCP01) and foaming defects (FJM01) are notable for their small size and the inherent difficulty in decting them during inspection.

3.3. SOTA methods for target defects

To analyze the ability to detect tiny defects in cigarette appearance, we experimentally compared mainstream one- and two-stage target detection algorithms. The results are presented in Table 2. This comparative study evaluates the models' detection accuracy (mAP) and reveals that YOLO v7 has the highest accuracy, followed closely by following. Notably, the SSD512 model detects small defects better than the YOLO series of models. Both the one- and two-stage models are highly accurate; however, the latter tends to be slower in detection. Given its effectiveness in identifying small defects, the SSD model, particularly the SSD512 variant, is well suited to our research needs (i.e., small dataset conditions). For a comprehensive analysis, we also include the SSD300 model in our comparisons to assess the performance of the different SSD models.

Table 2. Comparison of AP and mAP metrics across models in cigarette appearance defect detection.

Model	Mask RCNN	Faster RCNN	YOLO v5	YOLO v6	YOLO v7	YOLO v8	SSD300	SSD512
FBK01	0.7050	0.6370	0.5490	0.4760	0.7960	0.7750	0.6950	0.7720
FCP01	0.9310	0.9140	0.7610	0.7120	0.9100	0.7300	0.9040	0.9570
FJM01	0.9520	0.9290	0.6380	0.7570	0.8950	0.7960	0.8860	0.8990
FJZ01	0.7870	0.7650	0.7580	0.7430	0.7680	0.7320	0.8020	0.7560
FZD01	0.9970	0.9620	0.9940	0.9940	0.9150	0.9890	0.9970	0.9870
FZC01	0.9690	0.9730	0.9930	0.9930	0.9880	0.9930	0.9870	0.9750
FZC02	0.9510	1.0000	0.8740	0.8660	0.9840	0.8760	1.0000	0.9960
FZZ01	0.2770	0.3590	0.4400	0.4410	0.6740	0.4950	0.2790	0.3790
mAP	0.8260	0.8170	0.7510	0.7480	0.8660	0.7980	0.8190	0.8400

4. Improved SSD model based on variational Bayesian inference

As depicted in Figure 1, the framework of the enhanced SSD model illustrates the updated architecture. In this refined configuration, the convolutional layer weights within the SSD's VGG16 backbone network are determined using a backpropagation Bayesian algorithm. Given the complexity of computing the true posterior probability of these weights, a variational inference approach is used as an approximation. This variational posterior is then subjected to a series of calculations before being reintegrated into the SSD model. Such an iterative approach significantly bolsters the accuracy with which the model detects the target.

Key advancements integrated into the updated SSD network include the following:

- implementation of the backpropagation Bayesian algorithm;
- application of variational Bayesian inference;
- optimization of the loss function using BIoU;
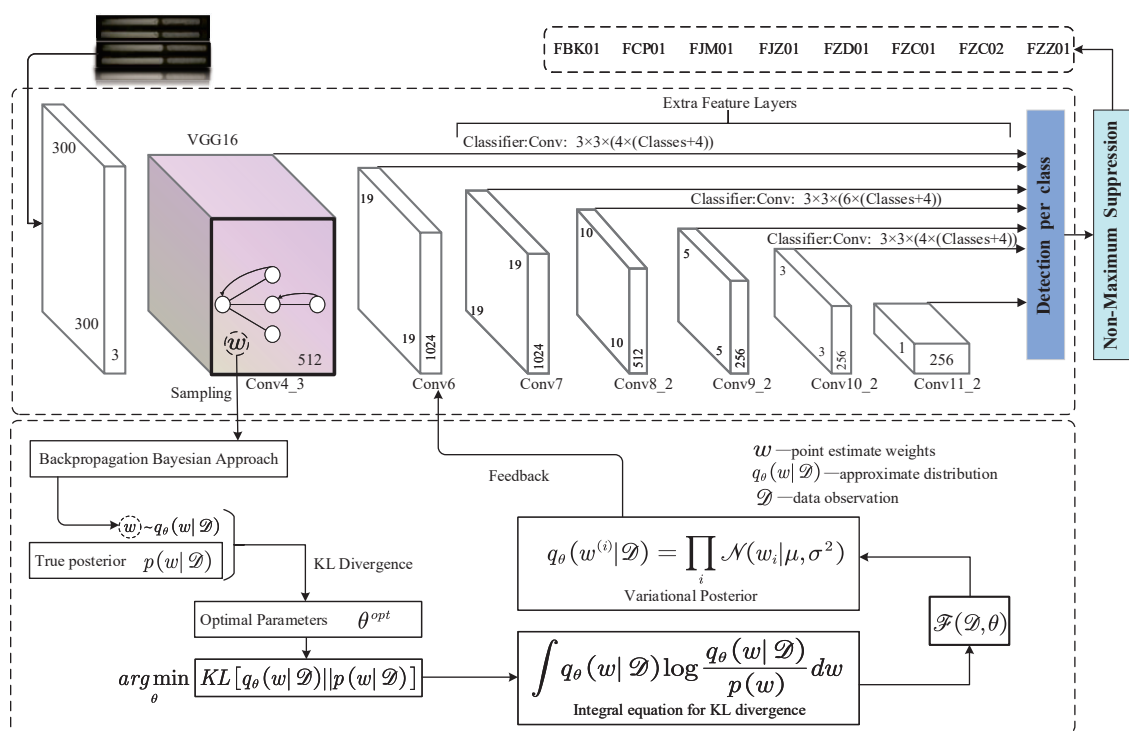- incorporation of ELU and leaky ReLU activation functions.

Figure 1. The structure of the improved SSD model, showcasing the integration of advanced techniques for enhanced defect detection.

4.1. Implementation of the backpropagation Bayesian algorithm

The Bayesian backpropagation method involves sampling neural network weights (denoted as w) during backpropagation. This method uses variational inference to approximate the posterior

distribution of these weights, represented as $w \sim q_\theta(w|\mathscr{D})$. Considering the computational complexity of obtaining the true posterior distribution $p(w|\mathscr{D})$, our approach finds distribution $q_\theta(w|\mathscr{D})$ that best approximates the true posterior.

