Journal of Building Engineering 61 (2022) 105154

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Indoor fire detection utilizing computer vision-based strategies

James Pincott [a,b,*], Paige Wenbin Tien [a], Shuangyu Wei [a], John Kaiser Calautit [a]

[a] Department of Architecture and Built Environment, University of Nottingham, Nottingham, UK
[b] Hoare Lea, 55 Aztec West, Almondsbury, Bristol, BS32 4UB, UK

ARTICLE INFO

Keywords:
Artificial intelligence
Deep learning
Vision-based approach
Fire detection
Smoke detection
Indoor detection

ABSTRACT

Fires are an ever-increasing risk in the world for both indoor and outdoor environments. Current technologies for detection in indoor environments are smoke and flame detectors. However, these detectors have several limitations during both the ignition phase of a fire and propagation. These systems cannot detect an exact position of the fire nor how the fire is spreading, or its size, all of which is necessary information for fire services when dealing with these incidents. A potential solution is to use artificial intelligence techniques such as computer vision, which has shown the potential to detect and recognize objects and activities in indoor spaces. This study aims to develop a vision-based indoor fire and smoke detection system. Existing models based on the Faster R–CNN Inception V2 and the SSD MobileNet V2 models were explored and adopted in this work. This study utilized small training and testing datasets (for indoor specific fire cases) of varying pixel density images. Initial evaluation of the approach was carried out by testing both models on videos, including a mock-up bedroom and living room and a CCTV video of office space. Both high-density smoke environments and flame density scenarios were recognized. The promising results were achieved from using only 480 training images. Despite the success achieved by the Faster R–CNN Inception V2, the SSD MobileNet V2 model showed low accuracy and missed detection results. Future works can focus on integrating this with our previously developed approach, such as occupancy detection, enhancing training data and models, using more advanced detection models, and integrating the proposed approach with the fire fighting and HVAC control systems.

1. Introduction

Fires are an ever-increasing global risk for indoor and outdoor environments. The indoor environment has multiple potential hazards, from electrical devices to flammable substances. These hazards can be the source of indoor fires [1]. The UK government in 2019 reported a total of 555,759 incidents attended by the fire service, with 28% (157,156) of these incidents being a fire-related event [1]. However, most incidents attended false fire alarms, accounting for 41% (229,882) of the total incidents attended [1]. Despite the figures by the UK government showing a decreasing number of cases, incidents are still high, thus demonstrating that fires a still a

Abbreviations: AI, Artificial intelligence; AIO, All-in-one; BES, Building energy simualtion; CCTV, Closed-circuit television; CNN, Convolutional neural network; COCO, Common objects in context; CPU, Central processing unit; DNN, Deep neural network; FN, False negative; FP, False positive; GPU, Graphics processing unit; HRR, Heat release rate; HVAC, Heating, ventilation, and air-conditioning; mAP, Mean average precision; ILSVRC, ImageNet Large Scale Visual Recognition Challenge; IoU, Intersection over union; RGB, Red, green, and blue; ROI, Region of interest; R-CNN, Regional-based convolutional neural network; SSD, Solid-state drive; SVM, Support vector machines; TN, True negative; TP, True positive; UK, United Kingdom.

* Corresponding author. Hoare Lea, 55 Aztec West, Almondsbury, Bristol, BS32 4UB, UK.
E-mail address: jamesgpincott@gmail.com (J. Pincott).

https://doi.org/10.1016/j.jobe.2022.105154
Received 28 March 2022; Received in revised form 13 June 2022; Accepted 16 August 2022
Available online 30 August 2022

significant risk. Current technologies for detection in indoor environments are smoke and flame detectors [2]. However, these detectors have several limitations during the ignition phase of a fire and propagation. These current systems cannot detect an exact position of the fire nor how the fire is spreading, or its size, all of which is necessary information for fire services when dealing with these incidents.

However, these drawbacks of current systems also mean it is a non-invasive detection method, unlike that of computer vision. Computer vision can be considered invasive due to the constant video feed of every space in a building. Current CCTV systems could be used as an initial integration of this technology. However, these systems don't always cover every corner of a room in a building. A separate camera system would need to be implemented to cover sensitive areas or areas of disinterest to security. To ensure privacy would be maintained, the feed will need to be filtered immediately so that no invasion of privacy is incurred. This will mean that no raw footage will be seen by anyone or stored. Hence, approaches such as [3–5], acknowledges such issues. The collected data from the detection made forms the deep learning influenced profiles (DLIP) that correspond to the real-time detection results within an indoor building space.

Since 2011, there has been an increase in published studies each year about artificial intelligence [6]. There has been a significant push to utilize this developing area of technology to improve building systems, from security to heating and cooling management [7]. A major development area for artificial intelligence is vision-based systems, as they can be used for multiple purposes [8,9]. This was immediately taken advantage of with multiple studies utilizing such developments for fire detection [10]; however, the detection was initially completed using hand-designed feature extraction techniques [11]. This meant that they were not robust enough to be accurate, especially in such a diversity of scenarios [12].

1.1. Novelty and contributions to knowledge

Convolutional neural networks (CNN) development allowed faster and generally more accurate detection using deep learning applied to computer vision. This is due to a CNN being able to extract features and classify them in one step. This saves time by replacing the need to create hand-designed feature extraction programmes and means that training time is decreased. Several studies have focused on utilizing this ability for fire detection and smoke detection [13], demonstrating that CNN can yield better performance than some relevant conventional video fire detection methods [14]. However, such studies focus on the outdoor environment, particularly forest fires [14]. Very few studies have focussed on indoor fire detection, especially in office spaces. Indoor spaces such as offices provide a number of challenges when using vision-based systems, such as obstacles blocking the view to the desired detection area and reflections that could interfere with vision-based fire detection.

This study will build on the previous works by Tien et al. [16,17] and Wei et al. [18] where a vision-based artificial intelligence (AI) approach was used to detect and recognize the usage of indoor spaces for aiding demand-driven control systems. The present study aims to develop a fire and smoke detection and recognition approach for buildings based on a similar approach so that a faster and more localized detection may be achieved. In order to address the gaps identified, an initial evaluation of the approach will be carried out by testing the detector on videos of indoor spaces such as a bedroom, living room, and office space. The selected videos will feature realistic indoor spaces, which will allow the evaluation of the capabilities of the proposed approach in detecting fire and smoke in complex indoor environments with multiple objects such as furniture and appliances. Furthermore, it will also include situations when the fire starts or ignites behind a piece of furniture or below a desk. It is envisioned that an all-in-one (AIO) system will be developed in the future that could have multiple functions such as control of heating, ventilation, air conditioning systems (HVAC), and fire safety. Such systems can offer alerts to scenarios that could not be detected before, such as security incidents or work alongside existing sensors to provide a cross-checking ability for systems such as indoor fire detection. This could improve overall safety in workplaces and at home and save emergency services time and money by reducing the number of false alarms. Fig. 1 shows an overview of the research method.

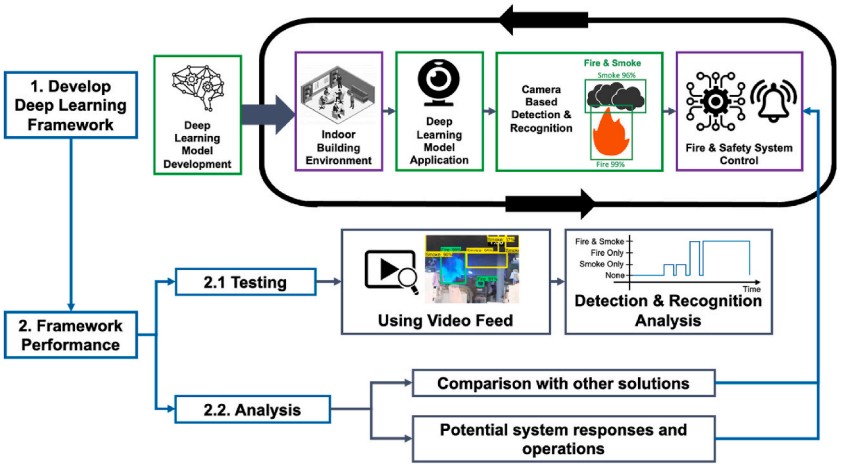


Fig. 1. Overview of the research method for the development of indoor fire and smoke detector using a computer-vision approach.

A literature review of both hand-made networks and more up-to-date systems using Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs) was carried out. Training data were generated by collecting and labelling a suitable size data set consisting of images including fire and smoke inside an indoor or outdoor environment. Training data was assessed through the evaluation of the importance of indoor vs outdoor training images for an indoor environment. TensorFlow was employed to select, train, and evaluate pre-existing open-source deep learning-based models. Assessment of the accuracy of detection, false detection and missed detection of fire and smoke, as well as the system's response time utilizing several case studies, was conducted.

2. Literature review of the state-of-the-art

This section presents a comprehensive review of literature that demonstrates the rationale behind the study's intentions and provides information as to why specific methods were chosen and the gaps in the research.

