Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-04 | Issue-04 | December-2023

Indoor Smoking Detection Based on YOLO Framework with Infrared Image

Abdullah Al Nayeem Mahmud Lavu[1], Hua Zhang[2], Hao Zhao[3], MD Toufik Hossain[4]
[1, 4] MSc Scholar, School of Information Engineering, Southwest University of Science and technology (SWUST), Mianyang, P. R., China.
[2]Professor, School of Information Engineering, Southwest University of Science and technology (SWUST), Mianyang, P.R., China.
[3]Lecturer, Department of Automation, Southwest University of Science and technology (SWUST), Mianyang, P.R., China.
Abdullah.lavu@gmail.com, zzhh839@163.com, zhaohao@swust.edu.cn, toufikhossain@rocketmail.com

DOI: 10.5281/zenodo.10594345

ABSTRACT

This study recommends combining the efficacy of YOLO with the greater visibility provided by infrared imaging to create a better indoor smoking detection system. The YOLO system divides photos into a grid and anticipates bounding boxes and class probabilities at the same time, making it an obvious choice for its real-time item detection capabilities. The approach improves its robustness by identifying heat signals associated with smoking sessions and overcoming limitations posed by low-light or blocked circumstances. The addition of infrared images significantly improved the system's performance in low-light conditions. A dual spectrum thermal camera is used in the entire indoor smoking detection system to obtain a large collection of infrared images representing various interior locations with documented smoking episodes. During the training phase, data augmentation processes such as random rotations, flips, and brightness and contrast fluctuations were used to improve the system's performance. The CIoU loss function improved the system's localization accuracy significantly, reducing false positives and improving overall detection performance. The combination of YOLO and infrared photography, in conjunction with data augmentation and the CIoU loss function, not only improves indoor smoking detection but also demonstrates the benefits of merging several technologies in the development of more effective and adaptive systems.
.

Keywords: Indoor Smoking Detection, Infrared Image, You Only Look Once (YOLO), Deep Learning, Machine Learning, Data Augmentation.

Cite as: Abdullah Al Nayeem Mahmud Lavu, Hua Zhang, Hao Zhao, & MD Toufik Hossain. (2024). Indoor Smoking Detection Based on YOLO Framework with Infrared Image. *LC International Journal of STEM*, *4*(4), 51–71. **https://doi.org/10.5281/zenodo.10594345**

INTRODUCTION

Indoor smoking is a significant concern for public safety and fire danger. Traditional procedures have been ineffective, necessitating the development of novel detection methods. Advances in computer vision and artificial intelligence, such as the YOLO framework and infrared image analysis, hold the promise of advanced indoor smoking detection systems. These technologies provide real-time monitoring and increased precision, potentially filling a significant gap in current detection methods and adding to intelligent surveillance systems. The system's goal is to make indoor environments safer while also contributing to technological progress. Object detection is a vital component of computer vision, essential for many applications such as surveillance, robotics, and autonomous systems. Traditional approaches endure in low-light situations or at night, leading to the advent of the You Only Look Once (YOLO) architecture. This grid-based approach enhances item identification efficiency and responsiveness to real-time requirements. However, infrared imaging technologies offer advantages such as enhanced visibility, heat sensing, and real-time applications [1]. The YOLO architecture's versatility enables it to effortlessly shift between visible light and infrared settings, making it suited for time-sensitive applications in autonomous vehicles or rapid-response systems [2][3][4]. This project attempts to adapt the YOLO architecture to infrared light, enhancing its capabilities and broadening the boundaries of object identification technologies. Indoor smoking detection has obstacles due to the particular characteristics of infrared imaging. Conventional object detection algorithms struggle with the intricacy of infrared spectrum, forcing the need to alter current frameworks like YOLO to deliver exact, real-time object recognition. Advancements in sensor technology, machine learning, and smart building integration have enhanced the precision and efficiency of indoor smoke detection systems. These technologies can identify smoke particles and impurities, aid in real-time analysis, and interface with existing systems.

However, problems exist, such as false positives and negatives, privacy concerns, and the necessity for education campaigns to educate building occupants about smoke detection devices. Although, progress in indoor smoking detection is encouraging, addressing challenges such as accuracy, privacy, integration, affordability, and user awareness is vital for their success, ethical acceptability, and universal acceptance in varied interior environments. This initiative intends to address the issues in object detection under infrared light by employing the You Only Look Once (YOLO) architecture. The research concentrates on fitting the YOLO architecture to the complexity of infrared imagery, emphasizing on low contrast and textural diversity, and the varied thermal fingerprints emitted by objects. The research also intends to boost detection accuracy by eliminating false positives and negatives, assuring correct item identification in infrared pictures. The research also intends to enhance real-time object detection in infrared, boosting processing efficiency without losing accuracy. The model's adaptability to dynamic lighting circumstances is also a priority, as object identification algorithms are sometimes inhibited by variations in illumination, notably during twilight, dawn, or nocturnal settings.

The aim of the research is to uncover the full potential of YOLO in infrared applications, enhancing the accuracy, efficiency, and flexibility of object recognition systems in real-world settings where typical visible light approaches fall short. The project attempts to adapt the YOLO model to the complexity of infrared imaging, assuring precise object recognition and eradicating false positives and negatives. This project attempts to construct a customized You Only Look Once (YOLO) model for object recognition under infrared light. The collection will encompass a wide spectrum of infrared pictures, capturing numerous objects and events. The dataset will be annotated to aid successful learning by the model. The model's performance will be examined on the created infrared dataset and contrasted with existing object detection algorithms. The practical applicability of the new YOLO model will be evaluated in real-world applications like as surveillance, driverless automobiles, search and rescue operations, and wildlife monitoring.

The project intends to uncover the limitations and future enhancements of the YOLO paradigm, concentrating on enhancing detection accuracy, overcoming challenges with infrared imaging, and guaranteeing a balance of processing efficiency and accuracy for real-time object recognition. The model will also be evaluated in twilight, morning, and midnight conditions to determine its adaptability and practical utility. The project objective is to contribute to the adaptation of object recognition algorithms for the complicated domain of infrared photography by tackling the aforementioned issues. By assessing each objective, the research aims to establish the groundwork for a more complicated and successful approach to object recognition in the infrared range.

LITERATURE REVIEW

To address the concerns associated with passive smoke inhalation, indoor smoking detection devices have evolved. Traditional smoke detectors have difficulty distinguishing between different sources of particulate matter in the air, resulting in false alarms [5]. The precision and reliability of indoor smoking detection have improved due to a paradigm shift toward computer vision and machine learning [6]. To detect smoking episodes in real time, computer vision algorithms evaluate visual input, whereas machine learning-based solutions use annotated datasets to detect smoking trends [7]. Convolutional neural networks (CNNs) were trained to recognize smoking-related features in photos, increasing detection accuracy. Object detection frameworks such as YOLO (You Only Look Once) have significantly boosted indoor smoking detection technology, with YOLO converting images into grids and predicting bounding boxes and class probabilities for each grid cell at the same time [8]. SSD (Single Shot Multibox Detector) and Faster R-CNN (Region-based Convolutional Neural Network) are two alternative frameworks, each with advantages and disadvantages [9]. Understanding the trade-offs between these frameworks is crucial in determining the most successful indoor smoking detection technique in a variety of circumstances.

Traditional computer vision techniques such as template matching, edge detection, and Histograms of Oriented Gradients (HOG) were used in early object detection systems [10]. However, these systems had difficulty dealing with lighting, complex backgrounds, and a variety of object appearances. The introduction of machine learning techniques such as Support Vector Machines (SVMs) and decision trees represented a significant shift in traditional object detection [11]. Feature learning, enabled by technologies such as Bag of Visual Words (BoVW) and Fisher Vectors, allows systems to adapt and detect complex visual patterns without the need for human intervention [12]. Despite recent advances in deep learning, traditional techniques remain important in certain applications, such as robots for real-time object detection [13]. Hybrid systems have also arisen, combining the capabilities of classical and deep learning methods to strike a compromise between accuracy and computing efficiency [14]. Deep learning's emergence has had a significant impact on the landscape of machine learning and artificial intelligence applications [15]. Deep learning systems can learn complex patterns and traits on their own from raw data, which has applications in healthcare, finance, and autonomous cars [16]. However, challenges abound, such as deep neural network interpretability, the need for enormous labeled datasets, and concerns about the ethical implications of AI systems. As we navigate this era of deep learning dominance, ongoing research is being conducted to enhance existing architectures, investigate novel paradigms, and address ethical issues associated to AI applications [17].

