Pertanika J. Sci. & Technol. 33 (6): 2797 - 2832 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

RGB vs. HSV for Kitchen Fire Detection with YOLOv5

Norisza Dalila Ismail, Rizauddin Ramli* and Mohd Nizam Ab Rahman

Department of Mechanical and Manufacturing Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ABSTRACT

Kitchen fires pose a significant challenge and threat to people and the environment. Prompt response and accurate classification of fire occurrences are crucial to ensure safety and reduce potential property damage. This study addresses the need for effective fire detection technologies by evaluating the performance of the You Only Look Once version 5 medium (YOLOv5m) model using both hue, saturation, and value (HSV) and visible light (red, green, and blue [RGB]) color spaces. To reduce false positives, the experiment was expanded by including background images in the training dataset. Two kitchen fire datasets, one in RGB and one in HSV, were used to train and evaluate the model. Based on the results, HSV color space offers higher recall and precision for fire-on-pan detection, achieving 0.882 and 0.931, outperforming RGB. The best overall mean performance was in the first experiment, RGB without background images, resulting in the highest mean average precision (mAP)@0.5:0.95 score of 0.651. This performance was limited by lower recall and precision compared to HSV in specific fire scenarios. A key limitation of this study is its focus on kitchen environments, for which the findings may not be directly generalizable to other fire scenarios with different environmental conditions or fire characteristics. Furthermore, the study is limited to the YOLOv5m architecture, where other detection models may yield different results. In terms of kitchen fire detection, this study provides a comprehensive comparison of the RGB and HSV color spaces, offering insights into their benefits and drawbacks. This research shows that the HSV color space is useful in certain fire detection scenarios, and that combining the two color spaces yields an improved detection model for real-time applications.

Keywords: Deep learning, HSV, kitchen fire detection, object detection, RGB, YOLOv5

ARTICLE INFO

Article history:
Received: 06 December 2024
Accepted: 03 July 2025
Published: 29 October 2025

DOI: https://doi.org/10.47836/pjst.33.6.17

E-mail addresses:
i.norisza@gmail.com (Norisza Dalila Ismail)
rizauddin@ukm.edu.my (Rizauddin Ramli)
mnizam@ukm.edu.my (Mohd Nizam Ab Rahman)
* Corresponding author

INTRODUCTION

A fire starting from the kitchen extended farther than the cooking space and could easily burn combustibles kept in other areas of the house. Flashover may occur if the fire

e-ISSN: 2231-8526

spreads swiftly. The appearance of flashover, which is a crucial factor in catastrophic fires, is typically interpreted as a sign of untenable conditions (Rasbash, 1991). The fire will spread more quickly and could endanger the entire building. This highlights the critical need for more sustainable and effective methods to prevent fires. Consequently, to address the fire risk issue, an innovative fire detection system must be developed.

Fire detection systems are designed to detect fires early, providing sufficient time to extinguish them and ensure the safe evacuation of residents. Early detection may protect the safety of emergency response personnel, reduce property losses, and minimize operational downtime since control measures are initiated (Graham & Roberts, 2000). Existing fire detection systems can include detectors for smoke, heat, or flame, which are connected to a fire alarm system that triggers an alarm to alert people to potential danger.

However, these systems have difficulties in accurately detecting fires in dynamic environments like kitchens, where smoke and steam are frequent byproducts of cooking activities (Choi et al., 2023; Milke & Zevotek, 2016). Due to the complex conditions in kitchen settings, such as variable lighting, reflective surfaces, and background clutter, it is becoming increasingly challenging for conventional sensors to differentiate between real fire occurrences. Therefore, there is a pressing need for more advanced and reliable fire detection systems capable of detecting genuine fire events early in typical kitchen environments. Transitions of various innovative methods have been introduced to improve fire detection systems, such as vision-based fire detection (Bu & Gharajeh, 2019; Dilshad et al., 2023). Vision-based systems use image texture and color to identify flames. RGB and HSV are two of the most used color spaces in fire detection because they facilitate the easy distinction between fire and other elements in a kitchen environment. Giwa and Benkrid (2024) suggested systems vary in complexity; to recognize distinguishing fire characteristics in a video series, including color, velocity, temporal variations, spatial variations, and dynamic texture, they use a variety of image processing, machine learning techniques, and convolutional neural network (CNN) structures.

The RGB color space is mostly used in fire detection, which presents a significant issue because it can be extremely sensitive to changes in illumination and reflections, resulting in false alarms (Bakri et al., 2018). On the other hand, the HSV color space is more resilient to changes in illumination since it distinguishes between color intensity and chromatic content. However, it is not completely error-proof (Mutar, 2019).

In machine learning, object detection is the process of identifying and categorizing objects that meet predetermined criteria. Examples include drawing boxes around the objects and labeling them with their assigned class. Modern object identification techniques like You Only Look Once (YOLO) have drawn much attention lately because of their dependability, speed, and accuracy (Redmon et al., 2015). Numerous studies and real-world applications have consistently demonstrated the efficiency of this novel approach, which has shown great promise across a range of applications.

YOLO was built on a CNN that predicts several bounding boxes at once and assigns class probabilities to each box in an image using a single neural network and a single forward propagation process. The evolution of YOLO architecture with the development of numerous iterations, including You Only Look Once version 1 (YOLOv1) to You Only Look Once version 11 (YOLOv11), which are YOLOv1 (Redmon et al., 2015), You Only Look Once version 2 (YOLOv2, Redmon & Farhadi, 2017), You Only Look Once version 3 (YOLOv3, Redmon & Farhadi, 2018), You Only Look Once version 4 (YOLOv4, Bochkovskiy et al., 2020), You Only Look Once version 5 (YOLOv5, Jocher, n.d.b), You Only Look Once version 6 (YOLOv6, Li et al., 2022), You Only Look Once version 7 (YOLOv7, C.-Y. Wang et al., 2022), You Only Look Once version 8 (YOLOv8, Jocher, n.d.a), You Only Look Once version 9 (YOLOv9, C.-Y. Wang et al., 2024), You Only Look Once version 10 (YOLOv10, A. Wang et al., 2024), and the latest in the year 2024 is YOLOv11 from Ultralytics. This version represents a significant improvement over the previous versions. Still, the official version exhibits some instability due to network-related updates. In this study, the YOLOv5 version (Jocher, n.d.b) was selected due to its network structure, which is more efficient than that of YOLOv8 (Geng et al., 2024), resulting in a smaller model size and improved deployment and operating efficiency. Furthermore, YOLOv5 enhances object detection performance by employing multiple optimization techniques and a less computationally intensive backbone, making it an excellent choice for real-time applications (Geng et al., 2024). With a pre-trained model and an intuitive application programming interface (API), YOLOv5 is a convenient and user-friendly choice.

Existing fire detection methods often struggle with the diverse and dynamic conditions prevalent in kitchen environments, including variations in lighting, the presence of steam and smoke from cooking, and the wide range of background objects. This paper aims to address these limitations by investigating the efficacy of a YOLOv5-based fire detection system that uses both RGB and HSV color spaces for early kitchen fire detection. Specifically, the primary objective of this study is to evaluate the performance of RGB and HSV color spaces when integrated with the YOLOv5 object detection model for detecting early-stage kitchen fires.

This work focuses on detecting visible kitchen fires in images taken in realistic kitchen environments. Although the initial plan included addressing the problem of early fire detection, the available dataset obtained so far contains mainly active fire images. For this reason, pre-flame states, including those for smoke-only and smoldering conditions, are not considered in this paper. The comparative study presented in this article focuses on the impact of color space representations (RGB versus HSV) on YOLOv5m's object detection performance in cluttered kitchen environments.

Two distinct datasets were created: one containing images in the RGB color space and the other in the HSV color space. This study systematically compares the performance of these two color spaces within the YOLOv5 framework to identify an optimal approach for accurately distinguishing fire from non-fire events in real-time kitchen scenarios.

RELATED WORKS

YOLOv5 has achieved significant gains in both accuracy and speed for real-time applications, and numerous works have focused on improving its architecture for fire and smoke detection. However, most of these models operate primarily in the RGB color space, without assessing how color representation may influence detection performance under challenging indoor environments, such as kitchens. Table 1 shows advances in object detection models, particularly in architectural improvements for vision-based fire detection. These include integrated channel attention and transformer mechanisms from Liang et al. (2023), while Geng et al. (2024) presented YOLOFM, which integrates the FocalNext block, a quantization-aware hybrid attention-reparameterized feature pyramid network (QAHARep-FPN), and the Focal-scaled intersection over union (SIoU) loss function.

Another project has been developed, which involved You Only Look Once version 5 small (YOLOv5s) backbone replacement with EfficientNetV2 and introduction of a Smooth L1 loss function (S. Wang & Wang, 2023). Enhancement of the YOLOv5 framework by D. Zhang and Chen (2024), which incorporates the FasterNet network into its backbone, while the neck network uses Ghost-Shuffle Convolution (GSConv). They also introduced a one-time aggregation cross-stage partial network module, a one-shot aggregation network-based ghost shuffle cross-stage partial (VoV-GSCP) in the modified model.

A modified version of the YOLOv5m with real-time auditory notifications was introduced by Abdusalomov et al. (2022), which was designed to assist visually impaired individuals, especially during a fire. Kobeissi and Boulos (2023) explored the use of a combination of YOLOv5 and MiDaS depth estimators to enhance indoor fire detection.

However, these enhancements often remain bound to the RGB color domain, potentially overlooking the benefits offered by alternative color spaces. There is a distinct dearth of targeted research addressing fire detection specifically inside kitchen contexts, despite strong evidence suggesting that a significant number of fires originate in kitchens (Chow & Liu, 2016; J. Liu et al., 2019; Ngai et al., 2022; Spearpoint & Hopkin, 2020; Zulkiffli, 2018). Although a great deal of research has been conducted on fire detection devices, relatively little has focused on the characteristics and challenges of kitchen fires. This gap in the literature underscores the need for further research into fire detection techniques specifically designed for kitchens to enhance early warning systems in high-risk environments.

Table 1
Summary of related work

Authors	Models' enhancements	Dataset	Color space used	Performance metrics	Findings	
D. Zhang and Chen (2024)		The backbone network integrates YOLOv5 with the FasterNet network. The neck network uses GSConv and VoV-GSCSP.	Public dataset	RGB	Average accuracy = 98.3%, reduction in memory usage = 31.4%, increase in detection speed = 13%.	The model is lightweight, maintaining accuracy. Reduce model memory usage, model parameters, and computational costs. Improve detection speed.
Geng et al. (2024)	Improved YOLOv5n with QARepNeXt for feature extraction, Integration of QAHARep-FPN, and Focal-SIoU loss function.	FM-VOC Dataset18644	RGB	mAP50-95 = 7.9% improvement.	Improved feature extraction for better fire and smoke detection.	
Liang et al. (2023)	Channel attention and transformer in YOLOv5s.	Self-built internet fire image	RGB	14% mAP improvement and 68 FPS.	Enhanced detection speed and accuracy for fire and smoke.	
S. Wang and Wang (2023)	YOLOv5s backbone replaced with EfficientNetV2 and SIoU loss.	PASCAL VOC2007	RGB	mAP = 20% improvement, 15% recall improvement.	Lightweight and real-time fire detection with improved accuracy.	
Kobeissi and Boulos (2023)	YOLOv5 and MiDaS depth-mapper, depth estimator.	Fire and smoke dataset 2) Fire and gun on Kaggle	RGB + Depth	98.26% accuracy was achieved in the larger MiDaS version.	Locating and detecting indoor fires with the use of a single camera.	