2.1. Indoor fire cycle

Every fire has a lifecycle from ignition to being extinguished, be it through human intervention or natural means. Fire experiences 4 stages during this cycle, shown in Fig. 1a where HRR stands for heat release rate.

The inception stage determines whether the fire continues or not. The presence of fire retardants or insufficient fuel supply can lead to extinguishment; however, if the conditions are favourable, the fire can develop and move into the growth stage. If the flame reaches the growth stage, it becomes a feedback loop with fuel being burnt, increasing the flame heat, thus leading to more fuel being burnt leading to further temperature increases [19]. The growth stage is the most important stage as it dictates the rate of propagation rate of the flame. The propagation is dependent on fuel type and quantity (highly flammable vs slow burning) as well as airflow. If there is a slow burning fuel along with poor air supply, the fire will take a greater amount of time to grow and develop. However, if a flammable substance involved with adequate air supply the fire will develop quickly. The latter scenario is where fast detection of the fire is important, as it can prevent loss of life and minimize damage. When the fire has developed enough, a flashover will occur. This is where the temperature of the flame has reached a critical point where exposed surfaces ignite, and the fire spreads rapidly through space [19]. The flashover time is when it is most hazardous for fire fighters.

If a vision-based system were to be in place during a large fire, it could inform the fire fighters not only about the stage of the fire but also the areas affected. This can help reduce the risk fire fighters are exposed to, along with helping them to make informed and strategical decisions on how to tackle the blaze. After the flashover point, the flame is fully developed and will continue to steadily burn sufficient fuel if available along with air supply. As the fire burns heat and smoke are released, reducing the amount of oxygen in the area. As the fuel or air supply becomes insufficient to sustain the fire, the decay phase is entered. This decay phase will lead to the eventual extinguishment of the fire unless air supply or fuel becomes abundant again.

Another potential use for vision-based systems, aside from detection and alarm sounding is the response. If a fire is properly controlled through fire ventilation systems, the decay stage can be reached faster. The solid line shows the normal fire development shown in Fig. 1b while the dotted line shows the fire development if controlled ventilation is in place [19]. The latter divergent dotted line shows the possible fire development if uncontrolled ventilation is encountered, i.e., window failure [19]. Tien et al. [3,4], which is

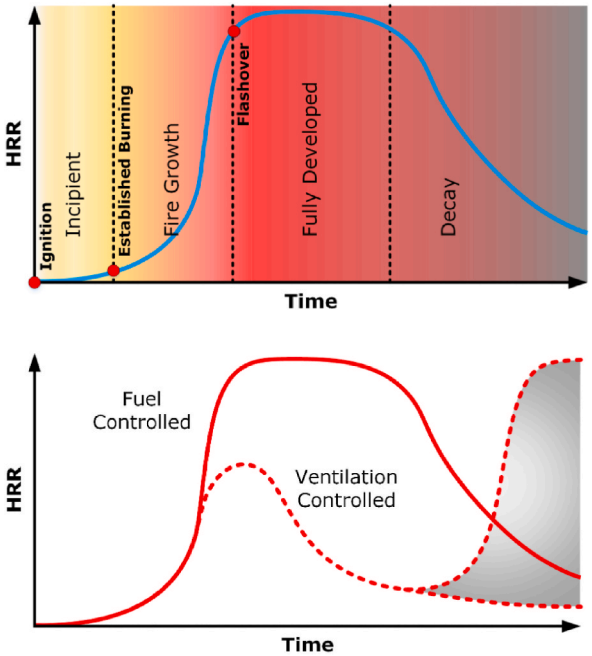


Fig. 2. (a). Graph showing the lifecycle of a fire [19]. (b) Graph showing possible fire lifecycle response to controlled airflow [19].

what this work aims to build off assessed a person and window detection system to aid in the reduction of over or under conditioning in spaces. This aimed to control the ventilation system depending on the occupant's presence and whether the windows were open or closed. If a fire detection system was also integrated into this, then controlled ventilation to minimize the development of the fire could be achieved. This is one of the key reasons to build off Tien et al. works [3,4].

2.2. Current sensors in use

Indoor environments in the UK, by law must have smoke and heat alarms to detect the presence of fire, however, the layout of these sensors is dependent on the risk associated with the environment [1]. Conventional fire alarms divide a building into broad zones and will identify the zone but not the specific area, on a fire alarm panel in case of fire [1]. This is usually installed in smaller or lower risk areas. Addressable fire alarms are more useful for detecting where the issue is, as each has a unique electronic address and can show where the problem is on the fire alarm panel [1]. These are more suitable for larger, high-risk environments such as schools and hospitals. Some current fire detection systems offer direct contact to the local fire service to take an immediate response. These sensors provide early warning signals to occupants in the case of a fire, allowing occupants to evacuate. However, this is all those current sensors offer. As mentioned in the introduction, the fire service responds to a greater proportion of false alarms than actual fire situations. This is partly due to fire alarms, often sensing smoke from non-threatening scenarios such as toasters burning toast, and cigarette smoke. These false alarms are due to a sensor not being able to differentiate dangerous and not dangerous scenarios or a second sensor able to cross-check the detection.

These traditional physical sensors have a number of limitations. They require proximity to fire sources so that they cannot work for the outdoor scenes [20]. The detection time of the sensors is dependent on their placement as well as the speed of development of the fire, meaning that the speed of the alarm sounding is situation dependant. The other key limitation of traditional sensors is them failing to alert occupants to danger. In 2018, 38% of alarms failed to sound when a fire was present, with 45% of these cases being due to incorrect positioning of the system [21]. The likelihood of the system failing depends on how well the alarms are maintained, just like any piece of electrical equipment. However, there is a higher chance of battery-powered detectors failing than that of mains powered sensors with 38% of investigated battery-powered sensors failing to sound while mains powered sensors failed in 21% of cases [21]. These values give a good failure detection rate to compare against artificial intelligence (AI) based systems.

2.3. Algorithms for fire detection

Bespoke or hand-designed algorithms were the initial development stage for artificial intelligence-driven fire detection systems, with some papers presenting promising results [22,23]. This development eventually led to the creation of deep learning neural networks. However, a basis of their predecessors must be established to compare architecture-based papers.

Chen et al. [15] proposed a fire alarm system based on video cameras through a red, green, and blue (RGB) model, which used variables such as saturation (especially in the red region) as well as fire dynamics. The initial decision about a fire region is based on colour thresholds before being verified through iterative checking of growth [15]. However, there were 2 discernible scenarios noted that caused false positives the first was 'non-fire objects with the same colours as fire and background with the illumination of fire-like light sources', and the second was 'background with the illumination of burning fires, solar reflections, and artificial lights' these were highlighted during the extraction phase rendering the algorithm unreliable. It is worth noting these issues as elements to explore within this paper to ensure minimal false positives in difficult scenarios.

Chen et al. [15] considered utilizing the disorderly characteristics of flames in order to verify the fire region further; however, this was exceedingly difficult to achieve due to the random flickers seen in fire due to environmental variables such as airflow paths and combustion materials. The pixel quantity of the potential flame region was compared between 2 consecutive images, and the difference must be over a certain threshold to trigger the alarm. This causes more issues, though, as the setting of each camera put in place will change, meaning there will be a difference in depth of field seen by the camera. With larger spaces, a fire far away from the camera may not reach the fire pixel quantity threshold as it will appear small, thus not triggering an alarm until the fire is exceptionally large. Another method considered was the growth rate of the fire, again using pixel quantities; however, growth thresholds also suffer from the distance a fire is from the camera. Despite this Chen et al. [15] decided that an iterative growth rate would be the better option. Unfortunately, no exact values were stated surrounding accuracy and false positives, so a comparison cannot be made; however, from the abovementioned issues, the model can be considered problematic.

Horng et al. [23] also based their model on colour detection, basing the detection threshold off 70 analyzed images which contained flames. This model enabled the removal of flame like regions in the image through the consideration of colour shift caused through reflections. This is an immediate improvement on Chen et al. [15] colour-based model, as Horng et al. [23] can remove areas of an image that Chen et al. [15] could not, leading to less possible false detections. Furthermore, Horng et al. [23] implemented a colour masking technique to further aid the removal of spurious fire regions. A final layer was added to the model in order to inform occupants of the severity of the flames. This model was tested and run on a Pentium I1 350 processor with 128 MB RAM (unknown speed). This vast amount of RAM meant that the model could be run at 30 frames per second (fps), however, this processing power seems unjustified with no mention of training time or minimum computational power to run the algorithm. Despite this oversight, a detection rate of 96.97% was achieved with the detection of a fire within 1 s [23]; however, no accuracy values were stated. This hand-design algorithm has exceeding positive results especially surrounding indoor fire detection; however, with no further supporting values, it is difficult to fully assess the model.

