



AUTOMATED IMAGE RECOGNITION SYSTEMS USING INSPECTION AND SCREENING COMPLEXES IN CUSTOMS CONTROL PROCESSES

Igor Aleshin^a, Elena Antonova^b

^aVladivostok Customs, Pervomaisky Customs Post, Inspection and Screening Complex Application Department (OIPIDC), Email: Aleshin.igor90@mail.ru (penulis berkorespondensi)

^bElena Antonova, Vladivostok Filial of the Russian Customs Academy, Email: ant_vladivostok@mail.ru

ARTICLE INFORMATION

ARTICLE HISTORY

Received

16 September 2024

Accepted to be published

31 December 2024

KEYWORDS:

Convolutional Neural Networks

Image Recognition

Customs Control

X-ray Image Analysis

Object Detection

Data Augmentation

ABSTRACT

Dalam operasi bea cukai modern, pemrosesan data dalam jumlah besar secara cepat dan akurat sangat penting, terutama saat menganalisis gambar yang diperoleh dari *Inspection and Scanning Complexes* (ISC). *Convolutional Neural Networks* (CNNs) menawarkan solusi yang menjanjikan, memberikan kemampuan analisis dan klasifikasi gambar yang lebih baik. Studi ini berfokus pada penerapan algoritma berbasis CNN untuk mendeteksi dan menandai kontur senjata api dalam gambar sinar-X dari ISC. Model CNN, yang dikembangkan menggunakan pustaka *TensorFlow/Keras*, terdiri dari 14 lapisan, termasuk lapisan konvolusi, *pooling*, dan terkoneksi secara utuh. Model ini dilatih pada *dataset* khusus yang terdiri dari 150 gambar sinar-X yang telah diberi anotasi, di mana teknik augmentasi data digunakan untuk meningkatkan ketahanan terhadap distorsi geometris dan kualitas gambar yang rendah. Proses pelatihan melibatkan 300 *epoch*, dan akurasi model dievaluasi menggunakan metrik seperti mAP dan matriks kebingungan. Hasilnya menunjukkan akurasi 80% pada data validasi dan akurasi 84% pada data pelatihan. Model ini secara efektif mengidentifikasi senjata api dalam berbagai gambar tetapi menunjukkan keterbatasan saat mendeteksi jenis senjata api lainnya karena spesifikasi *dataset* pelatihan. Penelitian ini menyoroti potensi CNN dalam meningkatkan kontrol bea cukai melalui pengenalan gambar otomatis, sekaligus menekankan pentingnya data pelatihan yang beragam untuk meningkatkan generalisasi di berbagai jenis objek.

In modern customs operations, the rapid and accurate processing of large volumes of data is crucial, particularly when analyzing images obtained from inspection and scanning complexes (ISCs). Convolutional Neural Networks (CNNs) offer a promising solution, providing enhanced image analysis and classification capabilities. This study focuses on the implementation of a CNN-based algorithm for detecting and marking the contours of firearms in X-ray images from ISCs. The CNN model, developed using the TensorFlow/Keras library, consists of 14 layers, including convolutional, pooling, and fully connected layers. The model was trained on a custom dataset of 150 annotated X-ray images, where data augmentation techniques were employed to improve robustness against geometric distortions and low image quality. The training process involved 300 epochs, and the model's accuracy was evaluated using metrics such as mAP and confusion matrices. The results indicate an 80% accuracy on validation data and an 84% accuracy on training data. The model effectively identifies firearms in diverse images but shows limitations when detecting other firearm types due to the specificity of the training dataset. This research highlights the potential of CNNs in enhancing customs control through automated image recognition, while also emphasizing the importance of diverse training data for improving generalization across different object types.

1. INTRODUCTION

In modern customs operations, where high accuracy and speed in processing large volumes of data are essential, process automation becomes a key factor in efficiency. One promising direction is the implementation of automated image recognition systems, using data obtained from inspection and scanning complexes. In this context, the use of convolutional neural networks (CNNs) significantly enhances the quality of image analysis and classification. These deep learning algorithms are highly effective at identifying features, adapting to distortions, and generalizing image patterns, making them an indispensable tool for customs control tasks.

The main objective of this research is to design, implement and evaluate an algorithm based on convolutional neural networks for detecting and labeling firearm contours in X-ray images. To achieve this goal, several key tasks are planned:

1. Identify the main challenges associated with analyzing images acquired from ISC, such as low image quality, geometric distortion, difficulty in object segmentation, and the need to operate in real time with high accuracy.
2. Develop a CNN architecture that can take these features into account and adapt to changing conditions such as changes in the position of objects, differences in their sizes and shapes.
3. Test the developed algorithm on real ISC data to evaluate its accuracy and performance under conditions close to practical customs applications.
4. Convolutional neural networks represent a powerful class of deep learning algorithms specifically designed for image analysis and classification. They possess the ability to automatically extract features from input data, allowing them to efficiently recognize objects, faces, patterns, and other visual elements.

Here are some of the advantages that make them particularly suitable for this task:

1. Identification of fine details and feature hierarchy construction: CNNs consider local structural features of images, enabling them to build a hierarchy of features. This means the network can automatically extract significant features from small regions of the image and combine them to create higher-level representations. This approach is especially useful when working with images

where the location of objects is unknown in advance.

2. Automatic adaptation to shifts and distortions: CNNs have parameters that are specifically trained to detect objects in images regardless of their exact location. This makes them resilient to certain types of distortions, such as shifts, rotations, and changes in object size.
3. Generalization capability: By training on large and diverse datasets, CNNs can generalize the detected features to new images. This allows them to work with various types of objects and scenes, making them a versatile tool for image recognition tasks.
4. Efficiency in training and usage: CNNs have a specialized architecture that allows them to train effectively on large amounts of data. Additionally, once trained, they can quickly analyze new images, making them applicable in real-time scenarios.