We use Kullback–Leibler divergence to assess the similarity between the approximate and true posterior distributions. This measure quantifies the difference between the two distributions. The primary objective of the proposed method is to determine the optimal parameter set θ^{opt} that minimizes the Kullback–Leibler divergence between the approximate distribution $q_\theta(w|\mathscr{D})$ and the true posterior distribution $p(w|\mathscr{D})$. This optimization problem is formulated as Eq (4.1).

$$\theta^{opt} = arg\min_\theta KL\left[q_\theta(w|\mathscr{D})\,\|\,p(w|\mathscr{D})\right]$$

$$= arg\min_\theta KL\left[q_\theta(w|\mathscr{D})\,\|\,p(w)\right] - \mathbb{E}_{q(w|\theta)}\left[\log p(\mathscr{D}|w)\right] + \log p(\mathscr{D}) \tag{4.1}$$

where,

$$KL\left[q_\theta(w|\mathscr{D})\,\|\,p(w)\right] = \int q_\theta(w|\mathscr{D})\log\frac{q_\theta(w|\mathscr{D})}{p(w)}dw \tag{4.2}$$

The optimization problem of Eq (4.1) is centered around the variational free-energy cost function, as delineated by Kullback and Leibler [40]. This cost function comprises three terms: the first term is the complexity cost $KL[q_\theta(w|\mathscr{D})|p(w)]$, which depends on the prior distribution $p(w)$. The second term is the likelihood cost $\mathbb{E}_{q(w|\theta)}[\log p(\mathscr{D}|w)]$, which depends on the likelihood of the data given the weights $p(\mathscr{D}|w)$. The third term $\log p(\mathscr{D})$ is a constant and is thus omitted from the optimization process.

Because of the complexity involved in computing the exact value of the Kullback–Leibler divergence, an exact solution is not trivial. To address this problem, we use the stochastic variational approach proposed by Graves [34], which involves optimizing a more tractable cost function that can be minimized during the training process by determining the optimal parameters θ. The equation for determining the optimal parameters θ is further expounded in Eq (4.3).

$$\mathscr{F}(\mathscr{D},\theta) \approx \sum_{i=1}^{n}\log q_\theta\left(w^{(i)}|\mathscr{D}\right) - \log p\left(w^{(i)}\right) - \log p\left(\mathscr{D}|w^{(i)}\right) \tag{4.3}$$

In Eq (4.3), the first term in the logarithm is a Gaussian distribution with mean μ and variance σ^2; it represents the variational posterior. This may be expressed as Eq (4.4).

$$q_\theta\left(w^{(i)}|\mathscr{D}\right) = \prod_i \mathscr{N}\left(w_i|\mu,\sigma^2\right) \tag{4.4}$$

Taking the natural logarithm of Eq (4.4) results in Eq (4.5).

$$\log\left(q_\theta\left(w^{(i)}|\mathscr{D}\right)\right) = \sum_i \log\mathscr{N}\left(w_i|\mu,\sigma^2\right) \tag{4.5}$$

which is the natural logarithm of the posterior distribution. The second term in Eq (4.3) is associated with the prior weight and is a product of individual Gaussian distributions, as detailed in Eq (4.6).

$$p\left(w^{(i)}\right) = \prod_i \mathscr{N}\left(w_i|0,\sigma_p^2\right) \tag{4.6}$$

The natural logarithm of Eq (4.6) provides the logarithmic expression of the prior, as shown in Eq (4.7).

$$\log\left(p\left(w^{(i)}\right)\right) = \sum_i \log \mathcal{N}\left(w_i|0, \sigma_p^2\right) \qquad (4.7)$$

The final term in Eq (4.3) is the likelihood, which is computed using the Softmax function.

4.2. Application of variational Bayesian inference

This paper tackles the intricate challenge of applying Bayesian methods to SSD networks, which are characterized by numerous parameters and complex functional behavior. These characteristics often preclude the possibility of exact integral solutions. Our approach involves integrating variational Bayesian inference within the VGG16 backbone of the SSD network, which replaces deterministic weights in the convolutional filters with probabilistic distributions. This implementation of variational inference leads to precise estimates of the posterior probabilities of these weights. Figure 2 shows the computational workflow of this process. This technique addresses the network's shortcomings in expressing uncertainty and circumvents overconfidence in decision-making, especially in regions of sparse data. Consequently, this approach functions in situations characterized by small data samples, where traditional models might struggle because of insufficient training data.
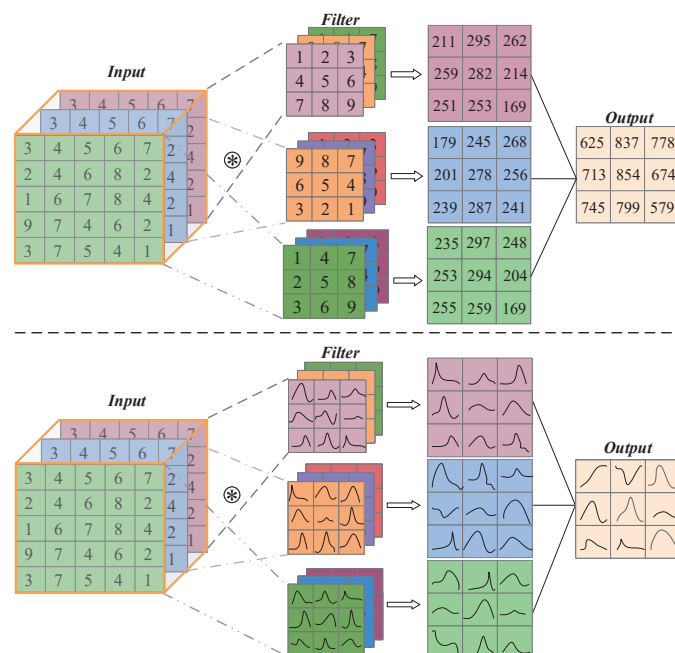

Figure 2. Traditional convolutional weights vs. probabilistic weights in VGG network architecture.

To develop Bayesian SSD networks, we use the local reparameterization technique proposed by Kingma et al. [41]. This technique involves sampling from the activation layer b instead of directly from the weights w. Implemented in the convolutional layer, this approach uses the variational posterior

distribution $q_\theta\left(w_{ijhw}, \alpha_{ijhw}\mu_{ijhw}^2\right)$, where i, j, h, w are the dimensions of the input and output layers and the filter size, respectively. Consequently, the activation layer b succeeds the convolutional layer as Eq (4.8).

$$b_j = A_i * \mu_i + \epsilon_i \odot \sqrt{A_i^2 * \left(\alpha_i \odot \mu_i^2\right)} \tag{4.8}$$

where $\epsilon_i \sim \mathcal{N}(0, 1)$ is a random variable that follows a standard normal distribution, A_i is the receptive field of the convolutional layer, $*$ is the convolution operation, and $\odot$ indicates component-wise multiplication.

4.3. Optimizing the loss function using BIoU

The BIoU loss function proposed by Lachlan et al. [42] is a key advancement in bounding box prediction optimization. It maximizes the overlap between the region of interest and the ground truth bounding box, thereby enhancing the accuracy of bounding box predictions. This loss function is specifically designed to improve the convergence of gradient descent optimization techniques, making it particularly beneficial for two-stage target detection algorithms.