Çelika and Demirel [22] in 2009 used a rule-based generic colour model for flame pixel classification with the aim to reduce the effects of changes in illumination and improve overall performance. Using their model, a 99% accurate detection was achieved; however, also incurred a 31.5% false alarm rate. This is low compared to other pixel-based works [15,24] however is significantly

higher than works using CNN training. Though Çelika and Demirel [22] claim the model proposed is more robust to illuminance changes; however, this work is focused on the external environment with a focus on forest fires. It should be noted, however, that it is recommended in the work that dense sensor distribution of sensors is needed for a high accuracy detector system.

Çelik [24], in 2010, 5 and 6 years after the works produced by Chen et al. [15] and Horng et al. [23] also proposed a bespoke algorithm based on a colour-based model building on previous works [22,24]. The algorithm can be broken down into 3 phases: fire pixel detection using colour information, detecting moving pixels and analyzing dynamics of moving fire pixels in consecutive frames.

The system was trained and tested on a frame size of 320 × 240 pixels achieving up to 30 fps (dependent on the amount of fire in the frame) [25]. It is unclear how long the model took to train and the computational power needed to run it; however, the testing process covered indoor and outdoor environments consisting of dark and light scenes. This is a more rigorous testing procedure compared to previous papers, clearly demonstrating the system's capability in all environments. Unfortunately, the indoor sequences in question aren't clear with how the fire develops, thus the potential of not testing extreme circumstances, i.e., ignition and propagation through space (most likely due to the inability to produce and record a full fire and lack of access to data). Despite this, a detection rate of 99% was achieved however the method of working out the detection rate combines fire and non-fire video segments, so it is inconclusive what the accuracy for fire detection is. Çelik [25] notes that the system could be used along with pre-existing fire detection systems to minimize false alarms but can be used as a standalone system.

Kong et al. [26] focus on the issue that many existing fire detection methods based on computer vision technology have achieved high detection rates, but often with unacceptably high false-alarm rates highlighted through Çelika and Demirel [24]. This paper uses logistic regression and temporal smoothing for fast fire flame detection [26]. A potential flame area is highlighted through colour and motion, after background subtraction. Logistic regression is then implemented to determine whether the region contains fire through size, colour, and movement features. This finally has temporal smoothing applied. The smoothing procedure restricts the region of interest (ROI) from large or irregular movements from frame to frame [27]. Temporal smoothing reduces the model's accuracy however, it reduces the false alarm rate caused by the model [26]. This method demonstrated that the proposed method was successful at determining fire regions in indoor and outdoor scenarios with a fire detection accuracy of up to 98% with an average detection speed of 1.8s improvement over Wang et al. [28] detection speed of 6.58s. However, issues were noted with false alarms, but no false detection rate was stated. It is also worth noting that the detection speed is scene specific and so will not be comparable to the model proposed through this paper.

Marbach et al. [29] presented a bespoke model in order to detect fire in video clips irrespective of setting. The image is assessed for a flame region, with any fire features extracted. This is then assessed for any fire like patterns. However, the interesting feature is that the fire alarm is only triggered if the fire pattern in the flame region persists for a specified period of time (around 60s though this was increased based on the false alarm rate [29]). This in-built threshold would mean that any false alarms should be exceptionally low in occurrence; however, a minute for an alarm response can be a significant period of time during the ignition stages of fire for highly flammable substances. The algorithm was run on a digital signal processor connected to CCTV cameras running at 25 frames per second (fps). This high fps is only achieved through a low resolution of 288 × 352 pixels [29]. The training took 15–20 days. The algorithm was then tested for 40 days with only 1 false alarm in the first week. It was noted that certain camera positions lead to non-optimal conditions and so increasing the false alarm rate. This testing was completed in controlled lab environments, demonstrating the algorithm's potential, but real-world testing could significantly change the accuracy and false alarm rate due to fire-like regions such as red t-shirts.

Despite Marbach et al. [29] implementation of a persistent flame detection before alarms were triggered, the solution of increasing flame detection time to reduce false alarms would mean that the detection time was increased to the point of the system being obsolete. It should be noted that the time could be adjusted to optimally suit the chosen environment; however, this could imply that it may not be suitable in some environments. This, alongside no accuracy figures or false alarms rates were stated only as a statement of robustness, shows the limitations of the work. It is clear from this that indoor detection could be possible. However, a suitable system needs to have consistency in terms of accuracy when exposed to a variety of environments in order to consider the system useable.

Töreyin et al. [30] proposed another colour-based detector utilizing methods from various papers previously published. Initially, moving pixels/regions are found through background estimation taken from [31]. These locations are bundled together and labelled using an algorithm from [32]. This process produces a pixel map. Pixel colour values are then taken and cross-referenced against a predetermined distribution. This will alert if a value represents a possible flame region. This predetermined distribution was taken from fire regions in multiple images. Possible colour values are plotted on a three-dimensional point cloud in the RGB colour space [33]. The frequency history of each pixel in the fire-coloured region is assessed, meaning that flicker can be determined, but a minimum of 20 fps is needed in order to achieve this. A final step of spatial wavelet analysis is done on a rectangular frame containing fire-coloured moving regions, before using fusion decision strategies which determine if a region contains fire or not by assessing each stage of the algorithm [30]. Töreyin et al. [30] stated that the algorithm could detect fire in the 19 sequences containing fire out of 61 sequences. However, accuracy and false alarm rate value were never stated.

Most of the bespoke algorithms assessed in this section focus around an RGB model to detect fire regions, some with background subtraction beforehand and some with further analysis after; however, despite achieving high accuracy rates of detection (for the papers that found this value [22,25,26] a common theme is a high false alarm rate. To reduce this, various strategies were employed, but those investigated would lead to slow detection. Furthermore, several papers never stated values allowing comparison or baselines to be assessed.

2.4. Deep learning driven models

As mentioned previously, the initial stages of object detection were bespoke or hand-crafted algorithms shown in Section 2.3. These

algorithms often extract shallower features compared to deep learning networks [34]. As stated by Zhao et al. [34], bespoke algorithms often combine multiple low-level image features with a high level-context; however, this often leads to performance only being maximized to a certain extent. Object detection evolved through deep learning networks, which can learn deeper features than that of hand-crafted algorithms. This led to the possibility of a high accuracy detection and, eventually, transfer learning which removes the significant time investment needed to optimize the network. This change was brought about in 2012 when a CNN architecture called AlexNet [35,36] won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competition, achieving half the error rate of the second-best model [36]. CNN, since then, has become the most recognizable architecture in the field and has been developed into Faster R–CNN, which is one of the most accurate models that exist in the field of object detection.

Jadon et al. [12] note that fire detection has been an active area of research for both bespoke models and deep learning approaches. Papers demonstrating deep learning technologies have been shown to outperform hand-crafted algorithms due to the vast potential of CNNs and variants but also offer the possibility of automatically extracting meaningful information. Jadon et al. [12] assessed a modified MobilNetV2 model to demonstrate that the current CNN technology can further outdo hand-crafted frameworks due to its low complexity, low cost, and high performance. A number of networks were considered for the assessment, such as MobileNets, GoogleNet and SqueezeNet. However, MobileNetV2 was chosen over the others due to its lower number of parameters in comparison (3.4 million parameters). It possesses a faster speed, especially for computer vision tasks. This lightweight network retains the same accuracy as that of the other mobile models while reducing the number of operations and memory required. Jadon et al. [12] used other works [37,38] as baselines for comparing how the modified MobileNetV2 performed. The results noted show that this model outperforms the baselines in accuracy, precision, false positives, and F-measure metrics. It also achieved 5 fps which was also higher than Muhammad et al. [39,40] and Foggia et al. [38], who both ran their networks on Raspberry Pi.

Muhammad et al. [39] used a fine-tuned CNN for CCTV cameras to detect various indoor and outdoor environments. Jadon et al. [12] take the detection one step further by proposing an adaptive prioritization mechanism for cameras in the system. Unfortunately, the results in the paper demonstrate fire detection, however, only at a minimal level. This system proposed purely would act as an alert system to emergency services if it were implemented; however, it does not provide further information about the propagation and development of the fire. Despite Muhammad et al. [40] claiming that the system can detect fire in indoor and outdoor environments, there is little to suggest in the paper that indoor situations achieved significant accuracy with average accuracy values excluded. Similarly, the region where fire occurs is not highlighted by the network. Muhammad et al. [40] highlight the necessity to demonstrate different lighting conditions, reflections, and other real-life occurrences to be tested and demonstrated to prove that such a system is accurate.

Zhang et al. [13] proposed a CNN-driven patch classifier for outdoor fire detection. This functioned in a step-by-step process, initially testing the entirety of an image (global level) to see whether the frame contained any possible fire regions before either moving onto the next frame or further filtering of the image is completed. The filtering consists of a fine grain patch classifier that detects the location of the fire. Zhang et al. [13] took several videos and sampled an image every 5 frames. This meant that all stages of the fire would be taken into account through deep learning. Thus, the network should be able to identify a fire from ignition to substantial size and propagation. This approach is needed for a deep learning approach so that the network can understand the full variation in scale and shape of a flame region. The frame size used was 240 $\times$ 320, similar to bespoke model tests [25,29]; however, this primarily focussed on outdoor environments.