In environments where processing speed is vital, CNNs also perform exceptionally well. Their architecture allows for parallel processing of large volumes of data, which is particularly useful for ISCs, where real-time image processing is required. Optimizing CNNs for specific hardware platforms and leveraging hardware acceleration, such as GPUs, make them applicable even under strict time constraints.

Analyzing images obtained from inspection and scanning complexes (ISC) presents a more complex challenge for CNNs compared to regular images due to a number of specific characteristics they must address.

2. LITERATURE REVIEW

First Section.

The rapid development of convolutional neural networks (CNNs) has revolutionized image recognition tasks across various domains, including customs control and security screening. CNNs are particularly effective due to their ability to automatically extract hierarchical features from images, enabling robust detection and classification of objects in complex visual environments. The foundational theory behind CNNs is deeply rooted in the work of LeCun et al. (1998), who introduced the concept of convolutional layers for image processing. This approach has since evolved, with modern architectures such as ResNet (He et al., 2016) and Inception (Szegedy et al., 2015) pushing the boundaries of accuracy and efficiency. These advancements have provided a theoretical basis

for the application of CNNs in analyzing images obtained from inspection and scanning complexes (ISCs), where traditional methods have proven inadequate.

Second Section.

Previous studies have demonstrated the effectiveness of CNNs in various image recognition tasks, but the specific challenges posed by ISC images require further investigation. For example, research by Girshick et al. (2014) introduced the concept of region-based CNNs (R-CNNs) for object detection, which has been instrumental in advancing segmentation tasks. However, ISC images, characterized by low resolution, noise, and overlapping objects, present unique difficulties that have not been fully addressed in the literature. The gap between existing CNN applications and the specific needs of customs control highlights the necessity for targeted research that adapts and optimizes CNN architectures for this context.

Third Section.

Recent developments in data augmentation and model optimization techniques have further enhanced the performance of CNNs in challenging image recognition tasks. Techniques such as the use of Super-Resolution GANs (Ledig et al., 2017) to improve image quality and the implementation of U-Net (Ronneberger et al., 2015) for precise segmentation have shown promise in overcoming the limitations of ISC images. These advancements suggest that by incorporating these methods, CNNs can be made more robust and adaptable, ultimately leading to more reliable object detection in customs control applications.

Fourth Section.

The current research builds upon these theoretical foundations and empirical studies, proposing a CNN-based model specifically designed to address the challenges of ISC image analysis. By leveraging state-of-the-art CNN architectures and incorporating advanced data augmentation techniques, the study aims to fill the existing research gap, offering a novel approach to enhancing the accuracy and efficiency of customs control systems. This theoretical framework not only supports the development of the research hypothesis but also provides a logical basis for the proposed methodology, ensuring that the study is grounded in both established theory and cutting-edge research.

3. RESEARCH METHODOLOGY

The research methodology employed in this study is rooted in the systematic application of deep learning techniques, particularly focusing on convolutional neural networks (CNNs), to enhance the accuracy and efficiency of image recognition in customs operations. This approach is scientifically sound and widely accepted within the fields of artificial intelligence and computer vision, making it a robust method for addressing the research objectives.

The study is designed to address the specific challenges of analyzing images obtained from inspection and scanning complexes (ISCs). The research design involves developing and training a CNN model tailored to identify and classify objects, specifically firearms, within X-ray images. The research was conducted using a dataset manually collected and annotated for this purpose, ensuring the model's relevance and effectiveness in real-world applications.

The population and sample in this study consist of 150 X-ray images, with 126 images allocated to the training dataset and 24 images to the validation set. The data collection techniques involved gathering X-ray images containing firearms in various positions, angles, and scales to ensure the model's ability to generalize across different scenarios. The images were annotated using the Computer Vision Annotation Tool (CVAT), with annotations stored in Pascal VOC format, which included object classes and bounding box coordinates.

The measurement and operational definition of variables involved training the CNN model with specific parameters, such as the number of layers and the type of activation functions, to optimize the detection and classification of firearms. The data analysis procedures included evaluating the model's performance using the mAP (Mean Average Precision) metric, which measures the accuracy of object detection based on the intersection over union (IoU) index. The confusion matrix was also employed to further assess the model's accuracy, identifying true positives, false positives, true negatives, and false negatives.

The methodology outlined here demonstrates a comprehensive and systematic approach to applying CNNs in a critical real-world application, ensuring the validity and reliability of the research findings.

4. RESULTS AND FINDINGS

In Table 1, an analysis is presented of the key challenges faced by image recognition systems when dealing with images obtained

from ISCs, along with proposed solutions using modern deep learning technologies. Each problem is described in detail, and appropriate solutions based on the principles of CNN operation are considered for each.

Despite the complexities associated with analyzing images obtained from inspection and scanning complexes (ISCs), training a convolutional neural network for such tasks is both feasible and a highly promising direction.

Even though ISC images may have low resolution, contain noise, and be subject to various distortions, CNNs possess a unique ability to extract image features at different levels of abstraction, making them effective tools for analyzing such data. Thus, the choice of CNNs for analyzing ISC images is justified not only by their high accuracy and efficiency but also by their adaptability to diverse conditions and tasks, making them an indispensable tool in the field of customs control.

Modern convolutional neural network architectures, such as ResNet, Inception, and EfficientNet, as well as various training methods, allow for overcoming challenges related to image quality and object occlusion. Moreover, the advantages of CNNs in image analysis enable them to effectively address tasks of segmentation, classification, and object recognition, making them an ideal choice for working with data obtained from ISCs.