In the context of a given sampling frame $b_s = (x_s, y_s, w_s, h_s)$ of the region of interest, the associated true target $b_t = (x_t, y_t, w_t, h_t)$, and the estimated bounding box $\beta = (x, y, w, h)$, the cost functions used in RCNN are defined as Eqs (4.9) and (4.10).

$$Cost_x = L_1\left(\frac{\Delta x}{w_s}\right) \tag{4.9}$$

$$Cost_w = L_1\left(\log\left(\frac{w}{w_t}\right)\right) \tag{4.10}$$

where $\Delta x = x - x_t$ is the displacement, and $L_1(z)$ is the Huber loss [43]. The Huber loss is defined within a restricted range of X and Y, as shown in Eq (4.11).

$$L_\tau(z) = \begin{cases} \frac{1}{2}z^2 & |z| < \tau \\ \tau|z| - \frac{1}{2}\tau^2 & otherwise \end{cases} \tag{4.11}$$

Although bounding box regression minimizes the IoU [expressed as $Cost = L_1(1 - IoU(b, b_t))$], various challenges impede CNNs from efficiently minimizing this loss during gradient descent. The following cost function addresses these challenges as Eq (4.12).

$$Cost_i = 2L_1\left(1 - IoU_B(i, b_t)\right) \tag{4.12}$$

In Eq (4.12), $IoU(b, b_t) \leq IoU_B(i, b_t)$ signifies an upper bound on the IoU function with free parameters $i \in x, y, w, h$. These unconstrained free parameters can be used to establish a maximum upper bound for the IoU. The lower bound for the IoU function is defined as Eqs (4.13) and (4.14).

$$IoU_B(x, b_t) = \max\left(0, \frac{w_t - 2|\Delta x|}{w_t + 2|\Delta x|}\right) \tag{4.13}$$

$$IoU_B(w, b_t) = \min\left(\frac{w}{w_t}, \frac{w_t}{w}\right) \tag{4.14}$$

4.4. Incorporation of ELU and leaky ReLU activation functions

One limitation of the ReLU activation function is that its gradient is zero when the input is negative, which leads to problems such as vanishing gradients and "neuron death" in deep neural networks. To address these issues, we introduce two alternative activation functions: ELU and leaky ReLU.

The leaky ReLU activation function mitigates the vanishing gradient problem by introducing a small, nonzero slope for negative inputs. This feature ensures that the gradient does not vanish and accelerates convergence during training. In contrast, the ELU activation function prevents the gradient from diminishing excessively for negative inputs, thereby overcoming the vanishing gradient problem. Moreover, ELU offers a smooth transition across the entire input range, including zero. This continuity of the gradient, which characterizes the ELU function, simplifies the optimization and enhances the performance of deep neural networks.

The leaky ReLU (f) and ELU (h) activation functions are expressed as Eqs (4.15) and (4.16), respectively.

$$f(x) = \max(\alpha x, x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \le 0 \end{cases} \tag{4.15}$$

$$h(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \le 0 \end{cases} \tag{4.16}$$

These functions provide alternatives to address the challenges posed by ReLU, thereby improving the robustness and efficiency of neural network learning.

5. Experimental results and discussion

5.1. Configuration

The experimental setup for this study consisted of a 64-bit Windows Server 2019 operating system, 256 GB of RAM, an NVIDIA Tesla P100 graphics card with 12 GB of graphics memory, and dual Intel Xeon E5-2650 v4 processors clocked at 2.20 GHz. The experiments were conducted using Python 3.7.0 and the PyTorch 1.8.2 framework, supported by CUDA 11.1.

5.2. Dataset description and training parameters

This study used a dataset comprising 2128 defect images, each of $1,280 \times 280$ pixels, obtained from an operational cigarette factory. This dataset is notably smaller in scale than datasets such as Cifar [44] and ImageNet [31], situating this research within the realm of small data analysis. Each bmp image in this dataset shows two cigarettes and is time-stamped to ensure accuracy.

For our analysis, the dataset was partitioned into training and validation sets in an 80:20 ratio, allocating 1722 images for training and 406 for validation. We used an SSD network enhanced with a VGG16 backbone for defect detection. The network was trained using stochastic gradient descent as the optimization algorithm. The stochastic gradient descent configuration was carefully calibrated, starting with an initial momentum of 0.9 and a learning rate of 0.02, in conjunction with a weight decay rate of 0.0001. To further refine the training process, we gradually increased the learning rate from its initial value. This approach enhances the training efficiency and consequently the overall performance of the model.
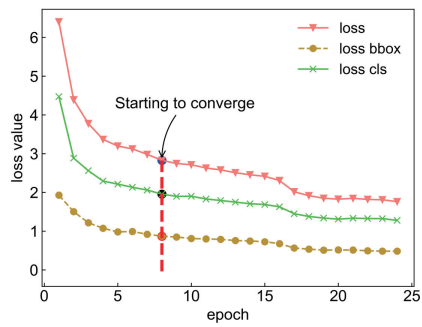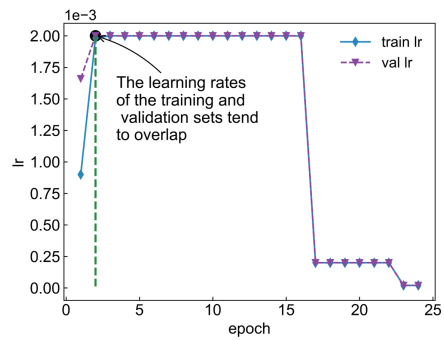
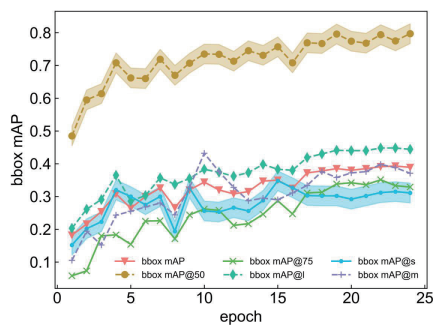(a) Loss function for improved SSD300
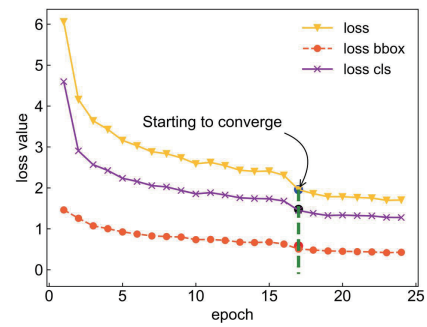
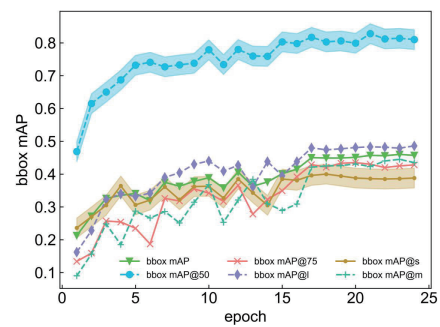(b) Loss function for improved SSD512
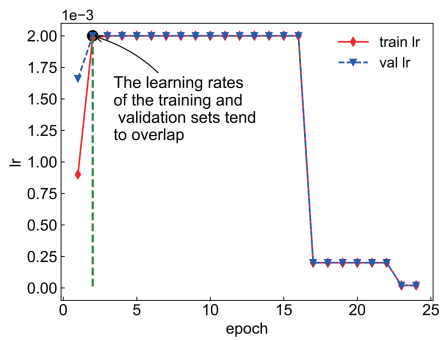
(c) Learning rate for improved SSD300

(d) Learning rate for improved SSD512

(e) Object detection mAP for improved SSD300

(f) Object detection mAP for improved SSD512

Figure 3. Training process of in improved SSD300 and SSD512 models.