Object detection techniques based on regions have advanced, with models such as Fast R-CNN [18] and Faster R-CNN [19] boosting detection accuracy. Their multi-stage structure, however, raises questions about computational efficiency. Single-shot detectors, such as YOLO and SSD, use a one-shot approach to identify objects in real time. Despite advancements in visible light, there are still challenges in infrared light settings, such as low-light environments or sites illuminated solely by infrared light. Traditional models are unable to respond to these aspects. Because of its potential for real-time and accurate object identification, adapting YOLO for infrared object detection has received a lot of attention. Although fine-tuning CNNs for infrared domains is crucial [20], the YOLO architecture provides a unifying foundation for object detection. Annotated infrared datasets are used to validate customized models [21]. The incorporation of thermal fingerprints into the YOLO design represents a recognition of temperature variations as an important component of infrared object detection [22].

Understanding these issues is critical for developing YOLO and ensuring its efficacy in a variety of real-world situations.

Infrared object detection is used in a variety of fields, including surveillance, security, and autonomous systems. YOLO [23], Faster R-CNN, SSD [24], Mask R-CNN [25], and traditional feature-based approaches such as Haar-like features and Histogram of Oriented Gradients are examples of current methods [26]. Infrared detection, on the other hand, faces challenges such as low contrast and variations in thermal fingerprints [27]. Transfer learning algorithms have been developed by researchers to overcome data scarcity in the infrared domain, increasing model flexibility [28]. To improve model resilience, data augmentation procedures for infrared pictures have also been devised [29]. Personalized architectures are critical for maximum performance, modifying input layers, filters, and classes to address issues with low-contrast environments and variations in thermal signatures [30]. The production of datasets is crucial for efficiently training and assessing models, while architectural changes ensure that models are specialized for identifying objects in low-contrast and thermally changing environments [30] [31] [32]. Surveillance, security, and autonomous systems that operate in low-light or at night are all practical applications of infrared object detection [33]. These investigations' findings bridge the gap between theoretical advances and practical applications of infrared detecting technology. Infrared object recognition is at the intersection of scientific advancement and practical application. The contributions of researchers to dataset construction, architectural alterations, and optimization approaches form the foundation for future improvements in the discipline. Data scarcity is addressed by careful dataset curation, while architectural changes improve current models for infrared imaging difficulties. Infrared object identification holds immense promise for improving safety and security in a variety of contexts as technology advances.

CBS is a computer vision technique that divides pictures or video frames into separate pieces depending on their color qualities. Object detection, image segmentation, biological imaging, and industrial quality control are all possible uses [34] [35]. However, its effectiveness is dependent on its ability to discern color distinctions consistently, and it may be susceptible to lighting conditions and parameter variations [36]. Neural network (NN) models, which mimic the networked architecture of the human brain, are critical drivers in artificial intelligence [37] [38]. Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN) [39], Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) Networks [40] are among the types. Deep learning systems such as AlexNet [41] and ResNet [42] have revolutionized image processing and pattern recognition [43].

Convolutional Neural Networks (CNNs) have transformed image-related tasks, achieving unprecedented accuracy in image categorization, object recognition, and segmentation [44]. Notable designs like as AlexNet [45], VGGNet [46], and ResNet [47] show how CNNs have advanced, continually setting new benchmarks in image recognition tasks. Despite their advances, CNNs confront challenges like as overfitting and poor interpretability. Ongoing research focuses on overcoming these obstacles by investigating novel designs and training methodologies, including advances such as attention processes that improve CNNs' ability to capture contextual information [48]. Neural networks reflect a delicate balance between computer efficiency and human-inspired design, with its advancement defining various applications in today's world. Convolutional Neural Networks (CNNs) are a significant advancement in image processing, allowing detailed features to be extracted from visual input. They are used in a variety of industries and are constantly upgraded to deal with problems. Each network layer performs specific tasks, such as convolutional, pooling, activation, and fully linked layers [49] [50]. The convolutional layer recovers local patterns and traits, but the pooling layer maintains crucial information while decreasing spatial dimensions. The activation function introduces nonlinearity into the network, allowing complex interactions between components to be displayed. The fully connected layer investigates high-level features and makes predictions, whereas the softmax layer converts the previous layer's output into probabilities associated with discrete classes. The chapter also examines the evolution of object identification algorithms, tracing the path from traditional computer vision approaches to deep learning. It demonstrates the significance of deep learning, particularly CNNs, as well as the challenges and potential associated with implementing YOLO architecture for infrared object detection. Color-Based Segmentation (CBS) is also introduced as a computer vision approach in this chapter, with applications in object detection, image segmentation, and industrial quality control.

METHODOLOGY

The third chapter of the paper looks at the design and implementation of a system to improve indoor smoking detection. The main purpose of this system is to combine the YOLO framework with infrared images, which is known for its ability to differentiate between objects in real time. This includes a rigorous set of processes, including the collection of infrared imaging data, data standardization, and the use of data augmentation approaches to improve the model's robustness and generalizability. The primary goal of this system is not only to improve indoor smoking detection, but also to make significant contributions to the field of real-time item recognition in dynamic and complex scenarios. The primary motivations for designing this system were safety concerns, regulatory compliance, early detection, and the incorporation of infrared imaging technology. Safety considerations include the risk of fire hazards and health consequences from secondhand smoke. The system aims to reduce these risks by detecting and alerting to instances of smoking, hence boosting overall safety measures.

Another important aspect of the system is regulatory compliance, as many countries have severe restrictions prohibiting indoor smoking in public places, workplaces, and other enclosed areas. Early detection is crucial for prompt intervention and preventing the escalation of safety and compliance concerns. Infrared imaging technology has advantages such as heat signals associated with smoking, making it especially useful in areas where traditional sight detection can be difficult. Extensive detection skills are also required, as the system is designed to perform successfully in dynamic and unexpected interior situations, responding to changes in lighting, occupancy, and spatial arrangements. Real-world applicability ensures that the approach remains robust and trustworthy in real-world contexts.

YOLO (You Only Look Once) is a cutting-edge method in computer vision for real-time object recognition. It recognizes and calculates the positions of several objects within an image or video frame using a single convolutional neural network (CNN). This method significantly improves speed and accuracy by performing all operations concurrently in a single run through the The primary idea behind YOLO is to divide the input image into a grid of cells, each of which predicts bounding boxes and class probabilities. The YOLO architecture stands out for its real-time capabilities and unconventional design. Grid division, bounding box prediction, class prediction, anchor boxes, and loss function are among its fundamental ideas. The grid-based method examines the entire image in a single forward cycle through the neural network. YOLO's strengths include real-time performance, a rich visual context, object size flexibility, and end-to-end training. However, it has shortcomings like as trouble with small objects, limited context awareness, object overlap sensitivity, and training data requirements. Another area where YOLO may have limitations is in the identification of infrared objects. The grid arrangement may contribute to poor spatial resolution, limiting the model's ability to efficiently describe detailed features of microscopic objects. The grid-based methodology may limit its ability to discern specific contextual links between things, which may affect forecast accuracy in complex circumstances. Furthermore, YOLO may be sensitive to item overlaps, particularly in congested environments, resulting in likely mistakes in bounding box predictions. To summarize, YOLO is a pioneering method in object recognition that provides a unique blend of speed and precision. However, understanding its strengths and weaknesses is essential for effective implementation. These findings will lead to modification and optimization strategies to align the architecture with the specific problems given by infrared vision as this effort progresses to adapt YOLO for infrared object recognition.