Table 1 *(continue)*

Authors	Models' enhancements	Dataset	Color space used	Performance metrics	Findings
Abdusalomov et al. (2022)	YOLOv5m with real-time auditory notifications for the visually impaired.	Publicly available image datasets and Google images	RGB	High-speed, high-accuracy fire detection in indoor scenes.	Real-time fire detection for indoor environments.

Note. YOLOv5 = You Only Look Once version 5; GSConv = Ghost-shuffle convolution; VoV-GSCSP = One-shot aggregation network-based ghost shuffle cross stage partial; RGB = Red, green, and blue; YOLOv5n = You Only Look Once version 5 nano; QARepNeXt = Quantization-aware re-parameterized network with aggregated transformations; QAHARep-FPN = Quantization-aware hybrid attention re-parameterized feature pyramid network; Focal-SIoU = Focal scaled intersection over union loss function; FM-VOC Dataset 18644 = Fire and smoke visual object classes dataset with 18,644 images; mAP50-95 = Mean average precision at IoU thresholds from 0.50 to 0.95; YOLOv5s = You Only Look Once version 5 small; mAP = Mean average precision; FPS = Frame per second; EfficientNetV2 = Efficient convolutional neural network version 2; SIoU = Scaled intersection over union; PASCAL VOC2007 = Pattern analysis, statistical modelling and computational learning – Visual object classes challenge 2007; MiDaS = Mixed depth and scale

In this study, three significant key contributions are presented:

1. A comparative analysis of RGB and HSV color spaces specifically for kitchen fire detection, an area that has received limited attention.
2. The use of YOLOv5m, a state-of-the-art object detection model, for this task, which offers a balance between speed and accuracy.
3. An evaluation of the impact of background complexity on fire detection performance, providing insights into the robustness of the proposed approach in real-world scenarios.

Although previous work has examined RGB images and various deep learning models for fire detection, little research has investigated discriminative analysis between the RGB and HSV color spaces within a YOLOv5 architecture. To the best of our knowledge, there is no systematic work evaluating RGB vs. HSV color spaces for fire detection in kitchens using YOLOv5m, particularly about background variation and clutter, a crucial issue in real-world kitchen environments.

This study addresses these limitations by evaluating and comparing the performance of RGB and HSV for the YOLOv5m object detection model in kitchen fire scenarios. With this motivation, the research is directed by the following questions:

RQ1: How does the HSV color space perform with respect to detection (i.e., precision, recall, and mAP) in comparison with the RGB color space when YOLOv5m is used for kitchen fire detection?

RQ2: What is the impact of including realistic kitchen background images in the training dataset on the detection performance and the number of false positives of a YOLOv5m-based kitchen fire detection system?

These questions were explored through a set of experiments on RGB and HSV datasets, with identical model settings for fair comparisons.

PROPOSED METHODS

Research Overview

The suggested method for creating an indoor kitchen fire detection model is shown in Figure 1. A thorough method for creating a fire detection model is described in the suggested flowchart. The process starts with image database acquisition, followed by data pre-processing, which involves several steps such as data cleaning, cropping, resizing, removing background, and annotation. To verify the model's applicability, the pre-processed data is then split into three datasets, which are training, validation, and testing. Using the training and validation datasets, respectively, the model is trained and validated during the training phase. Next, the testing dataset is used to evaluate the refined fire detection model's performance on an unknown dataset. The model's recall rate, precision rate, mAP@0.5, and mAP@0.5-95 are all evaluated as part of the evaluation process. With the potential to significantly reduce environmental fires, this suggested methodology offers a methodical and efficient approach to building kitchen fire-detection models.

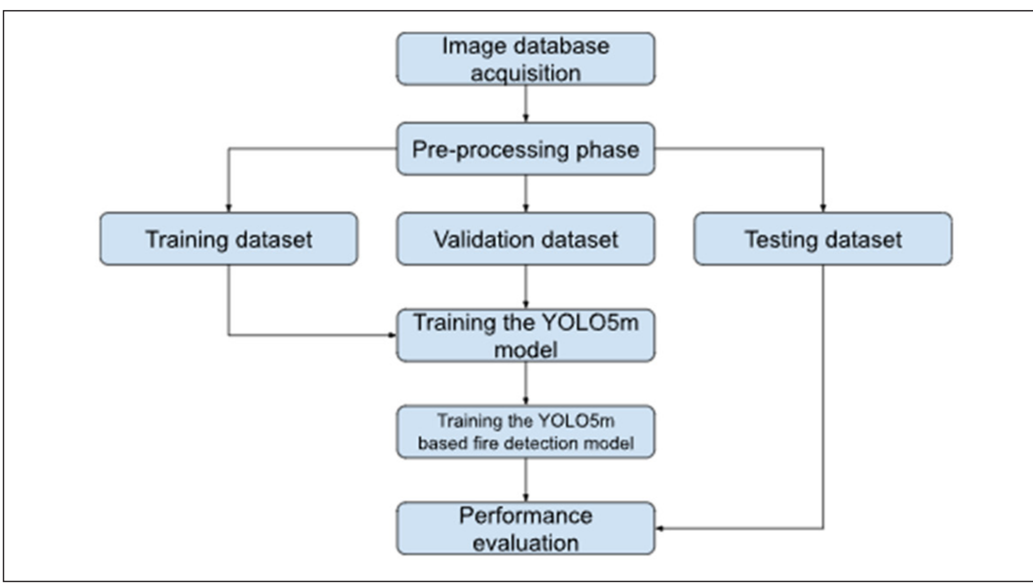

Figure 1. Flowchart of the proposed method
Note. YOLO5m = You Only Look Once version 5 medium

Object Detection Model

One object detection algorithm technique that is popular for its reliability, simplicity, and accuracy (Johnston et al., 2022). It is called YOLOv5. The YOLOv5 (Jocher, n.d.b) was released by Ultralytics and consists of three main components: the backbone, neck, and head.

The backbone depends on the CSP-Darknet53 convolutional network and employs the cross-stage partial (CSP) strategy to expedite information flow while alleviating issues linked to redundant and vanishing gradients (S. Liu et al., 2022; Wu et al., 2023). A version of the spatial pyramid pooling (SPP) is incorporated into the neck of the YOLOv5 model, and the Bottle-NeckCSP is integrated into the Path Aggregation Network (PANet) (Jocher, n.d.a). A combination of these techniques enhances the receptive field, allowing the network to retain speed while separating important context features. The CSPNet strategy (Wu et al., 2023) improved PANet in YOLOv4 as a feature pyramid network, providing better pixel-level localization accuracy in YOLOv5. The neck of architecture is crucial for managing object scaling and enabling the model to perform remarkably well on unseen data.

The head of YOLOv5 consists of three convolutional layers, similar to its predecessors (Wu et al., 2023). These layers, which differ slightly from the preceding versions in the computation of target coordinates for bounding boxes, predict the coordinates of bounding boxes, score, and object classification (Kim et al., 2023; Wu et al., 2023).

According to Johnston et al. (2022), there are four different levels of neural network models in YOLOv5, from simplest to most complex including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. With the help of an effective inference strategy and a lightweight network architecture, YOLOv5s is a single-stage detector that enables quick and precise object recognition (Vats & Anastasiu, 2023). Besides, YOLOv5l and YOLOv5x offer higher accuracy due to their deeper and wider architecture, but require significantly more computational resources, which can be a limitation for deployment on hardware with constrained processing power. YOLOv5m represents a balanced compromise between these extremes, offering improved detection precision over YOLOv5s while maintaining a reasonable inference speed and manageable resource consumption. Given the demand for quick, medium-level performance, and high precision without overwhelming computational cost, YOLOv5m is selected as the model for this paper.

Performance Metrics

In this study, mAP, precision, recall, F1-score, and latency are selected as the primary evaluation metrics to assess the fire detection model's performance. These metrics are widely used in object detection tasks because they capture both localization and classification performance. Precision indicates the proportion of correctly identified fire instances among all instances predicted as fire, while recall measures the model's ability to detect all actual fire instances. mAP, particularly at intersection over union (IoU) thresholds such as mAP@0.5 and mAP@0.5:0.95, provides a comprehensive summary by integrating precision and recall across multiple confidence levels and IoU thresholds. F1-score, the harmonic means of precision and recall, balances these two metrics, offering a single measure of a model's overall detection performance.

Since this study is limited to fire as a class, the computed mAP approximates the average precision (AP) across all fires. To compute the AP, the IoU must be provided first. The overlap between the ground-truth and predicted bounding boxes is measured using a metric called IoU. It is computed as the ratio of the intersection area to the union area of the two bounding boxes. This metric is frequently used to evaluate object recognition models, as it assesses how well the model can locate and identify objects in an image. A threshold value is compared with the IoU value for the anticipated and ground-truth bounding boxes to determine whether a detection is correct or incorrect. This study selected a threshold value of 0.5, which is commonly used in evaluating object detection models.

The link between the IoU value and the threshold value must be considered to assess an object detection model's performance properly. The terms listed below are introduced specifically to explain this relationship: For an IoU of 0.5 or greater, the results are considered true positives (TP); meanwhile, an IoU below 0.5 is classified as a false positive (FP), and a false negative (FN). TP denotes a ground truth bounding box that has been correctly detected, while FP denotes an item that has been incorrectly detected. FN denotes that the model cannot identify an object within the image. Since a true negative (TN) implies that there is no targeted object in the detected image, it was not considered while calculating performance (Padilla et al., 2020).

TNs are a crucial component of a model's performance, but they are usually excluded from the mAP computation for object detection since mAP is intended to assess the model's object recognition capabilities rather than its detection accuracy. Thus, TN is not relevant for this study, given that every fire image has previously been labelled during the pre-processing stage. Consequently, the evaluation performance will involve TP, FP, and FN to calculate the model's accuracy and reliability using precision (P) and recall (R). P will calculate the proportion of correct predictions to the model's accuracy in the detection process. The ratio of accurate positive forecasts to all ground facts is known as R. The P and R formulas, as indicated by Equations 1 and 2, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \qquad [1]$$

$$Recall = \frac{TP}{TP + FN} \qquad [2]$$

A precision-recall curve was used to identify the optimal threshold that maximizes both metrics. This curve is a valuable gadget for enhancing the performance of a detection model. The metric of AP in Equation 3 is a summary of the precision-recall curve to indicate the mean of all precisions. The weighted mean of precisions at each threshold, which represents the increase in recall in comparison to the previous threshold, is used to calculate the AP.

$$\mathrm{AP} = \sum^{k=i=1} [R\ (\mathrm{k}) - R\ (\mathrm{k}+1)\ ] \times P\ (\mathrm{k}) \qquad [3]$$

where i = number of thresholds, R = recall, k = index, and P = precision.