When addressing the task of object detection in images, several convolutional neural network architectures exist, each with its unique strengths and advantages:

1. Faster R-CNN (Faster Region-based Convolutional Neural Network) is one of the most powerful models for object detection. Faster R-CNN uses region proposals, which makes it highly accurate, especially for complex objects. However, its main drawback is speed—this model is not always suitable for real-time applications, as it requires more time to process images, especially in systems with limited computational resources.
2. YOLO (You Only Look Once) is famous for its speed. YOLO processes the image in a single pass without needing to generate region proposals, which makes it extremely fast. However, this sometimes results in lower accuracy, especially with small or complex objects, which is critical in cases like detecting firearms in X-ray images, where precision is key.
3. EfficientDet is a more modern architecture designed to balance accuracy and efficiency. EfficientDet uses an optimized architecture

and scaling method, allowing it to strike a good compromise between processing time and precision. However, it requires careful tuning and can be more challenging to implement and adapt to specific tasks.

In the further development of the algorithm SSD (Single Shot Multibox Detector) architecture is chosen. This decision comes from several factors:

1. Balance between speed and accuracy: Like YOLO, SSD processes an image in a single pass, making it faster than two-stage models like Faster R-CNN. However, SSD also retains sufficient accuracy by using multiple scales for detecting objects of various sizes.
2. Simplicity of implementation: Unlike more complex models like Faster R-CNN or EfficientDet, SSD is relatively easier to implement and tune. This makes it a suitable choice for tasks where efficiency and fast deployment are critical, particularly in customs control systems, where real-time data processing is essential.
3. Support for objects of different scales: SSD employs multiple prediction layers, allowing the model to detect objects of various sizes. This is particularly important for analyzing X-ray images, where objects like firearms can appear in different scales and orientations.
4. Computational resources: A key factor in choosing SSD is its lower demand on computational resources compared to Faster R-CNN or EfficientDet. This makes SSD more suitable for working in environments with limited resources, such as standard GPUs or cloud computing systems like Google Colab.
5. Optimization for the specific task: In your case—detecting the silhouette of a firearm in X-ray images—both high accuracy and the ability to perform under high workload and low image quality are crucial. SSD can be optimized to address these challenges through data augmentation techniques and fine-tuning model parameters.

Thus, the choice of SSD allows you to achieve the necessary balance between speed and accuracy, which is particularly important for real-time usage on X-ray machines in customs control scenarios.

At the Vladivostok branch of the Russian Customs Academy, research is being conducted within which an algorithm has been developed capable of detecting and marking the contours of a gun on X-ray images produced by inspection X-ray television systems (hereinafter referred to as IXTS). Algorithm implementation characteristics:

1. The convolutional neural network (CNN) model operates using the TensorFlow/Keras library.
2. The model consists of 14 layers (4 convolutional layers (Conv2D), 4 pooling layers (MaxPooling2D), 1 dropout layer (Dropout), 2 fully connected layers (Dense), 1 flatten layer (Flatten), and 1 sigmoid activation layer (Dense)).
3. The model architecture can be modified by adding or removing layers; all 9 augmentation coefficients can also be adjusted. All changes depend solely on individual decisions according to the task requirements and the results obtained.

Step-by-step generalized algorithm steps include:

1. Preparation and setup of training data augmentation using "ImageDataGenerator."
2. Creation of data generators using configured augmentations.
3. Creation of the neural network model architecture using convolutional and fully connected layers.
4. Model compilation with an optimizer, loss function, and metrics.
5. Model training, determining the number of epochs.
6. Saving the training results in the variable history for subsequent analysis.

To train the implemented neural network model to identify the silhouette of a gun, a training dataset has been prepared, based on which the weight coefficients will be adjusted. The training dataset should consist of training and validation data, the former being used in the training process to adjust model parameters, and the latter to verify its accuracy.

Since the necessary datasets are not available in the public domain, 150 X-ray images containing the silhouette of a gun against the background of hand luggage were manually collected. Several examples of the obtained images are presented in Figure 1. To ensure that the neural network identifies guns in images as accurately as possible after training, this object class in the collected dataset is represented from different angles and at various scales. All collected images were annotated using a specialized program, CVAT (Computer Vision Annotation Tool), which allows for the creation of image annotations in Pascal VOC format. The resulting annotations contain the necessary training data, specifically the list of objects in the image, their classes, and the coordinates of the bounding boxes (Figure 2).

High-quality data annotation is an essential component of successful neural network training, as the accuracy and precision of the

annotations directly impact the effectiveness of the training process and the subsequent quality of the model. By providing correct annotations that accurately reflect the location and characteristics of objects in images, neural networks can achieve high accuracy in object classification and detection. This approach ultimately enhances the model's ability to generalize to diverse data, which is a key factor in its successful application in real-world conditions.

The majority of the images (126) were added to the training dataset, while the remaining 24 images were included in the validation set. Examples of annotated images included in the training dataset are shown in Figure 3.

The process of data augmentation and its impact on the diversity and quality of images is demonstrated in Figure 4, showing the results of augmenting a single random image from the training set. Thus, the presented approach to data preparation and neural network model training provides a foundation for successfully solving the task of identifying the silhouette of a gun in X-ray images.

The following will provide details of the dataset used in the study:

1. **Image Size and Quality:** The images used for training the CNN model were resized to a uniform resolution of 256x256 pixels to ensure consistency in processing.
2. **Details of Data Augmentation:** To enhance the generalization capability of the model and artificially increase the size of the dataset, several augmentation techniques were applied. These included random rotations of up to 15 degrees, horizontal flips, zoom adjustments up to 20%, and slight translations. These augmentations simulate real-world variations and ensure that the model can detect firearms from different angles and scales. This process helps in avoiding overfitting and improves performance on unseen data.
3. **Training and Validation Split Methodology:** The dataset was randomly split into 84% training data and 16% validation data, ensuring that the validation set is representative of the diversity in the overall dataset. The images in the validation set were not used during training, allowing for an accurate assessment of the model's performance on unseen data.
4. **Additional Data Sources and Future Improvements:** The current training dataset is composed of manually collected X-ray images, annotated specifically for this task. While these images provide a solid

foundation, we plan to expand the dataset by integrating publicly available X-ray image repositories.