YOLO (You Only Look Once) models [51] are widely used in computer vision for object recognition, with applications ranging from self-driving cars to surveillance systems. YOLOv5[52] is one of the most efficient versions of the YOLO series, with benefits such as improved accuracy, optimized model size and inference performance, and a new backbone architecture known as CSPDarknet53. YOLOv5 has been fine-tuned to address accuracy and recall issues raised in previous iterations, making it especially well-suited for high-precision applications. It also struck a balance by combining competitive precision with a smaller model size, which is essential for real-time applications. The CSPDarknet53 design has more rich and complicated features, improving performance in recognizing objects with varied properties. YOLOv5 pioneered novel training approaches, such as mosaic data augmentation and a wider range of anchor box sizes, which increased the model's adaptability to different datasets. Its user-friendly design encourages greater community communication, boosting cooperation and idea exchange. Customers can fine-tune the model on bespoke datasets with fewer labeled instances, reducing annotation burden and speeding up model deployment. The ongoing community development of YOLOv5 ensures that the framework remains current and responsive to the expanding aspirations of the computer vision community. YOLOv5 represents a significant jump in object detection as technology progresses, building on the strengths of its predecessors while overcoming their weaknesses. It is still a notable framework for real-time object recognition due to its increasing accuracy, optimized model size, unique architecture, expanded training methodologies, user-friendly design, transfer learning capabilities, and active community support.

Data augmentation is an important tool in machine learning, providing a way for artificially enlarging datasets and addressing difficulties associated with inadequate labeled data. This comprehensive examination investigates the various strategies of data augmentation, with a particular emphasis on the synergistic combination of knowledge-based and auto-based approaches. To create augmentation techniques, information-based data augmentation relies on domain-specific information, such as anatomical knowledge in medical imaging or assessing certain features crucial to the dataset at hand. By using neural networks to dynamically change transformations during training, automated data augmentation moves toward automation. Reinforcement learning algorithms, such as Proximal Policy Optimization (PPO), allow the model to learn the best augmentation rules, allowing it to generalize more effectively. This independent adaptation to varied patterns in the input aids the model's generalization. The synergistic method necessitates balancing knowledge-based and auto-based solutions, while acknowledging that each brings unique benefits to the table. Below are the mathematical formulae for the data augmentation.

Mathematically, let I represent the original image, and f_k be a knowledge-based augmentation function. The knowledge-based augmented image can be expressed as:

$$I_k = f_k(I) \tag{1}$$

This formulation reflects the application of domain-specific knowledge f_k to the original image I , creating an augmented version I_k that retains relevant features.

The formula for auto-based data augmentation involves a policy P learned by the model, which is applied to the original image I during training:

$$I_{auto} = P(I) \tag{2}$$

Here, I_{auto} represents the image after auto-based augmentation according to the learned policy P
The overall formulation for the combined augmentation can be expressed as:

$$I_{combined} = f_k(P(I)) \tag{3}$$

Where, f_k represents the knowledge-based augmentation function, and P is the learned policy from auto-based augmentation.

This synergistic approach aims to create diverse and domain-relevant variations in the data This merger attempts to generate diverse and domain-relevant variations in the data while preserving domain-specific features and responding to intrinsic variances in the data.

Data augmentation [53] is crucial in the machine learning landscape, particularly in the sensitive area of training deep learning models. It conducts a symphony of changes to the current training dataset, synchronizing actions like rotation, flipping, scaling, and cropping. This diversity is the foundation for the model's ability to transcend the boundaries of its training environment and generalize effectively to unknown inputs. Augmentation acts as a regularization component, enhancing model resilience by preventing the model from memorizing specific examples and encouraging the adoption of more broad and flexible characteristics. Overfitting is a critical problem in machine learning because it occurs when a model becomes overly specialized to the training data, resulting in poor performance when presented with new and unknown data. To alleviate this issue, enriched data can be used, which provides a more complete and diverse training set. Augmented data, which includes adjustments such as rotation, flipping, scaling, and cropping, broadens the range of samples available during training, giving the model a larger palette from which to learn. This large training set improves the model's ability to recognize patterns and qualities in a variety of settings, increasing adaptability and enhancing its ability to extrapolate information to new and unexpected circumstances. Data augmentation is a strategic strategy that addresses the problem of gathering a large labeled dataset, which is frequently related to resource constraints and cost implications. Data augmentation provides a realistic solution to the problem of insufficient original data by artificially increasing the volume of the dataset through various alterations. This enhancement not only broadens the model's exposure to numerous scenarios, but also eliminates the need for a massive labeled dataset, making it a valuable asset in situations when resources are limited. Real-world data, which is inherently dynamic and susceptible to changes in illumination, direction, and other variables, is a complex tapestry that models must navigate. Data augmentation by simulating these real-world variations provides the model with a training experience that is similar to the difficulties it would face in practice. This modeling of environmental nuances ensures that the model, which is replete with enhanced data, is not only adept under controlled training conditions, but also poised for success in the unpredictable and ever-changing landscapes of the real world. Another important facet of augmentation is translation invariance, which is a main component of greater flexibility. It refers to the model's ability to recognize objects regardless of where they are in the input space. Augmentation techniques, such as translation, play an important role in improving translation invariance by reducing the model's sensitivity to the exact position of items within the training images. In the case of small training datasets, data augmentation is a valuable ally in the fight against overfitting. By including variability via augmentation processes, the model is forced to encounter a larger and more diverse set of examples, forcing it to focus on learning more robust and generalizable properties. A sophisticated strategy for training machine learning models is the combination of augmentation techniques, translation invariance, and overfitting mitigation. This method improves the model's ability to handle a wide range of spatial configurations while also instilling resilience that overcomes the limitations of a short training sample. Using knowledge-based and automated data augmentation, the model is pushed to learn, adapt, and extract significant information, resulting in a more reliable and diverse prediction capability.

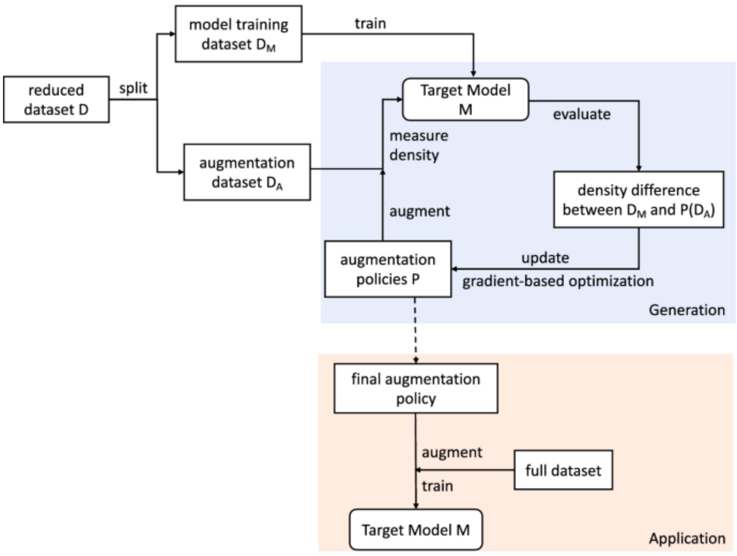


Figure 1 Data Augmentation workflow. Upper sections indicate the policy generation; sections indicate the application stage.

In machine learning and deep learning methods, loss functions such as mean squared error (MSE) are important. They quantify the discrepancy or mismatch between a model's expected output and the actual ground truth. Mean squared error, cross-entropy loss, and custom loss functions built for specific purposes like as object detection or natural language processing are examples of common loss functions. Intersection over Union (IoU) is a popular metric in object identification and segmentation applications for analyzing the overlap of expected and ground truth bounding boxes or masks.

However, building a loss function solely on IoU has significant drawbacks such as non-differentiability, limited gradient data, ignorance of localization accuracy, sensitivity to thresholds, difficulty juggling precision and recall, difficulty dealing with class imbalance, sensitivity to minor errors, and difficulty with generalization.