Two types of the classifier were tested in this paper; the first was a linear classifier, support vector machine (SVM) [41] and a nonlinear classifier, CNN [42]. Along with the 2 of these classifiers, a CIFAR 10 network was used. The SVM with the CIFAR 10 network, was noted that due to the linear working of the classifier, multiple patches had to be tested per frame due to the lack of global information. This meant that this model had a high computational cost [15]. The image could be initially identified as positive or negative (of containing fire) before the patch classifier identified the fire region in the image to solve this issue. To do this, a pre-trained network was implemented, this is known as transferred learning and is a good way to achieve high accuracy when using small datasets [15]. AlexNet was used as the pre-trained network consisting of 8 layers. The issue of using such a patch classifier is that the patches are a fixed size meaning the fire could not be fully constrained by these regions meaning there could be issues of the model not fully learning all the features associated with a fire region.

This shows that the CNN had a higher accuracy (in all cases) and a higher detection rate (apart from Pool-5 in training set 2). This, along with an, on average lower false alarm rate, demonstrates the benefits of using a CNN over an SVM. The use of transfer learning meant that CNN could perform better than the SVM and highlighted the benefits of using such a technique. Zhang et al. [14] and the results released the training files to provide benchmark data for other patch-wise detection systems that could be useful in the training of the model outlined in the methodology section of this report.

Zhang et al. [43] note that there are multiple benefits of using an R–CNN over traditional models, such as not needing to manually extract features as well as high processing speed. Along with an R–CNN, ZFNet [44] was chosen as the detection model. Zhang et al. [43] also investigated a method to detect forest fires by taking advantage of deep learning. A Faster R–CNN was used, and custom-made smoke images to train it. To achieve these custom images, videos of smoke were taken against a green screen. It should be noted that real videos were also used alongside the synthetic images to provide a more well-rounded data set.

Using a green screen meant it was far easier to extract the smoke regions in a frame. 2800 smoke frames were taken from 10 different videos recorded [43]. These frames then had the smoke region extracted through pixel probability of a pixel being smoke using a curve representing RGB space. This probability was then used alongside thickness (the concentration of smoke). This process extracts most of the smoke region; however, areas of less density are phased out and so does not capture every aspect.

The smoke frames were then applied to 12620 forest background images in various positions [43]. This method of creating training data means that an enormous sample size can be created. This means that the network training can fully adapt to various scenarios and

backgrounds. This is not normally achievable as raw footage of fire or smoke is often scarce with the small number available, meaning that in-depth training for various scenarios cannot be achieved. However, as mentioned previously, the issue surrounding the image processing not fully extracting all smoke particles from the image means that the network may not be able to detect thinner areas of smoke. Also, different substances burning produces different colours of smoke, and so without a range of substances burned, colour variation may not be dealt with well by this sensing network. The results gained from using this training technique alongside an R–CNN showed that the model was not sensitive to thin smoke but achieved high detection rates for most videos tested.

No accuracy or false alarm data was noted in the paper; however, the results showed that using synthetic images could be a very suitable training method to help with the lack of available training data and still achieve detection within the model.

Wu et al. [36] conducted a study looking at a deep learning vision-based system specifically looking at indoor detection for chemical plants such as a petroleum factory. These are high risk environments with highly flammable substances. In buildings such as this, the fire detection system was already vision-based, however, with a human operator. This is an unrealistic task for 1 person due to the huge number of video feeds, thus an automatic detection mechanism could remove any human error in this process. The method proposed can be broken down into 3 key steps: motion detection, fire detection and classification.

Wu et al. [36] state that other papers which utilize a combination of bespoke feature extractors for the ROIs and then use a network model to detect the fire in the ROI [45,46] destroy the purpose of utilizing a CNN as it won't learn the fire characteristics. Thus, utilizing a purely architecture driven approach should achieve more suitable results. The method utilized by Wu et al. [36] starts through utilizing background subtraction. This is done in order to determine the size of the remaining white region; only if the white region is over a threshold value will the image be processed by a CNN [36]. If the CNN processes the image, the fire region locations are determined. Issues were found surrounding false detection, so a region classification model was used to determine whether the region was a fire or not. With the fire detected, an immediate alert would be sent to staff. The dataset used consisted primarily of chemical plant fires and image datasets such as ImageNet. The dataset comprised 5075 images, with 80% used for training and the rest used for testing. This was run on a computer with an NVIDIA 1080 GTX GPU. The training was completed for 20000 steps. Wu et al. [36] achieved a detection rate of 98.4% of fire regions. No accuracy values of the detected regions were found. This paper [36], was the only one that specifically assessed an indoor environment using a deep learning architecture; however, the model may not be suitable for general purpose in other indoor environments due to the training being specific to heavy industrial environments. Further details about the CNN model and its application in detections systems are available in the later section (Section 3.1.2).

2.5. Transfer learning

This study intends to build on works completed and currently underway completed by Wei et al. [5,18] and Tien et al. [3,4,16,17]. The paper [3,4] proposed a new deep learning approach for energy management and optimization of HVAC systems. This focussed on predicting cooling and heating setpoints for HVAC systems using a computer vision system coupled with a deep learning algorithm. This was done to reduce the amount of over conditioning in buildings to help reduce the overall impact HVAC systems have on energy consumption and CO2 emissions. This method utilized a CNN network to detect occupants and their activities with class names of 'standing', 'sitting', 'walking', 'napping' and 'none'. Average model accuracy was reported at 76% [3].

From this, an occupancy profile was generated and fed into a building energy simulation (BES) model and sensible heat gain profiles were obtained and compared. The profiles showed that using this artificial intelligence-driven computer-driven system can more accurately predict the sensible heat gain than standard setpoint profiles [3]. This reduced the amount of cooling and heating provided to space. Tien et al. [16] built upon this initial work and improved the deep learning model's accuracy to 89.3%, where the increase in detection accuracy suggested a decrease in the predicted gains by 30.56%. These papers used transfer learning. Transfer learning utilizes pre-trained DNNs (Deep Neural Networks), which streamlines the optimization of the network. DNNs require a plethora of training samples along with significant computational power to initially train and optimize [47]. If the results gained from the model outlined in the methodology yield positive results, then the possibility of integrating the fire detection into Tien et al.'s [3,4,16,17] model would allow ventilation control as an immediate response to tackle fire as well as taking a step towards an all-in-one detection system which will be the inevitable evolution of building systems.

2.6. Literature gap

Fire poses a significant risk across industrial and domestic settings, especially to firefighters who must tackle the blaze. Through looking into the current systems in place in the U.K., baseline figures of missed detection were found to be between 21% and 45% [2]. This gives a target for vision-based systems to perform at, and so this study will aim to achieve a missed detection rate lower than 21%. All the papers assessed have varying reported values, with some papers only reporting detection rate while others provide accuracy and false alarm values. This inconsistency makes it difficult to draw comparisons. Along with this, a number of papers did not provide technical specifications of the computing system that was used for the training and testing.

Furthermore, only 2 papers report either training time [29] or global steps taken to train [36]. The combination of technical and training information would allow assessments of complexity and performance to be put into better perspective as a model trained for 500 steps would be expected to be inferior to that of the same model trained to 20,000 steps. This lack of consistency makes it difficult to gauge the performance of different models. Therefore, in this work, all technical specifications and training information will be clearly outlined along with assessment criteria of accuracy, false detection (false alarm rate), missed detection (inverse of detection rate) as well as the speed of detection where applicable. This will enable comparisons to be drawn across several papers as well as act as a versatile comparison to works in the future.

On top of this inconsistency, most of the papers based on bespoke or handmade algorithms assessed indoor environments; however, the deep learning papers mainly focussed on a wildfire or outdoor environments for fire detection. Only Wu et al. [36] focused on an

indoor environment; however, this was a specific industrial environment meaning the model produced may not be suitable for other environments. This is true of the vision-based fire detection research sector. Most papers focus on outdoor environments with little to no focus on indoor environments. Through this work, a deep learning vision detection system will be proposed (through transfer learning) for indoor environments. This will be tested for both domestic and commercial (offices) settings. As outlined in this section, there could be a number of benefits of using vision-based systems in indoor environments, such as fast detection, propagation tracking, informing fire fighters, and combining detection with an HVAC system to allow controlled ventilation to aid in the decay of the fire.

In order to achieve this, a full lifecycle of fire will be used rather than a number of non-consecutive fire images. This will show the capability of the proposed model by using a small dataset that will be gathered similar to [37].

3. Method

3.1. Deep learning method

To a great extent, traditional computer vision techniques have been continuously developed to become relevant to current solutions and applications. Computer vision has been recently adopted with deep learning methods to provide effective vision-based solutions such as object detection. This has been extensively developed due to its popularity for video surveillance and crowd counting applications. Based on the methods used to form the following types of object detectors, the deep learning method of CNN would be best suited and was adopted to propose a fire and smoke detector. CNN is a class of artificial neural networks that have become dominant in various computer vision tasks [48]. Unlike other neural networks, CNN requires data in the form of images which leads to its specialism in providing good performances within classification tasks that involve images and videos [48].

