



# An Efficient Approach to Human Security Screening Image Recognition Through a Lightweight CNN Utilizing Yolov5s and GhostNet



Qunli Zhao<sup>\*</sup>, Hesheng Cheng<sup>†</sup>

School of Computer and Artificial Intelligence, Hefei Normal University, Hefei 230061, China

Corresponding Author Email: [zql@hfnu.edu.cn](mailto:zql@hfnu.edu.cn)

<https://doi.org/10.18280/ts.400433>

## ABSTRACT

**Received:** 3 April 2023

**Revised:** 19 June 2023

**Accepted:** 28 June 2023

**Available online:** 31 August 2023

### Keywords:

*lightweight CNN model, human security screening, image recognition, Yolov5s*

Human security screening constitutes a vital component in public safety assurance across varied environments like airports, governmental edifices, and additional public spaces. Among the paramount challenges inherent in human security screening lies the immediate and precise discernment of prospective threats within X-ray images. Despite the potential exhibited by convolutional neural networks (CNNs) in image recognition tasks, including the detection of targets in X-ray imagery, the substantial computational burden and memory prerequisites often render real-time deployment impracticable on devices with limited resources. In the present study, a novel lightweight CNN approach, melding Yolov5s and GhostNet models with the coordinate attention mechanism, is introduced to alleviate the constraints found in existing techniques. By employing this combination, efficiency in computation and model accuracy has been augmented, thereby addressing the challenges of swift and accurate threat identification. Performance evaluation, conducted on a publicly accessible dataset comprising X-ray images pertinent to human security screening, demonstrated the superior detection accuracy and reduced storage footprint of the proposed model in comparison to prevailing alternatives. Overall, the approach delineated herein presents an efficacious and streamlined solution for real-time human security screening image recognition on resource-constrained devices, contributing a promising advancement in the field.

## 1. INTRODUCTION

In contemporary society, the evolution of security threats is marked by increasing complexity and sophistication. New methods of concealing weapons, explosives, and illicit goods are continually being developed. The limitations of traditional manual inspection methods in detecting these concealed threats can result in significant security vulnerabilities. Although certain enterprises have introduced human security screening systems utilizing X-ray imaging technology, these systems are not without drawbacks. Displaying the scanned human body's image, where carried items are made apparent, the need for manual checking is negated. Thus, human security screening systems have become essential in domains such as transportation, public events, and border control, striving to detect potential threats and bolster public safety. However, current methods are often reliant on human operators, making them potentially subjective, prone to error, and time-consuming.

The utilization of automated image recognition systems based on deep learning techniques has been explored, demonstrating significant potential for accurately and efficiently identifying security threats within X-ray images. Nevertheless, a complex background, comprising crowded scenes or cluttered environments, often present in human security screening images, raises the challenge of distinguishing between pertinent objects and irrelevant background elements. Potential threats or prohibited items may be obscured within the surrounding context, thus

underscoring the necessity of developing novel approaches for accurate segmentation and extraction.

While research related to X-ray image recognition in baggage screening abounds, direct application to human security systems is hindered by the distinctive mono-energy detector imaging methods used in the latter, resulting in grey-scale rather than color images. Several deep learning models, such as CNNs, have been applied to baggage screening [1-6], capitalizing on their capacity to learn hierarchical representations. Although their ability to discern intricate features and patterns has been leveraged to address various security threats, the high computational complexity and memory requirements limit their applicability in resource-constrained environments. The need for lightweight CNN models that balance accuracy with reduced computational and memory demands becomes evident.

In the present investigation, a lightweight CNN model-based approach for human security screening image recognition is introduced. By employing transfer learning and model compression techniques, the computational complexity and memory prerequisites of the CNN model are minimized without a commensurate reduction in performance. A novel dataset encompassing diverse human security screening objects and materials is employed to evaluate the approach, and a comparison is made with existing methods. The salient contributions of this study encompass:

(1) The introduction of a lightweight CNN model-based approach for human security screening image recognition, achieving unparalleled performance while significantly

curtailing computational complexity and memory demands.

(2) Evaluation utilizing a new dataset, indicative of various human security screening objects and materials, revealing the approach's applicability in realistic scenarios.