The mAP is formulated by averaging the AP across all classes, as shown in Equation 4.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} APk \qquad [4]$$

where n = number of classes, k = index, and APk = the Average Precision of class k.

For this study, the mAP formulation is based on a single class: kitchen fires. There are two variations of metrics obtained from the base formula of mAP, also known as mAP@50 and mAP50-95.

However, these metrics also have limitations. Precision and recall do not account for the spatial accuracy of the bounding boxes, which can be critical in real-time fire detection scenarios. Additionally, mAP may not fully reflect the model's performance in safety-critical situations, where false negatives (missed fires) carry more severe consequences than false positives. While F1-score provides a better balance between precision and recall, it does not address how well the bounding boxes align with the actual fire, which can be crucial for accurate fire detection. Latency, or the time taken for the model to process an image and make a prediction, is another important metric for real-time applications, ensuring that the model can provide timely alerts in dynamic environments like kitchens. Despite these limitations, these metrics remain the standard in object detection research, providing a meaningful basis for comparative evaluation with existing models.

Dataset Acquisition

The kitchen fire data employed in this study are the original ones developed by the authors. Real video clips of kitchen fire scenarios were filmed in controlled conditions. Next, OpenCV was used to split the videos into still image frames of size 1920×1080 pixels and to filter out poor-quality or simply blurred frames. The ultimate dataset is in two versions, one in RGB and another in HSV. The HSV dataset was built by converting each RGB image to HSV using OpenCV's cv2. cvtColor function. The dataset in this work contains images of visible kitchen fires, including flames on the stove and a pan under diverse lighting and background conditions. Thus, the focus of this work is on detecting flames during their visible state. Our experimental setup aims to study fire categorization in a kitchen setting, utilizing two different color spaces: RGB and HSV. A small, portable kitchen burner is used in the experiment as the primary source of fire, and data is collected from six different locations using two cameras: an iPhone X and a Samsung Galaxy Note 10 Lite. These smartphone models were deliberately chosen to reflect a realistic range of image capture devices available to everyday users. The iPhone X represents a high-end iOS device, while the Samsung Galaxy Note 10 Lite represents a mid-range Android device, allowing us to incorporate variations in hardware specifications, image processing algorithms, and camera sensors. This diversity helps enhance the generalizability and robustness of our vision-based fire detection model. These locations were carefully selected to ensure that our vision-based fire detection model would be robust, capturing thorough visual data from multiple perspectives.

The camera positioning is as follows: for the first round of recording, two cameras are placed at 45-degree angles relative to the stove's front view. Next is the placement of those two cameras, with one positioned directly in front and the other from a side view. These positions enable the capture of variations in fire dynamics and behavior, such as flame spread and intensity, that may occur during cooking activities. Furthermore, after the recording ended in the second round, the cameras were mounted at a top-down perspective, positioned 75 cm above the kitchen stove. This overhead view offers a crucial insight into the vertical spread of flames, which is vital for detecting fire patterns and understanding smoke evolution. The distance of the front camera from the stove is set at 50 cm, providing a clear, focused view of the fire's origin and progression.

The specific angles and camera arrangement for the experiment are shown in Figure 2. These camera placements are carefully configured to capture various angles and ensure comprehensive data collection, thereby improving the accuracy of fire detection systems. Due to the positioning of each camera, the experiment can examine flame properties such as color intensity and saturation, which are crucial for differentiating between fire and non-fire situations in actual kitchen settings. The experiment can achieve this by utilizing both RGB and HSV color spaces in combination.

By combining various viewpoints and color space analysis, this experimental design aims to minimize false positives while enhancing early fire detection and classification models in the kitchen context.

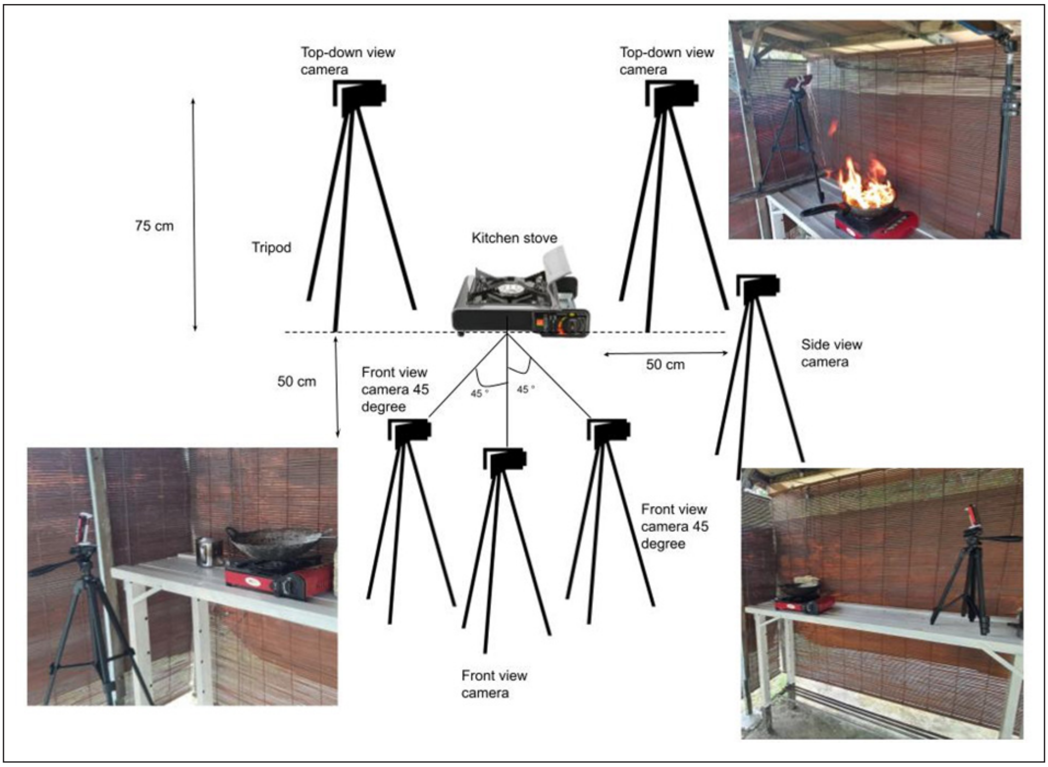

Figure 2. Camera positions in the experimental setup for fire detection

The images were taken from the six camera positions, as shown in Figures 3 and 4. In this experimental setup, the classification of two main types of fire: *stove flame* and *pan fire*. This classification is essential for differentiating between typical cooking flames, like those from a stove burner, and real fire incidents, like when a fire breaks out of a contained area and ignites objects like a pan. To improve the precision of fire detection in kitchen settings, the dataset produced by this experiment will concentrate on these two scenarios.

The term *stove flame* classification describes how a kitchen stove operates typically, with the flames contained within the burner and utilized for cooking. This class is crucial for determining safe cooking practices and lowering false alarms in fire detection systems. The *fire-in-the-pan* categorization, on the other hand, describes circumstances in which flames spread onto the pan, indicating a possible fire hazard. The identification of early fire occurrences that could result in uncontrolled kitchen fires depends heavily on this classification.

By capturing images from varying perspectives, from front, side, and top-down views, the dataset will provide comprehensive visual information for these two fire types. The combination of RGB and HSV color spaces will aid in identifying subtle differences in flame color, intensity, and spread between normal stove flames and hazardous pan fires.

This differentiation between *stove flame* and *pan fire* will play a vital role in refining the fire detection model, ensuring the system can accurately detect true fire hazards while minimizing false positives in kitchen environments.

Each position of the camera recorded a 10-minute video. The Google Colab platform's OpenCV package was then used to process these recordings. A total of 3,000 raw images were obtained by converting each video into a series of 500 images. Every image was thoroughly examined to guarantee the accuracy and caliber of the dataset. A final selection of 2,316 photos was made by eliminating blurry or poor-quality photos.

To evaluate the impact of color space on fire detection performance, the dataset was split into two versions: one in the original RGB color space and the other in HSV. The RGB images were used as-is for training the first model. For the HSV dataset, each RGB image was converted to HSV using OpenCV's cv2.cvtColor(image, cv2.COLOR_RGB2HSV) function in Python. The HSV-converted images were saved and labeled identically to the original RGB images to ensure consistency in the training process. Both the RGB and HSV image sets were then independently used to train the YOLOv5m model, using identical training parameters to facilitate a fair comparison. The YOLOv5 model was further tested on the selected dataset, with an emphasis on fire detection in both RGB and HSV color spaces.

Figure 3. A sample of six images represent six different positions of pan fire object from red, green, and blue (RGB) dataset: (a) 50 cm and 45° left, (b) 50 cm and 45° right, (c) 50 cm and side right, (d) 50 cm and center, (e) 75 cm and top left, and (f) 75 cm and top right, respectively

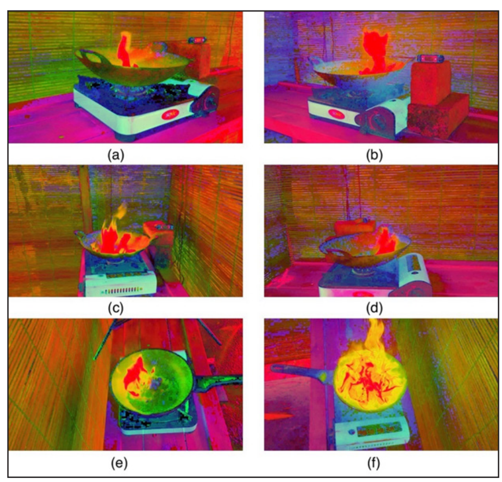

Figure 4. A sample of six images represent of six different positions of pan fire object from hue, saturation, and value (HSV) dataset: (a) 50 cm and 45° left, (b)50 cm and 45° right, (c) 50 cm and side right, (d) 50 cm and center, (e) 75 cm and top left, and (f) 75 cm and top right, respectively

Background Images

The background images with no fire indication from the stove and pan are shown in Figures 5 and 6. It has been suggested that including the background images in training might reduce false positives (Zhao et al., 2023). The six precise locations were used to get background images from the previous section.