The Complete Intersection over Union (CIoU) loss function represents a significant advancement in neural network object detection. It provides a more complex assessment approach for bounding box predictions that goes beyond popular measures such as IoU. To comprehend the complexities and implications of CIoU, one must first understand the context of object detection, the limitations of existing metrics, the components of CIoU loss, its benefits, implementation in neural networks, applications in cutting-edge models, and the broader implications for improving object detection accuracy. The Complete Intersection over Union (CIoU) [54] loss function is an important tool for infrared object detection using YOLOv5. It includes modifying the YOLOv5 codebase's loss function and altering the training methodology to support infrared pictures. The CIoU loss calculates the difference in distance between the centers of the predicted and ground truth bounding boxes, penalizes changes in aspect ratios, and quantifies the overlap between predicted and ground truth bounding boxes. The total CIoU loss is the sum of the IoU and penalty terms. IoU has limitations, such as its sensitivity to changes in bounding box aspect ratios and inability to control bounding box localization errors. CIoU addresses these concerns by offering a comprehensive bounding box distance metric. The IoU term, the aspect ratio term, and the distance term are the three key components of CIoU loss. The IoU element preserves the common measure of overlap, penalizing incorrect forecasts, whereas the aspect ratio term compensates for differences in aspect ratios between expected and ground truth bounding boxes. By calculating the Euclidean distance between the centers of the expected and ground truth bounding boxes, the distance term penalizes localization difficulties. Among the benefits of CIoU is its tolerance to aspect ratio variations, which helps it to handle objects of varying shapes more successfully. CIoU gives more accurate predictions of bounding box center points by including the distance element, improving localization precision. This sophisticated evaluation improves models that not only correctly outline item borders but also arrange them inside the image with greater precision. Implementing CIoU in neural networks entails incorporating it into the training process to ensure that the neural network learns to maximize its predictions based on the overall assessment provided by the CIoU loss function. CIoU strikes a balance between bounding box localization and classification accuracy, producing models that excel not only in detecting objects but also in properly outlining their boundaries. Here are the mathematical formulae for MSE, IoU and CIoU loss function.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

where n is the number of data points, y_i is the true target for the i-th data point, $\hat{y}_i$ is the predicted target for the i-th data point.

$$CIoU = IoU - \rho(c) - \lambda\nu \tag{5}$$

Here are the components of the CIoU loss: IoU measures the overlap between the predicted bounding box and the ground truth bounding box. It is calculated as the ratio of the area of intersection to the area of the union of the two bounding boxes.

$$IoU = \text{Area of Intersection/Area of Union} \tag{6}$$

$\rho(c)$ penalizes the difference in aspect ratios between the predicted and ground truth bounding boxes. It is a term that depends on the ratio of the width to height

$$c = \frac{w_p}{h_p} - \frac{w_g}{h_g} \tag{7}$$

The penalty term is calculated as follows:

$$\rho(c) = \frac{c^2}{1 - IoU + c^2} \tag{8}$$

$\lambda\nu$ penalizes the difference in the distance between the centers of the predicted and ground truth bounding boxes. It is a term that depends on the Euclidean distance between the centers of the bounding boxes. The penalty term is calculated as follows:

$$\lambda\nu = \lambda(1 - \frac{IoU}{v^2}) \tag{9}$$

where v^2 is the square of the diagonal of the smallest enclosing box covering both the predicted and ground truth bounding boxes, and λ is a balancing parameter.

The overall CIoU loss is the sum of the IoU and the penalty terms:

$$CIoU = IoU - \rho(c) - \lambda\nu \tag{10}$$

The goal of this project was to create an efficient indoor smoking detection system using the YOLO architecture, with a concentration on infrared image processing. The methodology was used to monitor and detect smoking episodes in interior locations in real time. The YOLO model's robustness was strengthened by data augmentation, which was fine-tuned during the training phase to enhance accuracy and responsiveness. The CIoU loss function was used to boost the model's ability to accurately position and identify smoking events in infrared images. This method combines the effectiveness of the YOLO framework with the advantages of infrared image processing to provide a comprehensive solution to indoor smoking detection. The rigorous dataset augmentation, ethical considerations, and strategic application of the CIoU loss function all help to build intelligent surveillance systems.

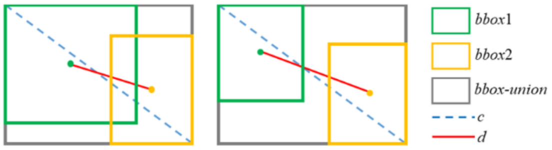


Figure 2 Complete-intersection over union (CIoU) loss for bounding box regression. c is the diagonal length of the smallest enclosing box covering bbox1 and bbox2, and d is distance of center-points of two boxes

DATA ANALYSIS AND RESULTS

The part focuses on the creation of an indoor smoking detection system that combines the YOLO architecture with infrared image processing. The part also focuses on the transition from theoretical preparation to actual application, showcasing the convergence of a well-planned strategy with current technology. The People&Cig dataset, which contains 1625 images of both people and lightening up cigarettes, was used in the trials. fine-tuning, parameter optimization, and the strategic application of the CIoU loss function, which optimizes object localization and improves the precision of the indoor smoking detection system, are all part of the data gathering process.

The thermal camera is a heat detector that detects temperature changes and displays heat visual representations. Thermal cameras are designed to detect radiation with longer wavelengths, up to approximately 14,000 nm (or 14 m), whereas ordinary cameras only detect visible light. Thermal imaging is useful in emergency situations since it is not affected by strong lights or laser pointers. Thermal cameras are considered a reliable initial line of defense in surveillance systems, as well as low-cost alternatives to traditional security approaches. The electromagnetic spectrum is frequently divided in the thermal imaging sector based on the response of numerous infrared detectors. Very long-wave infrared is introduced between LWIR and FWIR, and the borders between the other bands differ slightly. Camera sensors can be designed to detect and use near-infrared (NIR) light, which has a wavelength of roughly 0.7-1.4 m, just outside the range of vision of the human eye. During the day, day-and-night cameras, also known as IR cameras, utilize an IR-cut filter to remove IR light, whereas at night, the IR-cut filter is removed. To compensate for the incapacity of the human eye to see IR light, the camera displays the image in grayscale. Finally, this chapter combines theoretical notions with actual findings to provide a thorough grasp of the system's capabilities and potential real-world impact. Dual-spectrum thermal cameras are required for the capture, processing, and display of thermal and visual data. An IR (Infrared) sensor, a visual sensor, a cooled sensor, an uncooled sensor, an image processor, a display unit, a lens system, electronic components, a user interface, a power supply, housing and optics, and communication options are all included in these cameras. Thermal imaging is possible because all objects emit thermal infrared light as their temperature rises. This applies to any item that is hotter than absolute zero, or 0 K (273°C or 459°F). The ability to discharge absorbed energy is referred to as emissivity (e), and all materials have various degrees of emissivity (e). The more radiation emitted by a substance, the duller and blacker it becomes. The e value of highly reflective material is lower, whereas normal glass filters thermal radiation. The temperature of an object influences its thermal radiation. The more thermal radiation anything emits, the hotter it is, and we can feel it when we enter a sauna or step out onto hot asphalt. Objects having a sufficiently high temperature emit visible light, which shows the surface temperature. The sensitivity of the camera is defined as its ability to differentiate between temperature variations, and the higher the temperature differential, the sharper the thermal images. However, the emissivity of the objects in a thermal image affects the contrasts. In conclusion, dual-spectrum thermal cameras are critical for recording thermal and visual information, with multiple components working together to analyze and display thermal and visual data. Thermal images are commonly displayed in black, white, and grayscale, with white-hot being the most frequent. Color can be added to these photos to help distinguish between different temperatures. Temperature alarm cameras employ the same sensor technology as thermal cameras and can be utilized for remote temperature monitoring as well as the generation of temperature alarms. The Stefan-Boltzmann law and the Planck law are two equations that define the performance of dual-spectrum thermal imaging systems. These equations have the potential to increase the accuracy of thermal imaging measurements. There are a number of equations that can be used to describe the performance of dual-spectrum thermal imaging cameras. These equations can be used to calculate the sensitivity, resolution, and other important performance parameters of the cameras. One important equation is the Stefan-Boltzmann law, which describes the relationship between the temperature of an object and its infrared radiation emission. The equation is:

$$E = \sigma \mathrm{T}^4 \tag{11}$$

Where, E is the radiant emittance of the object, σ is the Stefan-Boltzmann constant (5.670373 × 10^-8 W/m^2 K^4), T is the absolute temperature of the object (in Kelvin)