(3) The implementation of a coordinate attention mechanism within the backbone network layer of the proposed lightweight CNN model.

(4) Insight into the trade-offs between model size and performance within human security screening image recognition.

The remainder of this article is organized to provide a comprehensive examination of the topic. Section 2 offers a review of the pertinent work on security screening systems and existing image recognition approaches. Section 3 delineates the proposed approach, including an in-depth discussion of the model architecture. Section 4 presents a thorough analysis of the experimental results, encompassing dataset description, baseline comparisons, and performance evaluation.

## 2. RELATED WORK

Research results specifically designed for human security screening applications are, at the time of writing, scarce. Nevertheless, the application area of human security screening can be broadly understood as a specific domain of image recognition, and as such, the existing research in the broader field of image recognition offers insights. This section elaborates on these existing approaches, subdivided into three main areas: traditional feature-based methods, deep learning-based methods, and lightweight CNN models.

### 2.1 Traditional feature-based methods

Early methods of image recognition within the context of human security screening were often reliant on handcrafted features combined with traditional machine learning algorithms. Discriminative features were extracted from security screening images using techniques such as Histogram of Oriented Gradients (HOG) [7], Scale-Invariant Feature Transform (SIFT) [8], and Local Binary Patterns (LBP) [9]. Subsequently, these features were coupled with classifiers like Support Vector Machines (SVM) [10] or Random Forests [11] for the detection and classification of objects.

While moderate success was attained by these approaches, significant challenges were encountered. Complex variations in image appearance, cluttered backgrounds, and the need for high-speed real-time processing were among these difficulties. Moreover, traditional feature-based approaches were found to lack end-to-end learning capability, resulting in sub-optimal solutions. The separated optimization of different stages, such as feature extraction, selection, and classification, often limited overall performance.

### 2.2 Deep learning-based methods

The introduction of deep learning, particularly CNNs, ushered in a new era in image recognition, applicable also to human security screening. Through the utilization of CNNs, remarkable performance enhancements were realized in various computer vision tasks, including but not limited to object detection and classification [12-16]. The inherent capability of CNNs to automatically learn hierarchical representations from raw image data eradicated the necessity

for handcrafted features.

Despite these successes, it must be noted that traditional CNN models were often found to be computationally taxing. This limitation constrained their real-time deployment, especially in environments with limited computational resources.

## 2.3 Lightweight CNN models

In response to the identified need for computational efficiency within human security screening applications, attention was directed towards the development of lightweight CNN models. These models, epitomized by Yolov5s [17], GhostNet [18-20], MobileNet series [21-23], ShuffleNet series [24, 25], EfficientNet series [26, 27], and SqueezeNet series [28, 29], were designed to balance recognition performance with reduced model complexity and computational demands.

A particular emphasis was placed on real-time object detection capabilities in models such as Yolov5s, while GhostNet focused on efficiency and model size reduction. Innovations were introduced across various models, such as depthwise separable convolutions in MobileNetV1, inverted residual blocks in MobileNetV2, and novel architecture search strategies in MobileNetV3. The ShuffleNet series reduced computational complexity through specific techniques, including pointwise group convolutions and channel shuffling. Meanwhile, the EfficientNet series offered a scalable range of models tailored to specific requirements and constraints.

A balance between model accuracy and computational efficiency has been sought in recent works, exemplified by optimizations to the Yolov5s model, resulting in a lightweight improved CNN model based on GhostNet and the coordinate attention mechanism [30]. Such efforts have contributed to enhancing both the efficiency and accuracy of image recognition, thereby reducing dependence on hardware capabilities.

The landscape of human security screening leverages a rich tapestry of methods drawn from the broader field of image recognition. From traditional feature-based techniques to cutting-edge lightweight CNN models, the evolving nature of this field continues to present opportunities and challenges. The ongoing search for a synergistic balance between accuracy, efficiency, and computational demands serves as a focal point for current and future research, reflecting the complex and dynamic nature of human security screening applications.

## 3. THE PROPOSED METHOD

### 3.1 Yolov5s and GhostNet

Two salient architectures central to the field of object detection and classification, namely Yolov5s and GhostNet, are to be discussed herein.