Figure 5. A sample of six different background images from the red, green, and blue (RGB) dataset

Figure 6. A sample of six different background images from the hue, saturation, and value (HSV) dataset

Pre-Processing

The images were cropped to eliminate any unrelated backgrounds, specifically non-fire, as shown in Figure 7. The images were then resized after cropping to be compatible with

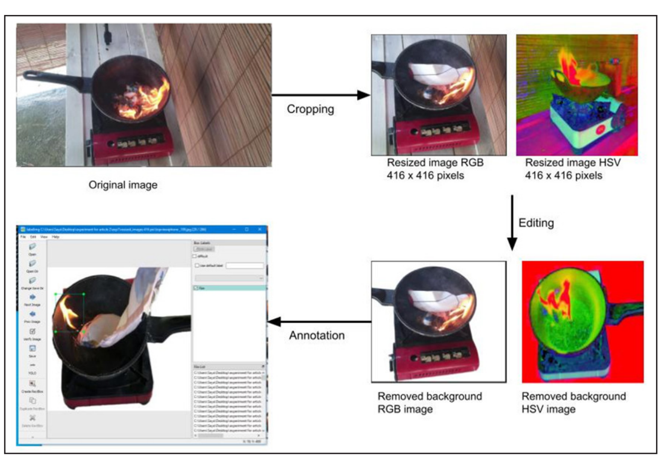

Figure 7. The flow of pre-processing steps for the fire image
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value

the YOLO model, which only accepts image sizes that are multiples of 32. The images were resized to a 416 × 416-pixel size.

In this study, no data augmentation techniques were applied during training. This decision was made to preserve the authenticity and raw visual characteristics of the kitchen fire scenarios as captured under real-world ambient lighting conditions. The aim was to evaluate the model's performance on naturally occurring variations in fire appearance without introducing synthetic transformations that could potentially distort fire features or misrepresent actual conditions. Additionally, no further normalization or pixel-intensity normalization was provided beyond the automatic internal pre-processing applied by the YOLOv5 framework to the input.

Dataset Partition

To ensure an exhaustive assessment of the model, the dataset was split into training, validation, and testing sets. The model was trained using the training set to identify the key features of the images. The validation set was used to monitor the model's performance throughout training, providing insights to tune its settings and parameters for improved performance. The performance of the trained model was assessed using the testing set. As Table 2 illustrates, the training, validation, and testing ratios in this article were set at 70:15:15. The datasets were shuffled using the scikit-learn shuffle function, with the random state parameter set to 1, to ensure reproducibility of the results. This ensures that the data is consistently and randomly ordered. There was no stratification splitting. However, randomization ensured that the two classes were equally distributed across the training, validation, and testing sets in the experiment. To investigate the effect of background context on model performance and the reduction of false positives, background images were integrated in Experiments 2 and 3 with a 5:1 fire-to-background ratio. Specifically, 244 background images were included in Experiment 2, and 486 background images were included in Experiment 3.

Table 2
Distribution of kitchen stove fire image dataset

Experiment	Color space	Number of images			Background images (5:1 ratio)
		Training	Validation	Testing	
1: Pan fire (No background)	RGB	200	43	43	0
	HSV	200	43	43	
2: Stove and pan fire with background	RGB	857	182	183	244
	HSV	857	182	183	244
3: Fused RGB and HSV	RGB and HSV	1,702	364	366	486

Note. RGB = Red, green, and blue; HSV = Hue, Hue, saturation, and value

Experiments

The experiments were conducted using a Lenovo ThinkStation P300 equipped with standard specifications, and all model training and testing were performed on Google Colab with Python 3.10.12 and PyTorch version 2.4.1+cu121. Training was conducted on an Nvidia (USA) Tesla T4 graphics processing unit (GPU) with 15 GB of video random access memory (VRAM). The central processing unit (CPU) and random-access memory (RAM) specifications were those provided by the Google Colab environment. The YOLOv5m model was selected for its balance between performance and computational efficiency. Required libraries included OpenCV, NumPy, PyTorch, and Matplotlib.

For YOLOv5m, it was trained for 200 epochs, using a batch size of 29, with the input image resolution of 416 × 416 pixels, which is similar to typical YOLO training settings. All training used the custom data set, and image preprocessing included resizing and normalization necessary for use with the YOLOv5 framework.

The model was optimized with the SGD optimizer with an initial learning rate of 0.01 and momentum of 0.937. An additional cosine learning rate schedule was applied, decaying the learning rate to 1% of its initial value at the end of training. Warmup was done with three epochs, initializing with a warmup momentum of 0.8 and a warmup bias learning rate of 0.1. These values correspond to the default YOLOv5 hyperparameters, which have been finely tuned for stable convergence.

The work from the first and second experiments is depicted in Figures 8 and 9. Figure 10 displays the block diagram for the third experiment. Before the training, validation, and testing stages, pre-processing was applied to every stove fire and pan fire, including the background images. To start, every fire on the pan was individually cropped to remove any non-fire items. In terms of the background images, any fire objects that appeared in them were removed by cropping off the area of the image showing the pan and stove on fire, resulting in a clean background. The bounding box surrounding the targeted items was then provided as ground truth using LabelImg to annotate the fire images. The background images were not labelled and were simply added to the training dataset.

In our approach, the original YOLOv5m model provided by Ultralytics was used without any architectural changes. The model was initialized with pre-trained common objects in context (COCO) weights to take advantage of transfer learning gains. The structure of the backbone, neck, and detection head was not modified based on the model. The standard version has been used with the intention that, in case of fire occurrence, it would be a fair comparison for the RGB and HSV scenarios.

The datasets were then divided into three sets of experiments:

1. In the first experiment, the study concentrated solely on pan fire images without any kitchen background. The purpose of this experiment was to determine whether color space (RGB vs. HSV) had an impact on fire detection accuracy. The RGB and HSV

datasets were trained separately on YOLOv5m, and the results were compared to determine the color space with superior performance in a controlled, background-free environment.

2. The second set of experiments consists of both stove fire and pan fire images. This experiment compares the performance of RGB and HSV on background images. The performance of the RGB and HSV datasets, which included photos of stoves and pan fires, was examined after each dataset was trained separately.
3. The third experiment uses the fusion of both the RGB and HSV datasets to create more training sets. This experiment aims to investigate whether the model performs better at fire detection when it learns features from both RGB and HSV datasets simultaneously.

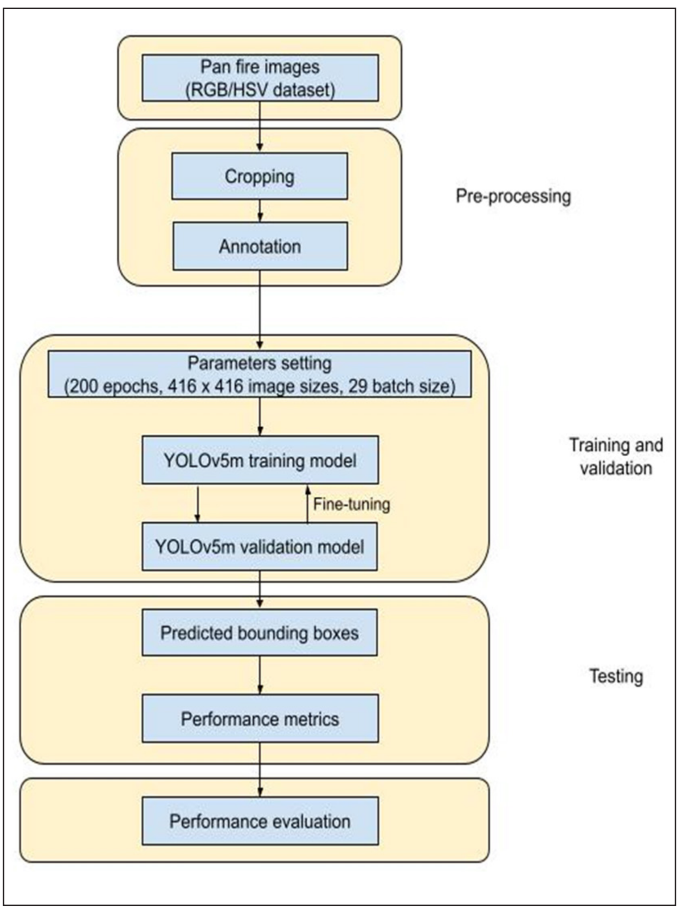

Figure 8. Flowchart of the first experiment
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value; YOLOv5m = You Only Look Once version 5 medium

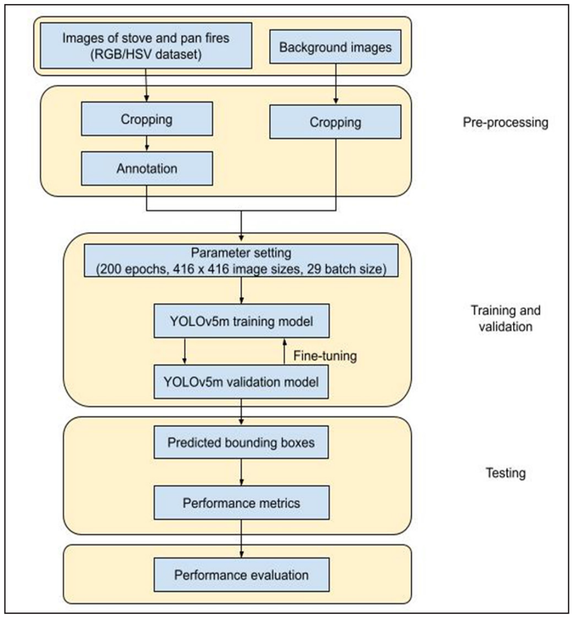

Figure 9. Flowchart of the second experiment
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value; YOLOv5m = You Only Look Once version 5 medium

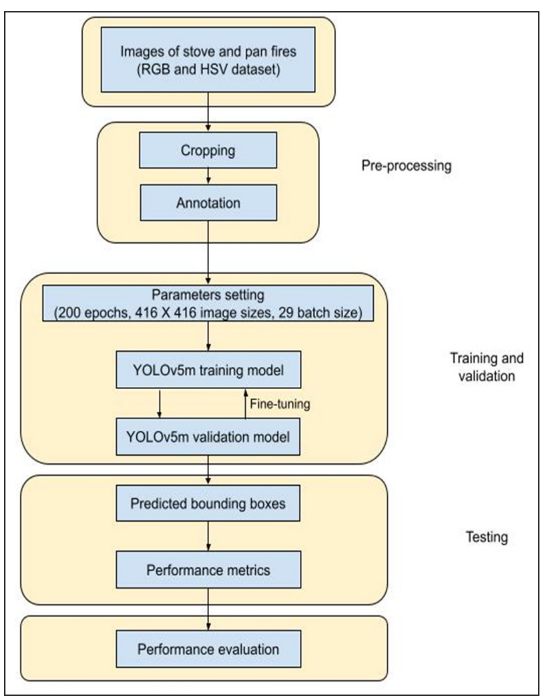

Figure 10. Flowchart of the third experiment
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value; YOLOv5m = You Only Look Once version 5 medium

RESULTS AND DISCUSSION

Results Overview

The performance of the RGB, HSV, and fused RGB-HSV datasets is compared across important assessment measures, including recall, precision, F1-score, mAP@50, and mAP@50-95, in the Table 3, which summarizes the experiment results. These metrics provide information about how well the model detects different scenarios and an overall sense of how effectively each model can detect fire and stove-related events across different kitchen settings. The calculations of the metrics were based on the validation functions provided in the YOLOv5 platform. A prediction was a true positive if the IoU of the predicted and ground truth bounding box was 0.5 or larger. mAP@50, which is the average precision at an IoU threshold of 0.5, and mAP@50–95 reflects the robustness of the model by averaging the performance over the IoU ranging from 0.5 to 0.95 with an increment of 0.05.