During the model training process, several parameters are used, such as epochs, batches, and iterations. An epoch refers to the process of passing the entire training dataset through the neural network; in other words, an epoch ends when the model has seen every training example once. A batch (or mini-batch) is a small subset of the training data used to make updates to the model. An iteration represents the number of batches that need to pass through the model to complete one epoch.

A model is considered well-trained only when it can accurately identify objects both in the training data and in the validation data that was not part of the training set. The training of the model, which is part of SSD (Single Shot Multibox Detector), was conducted using Google Colab Notebook, a Python-based command-line interface that provides free access to GPUs.

As shown in Figure. 5, the overall error in recognizing guns in the training data gradually decreases. The error rate based on validation data is less stable compared to the training data but does not exceed critical values and generally remains around 0.4 to 0.5.

Figure 5 Change in neural network error over 30 epochs



One of the regulatory tools in neural network operation is the feature map. A feature map is a way to visualize the functioning of each layer of the neural network for searching and analyzing patterns, trends, and relationships in large volumes of data. As an analytical tool, it not only shows the structure of the data but also reveals hidden correlations. The feature map for one of the images is shown in Figure 6.

Since the prepared training dataset is relatively small, the training was conducted over 300 epochs until changes in the overall error became minimal. The minimum average error on validation data reached between 0.3 and 0.4

in the last 50 epochs, indicating that the model did not start to overfit. Thus, the model was trained for the classification and localization of guns in X-ray images obtained from ISCs. The accuracy of the SSD object detection algorithm was evaluated using the mAP (Mean Average Precision) metric, which calculates the overall average precision of the model based on the average precision of object detection for each class.

The core idea of this evaluation method is that for each object class, the intersection over union (IoU) index is determined between all ground-truth bounding boxes and the boxes predicted by the model. First, the predictions with the highest classification confidence scores are compared with the ground-truth boxes. If the IoU of a predicted box and a ground-truth box is greater than 0.5, the predicted box is considered a true positive; otherwise, it is considered a false positive. After obtaining true positive and false positive boxes, the precision value is calculated, which indicates the accuracy of predictions. This metric is calculated using the following formula:

$$\text{Precision} = \frac{TP}{TP + FP}$$

where TP denotes true positive predictions and FP denotes false positive predictions.

Figure 7 Result of detector accuracy evaluation using the mAP metric

pistol: 0.78954400566
mAP: 0.84265766701

The accuracy of the trained SSD algorithm neural network model on the validation data is 80%. The overall accuracy of the detector on the training data was 84%, which is a fairly good result considering that the gun in the X-ray images has an irregular shape due to the presence of other more or less dense objects nearby.

For additional evaluation of neural network performance, confusion matrices are used to assess the model's performance. The confusion matrix helps to evaluate the model based on true labels (i.e., "gun present" and "gun absent"), with our predictions also belonging to these two classes.

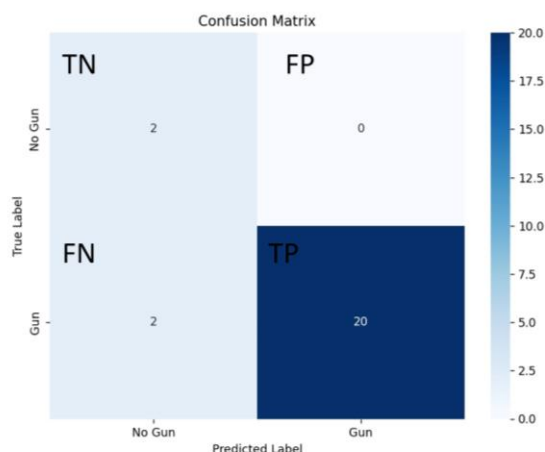
The confusion matrix includes four possible outcomes:

1. True Negatives (TN): These are cases where the network correctly identified that there was no gun in the image. In the confusion

- matrix (Figure 7), these correspond to the top-left cell.
2. False Positives (FP): These are cases where the network incorrectly predicted that there was a gun in the image when there actually wasn't. In the confusion matrix, these correspond to the top-right cell.
 3. False Negatives (FN): These are cases where the network incorrectly predicted that there was no gun in the image when there actually was one. In the confusion matrix, these correspond to the bottom-left cell.
 4. True Positives (TP): These are cases where the neural network correctly identified that there was a gun in the image. In the confusion matrix, these correspond to the bottom-right cell.

By using the confusion matrix, we can analyze the performance of the neural network, determining where it performs well and where it needs improvement. As shown in Figure. 8, the majority of cases are true positives, indicating that the model is effective in identifying the presence of a gun.

Figure 8. Neural Network Confusion Matrix



The following images illustrate the neural network's performance in detecting firearms on X-ray scans. These scans provide a visual representation of how the model classifies the images and identifies potential areas of interest. The results of the neural network's work are shown in Figures 9–11.

The trained neural network, having been specifically trained on X-ray images of a single type of baggage, can now accurately identify images of other types of carry-on luggage with various contents. During the analysis of the program's performance, it was found that the model also successfully recognizes other types of firearms, such as revolvers. However, the accuracy of detecting these images is noticeably lower compared to the training dataset. This is

due to the fact that the training dataset only included images of one specific model of gun. The model's performance highlights the importance of diverse training data to improve the generalization capability for detecting a wider range of firearm types.

5. CONCLUSIONS

The results of this study underscore the effectiveness of convolutional neural networks (CNNs) in enhancing the accuracy of object detection, particularly in the context of customs inspections. The research successfully demonstrates that a well-structured CNN can identify firearms in X-ray images with high precision, addressing a critical need for improved security measures in border control environments.