#### 3.1.1 Yolov5s

The Yolov5s, a subsequent variant of the You Only Look Once (YOLO) series, is distinguished for its real-time object detection competencies. Within this architecture, an equilibrium is struck between speed and accuracy, with a streamlined single-stage approach employed to predict bounding boxes and class probabilities in a single pass. A combination of backbone networks, inclusive of

CSPDarknet53 [31], is utilized to extract features at varying scales, a process that culminates in precise object localization and detection.

Four primary components encompass the architecture of Yolov5s: the Input, Backbone, Neck network, and Prediction (output) segments. Within the Backbone, the Focus, SPP, and CSP structures are incorporated.

### 3.1.2 GhostNet

The GhostNet model represents a lightweight CNN architecture, tailored to optimize both efficiency and size, without compromising accuracy. In this innovative design, the Ghost block is introduced, a construct that exploits branch replication alongside sparse connections to minimize computational intricacy.

By capitalizing on fewer parameters, efficient feature extraction is attained within GhostNet, an accomplishment evidenced by its commendable performance across various domains, such as image classification, object detection, and face recognition. The parsimonious nature of GhostNet's design renders it an apt solution for deployment on devices constrained by resources, thereby enabling adept inference in scenarios demanding edge computing.

The Yolov5s and GhostNet models exemplify modern achievements in object detection and classification, respectively. Yolov5s leverages a single-stage approach to balance efficiency and accuracy, whereas GhostNet focuses on reducing computational complexity through novel design features. These architectures lay the foundation for an array of applications, particularly where real-time processing and resource efficiency are paramount. The insights provided by these models may pave the way for future innovations in object detection and recognition, reinforcing the essential nature of continuous research and development within this dynamic field.

## 3.2 The lightweight model

### 3.2.1 Modification of Yolov5s structure

In this section, an alteration to the structure of Yolov5s is presented to lessen the quantity of parameters within the model and augment the speed of inference, while the accuracy of recognition is preserved. The specific modifications encompass the following:

- **Focus LM Block:**

A  $960 \times 960 \times 3$  image is first subjected to a down-sampling slicing operation, resulting in a  $480 \times 480 \times 12$  feature map. In a bid to mitigate the loss of image information during downsampling, the feature map is subsequently convolved with  $3 \times 3 \times 32$  convolution kernels, where a stride of 1 and padding of 2 are utilized, obtaining a  $480 \times 480 \times 32$  feature map. Maximum pooling is then employed to further truncate the feature parameters, ultimately leading to a  $240 \times 240 \times 32$  feature map. The output parameters are notably reduced in comparison to the original Focus block.

- **GBB (Ghost Bottleneck Block):**

The Ghost bottleneck block, specifically crafted for small CNNs, is explicated in literature [18]. This novel construction bears a resemblance to the Basic Residual Block found in ResNet, comprising an integration of multiple convolutional layers coupled with a shortcut connection. As depicted in Figure 1, the Ghost bottleneck block bifurcates into two distinct branches with strides of 1 and 2. Each branch is characterized by a pair of sequentially stacked Ghost blocks,

with their detailed structure elucidated in Figure 2.

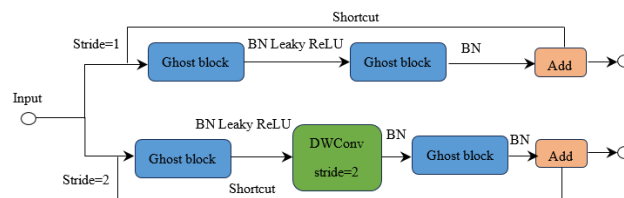


Figure 1. The Ghost bottleneck block

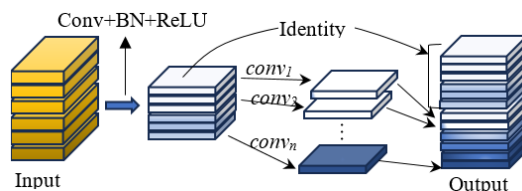


Figure 2. The Ghost block

Within these branches, the first Ghost block serves to augment the number of channels in the input feature map, thereby broadening the scope for subsequent computational processes. Conversely, the second Ghost block functions to curtail the number of channels in the output feature map, aligning with the network's diameter structure. This reduction is orchestrated to foster connectivity in information transmission between the two Ghost blocks through the diameter structure.