Fine-tuning the α parameter within the activation functions, specifically the leaky ReLU and ELU, is critical for optimizing model performance. Conventionally, the ELU uses a default α value of 1.0. In contrast, the leaky ReLU uses α to regulate its slope, thereby affecting the function's behavior for negative inputs. To empirically determine the optimal α value, we systematically varied α in a series of experiments. Figure 4 shows the empirical results, which demonstrate that $\alpha = 0.13$ significantly enhances the detection capability of the model. Notably, with this configuration, the performance metrics of both the SSD300 and SSD512 models surpass those of the established benchmark model.
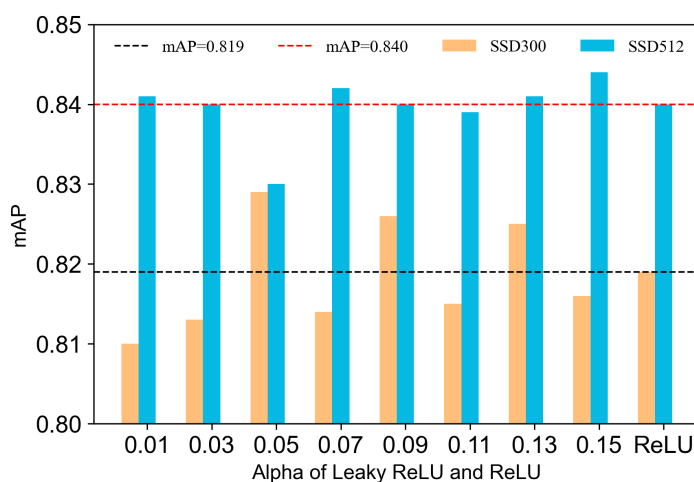

Figure 4. Performance enhancement of the leaky ReLU activation functions as a function of α. The graph shows the mean average precision achieved for different values of α. For comparison, we show the mAP value of the benchmark SSD300 model (black dashed line) and of the benchmark SSD512 model (red dashed line).

5.2.1. Analysis of the training process

Figure 3 presents a comprehensive analysis of the key training metrics for the enhanced SSD300 and SSD512 models, including the loss function, learning rate, and mAP. Figure 3(a),(b) reveal a decline in loss for both models with increasing training epochs, with SSD300 achieving lower loss quicker than SSD512. Figure 3(c),(d) detail the learning rate trends, where both models converge around epoch 2.5, followed by stability up to epoch 17, and then a decrease. Finally, Figure 3(e),(f) show the progression of the mAP, with both models producing similar results, although SSD512 attains a marginally higher mAP at the end of training.

5.2.2. Accuracy evaluation in improved SSD models

In Table 3, "VB" stands for variational Bayesian inference, "LReLU" for leaky ReLU activation function, and "BIoU" for bounded IoU loss function. The target sizes are categorized in "Area" as small ($<32^2$ pixels), medium (32^2 to 96^2 pixels), and large ($\geq 96^2$ pixels), and "maxDets" refers to

Table 3. Average precision of improved SSD and SSD.

Model	IoU	area	MaxDets	Original	VB	BIoU	ELU	LReLU
SSD300	0.50:0.95	all	100	0.477	0.542	0.522	0.539	0.546
	0.50	all	1000	0.819	0.831	0.817 ↓	0.803 ↓	0.807 ↓
	0.75	all	1000	0.455	0.506	0.487	0.524	0.538
	0.50:0.95	small	1000	0.256	0.463 ↑	0.430 ↑	0.476 ↑	0.466 ↑
	0.50:0.95	medium	1000	0.262	0.436	0.383	0.419	0.477
	0.50:0.95	large	1000	0.556	0.655	0.655	0.657	0.629
SSD512	0.50:0.95	all	100	0.513	0.571	0.528	0.564	0.574
	0.50	all	1000	0.840	0.851	0.838 ↓	0.841	0.871
	0.75	all	1000	0.480	0.569	0.505	0.548	0.565
	0.50:0.95	small	1000	0.297	0.458 ↑	0.418 ↑	0.481 ↑	0.462 ↑
	0.50:0.95	medium	1000	0.338	0.355	0.294 ↓	0.318 ↓	0.382
	0.50:0.95	large	1000	0.608	0.631	0.634	0.619	0.605 ↓

the maximum predicted bounding boxes per image. For IoU = 0.50 and maxDets = 1,000, the VB-enhanced SSD300 model achieves the highest average accuracy of 0.831, outperforming the original SSD300 model in all size categories. The ELU-activated SSD300 excels for small- and large-target detection, whereas the LReLU variant is superior for medium targets.

For the SSD512 model with the same IoU and maxDets, the LReLU-enhanced version attains a maximum average accuracy of 0.871. The BIoU-modified SSD512 model decreases slightly in overall accuracy compared with the original SSD512 model, yet all enhanced SSD512 variants excel in small-target detection. The LReLU-modified SSD512 surpasses the original SSD512 model in medium-target detection, and the BIoU version is more accurate for large targets.

Table 4. Average recall rate of improved SSD and SSD.

Model	IoU	area	MaxDets	Original	VB	BIoU	ELU	LReLU
SSD300	0.50:0.95	all	1	0.313	0.315	0.313	0.326	0.327
	0.50:0.95	all	10	0.614	0.631	0.609 ↓	0.634	0.632
	0.50:0.95	all	100	0.617	0.637	0.614	0.636	0.637
	0.50:0.95	small	100	0.572	0.564 ↓	0.530 ↓	0.572	0.554 ↓
	0.50:0.95	medium	100	0.490	0.510	0.455	0.479 ↓	0.547
	0.50:0.95	large	100	0.683	0.710	0.722	0.727	0.699
SSD512	0.50:0.95	all	1	0.315	0.320	0.304 ↓	0.331	0.326
	0.50:0.95	all	10	0.659	0.649 ↓	0.616 ↓	0.661	0.664
	0.50:0.95	all	100	0.665	0.637 ↓	0.626 ↓	0.667	0.672
	0.50:0.95	small	100	0.566	0.657	0.506 ↓	0.585	0.568
	0.50:0.95	medium	100	0.582	0.564 ↓	0.487 ↓	0.534 ↓	0.552 ↓
	0.50:0.95	large	100	0.711	0.707 ↓	0.723	0.715	0.694 ↓

5.2.3. Recall assessment of the enhanced SSD models

Based on Table 4, the enhanced SSD300 and SSD512 models achieve their highest average recall rates, particularly for large targets, with maxDets = 100 and with an IoU in the range 0.50–0.95. The ELU-activated SSD300 model and the BIoU-enhanced SSD512 model achieve peak average recalls of 0.727 and 0.723, respectively, outperforming the other models.