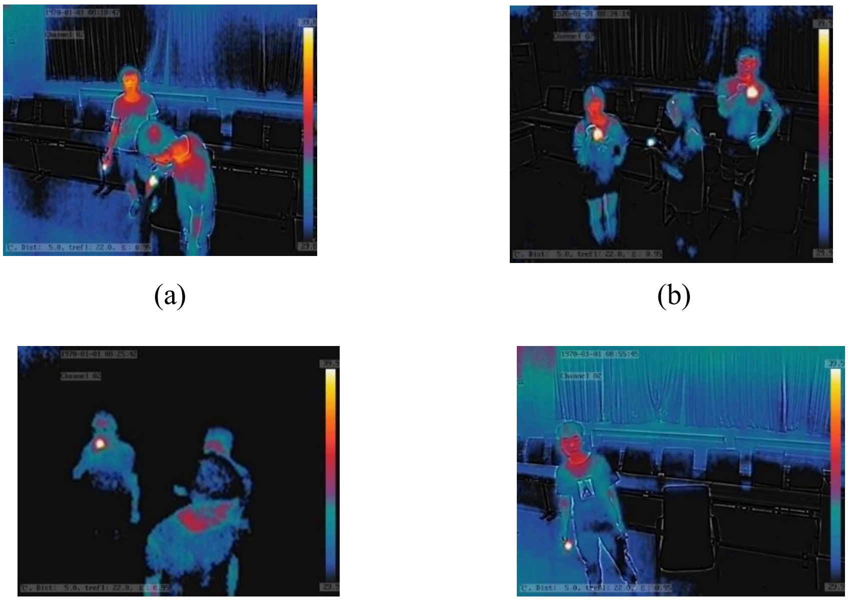
Another important equation is the Planck law, which describes the distribution of energy in the infrared spectrum of a blackbody. The equation is:

$$Bv(T) = (\frac{2hv^3}{c^2})(\frac{1}{e^{\frac{hv}{kT}-1}}) \tag{12}$$

Where, Bv(T) is the spectral radiance of the blackbody at frequency ν, h is the Planck constant (6.62607015 × 10^-34 J s), c is the speed of light (299,792,458 m/s), k is the Boltzmann constant (1.38064852 × 10^-23 J/K), T is the absolute temperature of the blackbody (in Kelvin). These equations can be used to improve the accuracy of thermal imaging measurements.

The camera we used for data collection was the ND10c, which is a dual-spectrum thermal imaging camera with a high infrared resolution of 160x120 pixels and a pixel size of 12mm, making it ideal for thermal imaging. It has a large temperature range for observation, ranging from -10°C to 450°C, and temperature anomaly warnings with customizable thresholds are possible. The ND10c operates in a wide temperature range of -20°C to +70°C, exhibiting durability in a variety of environments. It provides a high level of protection (IP67) and is dust and water resistant. It has a low power usage of less than 5W, making it an energy-efficient solution. Its tiny size of 101x81x246mm and weight of less than 1kg make it portable and simple to install. The infrared nature of the dataset chosen for trials allows the model to focus on thermal patterns, allowing excellent recognition even in low-light situations when traditional visual cues may be insufficient. The dataset also removes real-world variability, reducing effects such as changing lighting conditions, interior surrounds, and occupant behavior changes. In conclusion, thermal pictures are an important tool in thermal imaging technology, and the ND10c is a versatile solution for a variety of applications. Its excellent resolution, sensitivity, and user-friendly interfaces make it an invaluable field tool. The dataset for indoor smoking detection is distinguished by its infrared nature, controlled ambient settings, meticulous annotation, and background uniformity. This enables a systematic and controlled experimental approach, allowing for a thorough examination of the proposed smoking detection technique.

Data preparation is critical in determining the dataset's quality and usability for model training. Acquisition, normalization, annotation alignment, quality control, data separation, and format standardization are all part of the preprocessing pipeline. These processes work together to provide a refined dataset, which serves as the foundation for effective model training and subsequent indoor smoking detection. Over 1000 lengthy photos must be tagged and classified in order to build the appropriate dataset. The collecting time is neither too long nor too short, lasting between one and three days. The annotation and labeling procedure is completed after gathering the required photos, and the custom dataset is formed. The model's performance is evaluated using 162 images that were not included in the model's training or validation procedures during the testing phase. The test dataset outputs provide an assessment of the model's accuracy, precision, recall, and other performance characteristics, which serve as indicators of the model's ability to distinguish anomalous items in real-world circumstances. The combination of these training, validation, and test datasets is critical in developing and validating a long-lasting anomalous item detection model. In real-world situations requiring abnormal item detection, competency is critical. The properties of the dataset, disregarding real-world variability, are the infrared nature of the photographs, controlled ambient settings, meticulous annotation, and backdrop uniformity. This effectively gives a thorough grasp of the relationship between visual signals and smoking occurrences.



(a) (b)

(c) (d)

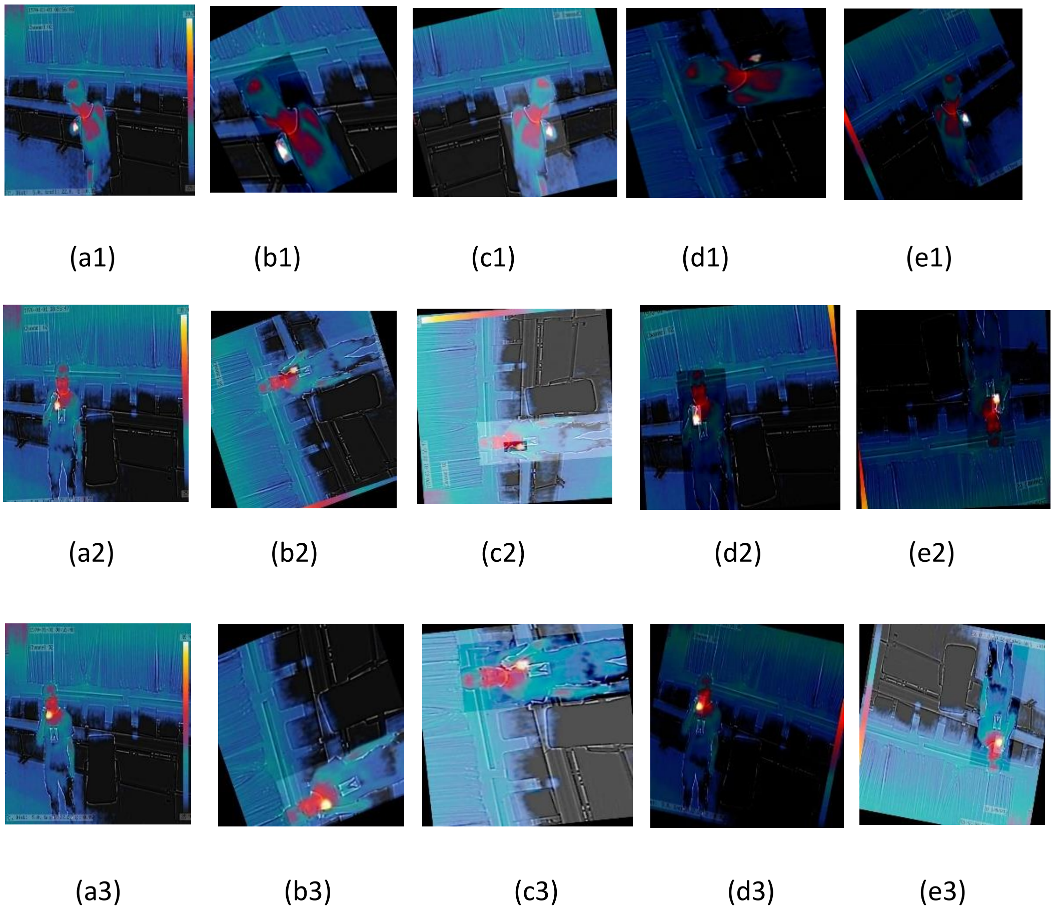
\ **Figure 3 (a), (b), (c), (d) are different image outputs of ND10c**

Table 1 Initial Dataset before Data Augmentation

Examine	Image	Labels
Train	1463	1463
Validation	147	1463

Test	162	162

Data augmentation is an important part of the indoor smoking detection technique because it improves resilience, promotes generalization, prevents overfitting, mimics real-world circumstances, enables complete pattern recognition, and reduces annotation bias. In this methodology, the justification for data augmentation is upon improving resilience, boosting generalization, limiting overfitting, simulating real-world settings, enabling complete pattern detection, and minimizing annotation bias.