A nuanced examination of the accompanying figure reveals a subtle disparity between the two Ghost blocks. Specifically, the first Ghost block employs the ReLU activation function, and the subsequent layers leverage the batch normalization (BN) process. Such an architectural decision facilitates an optimization of the feature map within the Ghost block. The resulting structure not only enhances the detection efficiency of the model but also concurrently minimizes the number of model parameters and computational operations.

The innovative design of the Ghost bottleneck block accentuates a finely-balanced interplay between efficiency and complexity, fortifying the model's agility without undermining its functional integrity. By carefully calibrating the number of channels and integrating activation and normalization processes, the GBB offers a promising pathway toward optimized object detection in constrained computational environments.

- **CAB (Coordinate Attention Block):**

The CAB, akin to the SE (Squeeze-and-excitation) [32] block, encodes channel relations and long-range dependencies with acute positional information. The delineation of coordinate information embedding and coordinate attention generation is structured in Figure 3.

- (1) Coordinate Information Embedding

In the initial phase of the process, the portion where coordinate information is embedded is identified. Typically, channel attention employs global pooling to universally encode spatial details as channel descriptors, an approach that often proves challenging in preserving the specific positional information. In an innovative departure from conventional methods, global pooling is decomposed into a duo of one-dimensional feature encoding operations to enable the attention block to apprehend spatial long-range dependencies with pinpoint location accuracy.

To elucidate, for a given input  $X$ , each channel is individually encoded along the horizontal and vertical coordinate axes using pooling kernels with dimensions  $(H, 1)$  and  $(1, W)$ , respectively. This encoding results in the output of the  $c^{th}$  channel of height  $h$ , as expressed by:

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < w} x_c(h, i) \quad (1)$$

In a parallel fashion, the output of the  $c^{th}$  channel of width  $w$  is articulated as:

$$Z_c^h(w) = \frac{1}{H} \sum_{0 \leq j < h} x_c(w, j) \quad (2)$$

These dual transformations collectively perform feature aggregation across two spatial orientations, culminating in a pair of direction-aware attention maps. Contrasting sharply with the SE block's method of generating a singular feature vector, these bifurcated transformations empower the attention module to not only capture long-range dependencies along a single spatial trajectory but also conserve precise location information along the other. Such a nuanced approach facilitates the network's ability to more accurately pinpoint targets of interest. The operational essence of this coordinate information embedding aligns with the  $X Avg Pool$  and  $Y Avg Pool$  segments in Figure 3, marking a significant advancement in the domain of spatial awareness and target recognition.

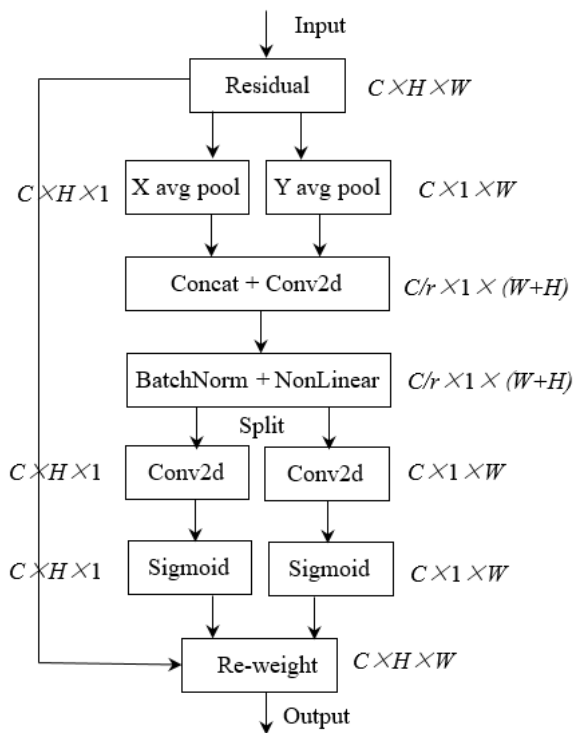


Figure 3. The coordinate attention block

(2) Coordinate Attention Generation

Subsequent to the coordinate information embedding, further enhancement of the spatial awareness and target recognition within the model is necessitated. This need gives rise to the design of a coordinate attention generation operation, which exploits the representation generated by the embedding