Table 3
Overall result of performance metrics for three experiments

Experiment	Color space	Recall	Precision	F1-score	mAP@50	mAP@50-95
Experiment 1 (Pan fire)	RGB	0.843	0.886	0.863	0.873	0.651
	HSV	0.882	0.931	0.905	0.938	0.632
Experiment 2 (Stove and pan fires)	RGB	0.933	0.941	0.937	0.950	0.632
	HSV	0.921	0.966	0.943	0.953	0.624
Experiment 3 (Fusion RGB-HSV)	Fused RGB and HSV	0.918	0.953	0.935	0.949	0.622

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Experiment 1: Training Dataset without Background Images

An extensive analysis of our suggested model's performance on datasets that only contain specific fires on pan images in Table 4 . Various performance metrics were used to analyze the capabilities of the model and provide an extensive discussion of the results.

Table 4
Performance metrics for Experiment 1

Training dataset	Recall	Precision	F1-score	mAP@50	mAP@50-95
RGB without background images	0.843	0.886	0.863	0.873	0.651
HSV without background images	0.882	0.931	0.905	0.938	0.632

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

In the first experiment, fire detection accuracy was tested in a controlled background-free environment by comparing RGB and HSV color spaces. The analysis was conducted based on the performance of YOLOv5m, which was trained separately on RGB and HSV datasets of pan fire images. The findings provide insight into how the two-color spaces differ in terms of processing and performance when it comes to detecting kitchen fires.

Processing Time and Epochs

For 200 epochs, or an equal number of iterations over the datasets, the RGB and HSV models were trained. Both color spaces required equal computational resources for training, as indicated by the consistency of processing times, which were approximately 15 minutes (0.249 hours for RGB and 0.248 hours for HSV). This equivalency shows that the amount of time needed to train the YOLOv5m model in this controlled setting was not significantly affected by the color space selection.

Inference Speed

Between the two models, there was a small difference in the inference speed during validation. While the HSV dataset processed photos at 2.32 iterations per second (it/s), the RGB dataset processed images at 2.41 it/s. Despite being negligible, this difference implies that the HSV photos were validated a little bit faster. This might be explained by the way the YOLOv5m model handles the color channels of RGB and HSV differently, albeit the differences are not great enough to imply that one has a computational advantage over the other.

Performance Metrics

The main objective of this experiment was to evaluate the accuracy of fire detection between the RGB and HSV color spaces. The performance metrics in Table 4 shows a considerable difference between the HSV and RGB models. In almost every metric used in this experiment, the HSV color space outperformed RGB. Compared to RGB, which obtained a recall of 0.843, a precision of 0.886, an F1-score of 0.863, and an mAP@50 of 0.873, HSV obtained a higher recall of 0.882, a precision of 0.931, an F1-score of 0.905, and an mAP@50 of 0.938. This shows that removing the background improves HSV's ability to identify fire from non-fire elements. This might be because the nature of the HSV color space allows less interference in color feature separation, such as the hue of fire in the presence of minimal distractions in the background. Nonetheless, it is intriguing to observe that, although the mAP@50-95 (which is a more rigorous measure and assessed over a range of IoUs) is a little lower for HSV (0.632) than RGB (0.651), the precision and recall metrics still favor HSV. This suggests that even though RGB may be constructed with greater leeway to varying levels of IoU, its effectiveness is inferior when it comes to detecting fire, in which case it is restricted to a narrow range of thresholds, whereby it facilitates detection more accurately than the RGB model.

Nonetheless, although the mAP@50-95, which is a more rigorous measure that assesses over a range of IoUs, is slightly lower for HSV (0.632) than for RGB (0.651), the precision, recall, and F1-score metrics still favor HSV. That means that HSV is more accurate in identifying fire when restricted to certain thresholds, even though RGB might have a wider tolerance at different IoU values. As demonstrated in Figures 11 and 12, the YOLO5m is utilized to detect the validation and testing image samples, respectively.

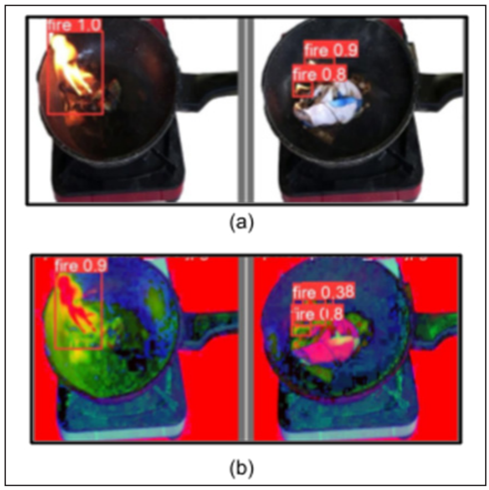

Figure 11. The YOLOv5m model was used to detect validation samples of (a) RGB and (b) HSV images, respectively
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

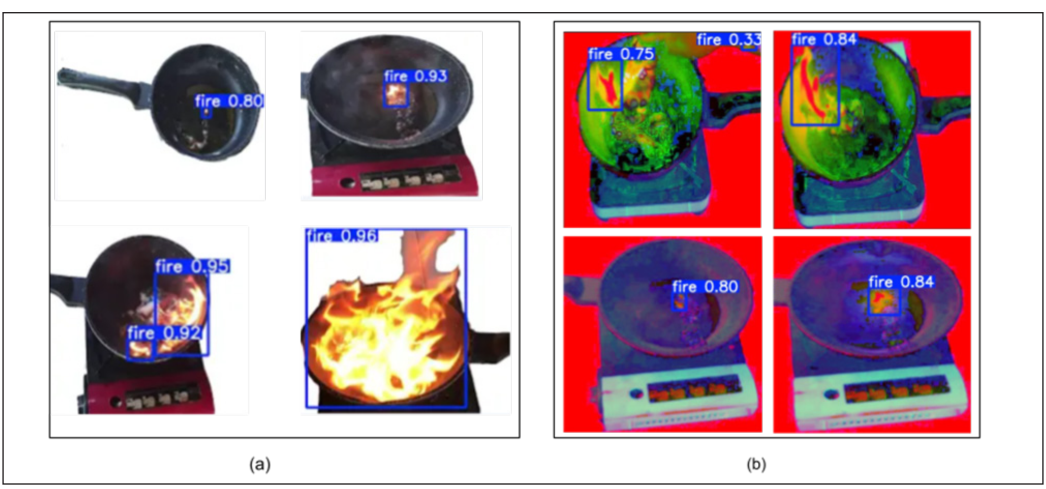

Figure 12. The YOLOv5m model was used to detect testing samples of (a) RGB and (b) HSV images, respectively
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Experiment 2: Training Dataset with Background Images

The second experiment was conducted with an expanded scope of the dataset by incorporating both stove flame and pan fire images, including the background surrounding, which increased the task complexity. The objective was to determine the performance matrix differences between RGB and HSV in the first experiment with more complex, real-world scenarios involving kitchen backgrounds. Both RGB and HSV color spaces were evaluated using the YOLOv5m model based on the detection of fire and stove elements, as shown in Table 5.

Table 5
Performance metrics for Experiment 2

Training dataset	Recall	Precision	F1-score	mAP@50	mAP@50-95
RGB with background images	0.933	0.941	0.937	0.950	0.632
HSV with background images	0.921	0.966	0.943	0.953	0.624

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Processing Time and Epochs

Both the RGB and HSV datasets required more time to train than in the first experiment, with RGB taking about 46 minutes (0.771 hours) and HSV taking a longer duration at 49 minutes (0.808 hours). Training was done for both datasets over 200 epochs. Given the complexity of the dataset and the inclusion of background features, the longer training time was anticipated. Given that the same number of epochs was used to train both models, the significantly longer processing time for HSV implies that the background complexity may have contributed to a slightly slower training process for this color space.

Inference Speed

Comparing RGB (1.28 it/s) to the first experiment, the inference speed during validation was noticeably slower. This is probably because there were more different kinds of images, including backgrounds, stove, and pan fire, which raised the computational processing demand. HSV model performance, on the other hand, was consistent with the first experiment, demonstrating an inference speed of 2.38 it/s, nearly twice as fast as the RGB model. This suggests that the HSV color space can handle backdrop clutter with lower processing time.

Performance Metrics

Both RGB and HSV color spaces performed well overall in terms of fire detection abilities. Nevertheless, a review of the performance measures revealed some significant variations. The HSV model outperformed RGB with a precision of 0.966 as opposed to

0.941. This indicates that the HSV model is more accurate in differentiating between fire and stove elements from the background, as it generates fewer false positives. For real-world applications, where lowering false positives is essential, this increased precision is noteworthy. With a recall score of 0.933 against HSV 0.921, the RGB model scored somewhat better than HSV. This indicates that, when compared to HSV, the RGB model was more accurate in recognizing true positives, catching a higher proportion of real fire and stove occurrences, but at the cost of more false positives. In terms of mAP@50, RGB scored 0.950 and HSV achieved 0.953, indicating equivalent performance for both models. According to this metric, the two-color spaces identify fire occurrences at the 50% IoU threshold nearly equally well. Nonetheless, the slightly elevated score of the HSV model bolsters its advantage in precision. In the mAP@50-95 metric, the RGB achieved a higher score of 0.632 than the HSV model with 0.624. This finding suggests that RGB may perform better across a range of IoU thresholds, potentially enhancing its robustness in identifying fires in congested environments with diverse background objects and surroundings.

Class-Wise Breakdown

More information about how each color space handles pan fire, and stove images can be found in the performance metrics broken down by class, shown in Table 6.

Table 6
Performance metrics for Experiment 2 by class

Model	Training class	Recall	Precision	F1-score	mAP@50	mAP@50-95
RGB	Pan fire	0.887	0.876	0.881	0.909	0.567
	Stove flame	0.994	0.989	0.992	0.991	0.697
HSV	Pan fire	0.939	0.852	0.893	0.919	0.558
	Stove flame	0.993	0.989	0.991	0.988	0.691

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

For pan fire images: The HSV model performed better than the RGB model (0.887) in terms of recall (0.939), indicating that it could identify more actual fires. HSV achieved mAP@50-95 (0.558) and precision (0.852), which were, however, lower than those of RGB (0.876, 0.567 for mAP@50-95). As a result, HSV provided a greater range of detection performance, even though it was more consistent in spotting fire. It also created more false positives. On the other hand, the RGB model was able to achieve a better balance between precision and recall metrics, because it was more adept at differentiating colors over complex backgrounds.