To develop such a CNN model-based fire and smoke detector, a general workflow process highlighted in Fig. 3 will be followed. It consists of selecting the suitable deep learning model and input data to processing the input data and the configuration of the model for training. Subsequently, the trained model would be deployed to form an AI-powered camera, ready to perform detection. Effectively, a deep learning method based on a neural network is employed to provide a computer vision-based approach that enables real-time fire and smoke detection within an indoor office-based environment. The following section presents further details corresponding to each stage of the workflow.

3.1.1. Data preparation: datasets and pre-processing stages

To form the computer vision-based detection model, input data in the form of RGB images were gathered. These images established both the training and testing image datasets. Pre-processing of the gathered images were performed with data augmentation for each image in both the training and testing dataset. This includes ensuring all images were non-identical, or at least with slight variation between each other. Techniques based on the TensorFlow Object Detection API workflow process were employed. This enabled images within the dataset to not have a restricted size. Therefore, to enable the model to learn all aspects related to fire and smoke, a diverse array of images with high variation in the image pixel densities were collected. Furthermore, some images consisted of fire or smoke only regions, and some images consisted of the presence of both elements within indoor environments of various types of buildings such as office spaces and dwellings. Unlike Zhang et al. [43], the smoke regions of both heavy and light smoke were covered. Furthermore, unlike Jadon et al. [12] and Zhang et al. [43], the fire regions within the images covered all stages of ignition, propagation, and flashover (Refer to Fig. 2).

It is important to note that some training images contained regions of possible error, i.e., reflections in screens and light fittings. Through including these regions, this would ensure the training procedure would enable the formation of a more robust model. Overall, search engines provided limited image sources based on the search for indoor fire and smoke scenario-based images. Hence, this led to training the models on a small dataset. As given in Table 1, various sizes were collected, giving a total of 480 training images, along with 120 testing images. The total number of images was significantly lower than [49] but was slightly larger than the dataset of [12]. However, due to the coverage of all the stages of a fire's growth and the performance under different scenarios, there should be a certain amount of robustness built into the model through the training images used. Furthermore, utilizing images with a diverse range of pixel densities from greater than 1080p resolution to below standard CCTV resolutions suggests that the model should be able to function with the provision of a diverse range of video streams with different quality of videos. Effectively, based on the suggestions by [50] on the effects of image resolution on the accuracy of models, it suggests that the large variations in the image datasets should enable the provision of a balance between the fire and smoke detection model accuracy with the training speed achieved.

Fig. 4 presents an example of both fire and smoke images located within the training and testing image datasets, along with the

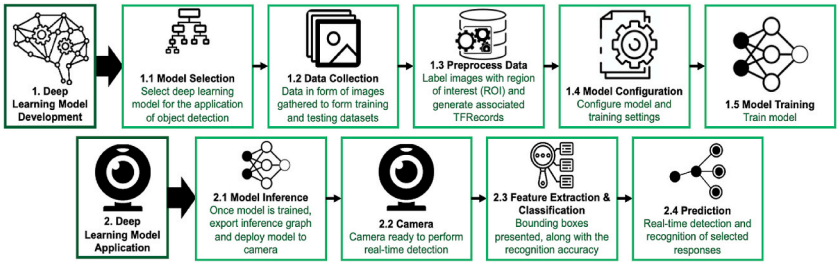


Fig. 3. Deep learning model workflow procedure for developing and applying the computer vision-based indoor fire and smoke detector.

Table 1
Description of the image training and testing dataset.

Category	Number of Images			Number of Labels		
	Training	Testing	Total	Training	Testing	Total
Fire	320	80	400	391	91	482
Smoke	160	40	200	237	65	302
Total	480	120		628	156	

process of how images were individually labelled using the software, LabelImg [51]. For each of the given images, labels in the form of bounding boxes were manually assigned entirely around each specific region of interest. For some images, multiple bounding boxes were assigned to enable the covering of as much area of the fire or smoke as possible. Furthermore, for some instances where images consisted of both fire and smoke at a close distance, it resulted in the assignment of overlapping of the bounding boxes.

3.1.2. Fire and smoke detection model: model selection and configuration

Based on the workflow process indicated in Fig. 1 and the selection of the CNN as the most suitable deep learning model to provide the fire and smoke detector, the TensorFlow Object Detection API [52] was employed as the framework platform in this study. It is an open-source framework used for constructing, training, and deploying deep learning models. Its detection and track-based application are based on the transfer learning approach, which stores and utilizes the knowledge obtained from one problem to cope with other similar but different issues. The transfer learning approach can effectively implement detection tasks with high detection performance while using reduced network training time and less requirement of the input dataset.

As the Model Zoo [52] mentioned, various models have been successfully pre-trained with some commonly used open datasets such as the COCO (Common Objects in Context) dataset and provided by the TensorFlow Object Detection API platform. As the architectures of the networks were predefined, it contributed to a fast training and deployment of desired detectors. For the present study, two models, the Faster R–CNN inception V2 and the SSD MobileNet V2 from the provided Model Zoo, were used to train two individual models separately.

In the Region-based Convolutional Neural Network (R–CNN) [53], the original model initially scanned the given image for possible objects and generated thousands of proposed regions. These regions then had a CNN run on top of them before finally feeding the output of the CNNs into a support vector machine (SVM) layer, which classifies the region [53]. The detected and classified region is then represented through bounding boxes [34]. The development of the original R–CNN through running feature extraction on the initial image before the proposed regions are selected leads to a massive reduction of the number of possible regions meaning that the speed of completing the detection is greatly increased for each image [53]. Along with this, the SVM was replaced with a SoftMax classifier, meaning that multiple SVMs are not required to be trained [34]. The performance difference between an SVM and SoftMax is minimal and often is considered comparable [54] and so is not considered a significant change for detecting one or two classes. With these augmentations to the R–CNN, the Fast R–CNN was produced [55]. With further development, the most up to date model is the Faster R–CNN. This is built on the Fast R–CNN by replacing the search algorithm, which generates the possible regions to a region proposal network (RPN) [56]. This once again improved the speed at which the detection process is completed. Single-shot detector (SSD) models [57] achieve detection in a different pipeline compared to that of the R–CNN. The SSD first passes the given image through multiple convolutional layers giving multiple feature maps of different scales [35]. Each of these feature maps has a convolutional filter applied to them to assess a set of bounding boxes [34]. Each box has an offset, and class probability predicted [33], with the best box being chosen and labelled. Compared to R–CNN and Fast R–CNN, the Faster R–CNN has a much faster running speed

Fig. 4. Example images gathered from Google Images of fire and smoke and the labelling using the software, LabelImg [51].

which can greatly benefit the live detection tasks.

As solely based on speed, the SSD types of networks perform faster than the Faster R-CNNs. However, at the cost of accuracy, the SSDs with Inception-v2 and MobileNet would become the most accurate of the fastest models. Furthermore, Tsang [50] suggests the later version of MobileNet (V2) shows slight improvements compared to the standard MobileNet model, and if the computational cost was not considered, then the Faster R–CNN architectures possess the highest performance in terms of accuracy [50]. These suggestions were derived from the evaluations made by Tsang [50], where both Faster R–CNN models were compared with SSDs in terms of the mean average precision (mAP) achieved across GPU time and with the performance under the operations of the models with a CPU and GPU.

Zhao et al. [34] noted that through their testing, a Faster R–CNN demonstrated improved performance in terms of classification over SSD as the SSD model is more prone to errors. SSD can also underperform in the detection of small objects. Therefore, it is expected that the Faster R–CNN will perform better. However, to what degree the SSD performs in fire detection is unknown. Hence, for this present study, two training models were selected to compare the model performance based on two different model configurations. The Faster R–CNN with Inception V2 was chosen to train the first model. This was indicated as Model A, with the model configuration shown in Fig. 5a, which was influenced by [58,59]. Specifically, the Faster R–CNN with the inception network of 'Inception V2' was selected as it offered a good middle ground between accuracy, memory usage by both CPU and GPU, as well as the speed of the model. The SSD MobileNet V2 was selected to train the second model. This was indicated as Model B, and from the influence by [60], the configuration is presented in Fig. 5b. This was chosen because the MobileNet V2 had a similar mean average precision (mAP) to that of the faster R–CNN inception V2 model [52], with little computational demand. It should be noted that we did not attempt to modify or improve the existing models in this study. However, it will be one of the strategies that can be considered in future works to enhance detection performance.