module. Characterized by global sensory fields and accurate location information, the newly generated attention graph is subject to three defining criteria. Firstly, the transformation is constrained to be as streamlined and efficient as possible, suiting applications within mobile environments. Secondly, it must be adept at leveraging the captured location data to precisely delineate the regions of interest. Lastly, an inherent capability to efficiently discern the relationships between channels is indispensable.

This aspect of coordinate attention generation is meticulously detailed in the remaining portion of the aforementioned diagram, exemplifying an elegant synthesis of simplicity, precision, and efficiency.

As corroborated by the insights presented in literature 29, an evaluation of three prevailing attentional mechanisms—SE, CBAM [33], and CAB—reveals a clear superiority of CAB in terms of both accuracy and computational velocity. Consequently, the CAB attentional mechanism is adopted, emblematic of a calculated decision to augment both the computational efficiency and accuracy of the model under discussion.

• Proposed Lightweight Model:

Upon the foundation of Yolov5s, a lightweight network model is crafted. The Backbone and Neck portion of this model witness the Focus block of Yolov5s supplanted by the Focus LM block, with the attention block CAB appended thereafter; the CSP block of Yolov5s is exchanged with the GBB block, followed by the convolution block Conv, and finally, the anterior and posterior feature maps are amalgamated by Concat, and the results are channeled to the output layer, as depicted in Figure 4.

The input and output parameters of each block in Figure 4 are listed in Table 1, from which it can be seen that the number of feature parameters of the input image is reduced a lot after a series of transformations are processed, which improves the computational speed of this lightweight model.

The modifications to the Yolov5s structure described herein underscore a strategic approach to enhancing computational efficiency without sacrificing recognition accuracy. Through innovative block designs and strategic implementation of existing methodologies, the lightweight model presents a promising avenue for object detection and recognition in constrained environments. Further exploration of these architectural changes may herald advancements in real-time processing capabilities, reaffirming the essential role of continuous innovation within the field.

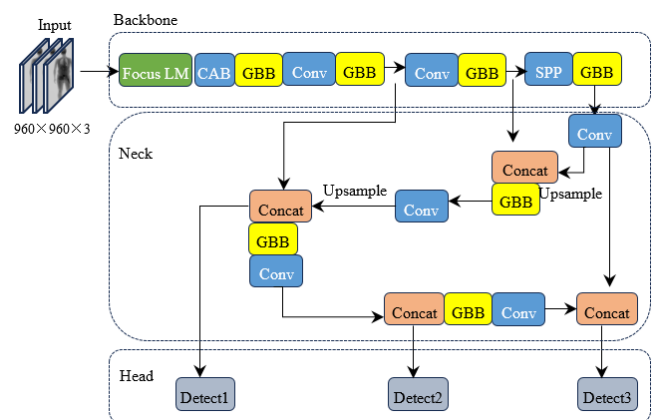


Figure 4. The lightweight CNN model

**Table 1.** The parameters of the proposed lightweight model

Input	Sub Module	Convolution	Stride	Output
960×960×3	Focus LM	32×3×3	1	240×240×32
240×240×32	CAB	64×3×3	1	120×120×64
120×120×64	GBB	64×3×3	1	120×120×64
120×120×64	Conv	64×3×3	2	60×60×64
60×60×64	GBB	128×3×3	1	30×30×128
30×30×128	Conv	128×3×3	1	30×30×128
30×30×128	GBB	256×3×3	1	30×30×256
30×30×256	SPP	256×1×1	1	30×30×256
30×30×256	GBB	128×3×3	1	30×30×128
30×30×128	Conv	256×1×1	2	15×15×256
15×15×256	Upsample			30×30×256
30×30×512	GBB	256×3×3	1	30×30×256
30×30×256	Conv	128×3×3	2	15×15×128
15×15×128	Upsample			30×30×128
30×30×256	GBB	128×3×3	1	30×30×128
30×30×128	Conv	256×3×3	1	30×30×256
30×30×256	GBB	512×3×3	2	15×15×512
15×15×512	Conv	512×1×1	1	15×15×512