The average recall increases as maxDets increases. For small-target detection, the ELU-enhanced SSD300 model leads with an average recall of 0.572, closely followed by the ELU-modified SSD512 model at 0.585. For medium targets, the leaky ReLU–activated SSD300 model performs the best. Overall, the enhanced SSD models consistently demonstrate superior recall across all target sizes, highlighting their effectiveness in various detection scenarios.

5.2.4. Comparative analysis of defect appearance types

Table 5. AP and mAP value comparison for various defect types: SSD and improved SSD models.

Model	Categories	Original	VB	BIoU	ELU	LReLU
SSD300	FBK01	0.695	0.815 ↑	0.783 ↑	0.745 ↑	0.697
	FCP01	0.904	0.854	0.875	0.874	0.894
	FJM01	0.886	0.910 ↑	0.849	0.870	0.864
	FJZ01	0.802	0.797	0.801	0.797	0.794
	FZD01	0.997	0.990	0.994	0.987	0.994
	FZC01	0.987	0.968	0.985	0.984	0.987
	FZC02	1.000	1.000	1.000	1.000	1.000
	FZZ01	0.279	0.300 ↑	0.353 ↑	0.374 ↑	0.372 ↑
	mAP	0.819	0.831 ↑	0.830 ↑	0.829 ↑	0.825 ↑
SSD512	FBK01	0.772	0.772	0.772	0.759 ↑	0.724
	FCP01	0.957	0.955	0.929	0.960 ↑	0.898
	FJM01	0.899	0.884	0.893	0.914 ↑	0.907 ↑
	FJZ01	0.756	0.800 ↑	0.783 ↑	0.805 ↑	0.780 ↑
	FZD01	0.987	0.987	0.992 ↑	1.000 ↑	0.997 ↑
	FZC01	0.975	0.982 ↑	0.973	0.981 ↑	0.981 ↑
	FZC02	0.996	0.991	0.978	0.997 ↑	0.996
	FZZ01	0.379	0.437 ↑	0.383 ↑	0.353	0.447 ↑
	mAP	0.840	0.851 ↑	0.832	0.846 ↑	0.841 ↑

Table 5 shows that the enhanced SSD300 models, which incorporate variational Bayesian inference, the BIoU loss function, and ELU and leaky ReLU activation functions, surpass the original SSD300 model in overall detection accuracy, registering gains of 1.2, 1.1, 1.0, and 0.6%, respectively. Although the original SSD300 model is more accurate for detecting small puncture defects, the variational Bayesian variant excels in identifying pinch foam defects. Moreover, all enhanced SSD300 models perform optimally in detecting crease and bursting defects and larger defects such as masking and sections of cigarette and splice paper in normal cigarettes.

In contrast, as shown in Table 5, the BIoU-enhanced SSD512 model is less accurate than the original SSD512 model. The variational Bayesian, ELU, and leaky ReLU-enhanced SSD512 models improve the accuracy by 1.1, 0.6, and 0.1%, respectively. Notably, the ELU-enhanced SSD512 model excels in detecting tiny puncture defects, surpassing both the other enhanced versions and the original SSD512 model. The ELU and leaky ReLU-enhanced SSD512 models also outperform the original SSD512 model in pinpointing tiny pinch foam defects. Overall, the improved SSD512 models offer optimal detection for various types of defects.

5.3. Ablation experiments

5.3.1. Accuracy

Section 5.2.1 details significant enhancements to the SSD model, including the integration of variational Bayesian inference, the adoption of the BIoU loss function, and the implementation of the ELU and leaky ReLU activation functions. These modifications notably enhance the model's performance compared with the original configuration. This section discusses ablation experiments that assess the individual and combined effects of these improvements, with a special focus on the ELU activation function. This emphasis is crucial for understanding issues such as vanishing gradients or explosions that may occur because of the synergy of ELU with other enhancements in the SSD model. The parameter settings for these ablation studies are the same as those in section 5.2.1.

Table 6. Comparative AP and mAP values for defect types: SSD vs. improved SSD models.

Model	Categories	Original	VB+BIoU	VB+LReLU	BIoU+LReLU	VB+LReLU+BIoU
	FBK01	0.695	0.806 ↑	0.700 ↑	0.703 ↑	0.577
	FCP01	0.904	0.812	0.887	0.898	0.858
	FJM01	0.886	0.839	0.890 ↑	0.888 ↑	0.882
	FJZ01	0.802	0.811 ↑	0.808 ↑	0.800	0.808 ↑
SSD300	FZD01	0.997	0.981	0.994	1.000 ↑	0.997
	FZC01	0.987	0.985	0.985	0.989 ↑	0.986
	FZC02	1.000	1.000	1.000	1.000	1.000
	FZZ01	0.279	0.284 ↑	0.370 ↑	0.399 ↑	0.344 ↑
	mAP	0.819	0.815	0.829 ↑	0.835 ↑	0.807
	FBK01	0.772	0.733	0.683	0.727	0.784 ↑
	FCP01	0.957	0.958 ↑	0.957	0.993 ↑	0.935
	FJM01	0.899	0.880	0.897	0.917 ↑	0.917 ↑
	FJZ01	0.756	0.804 ↑	0.784 ↑	0.777 ↑	0.791 ↑
SSD512	FZD01	0.987	0.975	1.000 ↑	0.990 ↑	1.000 ↑
	FZC01	0.975	0.983 ↑	0.980 ↑	0.982 ↑	0.981 ↑
	FZC02	0.996	0.996	0.993	0.995	0.993
	FZZ01	0.379	0.370	0.442 ↑	0.444 ↑	0.471 ↑
	mAP	0.840	0.837	0.842 ↑	0.845 ↑	0.855 ↑

Table 6 reveals that the ablation experiments improved the SSD300 model's overall detection accuracy. The incorporation of variational Bayesian inference and the leaky ReLU activation

function, especially when combined with the BIoU loss function, increases the accuracy by 1 and 1.6%, respectively. These modifications improve the detection of bursting, foaming, splicing, and crease defects. However, combining variational Bayesian inference with the BIoU loss function slightly reduces the overall accuracy, despite enhancing the performance in specific defect categories.

Regarding the SSD512 model, all combinations of enhancements, barring the pairing of variational Bayesian inference with the BIoU loss function, increase the overall detection accuracy. These combinations included variational Bayesian inference with leaky ReLU, the BIoU loss function with a leaky ReLU, and the integration of all three enhancements, which improved the accuracy by 0.2, 0.5, and 1.5%, respectively. These enhancements notably improve the detection of distinct single-class defects such as punctures, splice paper anomalies, and normal cigarette-paper defects.