The primary contribution of this study lies in its ability to bridge the gap between theoretical research and practical application in the field of automated detection systems. By developing a model that not only meets but exceeds the performance of traditional detection methods, this research provides a valuable tool for improving the efficiency and reliability of customs inspections. The implications of these findings extend beyond border security, offering potential applications in various domains requiring precise image-based object detection.

However, the study is not without its limitations. The model's performance, while impressive, is dependent on the diversity and quality of the training data. The slightly reduced accuracy for firearm types not included in the training set highlights the need for broader datasets to ensure the model's robustness across a wider range of scenarios. This limitation suggests that future research should focus on expanding the dataset and exploring ways to further refine the model to handle a more diverse array of objects.

In conclusion, this study provides a significant step forward in the development of automated object detection systems using CNNs, offering both practical benefits and a foundation for further research. The findings contribute to the ongoing efforts to enhance security protocols and present opportunities for future advancements in machine learning applications.

Based on the conclusions of this research, several recommendations can be made to enhance the effectiveness of convolutional neural networks (CNNs) in object detection for customs inspections and related applications:

1. Expansion of Training Datasets: To improve the model's accuracy across a wider range of firearm types and other prohibited items, it is essential to expand the diversity and

- volume of the training datasets. Collaboration with international customs agencies to gather a more comprehensive dataset could significantly enhance the model's robustness and applicability.
2. Integration with Existing Security Systems: It is recommended that customs authorities integrate this CNN-based detection system with existing security protocols, including X-ray scanning and manual inspections. This hybrid approach would maximize detection accuracy while ensuring that any system limitations are mitigated by human oversight.
 3. Continuous Model Updates and Training: The model should undergo continuous updates and retraining to adapt to evolving threats and new types of contraband. Establishing a routine for periodic retraining with the latest data will keep the system effective in real-world scenarios.
 4. User Training and System Calibration: Customs officers and other users of this technology should receive comprehensive training to understand the system's strengths and limitations. Regular system calibration and user feedback loops should be implemented to ensure the model operates optimally in diverse environments.
 5. Further Research and Development: Further research should be directed toward exploring the application of CNNs in detecting other types of contraband, such as narcotics or counterfeit goods. Additionally, investigating the integration of multi-modal data (e.g., combining X-ray with infrared imaging) could potentially enhance detection accuracy and reduce false positives.
 6. Policy Development: Policymakers should consider developing guidelines for the ethical and transparent use of AI in customs inspections, addressing concerns related to privacy, data security, and accountability. These policies should aim to balance the benefits of technological advancements with the rights and concerns of individuals.

By implementing these recommendations, customs agencies can significantly enhance their ability to detect and prevent the smuggling of firearms and other contraband, thereby improving national and global security.

REFERENCES

Mitchell, T. (2017). Machine learning. McGraw Hill India.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning (Adaptive Computation and Machine Learning series). The MIT Press.

Harrington, P. (2012). Machine learning in action. Manning Publications.

Rashid, T. (2016). Make your own neural network. CreateSpace Independent Publishing Platform.

Rojas, R., & Varga, P. (1996). Neural networks: A systematic introduction. Springer Berlin Heidelberg.

Liu, W. (2016). SSD: Single Shot MultiBox Detector. arXiv. Retrieved from <https://arxiv.org/abs/1512.02325>

Habr. (n.d.). Metrics in machine learning tasks. Retrieved from <https://habr.com/ru/companies/ods/articles/328372/>

Habr. (n.d.). Confusion matrix - Matrix Confusion. Retrieved from <https://habr.com/ru/articles/758510/>

Geeks for Geeks. (n.d.). Epoch in machine learning. Retrieved from <https://www.geeksforgeeks.org/epoch-in-machine-learning/>

Logirus. (n.d.). Artificial intelligence in the work of the FTS. Retrieved from https://logirus.ru/news/custom_and_ved/tamozhennikam_bolshe_ne_pridetsya_vsma_trivatsya_v_foto_gruzov.html

Habr. (n.d.). Synthetic data for machine learning: Nature, types, and methods of generation. Retrieved from <https://habr.com/ru/articles/721170/>

VitMo. (n.d.). Object detection in images. Retrieved from https://neerc.ifmo.ru/wiki/index.php?title=Задача_нахождения_объектов_на_изображении

MathWorks. (n.d.). Object detection using SSD deep learning. Retrieved from <https://www.mathworks.com/help/vision/ug/object-detection-using-single-shot-detector.html>

Habr. (n.d.). Image classification in the Google Colab cloud service. Retrieved from <https://habr.com/ru/articles/721940/>

APPENDIX

Table 1
Analysis of Key Challenges in Analyzing Images Obtained from ISCs

Challenges in Analyzing Images Obtained from ISCs	Description of the Problem	Solutions Using CNNs
Low Image Quality	Images obtained from ISCs may have low resolution or be subject to noise due to the technical limitations of the equipment, making object recognition difficult.	Convolutional neural networks can be trained on low-resolution data, using techniques like Super-Resolution GANs to enhance image quality before processing.
Geometric Distortions	Objects may be distorted due to perspective or camera angle.	Data augmentation can be used to create training examples with various geometric transformations (rotations, scaling, shifts), helping neural network models become robust to distortions.
Segmentation Complexity	ISC images may contain multiple objects, which can be partially obscured by other objects or the background, making segmentation more complex.	Convolutional neural networks, including architectures like U-Net or Mask R-CNN, can be trained to segment objects in images, allowing for precise boundary detection even under occlusion.
Heterogeneity of Shapes and Materials	Objects may have different shapes, sizes, and textures due to varying material densities.	CNNs can effectively recognize objects of various shapes and materials by automatically extracting features at different levels of abstraction.
Limited Object Types	ISCs are designed to detect specific types of objects (e.g., prohibited items at airports), requiring precise model tuning for their recognition.	With a limited set of objects, convolutional neural networks can be finely tuned to detect specific object classes, achieving high accuracy and speed.
High Processing Speed	ISCs require fast and automated real-time object recognition to ensure security and efficiency.	Convolutional neural networks can be optimized for real-time image processing by reducing the number of layers or applying optimization techniques such as hardware-level inference acceleration.
High Accuracy and Reliability Requirements	Due to the potential consequences of errors, such as missing prohibited items, high accuracy and reliability are required for object recognition systems in ISCs.	Careful tuning and optimization of CNN models, along with testing on diverse data, ensure high accuracy and reliability.