3.1.3. *Performance evaluation of the trained detection models*

Once the detection models were trained, model performance evaluation was conducted by applying the test images assigned from the test dataset (Table 1). This presents the results of the performance in terms of a confusion matrix, with the identification of the true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) for all fire and smoke labels. To further assess the detection performance of the proposed models, the measures including accuracy, precision and recall were employed, which can be obtained from the generated confusion matrix based on the expressions in Eqs. (1)–(3). Accuracy is defined as the percentage of correct predictions out of all the predictions. Precision, which calculates the number of correct predictions among the predicted positives, can measure the exactness of the model. Recall, which calculates the number of actual positives labelled as TP by the model, can measure the completeness of the model. In order to better quantify the detection performance, F_1 score, which combines precision and recall measures, was employed and expressed as Eq. (4). Respectively, the corresponding results are presented in Section 4.1. Furthermore, these common evaluation metrics were also used to analyze the detection performance under a series of tests. This is given in Section 4.3.3, with the performance results conducted under a series of video tests.

$$Accuracy = \frac{(TP + TN)}{(P + N)} \tag{1}$$

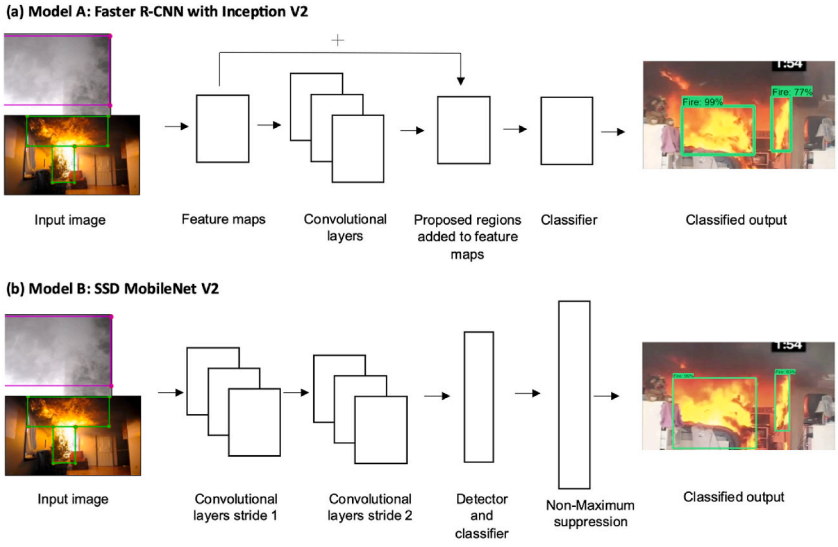


Fig. 5. General outline of the pre-existing convolutional neural network (CNN) architecture used for the training of the fire and smoke detection models based on a transfer learning approach. a) Model A: Using the Faster R–CNN with Inception V2 model [50] and b). Model B: Using the SSD MobileNet V2 model [58,59].

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F_1 \ \ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

3.2. Testing of the fire and smoke detection models

As presented in Fig. 3, a series of video feed tests were conducted to evaluate the performance of the trained models. As summarized in Table 2, five different videos were used with the two trained models, indicating a total of ten tests.

Video 1 shows a mock-up of a child's bedroom with many materials and basic fittings in the way of lights. It also contains an old television screen which allows for evaluating the potential effect of reflections. This scene then develops into a full flashover event, as shown in Fig. 6. The domestic lighting and reflections provide areas of potential false detections. Along with this, there is a variety of different coloured props, which will also test the robustness of the models.

Video 2 showed a scenario with a flashover demonstration. This allows the assessment of the models in a heavy smoke setting and in a large fire setting. The significant stages of the fire are shown in Fig. 7. The fire initially starts behind a sofa, and so obscuring the first moments of the fire. The fact that the fire is initially hidden will also test whether the speed of detection is superior even in disadvantaged scenarios (in comparison to smoke and fire alarms).

Video 3 was a CCTV video of office space; the video shows an electrical incident where a laptop explodes and sets fire to the space. Fig. 8 shows some of the main stages: before, during the explosion, the initial fire, fire with lighter smoke, fire with thick smoke, and developed fire with thicker smoke. The initial stages of the video are in black and white for night-time capture but turn into coloured recording as the light intensity increases from the fire. This process will test the models against a real-world scenario while detecting through a standard CCTV resolution.

Video 4 (Fig. 9) presents a typical residential setting of a living room. The fire started off from the curtains of the window, behind the two-seater sofa and spread towards the back corner of the room. For this experimental test, only the first 15 s of the original video [64] was used to test the trained fire, and smoke detector, as this segment presents the occurrence of the fire and smoke within the room.

Furthermore, Fig. 10 presents the key stages of the Test 5 video. This video represents a test scenario under controlled conditions developed by the BRE Trust [64] used to develop fire protection systems. Hence, this was selected to test the developed vision-based fire and smoke detector. This scenario involves a simulated working office workstation under an open ceiling. The fire started underneath the table and spread towards the whole region, whereby smoke was generated seconds after the fire.

As detailed in each description, all videos possess features that will aid in testing the robustness of each, such as reflections, movement, response time, and colour. These videos were taken from the links in the associated references [61–65].

4. Results and discussion

4.1. Detection model training results and evaluation

The models were trained using the graphics processing unit (GPU) NVIDIA GTX1080. Both models were trained using the labelled images in the training dataset detailed in Table 1. Due to the differences in the model configuration and processes (highlighted in Section 3.1.2), both models were trained for different durations until they reached a converged level. Hence, variations occur within the total training steps, with the average loss and minimum losses achieved. The results of the training of both models are summarized in Table 3.

Once the models were trained, the 120 images within the test dataset (Table 1) were used to assess the initial recognition performance of both models. A total of 156 labels were assigned to the images. The results were analyzed using a confusion matrix (Fig. 11). Model A provided a correct classification of fire up to 74.73% and 93.85% for smoke. The only drawback of Model A is that the response for fire is sometimes predicted as smoke (10.99%), and for some instances, predictions were not made. In comparison,

Table 2
Summary of video feed tests.

Test	Video	Model Used for Training
1a	Bedroom Fire Test – LancashireFire [61]	Faster R-CNN with InceptionV2 (Model A)
1b		SSD MobileNet V2 (Model B)
2a	Flashover Demonstration – OakRidgeFD [62]	Faster R-CNN with InceptionV2 (Model A)
2b		SSD MobileNet V2 (Model B)
3a	Laptop Explodes and Burns Down Office Building – 986613 – RM Videos [63]	Faster R-CNN with InceptionV2 (Model A)
3b		SSD MobileNet V2 (Model B)
4a	Living room fires with and without a fire sprinkler (Timecode) [64]	Faster R-CNN with InceptionV2 (Model A)
4b		SSD MobileNet V2 (Model B)
5a	Water Mist Fire Demonstration [65]	Faster R-CNN with InceptionV2 (Model A)
5b		SSD MobileNet V2 (Model B)

Fig. 6. Key stages in Test 1 video [61].

Fig. 7. Key stages of Test 2 video [62].

Model B provided lower performances with lower classification performances and a higher percentage of achieving false and no detections and classifications.

Table 4 presents the results based on the common evaluation metrics detailed in Section 3.1.3. The accuracies achieved reflect the results presented in the confusion matrix in Fig. 11. As given by the F_1 Scores, it indicates that both models should be capable of carrying out detection of both fire and smoke. Overall, Model A performed better than Model B.

4.2. Framework test results

This section presents the performance of the two trained models (Model A and Model B) using the five selected videos. Following the summary of the tests in Table 4, each model was applied separately to each video. Fig. 12 shows a preview of all the video feed tests performed, along with the links to Videos 1–5 showing the achieved detection results. Each test performed was based on the actual duration of the videos from the original sources. It should be noted that in practice, the device won't be storing or outputting images or

Fig. 8. Key stages of Test 3 video [63].

Fig. 9. Key stages of Test 4 video [64].

videos. It will only output real-time information on the number of instances based on the bounding boxes presented for fire and smoke.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.jobe.2022.105154

4.2.1. Test 1 detection and recognition results

Fig. 13 shows a more comprehensive overview of Test 1a and Test 1b, with key stages of the video captured. Before the fire starts, there are multiple camera shots around the room, with various items being put in place before the demonstration starts. These items being put into place in case some false detection from both the faster R–CNN and SSD models; however, the faster R–CNN false detections only last for fractions of a second while the SSD error is over a few seconds. Both the Faster R–CNN model and the SSD model are slow to react to the ignition with the faster R–CNN initial detection after 21 s for fire, while the SSD model took 24 s.