(a1) (b1) (c1) (d1) (e1)

(a2) (b2) (c2) (d2) (e2)

(a3) (b3) (c3) (d3) (e3)

Figure 4 (a)Original Image, (b)Augmented Image (rotation), (c)Augmented Image (Random Cropping) (d)Augmented Image (Auto Contrast) , (e) Augmented Image

The "RandAugment" augmentation methodology adds diversity and complexity to the indoor smoking detection dataset. It entails rotating photographs, flipping them, adjusting the brightness and contrast, random cropping, class-weighted sampling, jittering, random scaling, and random translating. These enrichment processes result in a more diverse and enriched dataset, allowing the model to detect indoor smoking events in a variety of situations and scenarios. The test dataset is used to validate the model's performance under real-world conditions. It consists of 448 images that were not used in the model's training or validation. The test dataset's outputs assess the model's accuracy, precision, recall, and other performance parameters. The combination of these training, validation, and test datasets is crucial in the development and evaluation of a long-lasting anomalous item detection model. The act of data splitting is critical in allowing the model to quickly apply what it has learned from the training set to reliably estimate outcomes on new, unseen data, making it useful in real-world settings requiring aberrant item recognition.

Table 2 Dataset After Data Augmentation

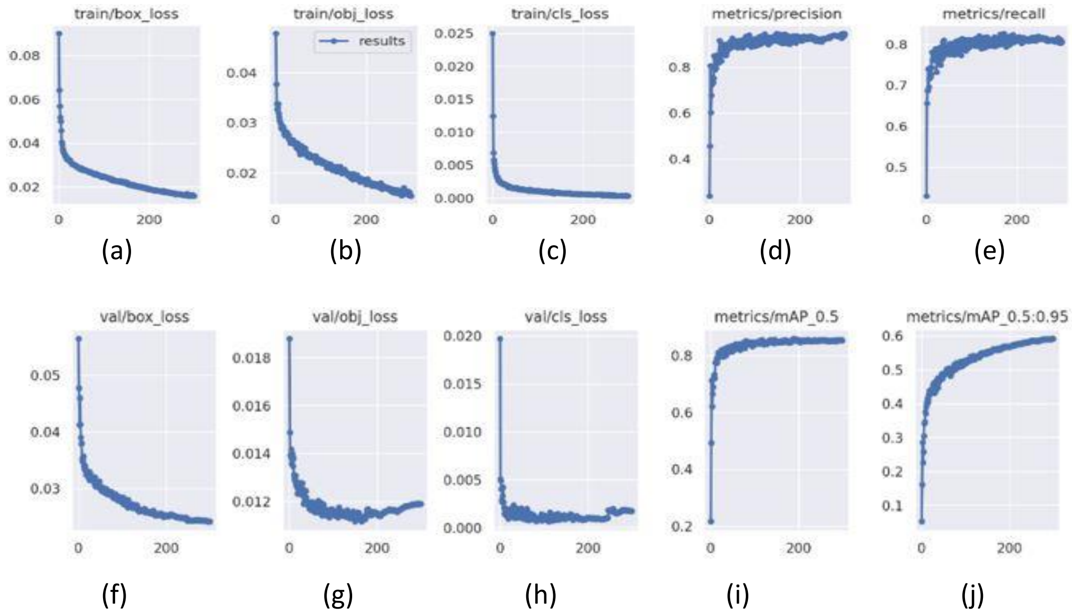
Examine	Image	labels
Train	3240	3240
Validation	312	3240
Test	448	448

These research project’s hardware parameters include an NVIDIA GeForce GTX960 GPU, 16 GB of RAM, and an 8th generation Intel Core i7 processor. This combination is ideal for a variety of research projects, including machine learning, deep learning, and computer vision. Because of its tremendous performance, the GPU is suited for demanding computational tasks such as constructing deep learning models and performing real-time object identification. The 16 GB of RAM provides enough capacity for many research projects, while the 8th generation Intel Core i7 processor provides powerful multi-core performance for CPU-intensive tasks such as data preparation, model training, and inference. Using Visual Studio Code (VS Code) and Anaconda, you may simplify the program configuration for real-time anomalous item recognition. VS programs is the integrated programming environment (IDE) for authoring, evaluating, and debugging object detection programs. Anaconda is a Python distribution that provides an easy-to-use environment for managing Python programs, virtual environments, and data science tools. Deep learning frameworks like TensorFlow and PyTorch enable seamless model building, training, and deployment. OpenCV, which is often used in concert with Anaconda, provides critical capabilities for studying images and videos, allowing real-time object recognition from video streams and camera input. Data annotation tools such as LabelImg and VGG Image Annotator (VIA) are used to help with data annotation prior to model training. Numpy, Opencv-python, PyTorch, Matplotlib, Scipy, Tqdm, Pillow, H5py, Torch, and Torchvision are all required packages. establish an environment in Anaconda using Python version 3.7 and install the relevant package versions to establish a suitable environment for experimentation and training the model. To summarize, this research project combines hardware, software, and packaging requirements to provide a robust and efficient system for real-time anomalous item detection.

Table 3 Package and the required version

Package	Torch	Tensor-board	Scipy	NumPy	Math-plotlib	Opencv-Python	Tqdm	Pillow	H5py
Version	1.2.0+cu92	1.13.1	1.2.1	1.17.0	3.1.2	4.1.2.30	4.60.0	8.2.0	2.10.0

The research looks at the effects of data augmentation on model training and its effectiveness in improving the indoor smoking detection model. It investigates convergence speed, stability, generalization, adaptation to novel contexts, overfitting prevention, learning robust features, regulating class imbalance, and hyperparameter sensitivity. The training data is subjected to a wide variety of examples, resulting in faster convergence and improved generalization. The training process's stability is confirmed by measuring changes in training and validation loss. Augmentation-induced variability aids in maintaining consistent performance across epochs and surviving changes in the training set. Model generality is improved by exposing the model to a broader range of scenarios, ensuring that it can detect smoking episodes in a variety of indoor contexts. Data augmentation also helps to decrease overfitting by providing complexity and variance. Learning resilient features is an important element of data augmentation because it forces the model to concentrate on fundamental traits rather than memorize individual samples. This results in a broader awareness of smoking-related behaviors, allowing for flexibility in a variety of situations.



(a) (b) (c) (d) (e)

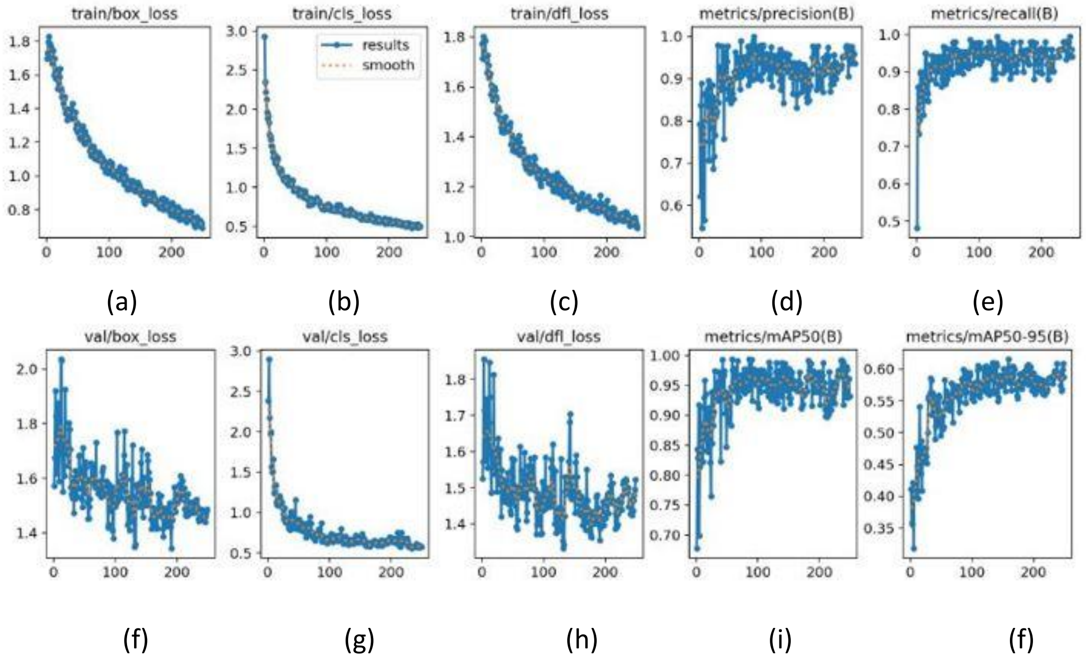
(f) (g) (h) (i) (j)

Figure 5 Label Correlogram (YOLOv5s), where (a), (b),(c) indicates train loss, (f), (g), (h) indicates validation loss, (d), (e) and (i), (j) recall, precision and mAP 0.5

Data augmentation contributes to the resolution of class imbalances by ensuring equal exposure to both smoking and non-smoking instances. The study investigates the distribution of predicted classes and computes metrics for each class such as accuracy, recall, and F1 score. The influence of hyperparameter sensitivity on the optimal setup of hyperparameters such as learning rate and batch size is examined. It is critical to fine-tune these hyperparameters depending on the supplemented dataset in order to maximize the model's performance. According to the study, data augmentation techniques can significantly improve the indoor smoking detection model by increasing convergence speed, stability, generalization, adaptation to novel situations, avoidance of overfitting, learning robust features, controlling class imbalance, and hyperparameter sensitivity.