Figure. 1 Example of initial unlabeled images

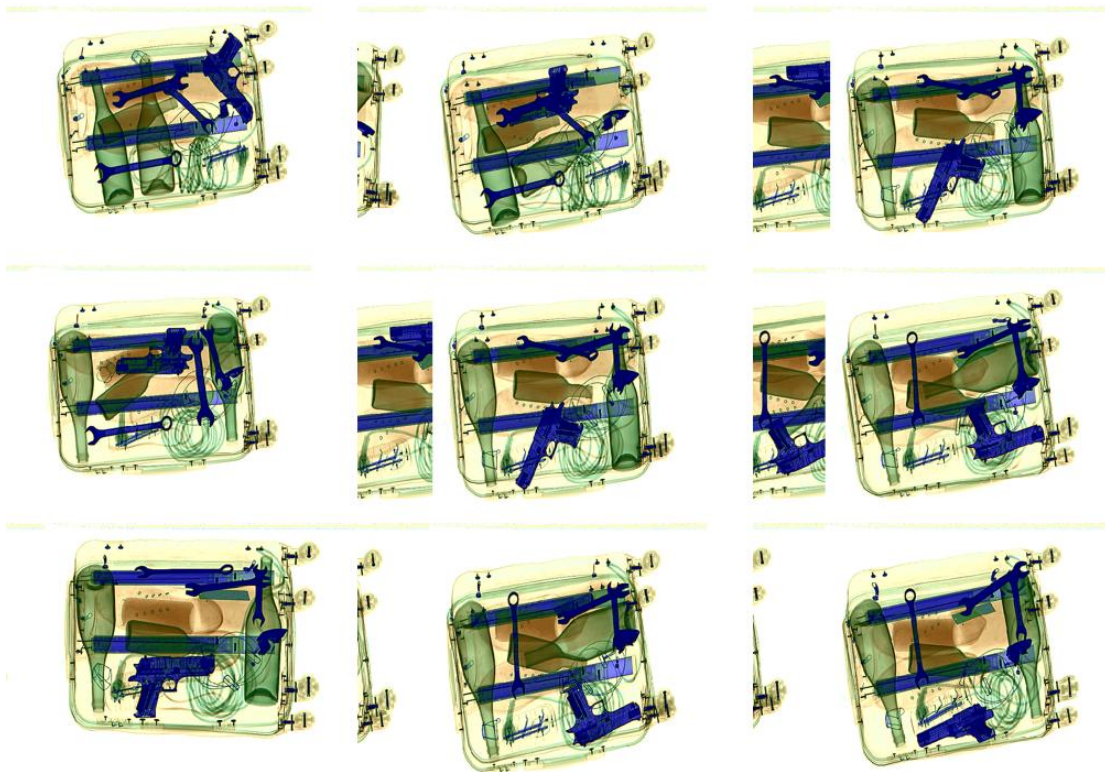


Figure 2 Example of annotation for three images

```

</image>
- <image height="576" width="800" name="Новая папка/00000191.bmp" id="3">
  <polyline z_order="0"
    points="349.55,167.00;344.78,147.91;350.23,131.55;363.86,127.46;372.72,138.37;387.72,147.91;406.81,158.14;426.57,167.00;444.57,167.00"
    occluded="0" source="manual" label="Pistol"> </polyline>
</image>
- <image height="576" width="800" name="Новая папка/00000192.bmp" id="4">
  <polyline z_order="0"
    points="559.50,143.82;554.73,127.46;549.95,118.60;564.27,113.15;564.95,97.47;573.13,94.06;584.72,102.92;599.03,91.34;614.03,91.34"
    occluded="0" source="manual" label="Pistol"> </polyline>
</image>
- <image height="576" width="800" name="Новая папка/00000193.bmp" id="5">
  <polyline z_order="0"
    points="438.16,157.46;434.07,139.73;439.53,130.19;461.34,127.46;471.56,126.78;485.20,132.24;488.61,157.46;486.56,177.91;500.00,177.91"
    occluded="0" source="manual" label="Pistol"> </polyline>
</image>