Stove flame images: Both color spaces performed well. While the HSV model produced similarly good metrics (recall = 0.993, precision = 0.989), the RGB model obtained a recall of 0.994 and a precision of 0.989. The nearly flawless detection of stoves in all models suggests that significant visual cues, including the stove's unique shape and structure, were crucial in achieving high detection accuracy across all color spaces.

Both classification detection utilizing YOLOv5m for both color spaces are shown in Figures 13 and 14.

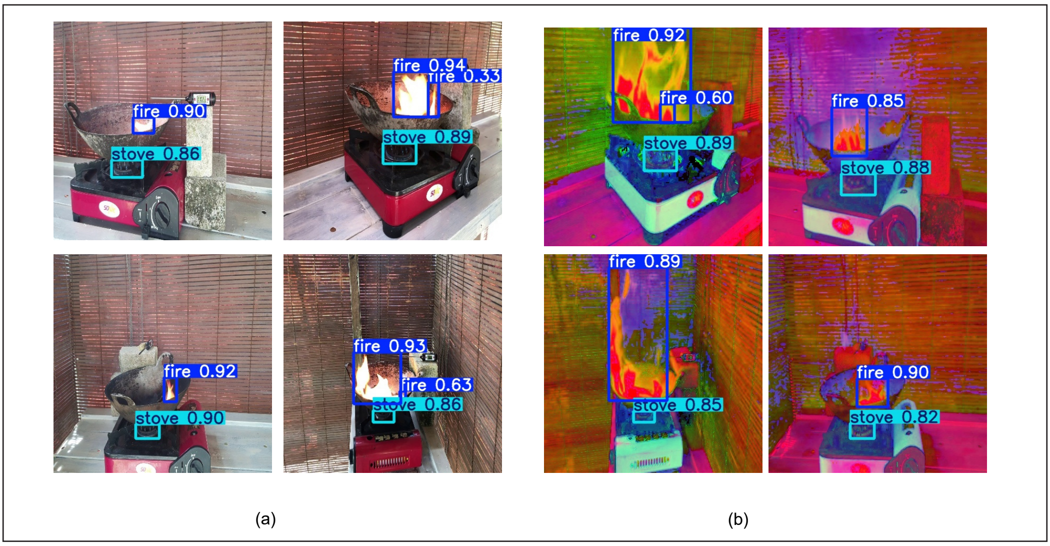

Figure 13. The YOLOv5m model was used to detect validation samples of (a) RGB and (b) HSV images, respectively

Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

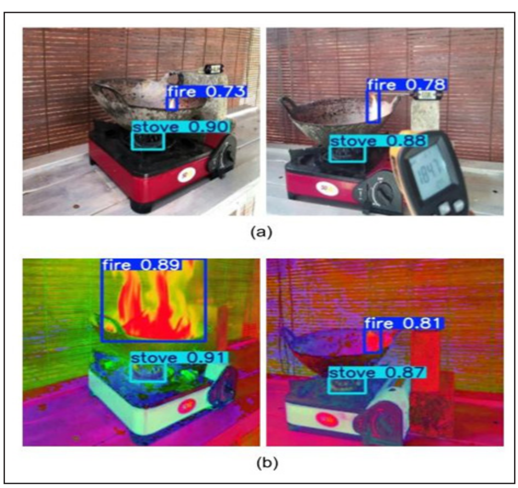

Figure 14. The YOLOv5m model was used to detect testing samples of (a) RGB and (b) HSV images, respectively

Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Experiment 3: Training Dataset Fused RGB-HSV with Background Images

The third experiment was done to evaluate the performance of a fusion of two different color spaces in one dataset. Through the combination of both RGB and HSV datasets, the experiment aimed to take advantage of RGB and HSV complementing capabilities, particularly in handling background clutter and color differentiation.

Training Process and Epochs

In contrast to the 200 epochs used in earlier experiments, the model was trained for only 153 epochs. The training period was completed in about 69 minutes, or 1.15 hours, with an early halt happening when no improvement was seen for the final 10 epochs. This early stopping mechanism indicates that the model reached its optimal solution more quickly. However, it also suggests that longer training times were not necessary to achieve peak performance, due to the added complexity introduced by combining the RGB and HSV datasets.

Inference Speed

During validation, the inference speed was 1.19 iterations per second (it/s), which was slower than in Experiment 2 (2.38 it/s) for HSV. The larger and more complex dataset produced by combining two color spaces is likely the cause of this speed reduction. More computational overheads may have been incurred by the model's need to process a more varied set of features.

Performance Metrics

The fusion model did not significantly exceed the RGB or HSV models individually, but it did maintain good performance, especially in precision. The following is a summary of key performance metrics as shown in Table 7.

Table 7
Performance metrics for Experiment 3

Training dataset	Recall	Precision	F1-score	mAP@50	mAP@50-95
Fusion of RGB and HSV datasets	0.918	0.953	0.935	0.949	0.622

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

As in the earlier experiments, a precision of 0.953 shows that the fusion model maintained its high level of accuracy in lowering false positives. This great precision indicates that the fusion did not negatively impact the model's capacity to accurately identify stove or real fire incidents, as it is consistent with the individual RGB and HSV models.

The recall value of 0.918 shown in this experiment is marginally less than in Experiment 2 (HSV = 0.921, RGB = 0.933). This implies that while the fusion model performed well, it was not as successful as the individual RGB and HSV models in detecting every fire occurrence. The enhanced complexity of the fused dataset may have added more variability and confusion in identifying specific fire patterns, which could account for the slightly lower recall.

The mAP@50 result with 0.949 shows a good overall detection capacity, nearly matching the results from the individual RGB (0.950) and HSV (0.953) models. This consistency indicates that, at the 50% IoU threshold, combining RGB and HSV does not impair the model's performance, suggesting that it remains capable of detecting fire and stove elements.

The mAP@50-95 value of 0.622 is marginally less than the mAP@50-95 values obtained from the separate RGB (0.632) and HSV (0.624) experiments. This indicates that the fusion model struggled at higher IoU thresholds, where bounding box predictions require greater accuracy. This could imply that some inconsistencies or overlapping features were introduced when merging features from the two-color spaces, which would have resulted in less accurate predictions in difficult circumstances.

Class-Wise Breakdown

Important details on how the fusion model processed fire and stove images may be found in the class-by-class breakdown shown in Table 8.

Table 8
Performance metrics for Experiment 3 by class

Model	Training class	Recall	Precision	F1-score	mAP@50	mAP@50-95
Fusion of RGB and HSV datasets	Pan fire	0.913	0.847	0.879	0.906	0.561
	Stove flame	0.993	0.989	0.991	0.991	0.684

Note. mAP = Mean average precision; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Pan fire images: In contrast to the RGB (0.887) and HSV (0.939) models from the second experiment, the recall for pan fire images was, on average, 0.913. This implies that the fusion model's ability to identify actual fire incidents was effective. The fusion model, however, generated more false positives, as evidenced by the decreased precision for fire images (0.847) compared to the RGB model (0.876). Furthermore, the mAP@50-95 for fire images (0.561) was on average of RGB (0.567) and HSV (0.558), suggesting a deterioration in the model's capacity to locate fire incidents at various IoU thresholds reliably.

Stove flame images: The recall and precision remained extremely high (recall = 0.993, precision = 0.989), in line with other trials. The fusion of RGB and HSV did not adversely affect the accuracy of stove detection, as evidenced by the mAP@50-95 for stove images of 0.684, which was slightly lower than the prior results in Experiment 2. This implies that the complexity added by the fused dataset had a lesser impact on stove flame photos, which are easier to distinguish and identify visually.

Both classification detection utilizing YOLOv5m for experiment 3 is shown in Figures 15 and 16 during the validation and testing process.

Figure 15. The YOLOv5m model was used to detect validation samples of fused RGB and HSV images, respectively
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

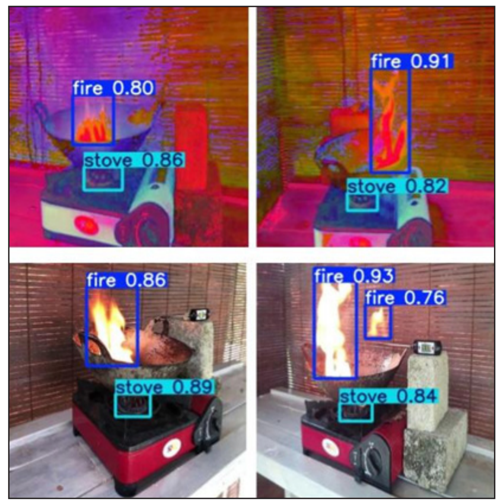

Figure 16. The YOLOv5m model was used to detect testing samples of fused RGB and HSV images
Note. YOLOv5m = You Only Look Once version 5 medium; RGB = Red, green, and blue; HSV = Hue, saturation, and value

Analysis of Confusion Matrices for Each Experiment

Confusion matrices are used to further evaluate model performance and understand the nature of classification errors in these experiments. These matrices demonstrate the model's ability to distinguish between 'fire', 'stove fire', and 'background' classes. A higher incidence of false positives, specifically, misclassifications of background as fire, is shown in Experiment 1 (false positive rate: 22% for HSV, 23% for RGB), where background images were excluded. In contrast, Experiments 2 and 3, which incorporated varied background contexts, show a marked reduction in false positives, underscoring the critical role of background information in improving model robustness and reducing misclassification rates.

In the first experiment, which involved no background, models were trained on images depicting pan fire, allowing isolated evaluation of the impact of color space on fire detection. The confusion matrices for RGB and HSV in Figure 17 show strong diagonal dominance, indicating that the models were highly accurate in classifying fire versus background. The HSV database model yielded fewer false negatives than the RGB model. This suggests that HSV color space provides superior color segmentation properties during the detection of fire pixels when minimal noise is involved.