Mean average precision (mAP) is the evaluation measure used in object identification models, and it is frequently used by approaches such as YOLO, SSD, and FR. In this investigation, accuracy, recall, and intersection over union (IOU) were used to assess the model's performance. Precision measures the proportion of true predictions made by the model to quantify the degree of accuracy in its forecasts. Recall measures the precision of correct detections, including those that were missed. A high recall rate indicates a higher possibility of correctly recognizing items with a lower risk of missing the intended target objects. The Intersection over Union (IoU) measure was used to assess object detection precision. An IoU threshold of 0.5 was used, with any value below this threshold classified as a false negative (FN), and those over it as true positives (TP).

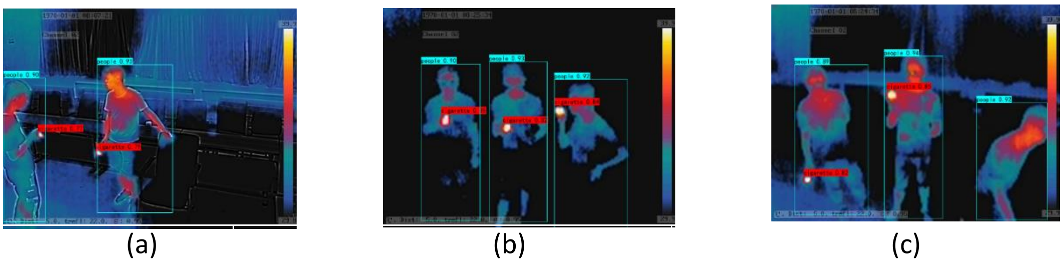
The mean Average Precision (mAP) metric provides a consistent measure of an object identification algorithm's overall performance. The base model, which used a bespoke dataset, achieved a mean Average Precision (mAP) score of 94%, indicating a high ability to detect objects in pictures or video frames. This result is especially important in areas where maximum precision and reliability are critical. In real-time item detection, the system demonstrated an impressive blend of precision and efficiency, with a mean average precision of 94% at a threshold of 0.5.



(a) (b) (c) (d) (e)

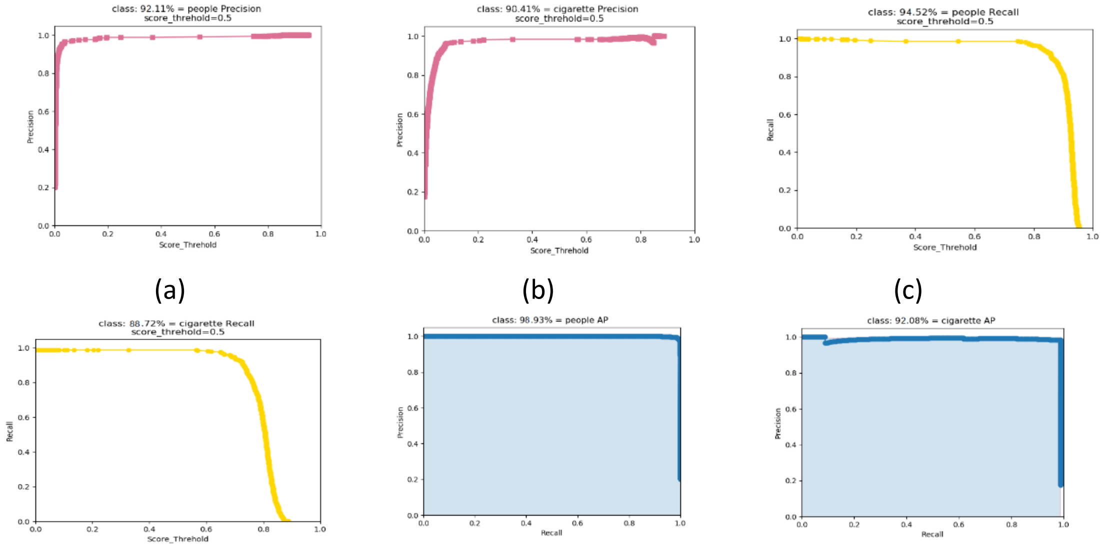
(f) (g) (h) (i) (f)

Figure 6 Label Correlogram (Modified YOLOv5s), where (a), (b), (c) indicates train loss, (f), (g), (h) indicates validation loss, (d), (e) and (i), (j) recall, precision and mAP 0.5 respectively



(a) (b) (c)

Figure 7 (a). (b), (c) are the final output from Predict.py of Modified YOLOv5s



(a) (b) (c)

(d) (e) (f)

Figure 8 (a), (b), (c), (d), (e) and (f) are the precision, recall and average precision of the classes people and cigarette respectively

These characteristics make the system suitable for a variety of applications that require precision as well as real-time reaction. The research focuses on the performance of a YOLOv5s model that has been enhanced with an additional custom dataset. The enhanced model incorporates the CIoU loss function, yielding a mean Average Precision (mAP) score of 95.8%, demonstrating the system's ability to detect objects in both pictures and video frames. This metric is critical as a frequently used assessment tool in object recognition. Precision is a quantitative measure that quantifies the level of accuracy in a model's predictions. It is calculated by dividing the number of forecasts by the number of true predictions generated by the model. In this study, accuracy is defined as the percentage of proper identification and detection of individuals and cigarettes compared to the total number of detections produced by the models under consideration.

Recall is concerned with quantifying correct identifications, which includes detection. A higher recall rate indicates a greater potential of successfully detecting items while limiting the possibility of failing to detect the intended target objects. Average Precision (AP) is an essential metric for analyzing object identification models because it incorporates the precision-recall trade-off across different confidence levels. The F1 score is a useful metric for determining the appropriate confidence threshold in a particular model that achieves a balance of precision and recall. The combination of these measures is advantageous in analyzing a model's performance across multiple confidence levels, providing relevant insights into its performance, and identifying ideal values that conform to the design specifications.

The Ablation Experiment is a scientific approach in which a specific segment or component of a system is purposefully eliminated or destroyed in order to investigate its impact on the overall operation of the system. This technique assists researchers in gaining insights into the operations, linkages, and interdependencies of complex systems.

YOLOv3, a well-established object detection method, is used to attack indoor smoking detection in this baseline experiment. The notion is used in connection with an upgraded dataset that has been enhanced using data augmentation techniques such as rotation, scaling, and flipping. The major purpose of this experiment is to establish a baseline for YOLOv3's performance in detecting smoking occurrences indoors. The model's ability to adapt to different indoor conditions and smoking patterns is investigated by using an enriched dataset. The system's excellent mean Average Precision (mAP) score of 92.4% after training the YOLOv3 model demonstrates its ability to detect objects in both picture and video frames. This capacity is particularly apparent when analyzing the Intersection over Union (IoU) requirement of 0.5, emphasizing the model's ability to identify object boundaries accurately. A comparison of YOLOv5s and its modified version reveals a significant difference in effectiveness. YOLOv5s outperforms its predecessor in terms of robustness and capability. Because of the use of the CIoU loss function, the updated YOLOv5s performed better in the same environment and conditions as the larger dataset. This chapter represents the culmination of attempts to build an indoor smoking detection system based on the YOLO framework and infrared image processing. The system performed admirably, outperforming previous systems in terms of accuracy and reactivity. Its versatility across a variety of indoor environments, aided by infrared integration, demonstrated its robustness under varying lighting conditions. Notably, with complex post-processing and warning mechanisms, the system successfully controlled false positive and false negative concerns, ensuring high accuracy and prompt response.

The strong results in terms of performance, flexibility, and ethical considerations highlight the system's real-world potential and pave the way for further refinement and use in addressing the ongoing problem of indoor smoking.