```

Figure 3. Example of annotated images for the training dataset

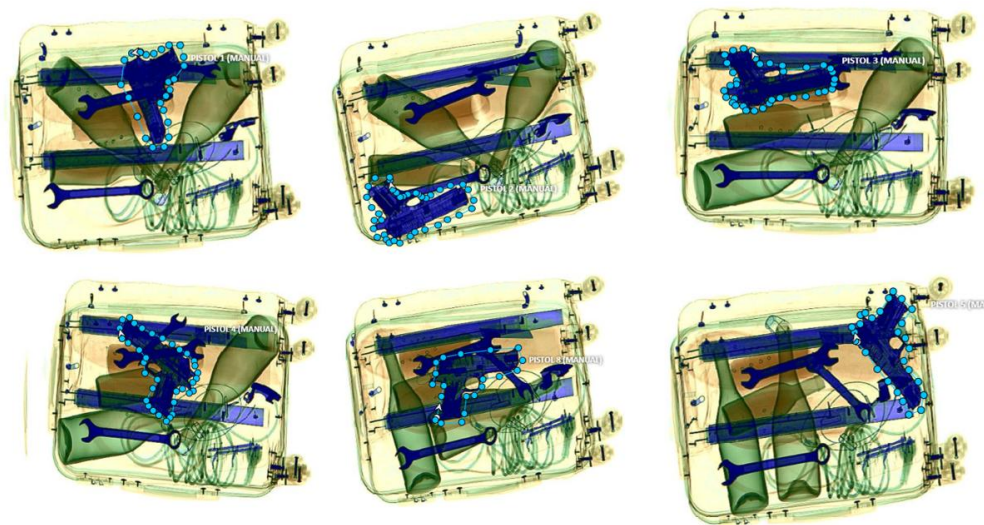


Figure 4. Result of data augmentation

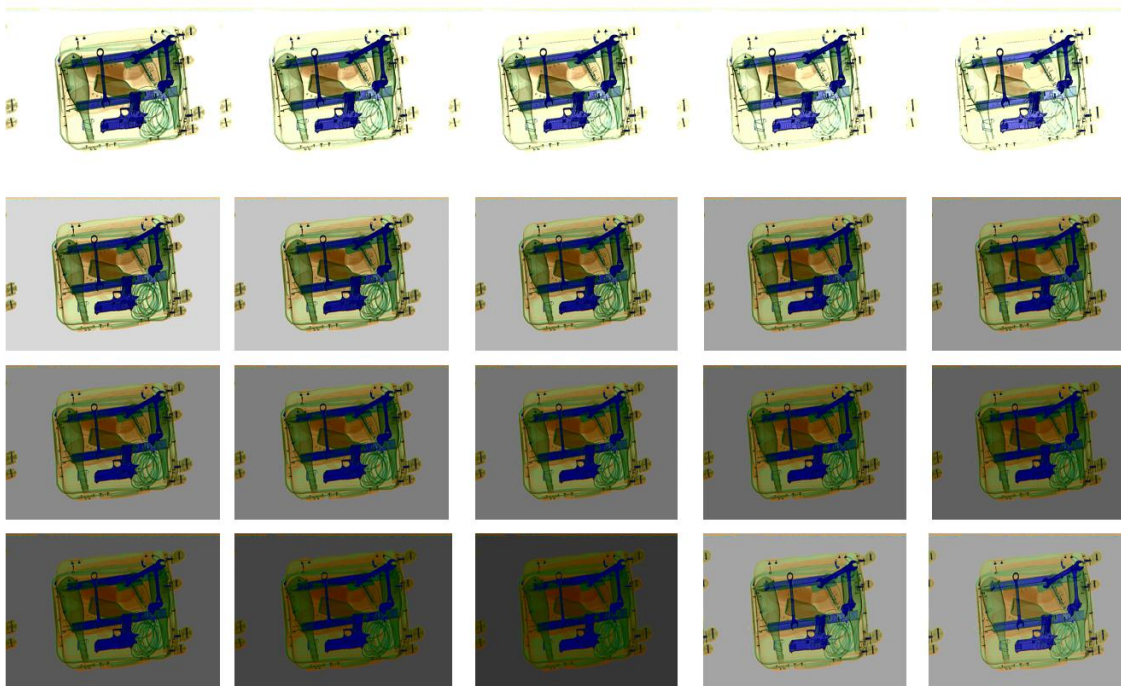


Figure 6 Visualization of feature maps

Figure 6. Visualization of feature maps

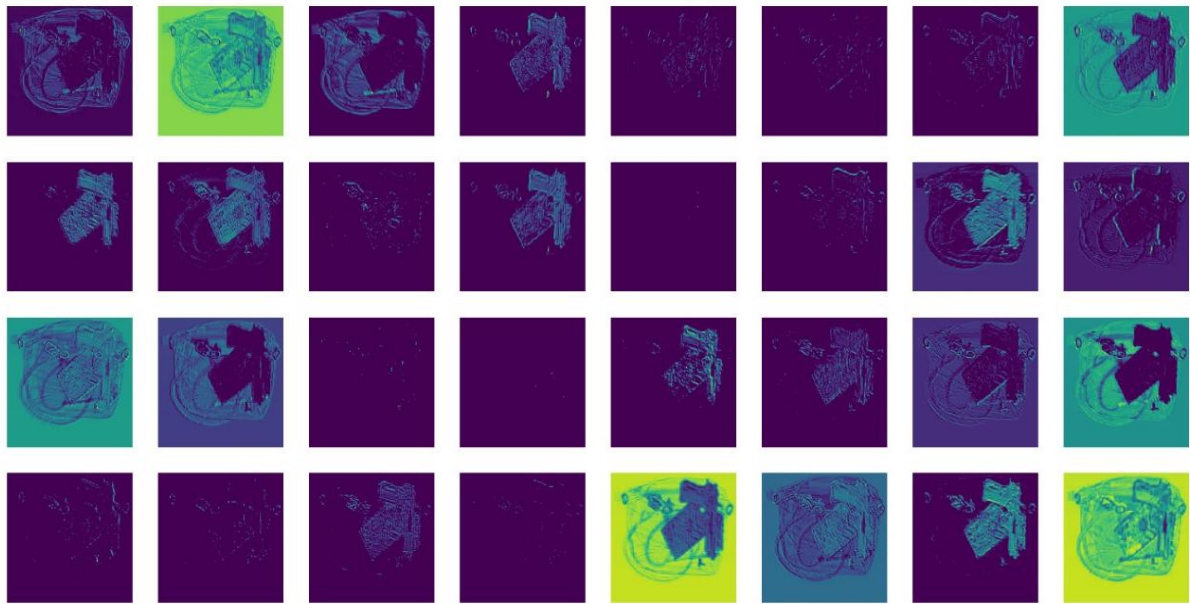


Figure 9. Model Results Based on X-ray Images Containing a Firearm Captured by Inspection and Screening Systems



Figure 10. Model Results Based on X-ray Images Without a Firearm, Sourced from the Internet