Through the video, the faster R–CNN and SSD perform similarly, with both overall detection rate as well as accuracy; however, the faster R–CNN did not achieve any correct smoke detection over the collection period. Despite these shortcomings, the reflection caused by the tv screen opposite the bed caused no false detection, which shows significant robustness of the models. Overall, the SSD

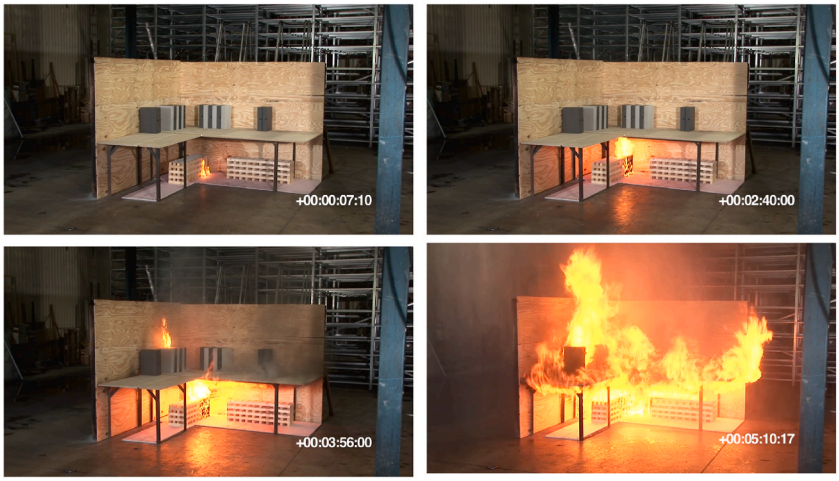


Fig. 10. Key stages of Test 5 video [65].

performed marginally better than the faster R–CNN in both detection rate and accuracy; however, the SSD did have a higher error in detection for both fire and smoke.

4.2.2. Test 2 detection and recognition results

Given in Fig. 14 for Tests 2a and b, the fireman sets up the scenario at the beginning of the video in which the faster R–CNN falsely detects this as smoke for brief moments followed by false detection of smoke in the background but disappears when the fire is detected. The faster R–CNN detects the fire within 4 s of appearing in the frame, while the SSD takes over 30 s before an initial detection; this is well after the SSD time as well as the fire alarm, with it sounding off after 16 s. This slower detection time of the SSD can be seen as a consistent trait through all of the tests and may show non-suitability towards fire detection where response time in the form of initial detection is key. The faster R–CNN model performed very well through this video with a high detection rate and accuracy, while the SSD did not perform nearly as well with significantly lower overall detection, a greater amount of false detection as well as lower overall accuracy of smoke and fire detection than that of the CNN.

4.2.3. Test 3 detection and recognition results

For Test 3a and b (Fig. 15), despite the video initially being in black and white due to a night-time capture mode along with flashing lights on the phone in the office, neither model had any false detection before the laptop explodes. The Faster R–CNN (Test 3a) initially detects the smoke from the explosion before clearing and taking under 10 s to detect the following fire. Unlike the SSD (Test 3b), which takes over a minute before the first fire detection is achieved. The Faster R–CNN achieves a good detection rate even through the increasingly smoke-filled environment, while the SSD is patchy with some periods of detection working well and others not through the smoke. The faster R–CNN performed with high detection rates and high average accuracy while the SSD had lower detection rate for both fire and smoke and lower average accuracy compared to the Faster R–CNN. This scenario proves viable night-time detection, i.e., greyscale images, with good resilience to high density smoke scenarios demonstrating detection flexibility and so is still comparable to current fire and smoke detectors.

4.2.4. Test 4 detection and recognition results

Fig. 16 presents the test results of catching fire in living room within a domestic building without fire sprinklers in daytime. This scenario presents a well-lit room with high contrast image between the light and window and the rest of the room. The fire started at the beginning of the video. At around 16s, fire presented next to the curtains behind the sofa and in front of the curtains on the left side of the living room, and at 22s, smoke also appeared within the scene. However, none of them was detected by the Faster R-CNN or SSD model. As time went by, fire kept rising towards the top of the curtains and was first detected by the Faster R-CNN at 27s. After that, the Faster R-CNN could continuously detect the fire, as shown in Fig. 16a (Test 4a). Yet, in the whole process of implementation, the SSD could not recognize fire in the video. It indicated a higher fire detection ability of the Faster R-CNN in comparison to the SSD. It should be noted that smoke was not identified by either model, suggesting that further improvements are required to enable a higher smoke detection rate in different circumstances for both models. It should be noted that this scenario is the only one with a window in the video frame. This bright element could impact the high missed smoke detection.

4.2.5. Test 5 detection and recognition results

For Tests 5a and b (Fig. 17), a scenario was set up by the fireman to demonstrate the water mist fire which started below the table, this is comparable to an industrial scenario. The implementation of detection and recognition was presented in Video 5. Within 5 s from the start of the video (1min 47s from catching fire), the Faster R-CNN began identifying fire in Test 5a. At around the same time, smoke started to appear within the scene. However, none of them was detected in Test 5b. At 54s in the video (3min 55s from catching fire), fire propagated and appeared on top of the table. The Faster R-CNN detected this situation straightaway, while the SSD took

Table 3
Summary of the model training results.

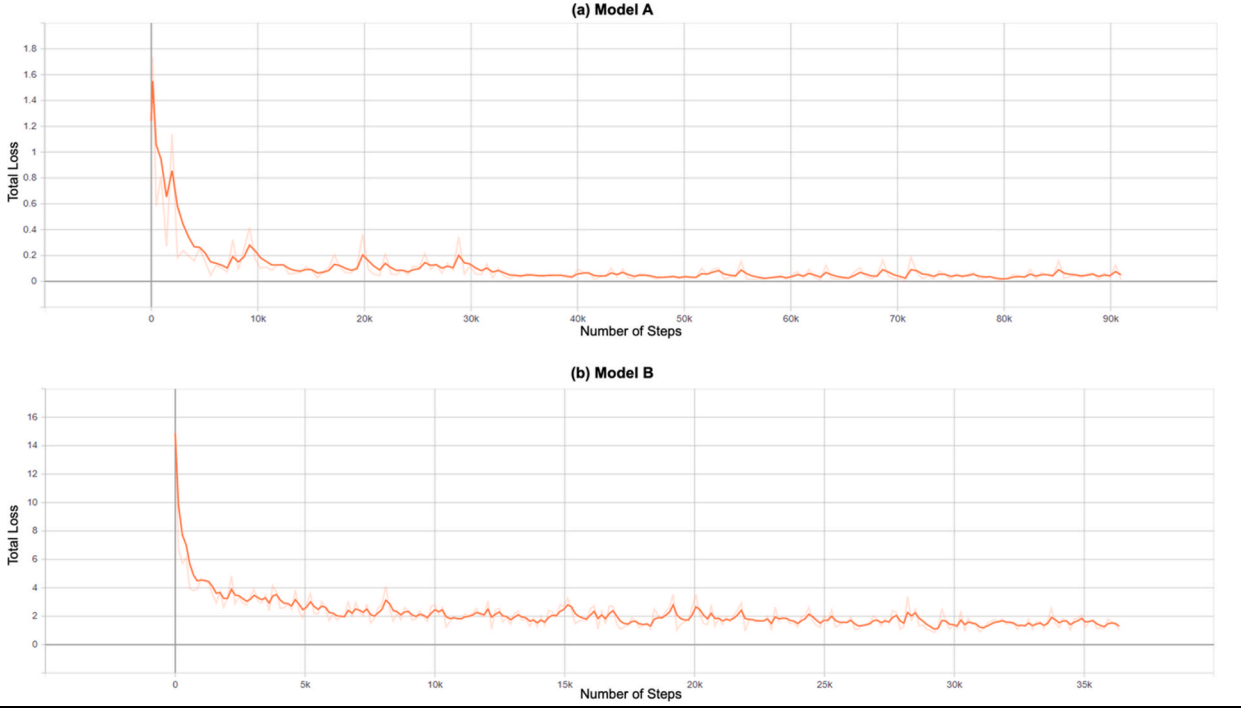
Training Conditions and Results	Model A	Model B
Model Used	Faster R-CNN with InceptionV2	SSD MobileNet V2
Total Steps	90,973	36,348
Training Duration	5 h 45 min, 54 s	8 h 19 min, 44 s
Average Loss	0.10534	2.14576
Minimum Loss	0.00498	0.85439
Total loss versus the number of training steps		

Fig. 11. Confusion matrix providing a visualization of the results of the classification of fire and smoke using the trained models, a). Model A – Faster R-CNN with InceptionV2, b). Model B - SSD MobileNet V2.

Table 4
Model performance results based on the common evaluation metrics to assess the trained fire and smoke detection models.

Model	Class	Category	Accuracy	Precision	Recall	F_1 Score
Model A	1	Fire	87.36%	1.0000	0.7472	0.8553
	2	Smoke	91.43%	0.8220	0.9531	0.9163
Model B	1	Fire	75.28%	1.0000	0.5055	0.6798
	2	Smoke	72.69%	0.9646	0.6000	0.7398

around 8s to identify it. As time passed, a large region of smoke appeared in the scene. However, it was not detected in either test until 4min 9s after the fire, at which the smoke was captured by the Faster R-CNN in Test 5a. It can be seen that the SSD has a longer response time during the whole test with the comparison to the Faster R-CNN, which performed better through this video in terms of the detection rate and accuracy. It suggested the suitability of the Faster R-CNN for fire detection. However, unstable detection of fire and minimal detection of smoke was achieved in Test 5, which indicated that further enhancements are necessary and crucial for future implementation of fire and smoke detection.