Overall, the SSD512 model, with its range of enhancements, offers superior detection accuracy for multiple defect types, such as bursting, foaming, splicing, masking, and crease defects, in addition to normal cigarette-paper and folding defects.

Table 7. Comparing mAP, Parameters, FLOPs, and FPS for SSD300 and SSD512 models.

Model	Ablation	mAP	Parameter(M)	FLOPs(GFLOPs)	FPS
	Original Model	81.90%	24.68	30.84	86.50
	VB	83.10%	30.45 ↑	28.75 ↓	88.00
	BIoU	83.00%	24.68	30.84	86.90
	LReLU	82.50%	24.68	30.84	89.20
SSD300	ELU	82.90%	24.68	30.84	82.30 ↓
	VB+BIoU	81.50% ↓	30.45	28.75 ↓	91.70
	BIoU+LReLU	83.50%	24.68	30.84	91.70
	VB+LReLU	82.90%	30.45 ↑	28.75 ↓	95.10
	VB+LReLU+BIoU	80.70% ↓	30.45 ↑	28.75 ↓	89.60
	Original Model	84.00%	24.68	88.71	81.70
	VB	85.10%	31.19 ↑	82.79 ↓	83.40
	BIoU	83.20% ↓	25.42 ↑	88.70	84.40
	LReLU	84.10%	25.42 ↑	88.70	83.10
SSD512	ELU	84.60%	25.42 ↑	88.70	83.30
	VB+BIoU	83.70% ↓	31.19 ↑	82.79 ↓	97.80
	BIoU+LReLU	84.50%	25.42 ↑	88.70	89.10
	VB+LReLU	84.20%	31.19 ↑	82.79 ↓	79.10 ↓
	VB+LReLU+BIoU	85.50%	31.19 ↑	82.79 ↓	83.70

5.3.2. Performance

Table 7 shows how various enhancements affect the SSD300 and SSD512 models. These results are summarized below:

1) **SSD300 model**

- Variational Bayesian inference combined with the BIoU loss function and the leaky ReLU

activation function: Reduced computational load and increased inference speed but no notable improvement in overall accuracy.
- Variational Bayesian inference alone: + 1.2% accuracy, −2.09 GFLOPs computational load, improved inference speed, and more parameters.
- BIoU loss function modification: + 1% accuracy, unchanged parameters and computational volume, decreased inference speed.
- Combination of variational Bayesian inference and leaky ReLU: Increased inference speed.
- BIoU loss function with leaky ReLU activation: + 1.6% accuracy, reduced computational load, slightly increased inference speed, and more parameters.

2) **SSD512 model**

- BIoU loss function and variational Bayesian inference: Slightly decreased overall accuracy, faster inference, reduced computational FLOPs, and more parameters.
- Variational Bayesian inference alone: + 1.1% accuracy, −5.92 GFLOPs computational FLOPs, slight improvement in inference speed.
- ELU and leaky ReLU activation functions: + 0.6% and + 0.1% in accuracy, unchanged computational FLOPs, and marginal increase in inference speed.
- Leaky ReLU activation with BIoU loss function: + 0.5% accuracy, minimal changes in parameters, the same computational volume, faster inference.
- All three enhancements (variational Bayesian inference, leaky ReLU, and BIoU loss function): + 1.5% accuracy, −5.92 GFLOPs computational load, + 6.51 million parameters, and slight improvement in inference speed.

In summary, the ablation study summarized in Table 7 highlights the varying impacts of different enhancements of the SSD300 and SSD512 models. Although some modifications to the SSD300 model do not significantly increase the overall accuracy, they reduce computational demands and improve inference speeds. Conversely, the modifications of the SSD512 model, particularly those involving the BIoU loss function, variational Bayesian inference, and activation functions, not only reduce computational load but also improve accuracy and inference speed. These results demonstrate the potential of these enhancements in balancing performance improvements with computational efficiency, indicating promising directions for future optimization in defect detection applications.

5.4. Practical detection of defects in cigarette appearance

Figure 5 shows the effectiveness of the proposed method for detecting various defects in cigarette appearance, revealing high overall accuracy, as indicated by the values in parentheses. However, certain defect types, such as foaming defects (FJM01) in Figure 5(c), are detected with notably lower accuracy (as low as 0.45 and 0.40). This decrease in accuracy is tentatively attributed to the visual similarities between foaming and puncture defects, making differentiation challenging. Furthermore, the limited presence of foaming defects in the training set may hinder the model's ability to effectively learn their distinct characteristics.

Similarly, fold defects (FZZ01), shown in Figure 5(e),(f),(i), consistently register low detection accuracy, ranging from 0.33 to 0.43. This suggests that the model's difficulty in detecting fold defects is not solely a consequence of sample size but might be relate to the inherent challenge of discerning
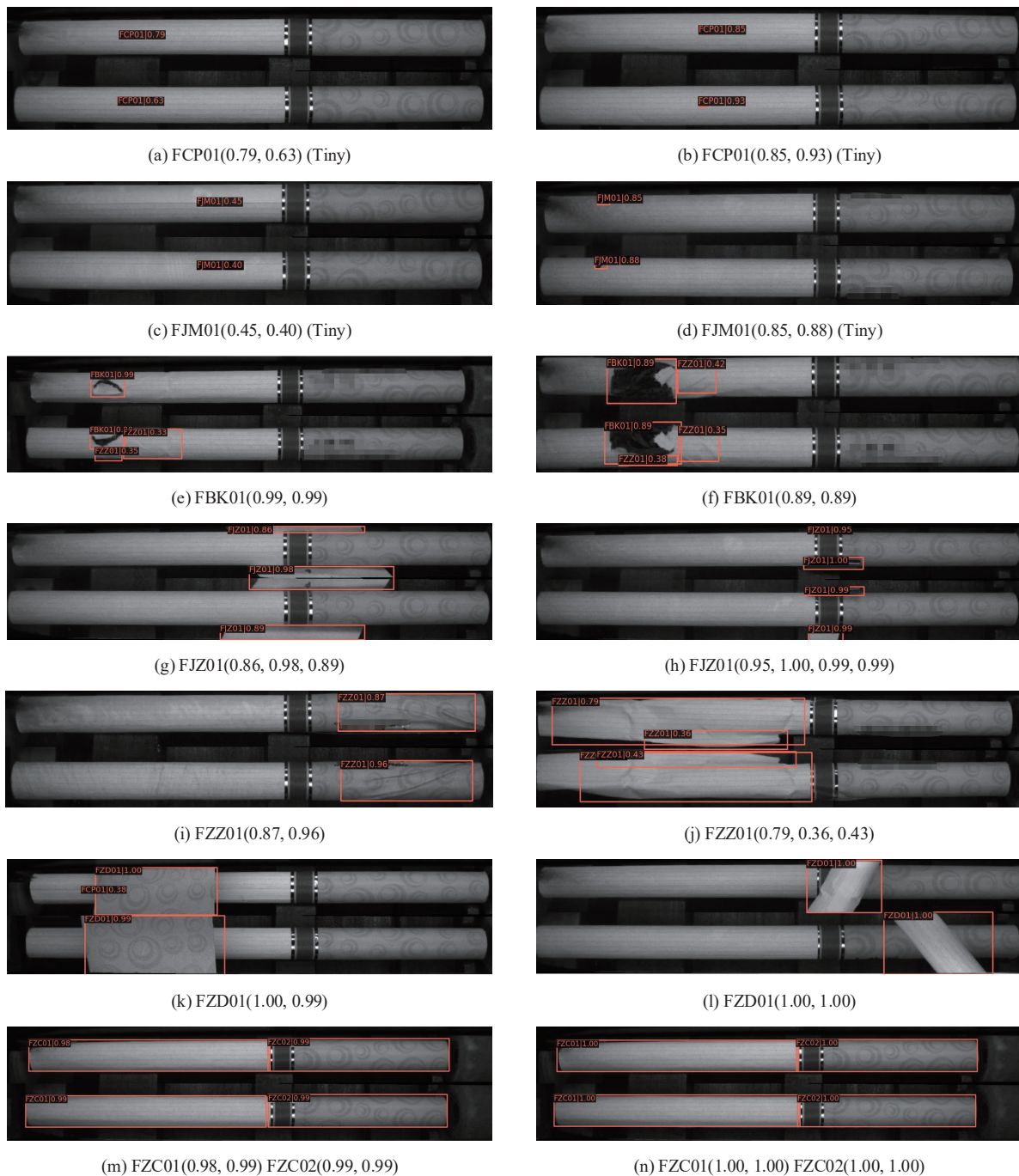
(a) FCP01(0.79, 0.63) (Tiny)