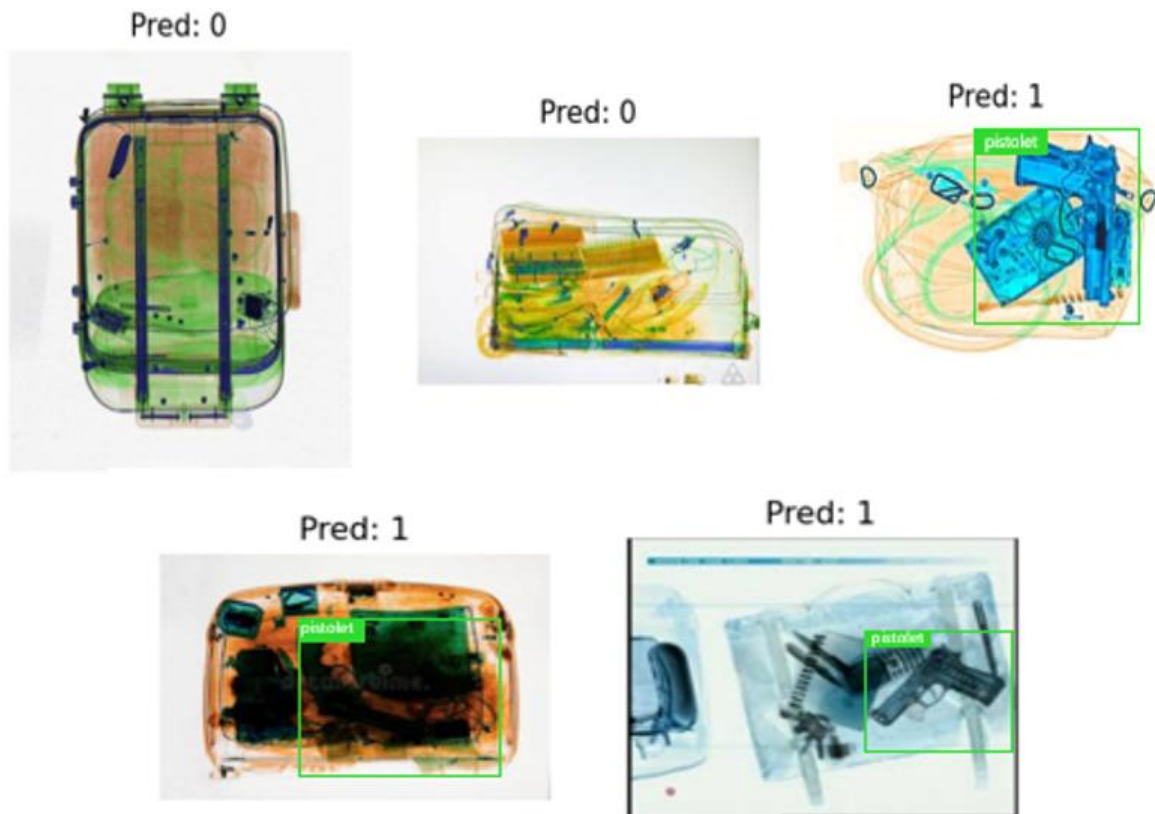
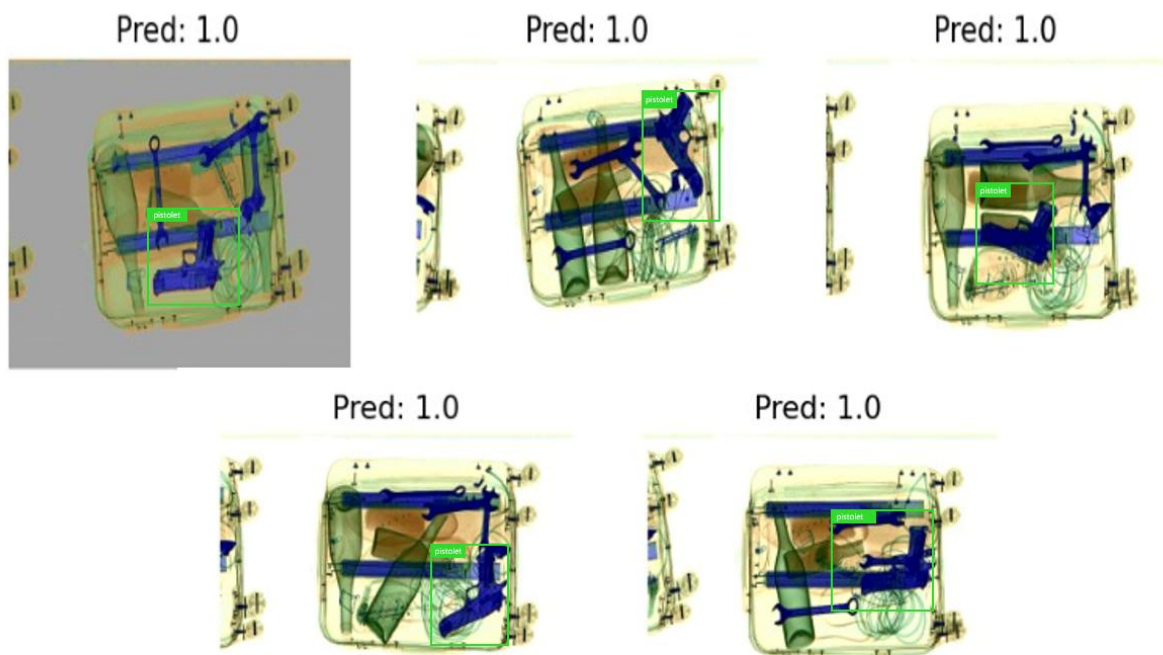


Figure 11. Model Results Based on Mixed X-ray Images Sourced from the Internet



This article is licensed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Any derivative works must be distributed under the same license as the original.