## 4. EXPERIMENT

### 4.1 Datasets

In the absence of publicly available datasets in the field of human security screening, a dataset was assembled from images actually collected by human security screening machines over a period exceeding five years. This collection comprises a total of 10,000 images, 7176 of which contain prohibited items (e.g., guns, knives). The remaining images feature items such as belt buckles, mobile phones, and keys, but are devoid of prohibited objects. A partitioning of the dataset into training and test sets was conducted at a ratio of 70% and 30%, respectively. A representative example from this dataset is depicted in Figure 5.



**Figure 5.** Sample images of the dataset

### 4.2 Experiment environment

The experimental setting employed in this study is delineated as follows: an operating system of Windows 11, CPU of Intel Core i9-13900K, memory capacity of 64GB, a graphics card comprising NVIDIA GeForce RTX4090, a CUDA version of 11.6, and a PyTorch version of 1.12.1.

### 4.3 Evaluation metrics

Precision rate, recall rate, and average precision were selected as the model accuracy evaluation indices to assess the

experimental outcomes. These metrics are mathematically represented below:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$AP = \int_0^1 p(r)dr$$

$$mAP = \frac{\sum_{i=1}^c AP_i}{C}$$

where, TP (True Positive), FN (False Negative), FP (False Positive), and TN (True Negative) are defined accordingly. The formulation of average accuracy and mean Average Precision (mAP) is also articulated.

### 4.4 Experimental results

Within the experimental framework delineated in Section 4.2, the proposed lightweight model was implemented for human security screening. Figure 6 exhibits the detection outcomes. A comparison was conducted using the following methodologies to authenticate the overall performance of the proposed models:

(1) Training of the dataset with the unoptimized Yolov5s model followed by validation on test data.

(2) Integration of the Ghostnet block with Yolov5s and substitution of the CSP block in Yolov5s with Ghostnet.

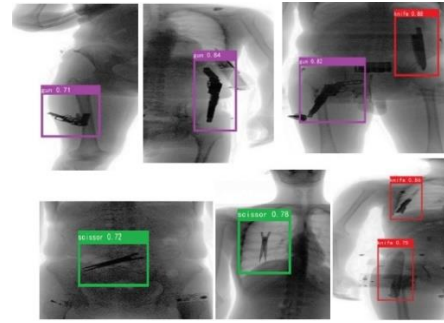
(3) Introduction of the CAB module based on (2).

(4) Implementation of the lightweight model proposed herein: Replacement of the Focus module in Yolov5s with Focus LM, addition of the CAB module, and substitution of the CSP module in Yolov5s with Ghost bottleneck.

Tables 2 and 3 demonstrate the results of the comparison, indicating competitive performance in mAP and AP50 values. Moreover, the proposed model exhibits significantly fewer parameters and lower computational complexity (FLOPs). These findings underscore the efficacy and efficiency of the

suggested lightweight CNN model for human security screening image recognition.

The detection outcomes of three types of prohibited items (namely, guns, knives, and scissors) carried by the human body are presented in Figure 6. Observations from this figure elucidate that, notwithstanding the placement of these prohibited items in various parts of the human body, accurate detection was achieved by the lightweight CNN model. The detection time for each image was found to be less than 40 ms, affirming that the lightweight model satisfies the real-time detection requirements for human security screening image recognition.



**Figure 6.** Image recognition results of the proposed model

**Table 2.** Performance comparison of different models

Model	Parameters (M)	FLOPs (G)	mAP	AP <sub>50</sub>
Yolov5s	7.3	16.97	0.782	0.811
Yolov5s+GhostNet	5.6	10.39	0.791	0.815
Yolov5s+GhostNet+CAB	3.9	7.12	0.802	0.824
Proposed Model	2.8	3.13	0.832	0.847

**Table 3.** Comparison of the AP value of different models

Items	Yolov5s	Yolov5s+GhostNet	Yolov5s+GhostNet+CAB	Proposed Model
gun	0.802	0.807	0.818	0.833
knife	0.798	0.808	0.82	0.845
scissor	0.722	0.716	0.724	0.764

## 5. CONCLUSION

In the present study, a lightweight CNN model uniquely tailored for human security image detection was introduced. This model's architecture was enriched by integrating the CAB attention mechanism module with Yolov5s, substituting the CSP module in Yolov5s with the Ghost bottleneck module, and modifying the Focus module.