(b) FCP01(0.85, 0.93) (Tiny)

(c) FJM01(0.45, 0.40) (Tiny)

(d) FJM01(0.85, 0.88) (Tiny)

(e) FBK01(0.99, 0.99)

(f) FBK01(0.89, 0.89)

(g) FJZ01(0.86, 0.98, 0.89)

(h) FJZ01(0.95, 1.00, 0.99, 0.99)

(i) FZZ01(0.87, 0.96)

(j) FZZ01(0.79, 0.36, 0.43)

(k) FZD01(1.00, 0.99)

(l) FZD01(1.00, 1.00)

(m) FZC01(0.98, 0.99) FZC02(0.99, 0.99)

(n) FZC01(1.00, 1.00) FZC02(1.00, 1.00)

Figure 5. Detection results for various of defects in cigarette appearance defects using our the proposed enhanced system. It effectively, which identifies the eight defect types discussed in the paper and shows marked improvement in detecting offers noticeably improved detection of small targets.

subtle crease features that vary in spatial distribution and degree of deformation, similar to the issue noted during detecting tiny targets. Both the experimental results and practical detection examples underscore the low detection accuracy of fold defects (FZZ01). A detailed statistical analysis revealed that fold defects constitute merely 10% of the total sample pool, which could contribute to the reduced accuracy observed in this category. Further analysis of images depicting cigarettes with crease defects reveals considerable variations in the extent of creasing. Moreover, these defects exhibit notable differences in terms of physical morphology and spatial distribution, posing additional challenges to the effective detection of crease defects.

In conclusion, although the proposed model is good at detecting a broad range of defect types, it faces certain limitations in accurately identifying defects that are either highly similar in appearance or characterized by subtle, complex features.

6. Conclusions

This study develops an enhanced SSD model that integrates variational Bayesian inference to accurately detect visual defects in cigarettes during manufacturing. Despite the limitations posed by a small dataset of 2128 images, the model detects defects with significant accuracy, particularly in its SSD300 and SSD512 configurations. The incorporation of Bayesian inference is instrumental in achieving this heightened accuracy. Future work will focus on expanding the dataset and further refining the model's capabilities. Such advancements should substantially improve the accuracy of automated visual inspection systems, thereby contributing to more stringent quality-control measures in manufacturing.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare there is no conflict of interest.

References

1. *China Tobacco Machinery*, Changde Tobacco Machinery Limited Liability Company, 2022. Available from: https://www.ccdtm.com/info/info/_intr.jsp.

2. *G.D 121P-20K Cigarette Manufacturing Machines*, Gulf Tobacco, 2023. Available from: http://www.cigarettemanufacturingmachines.com/gd-121p-20k-cigarette-manufacturing-machine/.

3. *121P Double rod cigarette maker*, Gulf Tobacco, 2023. Available from: https://www.gidi.it/en/solutions/product/121p.

4. Z. Y. Xiao, Research and implementation of cigarette defect detection algorithm, *Yunnan Univ.*, 2018.

5. Y. X. Yang, Design and implementation of an image processing based cigarette defect detection method, *Yunnan Univ.*, 2018.

6. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788. https://doi.org/10.1109/CVPR.2016.91

7. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, et al., Ssd: Single shot multibox detector, in *Proceedings of the European Conference on Computer Vision(ECCV)*, **9905** (2016), 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

8. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, *Adv. Neural Inform. Process. Syst.*, **30** (2017). https://doi.org/10.48550/arXiv.1706.03762

9. K. Shridhar, F. Laumann, M. Liwicki, Uncertainty estimations by softplus normalization in bayesian convolutional neural networks with variational inference, preprint, arXiv: 1806.05978.

10. R. Qu, G. Yuan, J. Liu, H. Zhou, Detection of cigarette appearance defects based on improved SSD model, in *Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering*, (2021), 1148–1153. https://doi.org/10.1145/3501409.3501612

11. Z. W. Du, H. Zhou, C. Y. Li, Small object detection based on deep convolutional neural networks: A review, *Comput. Sci.*, **49** (2022), 205–208. https://doi.org/10.11896/jsjkx.220500260

12. H. J. Yang, L. Meng, An improved algorithm for small target detection in aerial photography images based on YOLOv5, *Comput. Eng. Sci.*, **45** (2023), 1063–1070.