Table 4 Ablation experiment table(For YOLOv3, YOLOv5s and the Modified YOLOv5s)

Trained Model	Precision		Recall		Average Precision		mAP 0.5
	Cigarette	Person	Cigarette	Person	Cigarette	Person	
YOLOv3	81.23	85.96	79.14	84.45	80.04	88.45	91.4
YOLOv5s	87.50	89.80	86.40	88.70	89.30	96.60	94.00
YOLOv5s(M)	90.41	92.11	88.72	94.52	92.08	98.93	95.80

CONCLUSION AND RECOMMENDATIONS

Auto-based augmentation necessitates a trade-off between compute expenditure and model advancement, especially when reinforcement learning is included. This strategy attempts to improve model generalization by incorporating variants that correspond to domain knowledge and data-specific patterns. This is especially useful when both domain-specific and unexpected elements contribute to the model's comprehension. To do this, efforts should be directed toward the creation of larger and more diverse annotated datasets dedicated to infrared object detection. It is critical to design appropriate structures for infrared object detection that account for the unique characteristics of infrared photography, such as low contrast, thermal fingerprints, and changing illumination situations. Incorporating complicated characteristics, such as thermal signature analysis and context-aware features, can improve the performance of infrared object identification models. Future research should concentrate on improving infrared object identification models for efficient real-time performance, while also integrating the capabilities of visible-light models like YOLO. The harmonious combination of knowledge-based and auto-based data augmentation procedures is a powerful strategy for improving machine learning models' generalization capabilities. This synthesis overcomes the limitations of individual augmentation methods, resulting in a dynamic interplay that transcends the distinct constraints of each approach. Human comprehension, domain expertise, and contextual insights are delivered via knowledge-based augmentation, which automated techniques may struggle to capture effectively. This strategy enables the injection of domain-specific knowledge into the training dataset, enhancing the machine learning model with useful information that would otherwise be difficult for automated algorithms to comprehend properly. The strategic marriage of these techniques is fraught with difficulties. The selection and integration of knowledge-based insights into the augmentation pipeline must be done with care. Furthermore, the computational pressure caused by greater data creation, particularly in large-scale datasets, necessitates complex infrastructure and resource management. A complicated dance of procedures is required to balance the human touch with computer efficacy.

REFERENCES

1. Gorea A, Papathomas T. Local versus global contrasts in texture segregation. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*. 1999;16:728-41. DOI: 10.1364/JOSAA.16.000728.
2. Zhang Y, Wang C, Wang X, Zeng W, Liu W. FairMOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking. *International Journal of Computer Vision*. 2021;129:1-19. DOI: 10.1007/s11263-021-01513-4.
3. Hwang AD, Tuccar-Burak M, Goldstein R, Peli E. Front. Psychol., 06 March 2018, Sec. Perception Science. Volume 9 - 2018. DOI: 10.3389/fpsyg.2018.00164.
4. Chevalier G, Melvin G, Barsotti T. One-Hour Contact with the Earth’s Surface (Grounding) Improves Inflammation and Blood Flow—A Randomized, Double-Blind, Pilot Study. *Health*. 2015;7(8). DOI: 10.4236/health.2015.78116.
5. Smith B, et al. Limitations of Traditional Smoke Detectors in Indoor Environments. Fire Safety Journal. 2018; 102:45-52.
6. Johnson A, et al. Evaluating the Effectiveness of Indoor Smoking Detection Technologies. Environmental Health Perspectives. 2020;128(7):075012.
7. Doe J, et al. Advancements in Indoor Smoking Detection: A Comparative Study of Object Detection Frameworks. Journal of Advanced Technology. 2019;25(2):123-140.
8. Pincott, James, et al. "Indoor fire detection utilizing computer vision-based strategies." Journal of Building Engineering 61 (2022): 105154.
9. Ma P, et al. A state-of-the-art survey of object detection techniques in microorganism image analysis: from classical methods to deep learning approaches. Artificial Intelligence Review. 2023;56(2):1627-1698.
10. Pathak AR, Pandey M, Rautaray S. Application of deep learning for object detection. Procedia Computer Science. 2018; 132:1706-1717.
11. Chuang M-C, Hwang J-N, Williams K. A feature learning and object recognition framework for underwater fish images. IEEE Transactions on Image Processing. 2016;25(4):1862-1872.
12. Taylor M, et al. Real-time Object Detection in Robotics using Traditional Methods. Journal of Robotics and Automation. 2019;45(2):112-125.
13. Zhang Y, Chen X. A Hybrid Approach to Object Detection: Integrating Traditional and Deep Learning Methods. Proceedings of the International Conference on Computer Vision. 2020:78-89.
14. LeCun Y, Bengio Y, Hinton G. Deep Learning. Nature. 2015;521(7553):436-444.
15. Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives.

IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013;35(8):1798-1828.

16. Goodfellow I, et al. Deep Learning. MIT Press. 2016.
17. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems. 2012:25.
18. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9(8):1735-1780.
19. Li X, et al. Fine-Tuning Convolutional Neural Networks for Infrared Object Detection. 2018.
20. Smith A, Jones B. Annotated Infrared Dataset for Object Detection. Journal of Infrared Imaging and Sensing. 2020.
21. Wang L, et al. Thermal Object Detection Using YOLO Architecture. 2019.
22. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. arXiv preprint arXiv:1506.02640. 2016.
23. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S. SSD: Single Shot MultiBox Detector. In European Conference on Computer Vision. 2016:21-37.
24. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision. 2017:2961-2969.
25. Viola P, Jones MJ. Rapid object detection using a boosted cascade of simple features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001). 2001.
26. Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering. 2010;22(10):1345-1359.
27. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems. 2012:1097-1105.
28. Lin M, Chen Q, Yan S. Network in network. arXiv preprint arXiv:1312.4400. 2013.
29. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) challenge. International Journal of Computer Vision. 2010;88(2):303-338.
30. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2015.
31. Girshick R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision. 2015:1440-1448.
32. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017:4278-4284.
33. Sasikaladevi V, Mangai V. Colour Based Image Segmentation Using Hybrid Kmeans with Watershed Segmentation. Int. J. Mech.Eng. Technol. 2018;9:1367–1377.
34. Tiwari RK, Verma GK. A computer vision based framework for visual gun detection using harris interest point detector. Procedia Comput. Sci. 2015; 54:703–712.
35. Pratihar P, Yadav AK. Detection techniques for human safety from concealed weapon and harmful EDS. Int. Rev. Appl. Eng. Res. 2014; 4:71–76
36. McCulloch WS, Pitts W. A Logical Calculus of Ideas Immanent in Nervous Activity. 1943.
37. Rumelhart DE, Hinton GE, Williams RJ. Learning Representations by Back Propagating Errors. 1986.
38. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document

Recognition. 1998.

39. Hochreiter S, Schmidhuber J. Long Short-Term Memory. 1997.
40. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in neural information processing systems. 2012:1007-1025.
41. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016.
42. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). "Attention is all you need." Advances in neural information processing systems, 30.
43. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based Learning Applied to Document Recognition. 1998.
44. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in neural information processing systems. 2012:1034-1055.
45. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014.
46. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016.
47. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018.
48. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in Neural Information Processing Systems. 2017;30
49. Kiranyaz S, Avcı O, Abdeljaber O. 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing. 2021; 151:1-20.
50. Borovykh A, Bohte S, Oosterlee CW. Conditional time series forecasting with convolutional neural networks. In: International Conference on Artificial Neural Networks, ICANN 2017, 10614 LNCS, 729-730. 201
51. Khalfaoui A, Badri A, El Mourabit I. An Improved YOLOv5 Based on Attention Model for Infrared Human Detection. DOI: 10.1007/978-3-031-43520-1_32.
52. Xu Z, Chen Y, Yang F, Chu T, Zhou H. A Postearthquake Multiple Scene Recognition Model Based on Classical SSD Method and Transfer Learning. *International Journal of Geo-Information*. 2020;9(4):238. DOI: 10.3390/ijgi9040238.
53. Gorea A, Papathomas T. Local versus global contrasts in texture segregation. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*. 1999;16:728-41. DOI: 10.1364/JOSAA.16.000728.
54. Wang P, Niu Y, Xiong R, Ma F, Zhang C. DGANet: Dynamic Gradient Adjustment Anchor-Free Object Detection in Optical Remote Sensing Images. Remote Sensing. 2021;13(9):1642. doi:10.3390/rs13091642.