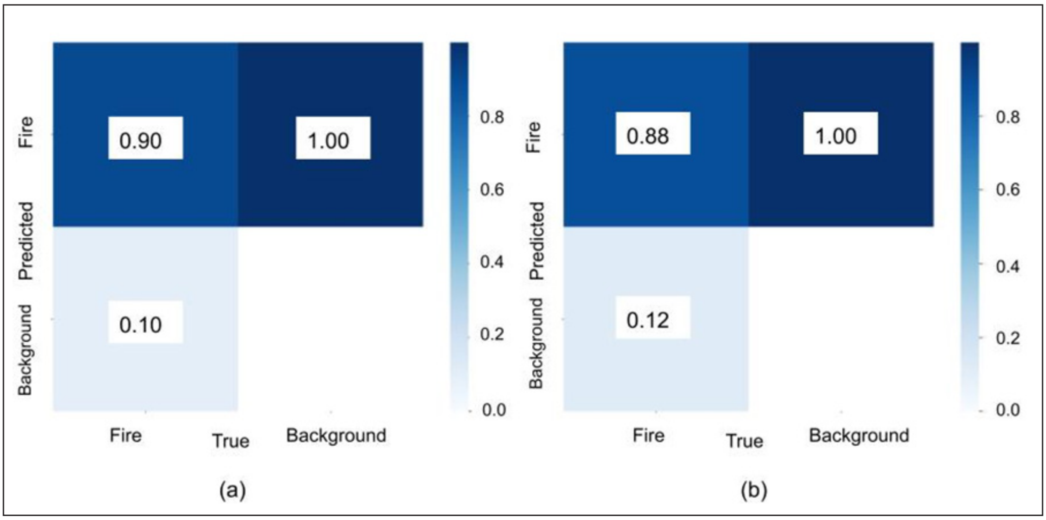

Figure 17. Confusion matrices for the first experiment using (a) RGB and (b) HSV images, respectively
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value

In the second experiment, the dataset included a mix of stove fires and pan fires, as well as actual kitchen environments. This is the challenging part of the model as it involves a complex background. As shown in Figure 18, both HSV and RGB models show increased misclassification rates. The increment of misclassified stove fire images as background in the RGB model suggests that its features were more susceptible to distraction from surrounding kitchen elements. However, the HSV model depicted improvement in these scenarios, maintaining clearer separation between fire-related classes and the background. The HSV color space shows better color contrast and stability under varying lighting conditions, which is normal in a real kitchen environment.

The third experiment was executed to assess the fused RGB and HSV dataset performance, which consisted of training samples drawn from both RGB and HSV formats. The result from the confusion matrix in Figure 19 demonstrates that the model trained outperformed individual color space models. A few pan fire and stove fire images were misclassified together with background images. This finding confirms that feature diversity obtained through fusion allows the model to learn a more comprehensive representation of fire characteristics accurately.

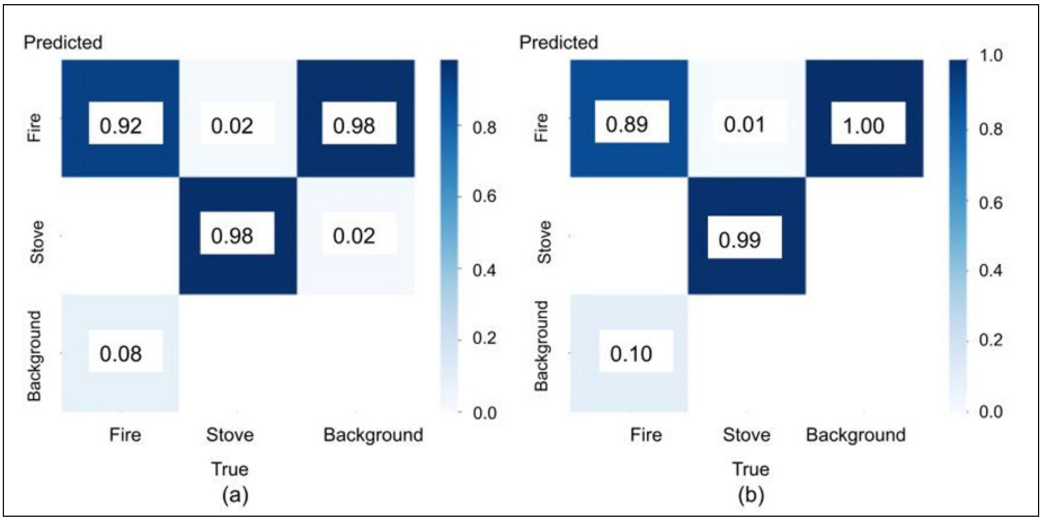

Figure 18. Confusion matrices for the second experiment using (a) RGB and (b) HSV images, respectively
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value

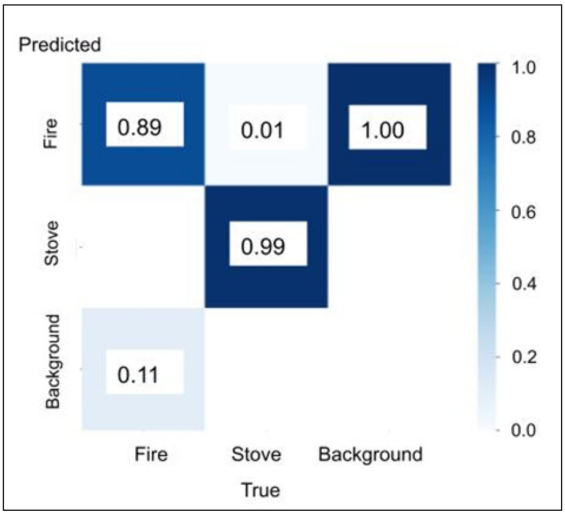

Figure 19. Confusion matrix for the third experiment using fused RGB and HSV images
Note. RGB = Red, green, and blue; HSV = Hue, saturation, and value

The results revealed that HSV was slightly better than RGB in some cases. This effect might also be related to the fact that HSV allows the direct separation of chromatic (hue and saturation) and intensity (value) information and is thus more resilient to lighting variations. This property likely helps HSV better localize fire features in kitchen scenes, where illumination changes are common due to flames, reflections, or shadows.

Thus, the preference for HSV over RGB found in Experiments 1 and 2 might be because it separates chromatic content (hue) from illuminant (value). Fire is usually very red, orange, and yellow, which is consistent in all lighting. Unlike RGB, where color and brightness are intertwined, HSV helps the fire model to concentrate more on the hue feature of the fire in distinguishing flames from typical kitchen backgrounds, such as metal surfaces, tiles, and utensils.

Furthermore, with background images in Experiment 2, the task became more complex, possibly including items such as stove knobs, cooking utensils, or reflections of lighting. The latter might introduce ambiguities to the model in RGB space, as similar colors can also exist in non-fire areas. Unlike HSV, the color information separation on the chromatic side of the HSV model that was applied in the framework subsequently suppressed the background clutter and made detection more robust.

The false positive rate, which is defined as the proportion of images without fire that were misclassified as containing fire, was significantly higher in Experiment 1 (0.22 for HSV, 0.23 for RGB), with reduced background variability. However, the false positive rate dropped dramatically in Experiments 2 and 3, using a variety of background images (0.13 for HSV, 0.08 for RGB, and 0.08 for fused RGB + HSV). These results validate that the use of background images enhances the model's ability to discriminate between fire and non-fire scenes, resulting in a decrease in the number of false alarms.

Comparison Between the Proposed Model and Other Approaches

A comparison between the proposed YOLOv5m model and other cutting-edge YOLOv5 techniques used in current fire detection studies is shown in Table 9. All the studies that were utilized for comparison involved RGB datasets, while one study used HSV and luminance, chrominance-blue, and chrominance-red (YCbCr) datasets as part of their training images. In general, the comparison illustrates the potential advantages of utilizing both RGB and HSV data for object detection tasks, particularly for fire detection. The results shown in Table 9, the greatest mAP@0.5 values, 0.950 and 0.953, respectively, are obtained by the proposed model YOLOv5m with background images (HSV and RGB), suggesting good precision for detecting fire events with background context. The performance of the proposed models demonstrates a trade-off between the use of background and non-background images. It is interesting to note that RGB and HSV with background images maintain greater mAP@0.5 but have difficulty with mAP@0.5:0.95, suggesting a tendency to over-rely on background information or difficulties generalizing across different IoUs. HSV-based models typically outperform RGB-based models or perform about the same. For instance, compared to the RGB equivalent (0.873 mAP@50), the suggested model with HSV, excluding the background (0.938 mAP@50), performs better. This implies that HSV, perhaps because of its more intuitive representation of color fluctuations in fire, might be more suitable for fire detection.

Table 9
Comparison between the proposed model and other approaches

Method	**Color space**	**mAP@050**	**mAP@50-95**	**Type of fire detection**
YOLOv5 and flame threshold segmentation (Zhao et al., 2022)	RGB	0.863	0.576	Fire, smoke, lamp, reflections, sunshine
YOLOv5s-CA (W. Liu et al., 2022)	RGB	0.946	-	Smoke and fire
YOLOv5 with SE attention mechanism (Fan & Zhan, 2023)	RGB	0.941	0.799	Flame
YOLOv5 and YOLOv8 (Mahmoud et al., 2024)	HSV and YCbCr	0.895	0.954	Fire and smoke
YOLOv5m (Proposed model)	RGB without background images	0.873	0.651	Fire
YOLOv5m (Proposed model)	HSV without background	0.938	0.632	Fire
YOLOv5m (Proposed model)	RGB with background images	0.950	0.632	Fire
YOLOv5m (Proposed model)	HSV with background images	0.953	0.624	Fire
YOLOv5m (Proposed model)	Fusion of RGB and HSV datasets	0.949	0.622	Fire

Note. mAP = Mean average precision; YOLOv5 = You Look Only Once version 5 ; RGB = Red, green, and blue; YOLOv5s-CA = You Look Only Once version 5 small with coordinate attention; SE = Squeeze-and-excitation; YOLOv8 = You Look Only Once version 8; YCbCr = Luminance, chrominance-blue, and chrominance-red; YOLOv5m = You Look Only Once version 5 medium; HSV = Hue, saturation, and value

Most models, including the proposed one, exhibit a rapid decline in mAP@0.5 to mAP@0.5:0.95, highlighting areas that require improvement for handling diverse object scales and ensuring reliable detection accuracy. This disparity implies that although the models are capable of simple fire detection, they are not as good at more complex detections across a range of IoU thresholds. The small benefits of combining RGB and HSV color spaces imply that the complementary qualities of these color spaces are not being fully utilized. Subsequent research could examine sophisticated fusion methodologies or examine the reasons behind the fusion's inability to produce appreciably superior outcomes.

Implications and Ethical Considerations

This work demonstrates that using the color space method along with YOLOv5 will lead to improved kitchen fire detection, thus promoting household safety by providing timely responses in fire emergencies. The method is applicable for low-cost, real-time use in the home, where conventional systems fail. Legally, the issue of privacy is paramount, especially when it comes to using cameras in private locations. Local data analysis and user consent are critical for protecting personal privacy and complying with data protection laws.

CONCLUSION

The performance of object detection models can be greatly influenced by the selection of the training dataset. In particular, the top-performing dataset changes based on the metric of interest. Combining datasets can improve overall performance, demonstrating the advantages of using diverse training data sources in object detection applications. To determine the best training datasets and methods for object detection models, these results highlight the significance of carrying out experiments and evaluations. However, in this study, the proposed model resulted in the best overall mean performance in the first experiment, RGB, without background images, generating the highest mAP@0.5:0.95 score of 0.651. Interestingly, the overall performance in terms of precision and recall still favors HSV, even while the mAP50-95 is lower when compared to RGB. This suggests that HSV is more accurate in detecting fire when restricted to thresholds, whereas RGB may have a wider tolerance at different IoU values. Nevertheless, the fused model demonstrated high performance in identifying stove images (recall = 0.993, precision = 0.989), suggesting that stove identification is less susceptible to changes in color space and can be consistently achieved using RGB, HSV, or both. The detection scenario determines which RGB and HSV to use. HSV gives better precision for more straightforward, controlled conditions, which is essential for reducing false positives. However, RGB can more clearly distinguish fire from surrounding objects in real-world kitchen settings with more complicated backgrounds, and it might offer more dependable detection in certain situations. To analyze the benefits of combining a color space, we adopted an early fusion strategy. More specifically, each image was transformed from RGB to HSV, and both (RGB and HSV) were stacked along the channel dimension to form a single multichannel input. The YOLOv5 model was then trained using this aggregated input. The aim was to get the best attributes of each color spectrum to describe fire objects. However, the results show that this embedding was not beneficial, possibly because the input is more complex.