Several advancements were realized through these modifications:

(1) Reduction in Parameter Complexity: By refining the Focus module, the number of parameters in the input data was notably reduced. This alteration led to an overall simplification in model architecture without compromising performance.

(2) Enhanced Accuracy: The introduction of the CAB module played a pivotal role in enhancing the precision of human security image detection. By efficiently leveraging spatial information and channel relationships, a more nuanced understanding of the detection targets was facilitated.

(3) Optimization of Computational Process: The employment of the Ghost bottleneck module served to streamline the calculation process within the model. By doing so, both the quantity of parameters required in the computational process and the overall calculation speed were significantly improved.

Through rigorous experimental comparisons, it was found that the augmented model maintained accuracy while demonstrating marked reductions in both parameters and computational quantities. This dual achievement of preserving quality while enhancing computational efficiency positions the proposed model as a potent solution that fulfills the stringent demands of human security image detection.

The findings of this research shed light on innovative methodologies for minimizing computational complexity without sacrificing detection accuracy. The incorporation of

the CAB attention mechanism and Ghost bottleneck module, along with modifications to the Focus module, underpins a powerful, lightweight architecture that can be adapted to other related fields beyond human security screening.

The present study marks a significant contribution to the existing body of knowledge in the realm of security image detection. It provides a robust foundation for future work, particularly in exploring further architectural enhancements, the application of these principles to other detection challenges, and in the real-world implementation in security screening scenarios.

In conclusion, the proposed lightweight CNN model, characterized by the judicious integration of several novel components, represents a promising avenue for advancing the efficiency and effectiveness of human security image detection. The insights gained from this research may pave the way for broader applications and continued innovation in the field.

## ACKNOWLEDGMENT

This study was supported by the Natural Science Foundation of Anhui Province in China (Grant No.: KJ2020A0113) and the Anhui Province Key Laboratory of Simulation and Design for Electronic Information System (Grant No.: 2020ZDSYSYB08).

## REFERENCES

- [1] Nie, X. (2019). The impact of conditional dependence on checked baggage screening. *European Journal of Operational Research*, 278(3): 883-893. <https://doi.org/10.1016/j.ejor.2019.04.034>