13. L. Li, M. Li, H. Hu, An algorithm for cigarette capsules defect detection based on lightweight faster rcnn, in *2021 40th Chinese Control Conference (CCC)*, (2021), 8028–8034. https://doi.org/10.23919/CCC52363.2021.9550392

14. E. Kim, J. Lee, H. Jo, K. Na, E. Moon, G. Gweon, et al., SHOMY: Detection of small hazardous objects using the you only look once algorithm, *KSII Trans. Int. Inform. Syst. (TIIS)*, **16** (2022), 2688–2703. https://doi.org/10.3837/tiis.2022.08.012

15. J. Diers, C. Pigorsch, A survey of methods for automated quality control based on images, *Int. J. Comput. Vis.*, **131** (2023), 2348–2356. https://doi.org/10.1007/s11263-023-01822-w

16. H. Q. Wang, Deep learning-based target detection of cigarette defects, *Shenyang Univ. Chem. Techn.*, 2022. https://doi.org/10.27905/d.cnki.gsghy.2022.000092

17. G. W. Yuan, J. C. Liu, R. Qu, H. Zhou, Classification of cigarette appearance defects based on ResNeSt, *J. Yunnan Univ. Natural Sci. Edition*, **44** (2022), 464–470. https://doi.org/10.7540/j.ynu.20210257

18. H. Y. Liu, G. W. Yuan, L. Yang, K. Liu, H. Zhou, An appearance defect detection method for cigarettes based on C-CenterNet, *Electronics*, **11** (2022), 2182. https://doi.org/10.3390/electronics11142182

19. Y. L. Li, S. Yang, L. F. Fan, Y. H. Xiong, Q. Zhu, L. H. Zhang, Online inspection of cigarette seam defects based on machine vision, *Tobacco Sci. Technol.*, **56** (2023), 93–98. https://doi.org/10.16135/j.issn1002-0861.2022.0474

20. G. W. Yuan, J. C. Liu, H. Y. Liu, Y. Ma, H. Wu, H. Zhou, Detection of cigarette appearance defects based on improved YOLOv4, *Electr. Res. Arch.*, **31** (2023), 1344–1364. https://doi.org/10.3934/era.2023069.

21. Y. H. Ma, G. W. Yuan, K. Yue, H. Zhou, CJS-YOLOv5n: A high-performance detection model for cigarette appearance defects, *Math. Biosci. Eng.*, **20** (2023), 17886–17904. https://doi.org/10.3934/mbe.2023795

22. H. Y. Liu, G. W. Yuan, Detection of cigarette appearance defects based on improved YOLOv5s, *Comput. Technol. Dev.*, **32** (2022), 161–167.

23. D. Feng, Z. G. Li, A. M. He, X. Yang, S. Wang, H. Dong, et al., Appearance quality inspection of cigarette products based on local characteristic similarity metric, *Tobacco Sci. Technol.*, **56** (2023), 82–90. https://doi.org/10.16135/j.issn1002-0861.2022.0807

24. R. Qu, Research on cigarette appearance defect detection based on improved SSD, *Yunnan Univ.*, 2022. https://doi.org/10.27456/d.cnki.gyndu.2022.001224

25. J. C. Liu, Deep learning-based cigarette appearance defect detection and classification, *Yunnan Univ.*, 2022. https://doi.org/10.27456/d.cnki.gyndu.2022.002164

26. Y. Peng, Cigarette appearance quality detection based on improved YOLO deep learning model, *Yunnan Univ. Fin. Econom.*, 2022. https://doi.org/10.27455/d.cnki.gycmc.2023.000862

27. Y. Peng, D. Jiang, X. Z. Lv, Y. Liu, Efficient and high-performance cigarette appearance detection based on YOLOv5, in *2023 International Conference on Intelligent Perception and Computer Vision (CIPCV)*, (2023), 7–12. https://doi.org/10.1109/CIPCV58883.2023.00010

28. X. M. Li, G. Q. Xie, Z. Huang, C. Yu, Cigarette appearance detection system based on cascaded convolution network, *J. Comput. Appl.*, **43** (2023), 346–350. https://doi.org/10.11772/j.issn.1001-9081.2022030364

29. M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The PASCAL visual object classes (VOC) challenge, *Int. J. Comput. Vis.*, **88** (2010),303–338. https://doi.org/10.1007/s11263-009-0275-4

30. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., Microsoft coco: Common objects in context, in *Computer Vision–ECCV 2014: 13th European Conference*, **13** (2014), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

31. J. Deng, W. Dong, R. Socher, L. Li, K. Li, F. Li, ImageNet: A large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (2009), 248–255. https://doi.org/10.1109/CVPR.2009.5206848

32. L. Zhang, B. W. Zhou, L. H. Wu, SSD network based on improved convolutional attention module and residual structure, *Comput. Sci.*, **49** (2022), 211–217. http://qikan.cqvip.com/Qikan/Article/Detail?id=7106717136

33. J. Leng, Y. Liu, An enhanced SSD with feature fusion and visual reasoning for object detection, *Neural Comput. Appl.*, **31** (2019), 6549–6558. https://doi.org/10.1007/s00521-018-3486-1

34. A. Graves, Practical variational inference for neural networks, *Adv. Neural Inform. Process. Syst.*, **24** (2011), 2348–2356. https://dl.acm.org/doi/10.5555/2986459.2986721

35. C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight uncertainty in neural network, in *Proceedings of the 32nd International Conference on Machine Learning*, **37** (2015), 1613–1622. https://dl.acm.org/doi/10.5555/3045118.3045290

36. N. D. Nguyen, T. Do, T. D. Ngo, D. D. Le, An evaluation of deep learning methods for small object detection, *J. Electr. Comput. Eng.*, (2020), 2348–2356. https://doi.org/10.1155/2020/3189691

37. Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, S. Hu, Traffic-sign detection and classification in the wild, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 2110–2118. https://doi.org/10.1109/CVPR.2016.232

38. A. Torralba, R. Fergus, W. T. Freeman, 80 million tiny images: A large data set for nonparametric object and scene recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, **30** (2008), 1958–1970. https://doi.org/10.1109/TPAMI.2008.128

39. C. Chen, M. Y. Liu, O. Tuzel, et al., R-CNN for small object detection, in *Asian Conference on Computer Vision(ACCV)*, **10115** (2014), 214–230. https://doi.org/10.1007/978-3-319-54193-8_14

40. S. Kullback, R. A. Leibler, On information and sufficiency, *Annals Math. Stat.*, **22** (1951), 79–86.

41. D. P. Kingma, T. Salimans, M. Welling, Variational dropout and the local reparameterization trick, *Adv. Neural Inform. Process. Syst.*, **2** (2015), 2575–2583. https://dl.acm.org/doi/abs/10.5555/2969442.2969527

42. L. Tychsen-Smith, L. Petersson, Improving object localization with fitness NMS and bounded IoU loss, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2023), 6877–6885. https://doi.org/10.1109/CVPR.2018.00719

43. P.J. Huber, Robust estimation of a location parameter, *Breakthr. Stat.*, (1992), 492–518. https://doi.org/10.1007/978-1-4612-4380-9_35

44. A. Krizhevsky, G. Hinton, Learning multiple layers of features from tiny images, *Handbook Syst. Aut. Dis.*, **1** (2009).

45. K. He, G. Gkioxari, N. Parmar, P. Dollar, R. Girshick, Mask r-cnn, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, (2017), 2961–2969. https://arxiv.org/abs/1703.06870

46. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in *European conference on computer vision (ECCV)*, (2020), 213–229. https://doi.org/10.1007/978-3-030-58452-8_13