The proposed datasets were assessed using the specific YOLOv5m architecture model. In addition to evaluating the dataset using other state-of-the-art models, apart from YOLOv5, future research should involve increasing the quantity and variety of images and testing the model's robustness under various settings to further improve its performance. Future research could further examine the robustness of the proposed model at various distances from the camera to distinguish between fire and other reddish objects. This study was limited to a single run of each experiment due to time and computational constraints. As a result, statistical measures such as mean, standard deviation, and significance testing between the RGB and HSV results were not performed. Future work will involve repeated experiments to facilitate more robust statistical analysis. Another limitation is that the study focused on controlled kitchen environments. Further research is needed to evaluate the performance of the proposed approach in more complex and cluttered real-world kitchens. The dataset used in this study, while carefully curated, may not fully represent the wide variety of fire scenarios that can occur in kitchens. Future work should explore the use of larger and more diverse datasets.

The research also suggested employing more advanced GPU specifications to reduce training times and investigating how various image resolutions impact the model's performance. To further improve object detection model performance, additional feature engineering can be investigated, in line with the indicated future work. Another suggestion is to investigate performance in real-world kitchens, exploring the use of temporal information, evaluating computational efficiency on embedded devices, and addressing the limitations of the current dataset. By offering more details about an object's form and texture, these characteristics can raise the precision of object detection algorithms. To further confirm the durability of HSV in practical applications, additional research may examine how these results apply to more complex scenarios with background noise.

ACKNOWLEDGEMENTS

The authors sincerely thank the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, for supporting this project.

REFERENCES

Abdusalomov, A. B., Mukhiddinov, M., Kutlimuratov, A., & Whangbo, T. K. (2022). Improved real-time fire warning system based on advanced technologies for visually impaired people. *Sensors*, *22*(19), 7305. https://doi.org/10.3390/s22197305

Bakri, N. S., Adnan, R., Samad, A. M., & Ruslan, F. A. (2018). A methodology for fire detection using colour pixel classification. In *IEEE 14th International Colloquium on Signal Processing and Its Applications* (pp. 94–98). IEEE. https://doi.org/10.1109/CSPA.2018.8368692

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection*. arXiv. http://arxiv.org/abs/2004.10934

Bu, F., & Gharajeh, M. S. (2019). Intelligent and vision-based fire detection systems: A survey. *Image and Vision Computing*, *91*, 103803. https://doi.org/10.1016/j.imavis.2019.08.007

Choi, H.-B., Hwang, E.-H., & Choi, D.-M. (2023). Indoor air quality sensor utilization for unwanted fire alarm improvement in studio-type apartments. *Fire*, *6*(7), 255. https://doi.org/10.3390/fire6070255

Chow, W. K., & Liu, J. (2016). Application of nonlinear dynamics in studying flashover fire in a small open kitchen. *Journal of Applied Mathematics and Physics*, *4*(5), 914–924. https://doi.org/10.4236/jamp.2016.45100

Dilshad, N., Khan, T., & Song, J. (2023). Efficient deep learning framework for fire detection in complex surveillance environment. *Computer Systems Science and Engineering*, *46*(1), 749–764. https://doi.org/10.32604/csse.2023.034475

Fan, J., & Zhan, X. (2023). Flame recognition algorithm for aerial cruising firefighting robot based on RGB and improved YOLOv5. In *4th International Conference on Computer Vision, Image and Deep Learning* (pp. 363–369). IEEE. https://doi.org/10.1109/CVIDL58838.2023.10167099

Geng, X., Su, Y., Cao, X., Li, H., & Liu, L. (2024). YOLOFM: An improved fire and smoke object detection algorithm based on YOLOv5n. *Scientific Reports*, *14*, 4543. https://doi.org/10.1038/s41598-024-55232-0

Giwa, O., & Benkrid, A. (2024). A new flame-based colour space for efficient fire detection. *IET Image Processing*, *18*(5), 1229–1244. https://doi.org/10.1049/ipr2.13022

Graham, T. L., & Roberts, D. J. (2000). Qualitative overview of some important factors affecting the egress of people in hotel fires. *International Journal of Hospitality Management*, *19*(1), 79–87. https://doi.org/10.1016/S0278-4319(99)00049-3

Jocher, G. (n.d.a). *Ultralytics/Ultralytics*. GitHub. https://github.com/ultralytics/ultralytics

Jocher, G. (n.d.b). *Ultralytics/Yolov5*. GitHub. https://github.com/ultralytics/yolov5

Johnston, J., Zeng, K., & Wu, N. (2022). An evaluation and embedded hardware implementation of YOLO for real-time wildfire detection. In *IEEE World AI IoT Congress* (pp. 138–144). IEEE. https://doi.org/10.1109/AIIoT54504.2022.9817206

Kim, J.-H., Kim, N., & Won, C. S. (2023). High-speed drone detection based on Yolo-V8. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1–2). IEEE. https://doi.org/10.1109/ICASSP49357.2023.10095516

Kobeissi, A., & Boulos, M. (2023). Innovative single-camera depth mapping for enhanced indoor fire detection and localization. In *4th International Multidisciplinary Conference on Engineering Technology* (pp. 278–282). IEEE. https://doi.org/10.1109/IMCET59736.2023.10368270

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). *YOLOv6: A single-stage object detection framework for industrial applications*. arXiv. https://doi.org/10.48550/arXiv.2209.02976

Liang, D., Bui, T., & Wang, G. (2023). Fire and smoke detection method based on improved YOLOv5s. In *8th International Conference on Communication, Image and Signal Processing* (pp. 293–300). IEEE. https://doi.org/10.1109/CCISP59915.2023.10355820

Liu, J., Wang, P., & Song, G. (2019). Numerical study of kitchen fires in high-rise residential buildings. In *9th International Conference on Fire Science and Fire Protection Engineering* (pp. 1–4). IEEE. https://doi.org/10.1109/ICFSFPE48751.2019.9055772

Liu, S., Feng, J., Zhang, Q., & Peng, B. (2022). A real-time smoke and fire warning detection method based on an improved YOLOv5 model. In *5th International Conference on Pattern Recognition and Artificial Intelligence* (pp. 728–734). IEEE. https://doi.org/10.1109/PRAI55851.2022.9904105

Liu, W., Quijano, K., & Crawford, M. M. (2022). YOLOv5-tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 8085–8094. https://doi.org/10.1109/JSTARS.2022.3206399

Mahmoud, O., Saad, A., & Nazih, N. (2024). Deep learning-powered vision system for fire detection and localization in harsh environments. In *International Telecommunications Conference* (pp. 300–305). IEEE. https://doi.org/10.1109/ITC-Egypt61547.2024.10620511

Milke, J., & Zevotek, R. (2016). Analysis of the response of smoke detectors to smoldering fires and nuisance sources. *Fire Technology*, *52*, 1235–1253. https://doi.org/10.1007/s10694-015-0465-2

Mutar, A. F. (2019). Study fire detection based on color spaces. *Al-Mustansiriyah Journal of Science*, *29*(4), 93–99. https://doi.org/10.23851/mjs.v29i4.414

Ngai, M., Fu, E., Tam, A., Yang, A., & Ngai, G. (2022). Finding the signal from the smoke: A real-time, unattended fire prevention system using 3D CNNs. *Journal of Student Research*, *11*(3). https://doi.org/10.47611/jsrhs.v11i3.2981

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. In *International Conference on Systems, Signals and Image Processing* (pp. 237–242). IEEE. https://doi.org/10.1109/IWSSIP48289.2020.9145130

Rasbash, D. J. (1991). Major fire disasters involving flashover. *Fire Safety Journal*, *17*(2), 85–93. https://doi.org/10.1016/0379-7112(91)90030-3

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). *You Only Look Once: Unified, real-time object detection*. arXiv. https://doi.org/10.48550/arXiv.1506.02640

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6517–6525). IEEE. https://doi.org/10.1109/CVPR.2017.690

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*. arXiv. https://doi.org/10.48550/arXiv.1804.02767

Spearpoint, M., & Hopkin, C. (2020). A study of the time of day and room of fire origin for dwelling fires. *Fire Technology*, *56*, 1465–1485. https://doi.org/10.1007/s10694-019-00934-5

Vats, A., & Anastasiu, D. C. (2023). Enhancing retail checkout through video inpainting, YOLOv8 detection, and DeepSort tracking. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 5530–5537). IEEE. https://doi.org/10.1109/CVPRW59228.2023.00585

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). *YOLOv10: Real-time end-to-end object detection*. arXiv. https://doi.org/10.48550/arXiv.2405.14458

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. arXiv. https://doi.org/10.48550/arXiv.2207.02696

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). *YOLOv9: Learning what you want to learn using programmable gradient information*. arXiv. https://doi.org/10.48550/arXiv.2402.13616

Wang, S., & Wang, X. (2023). ES-YOLO: A new lightweight fire detection model. In T. Zhang & T. Yang (Eds.), *Third International Conference on Computer Vision and Data Mining* (Vol. 12511, pp. 347–355). SPIE. https://doi.org/10.1117/12.2660040

Wu, S., Li, Z., Li, S., Liu, Q., & Wu, W. (2023). Static gesture recognition algorithm based on improved YOLOv5s. *Electronics*, *12*(3), 596. https://doi.org/10.3390/electronics12030596

Zhang, D., & Chen, Y. (2024). Lightweight fire detection algorithm based on improved YOLOv5. *International Journal of Advanced Computer Science and Applications*, *15*(6), 809–816. https://doi.org/10.14569/IJACSA.2024.0150681

Zhao, Y., Wu, S., Wang, Y., Chen, H., Zhang, X., & Zhao, H. (2023). Fire detection algorithm based on an improved strategy of YOLOv5 and flame threshold segmentation. *Computers, Materials and Continua*, *75*(3), 5639–5657. https://doi.org/10.32604/cmc.2023.037829

Zulkiffli, Z. (2018, June 4). Lupa tutup api dapur faktor kebakaran meningkat [Forgetting to cover the kitchen fire increases the risk of fire]. *Berita Harian*. https://www.bharian.com.my/berita/kes/2018/06/433790/lupa-tutup-api-dapur-faktor-kebakaran-meningkat