- [2] Seth, S., Feng, Q. (2022). A multi-level weighted alarm security system for passenger and checked-baggage screening. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236(5): 727-737. <https://doi.org/10.1177/1748006X211042710>
- [3] Subramani, M., Rajaduari, K., Choudhury, S.D., Topkar, A., Ponnusamy, V. (2020). Evaluating one stage detector architecture of convolutional neural network for threat object detection using X-ray baggage security imaging. *Revue d'Intelligence Artificielle*, 34(4): 495-500. <https://doi.org/10.18280/ria.340415>
- [4] Skorupski, J., Uchroński, P. (2020). Multi-criteria group decision-making approach to the modernization of hold baggage security screening system at an airport. *Journal of Air Transport Management*, 87: 101841. <https://doi.org/10.1016/j.jairtraman.2020.101841>
- [5] Mery, D., Saavedra, D., Prasad, M. (2020). X-ray baggage inspection with computer vision: A survey. *IEEE Access*, 8: 145620-145633. <https://doi.org/10.1109/ACCESS.2020.3015014>
- [6] Ponnusamy, V., Marur, D.R., Dhanaskodi, D., Palaniappan, T. (2021). Deep learning-based X-Ray baggage hazardous object detection – An FPGA implementation. *Revue d'Intelligence Artificielle*, 35(5): 431-435. <https://doi.org/10.18280/ria.350510>
- [7] Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) San Diego, CA, USA*, 1: 886-893. <https://doi.org/10.1109/CVPR.2005.177>
- [8] Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60: 91-110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [9] Heikkilä, M., Pietikäinen, M., Schmid, C. (2009). Description of interest regions with local binary patterns. *Pattern Recognition*, 42(3): 425-436. <https://doi.org/10.1016/j.patcog.2008.08.014>
- [10] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408: 189-215. <https://doi.org/10.1016/j.neucom.2019.10.118>
- [11] Speiser, J.L., Miller, M.E., Tooze, J., Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134: 93-101. <https://doi.org/10.1016/j.eswa.2019.05.028>
- [12] Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 39(6): 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. (2016). Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14*, Springer International Publishing, pp. 21-37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [14] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [15] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- [16] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [17] Jocher, G., Changyu, L., Hogan, A., Yu, L., Rai, P., Sullivan, T. (2020). *ultralytics/yolov5: Initial Release*. Zenodo. <https://doi.org/10.5281/zenodo.3908560>
- [18] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C. (2020). Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1580-1589.
- [19] Paoletti, M.E., Haut, J.M., Pereira, N.S., Plaza, J., Plaza, A. (2021). Ghostnet for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(12): 10378-10393. <https://doi.org/10.1109/TGRS.2021.3050257>
- [20] Yuan, X., Li, D., Sun, P., Wang, G., Ma, Y. (2022). Real-time counting and height measurement of nursery seedlings based on GHOSTNET–YoloV4 network and binocular vision technology. *Forests*, 13(9): 1459. <https://doi.org/10.3390/f13091459>
- [21] Al-Nabulsi, J., Turab, N., Owida, H.A. (2023). Enhanced facial recognition techniques for masked individuals amid the COVID-19 pandemic. *Mathematical Modelling of Engineering Problems*, 10(4): 1288-1296. <https://doi.org/10.18280/mmep.100422>
- [22] Thao, L.Q., Cuong, D.D., Nhi, N.N., Tam, N.D. (2022). A deep learning powered system to lie detection while online study. *Traitement du Signal*, 39(3): 893-898. <https://doi.org/10.18280/ts.390314>
- [23] Okokpujie, K., Okokpujie, I.P., Ayomikun, O.I., Orimogunje, A.M., Ogundipe, A.T. (2023). Development of a web and mobile applications-based cassava disease classification interface using Convolutional Neural Network. *Mathematical Modelling of Engineering Problems*, 10(1): 119-128. <https://doi.org/10.18280/mmep.100113>
- [24] Zhang, X., Zhou, X., Lin, M., Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848-6856.
- [25] Liu, C., Yang, J., Liu, Y., Zhang, Y., Liu, S., Chaikovska, T., Liu, C. (2023). A cervical lesion recognition method based on ShuffleNetV2-CA. *Information Dynamics and Applications*, 2(2): 77-89. <https://doi.org/10.56578/ida020203>
- [26] Tan, M., Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pp. 6105-6114.
- [27] Bodapati, J.D., Ahmed, S.F., Chowdary, Y.Y., Sekhar, K.R. (2023). A deep convolutional neural network framework for enhancing brain tumor diagnosis on MRI scans. *Information Dynamics and Applications*, 2(1): 42-50. <https://doi.org/10.56578/ida020105>
- [28] Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K.,

- Dally, W.J., Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. arXiv preprint arXiv:1602.07360. <https://doi.org/10.48550/arXiv.1602.07360>
- [29] Baskar, C., Govindasamy, G.P., Anbalagan, S., Roomi, S.M.M. (2022). Computer graphic and photographic image classification using transfer learning approach. *Traitement du Signal*, 39(4): 1267-1273. <https://doi.org/10.18280/ts.390419>
- [30] Hou, Q., Zhou, D., Feng, J. (2021). Coordinate attention for efficient mobile network design. *Computer Vision and Pattern Recognition*. <https://doi.org/10.48550/arXiv.2103.02907>
- [31] Wang, C.Y., Liao, H.Y. M., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 390-391.
- [32] Hu, J., Shen, L., Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132-7141. <https://doi.org/10.1109/CVPR.2018.00745>
- [33] Woo, S., Park, J., Lee, J.Y., Kweon, I.S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3-19.