4.3. Analysis of the detection performance

4.3.1. Intersection over union (IoU) detection accuracy

Fig. 18 presents the average IoU (%) detection accuracy achieved for fire and smoke across all video tests. Fire achieved higher IoU accuracy than smoke, with an average accuracy of 82.95%. Respectively, smoke achieved an IoU accuracy of 49.34%. For Model A, the average values of fire, smoke, and overall IoU accuracy were 95.73%, 52.34%, and 74.04%; for Model B, they were 70.17%, 46.34%, and 58.26%. However, for some cases, such as Test 1 using Model A, Test 4 and Test 5 using Model B, smoke was not detected or recognized. High variations were achieved across the IoU values achieved, suggesting the performance of the trained models were dependent on the scenarios or conditions of the detected space in the test videos.

Overall, the application of the detector in different indoor settings provided initial results indicating the ability of both models to enable viable fire and smoke detections and also showing its potential for further development. The low IoU accuracies for some tests suggest the need for improvement and further investigation of different aspects to improve the model performance. The detection accuracy could be influenced by the content of the indoor space in the video frame. This includes the environmental settings such as the lighting of the room/space recorded in the video, the amount of fire/smoke portrayed in each video, the position of where fire and/or smoke is initially started captured by the camera and also the distance from the fire/smoke with the camera. Improvements could be made through the training of two separate models for fire and smoke and combining them, along with modifications to the image datasets and model configurations. The following section provides an in-depth analysis of the detection performance during these tests.

4.3.2. Detection performance during each test

Fig. 19 presents the detection performance results in terms of the classification of fire and smoke as a correct detection, no detection, or incorrect detection. Similar to the previous evaluation, it also suggests that Model A performed better than Model B with an approximately 0.33% more of the time would achieve correct detections, along with a decrease of 11.38% achieving no/missed detection. However, results did indicate a potential for Model B, as compared with Model A, it provided less amount of incorrect detection with up to a decrease of 0.12% of the time.

Based on video tests for fire and smoke, the combined average percentage of time that achieved correct detection was 61.02%, 51.07%, 62.34%, 3.57% and 29.91%. Test 3 achieved the best performance, which indicates that it could carry out the detections even when the video feed has no colour (black and white and not colour (RGB)). Moreover, the recognition ability becomes more dependent on the shape of the selected response with high dependence on the position of the camera with the fire. Comparing the results based on each test, it presents the achievement of the given performances of both fire and smoke detection had many contributing factors that

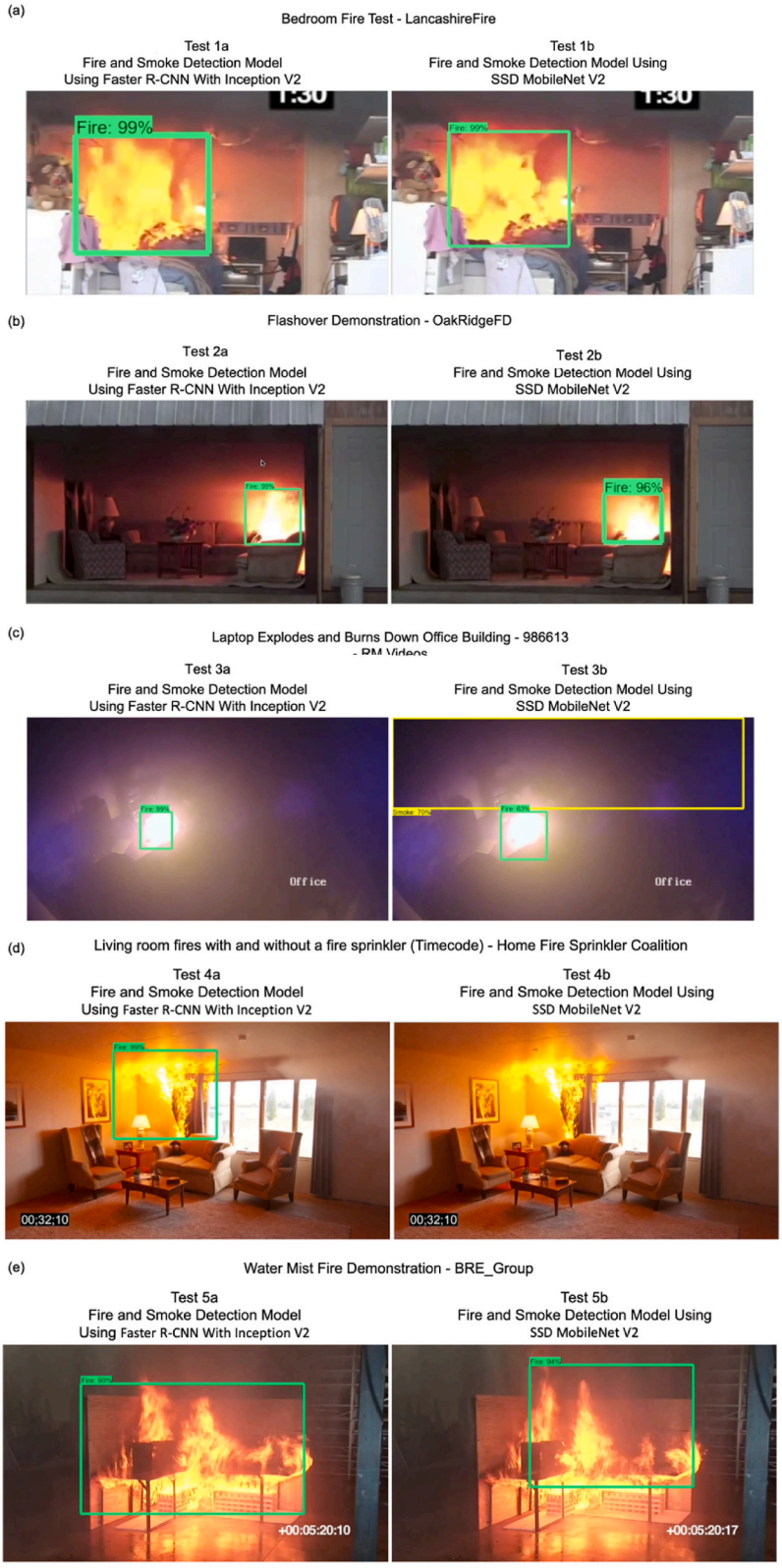


Fig. 12. Summary of the video feed detection and recognition performance test for both fire and smoke: (a). Video 1 – Tests 1a and 1b, (b). Video 2 - Tests 2a and 2b (c). Video 3 – Tests 3a and 3b. (d) Video 4 – Tests 4a and 4b. (e) Video 5 – Tests 5a and 5b.

Fig. 13. Test 1 detection and recognition results, **see Video 1.**

were individually based on each of the test videos.

For the detection and recognition of fire, Test 1A achieved the highest percentage of correct detections with 92.37%. The front facing position of the camera enabled accurate identification of the time when fire was present. Similarly, Test 2a and Test 2b of a living room set up, also consisted of a similar vision taken from the camera, achieving corrections up to 86.50% for Test 2a and 53.99% for Test 2b leading to the achievement of high percentage of time with correct corrections. However, such positioning of the camera may not represent the most common and ideal place as cameras or sensors placed in dwellings are usually placed towards the ceiling. Test 3A and 3B performed best for the detection of smoke, with correct detections up to 57.19% and 55.24%.

To enable further understanding of the two model performances, tests were performed using Tests 4a, 4b, 5a and 5b. Instead of the mock up scenario used in Test 2, Test 4 consisted of a situation where fire appeared in the living room. For this, the highest percentage of no/missed detections were achieved for both fire and smoke, indicating further improvement is required to increase the number of correct detections, while reducing incorrect detection and the number of no/missed detections. With Test 5a and 5b based on a common industrial scenario, it verified the current ability of both models and suggests further investigation is possible to enhance the performance in the development of a viable vision-based fire and smoke detector.

4.3.3. Further evaluation of the detection performance based on the classification evaluation metrics

The following provides a further evaluation of the detection performance based on the video feed tests using both models. The evaluation is based on the analysis using the classification evaluation metrics. The results presented in Fig. 20 were in the form of the confusion matrices based on the percentages of labelled responses. This was due to the unequal number of labels for each response as both fire and smoke detection were performed for various times during each video feed test.

For each of the tests, the confusion matrix enabled the identification of results in terms of the true positives, true negatives, false positives, and false negatives for both fire and smoke. Overall, high variation in the results was observed as the performance differs between each of the tests. For example, this was presented in the performance of fire recognition using Model A, as the percentage of true positive values ranged from 14.28% in Test 4A to 88% in Test 1A. Overall, the results indicated an average percentage of true positive values for the fire was 61.97% when Model A was applied and 31.47% for the application of Model B. For smoke, a percentage

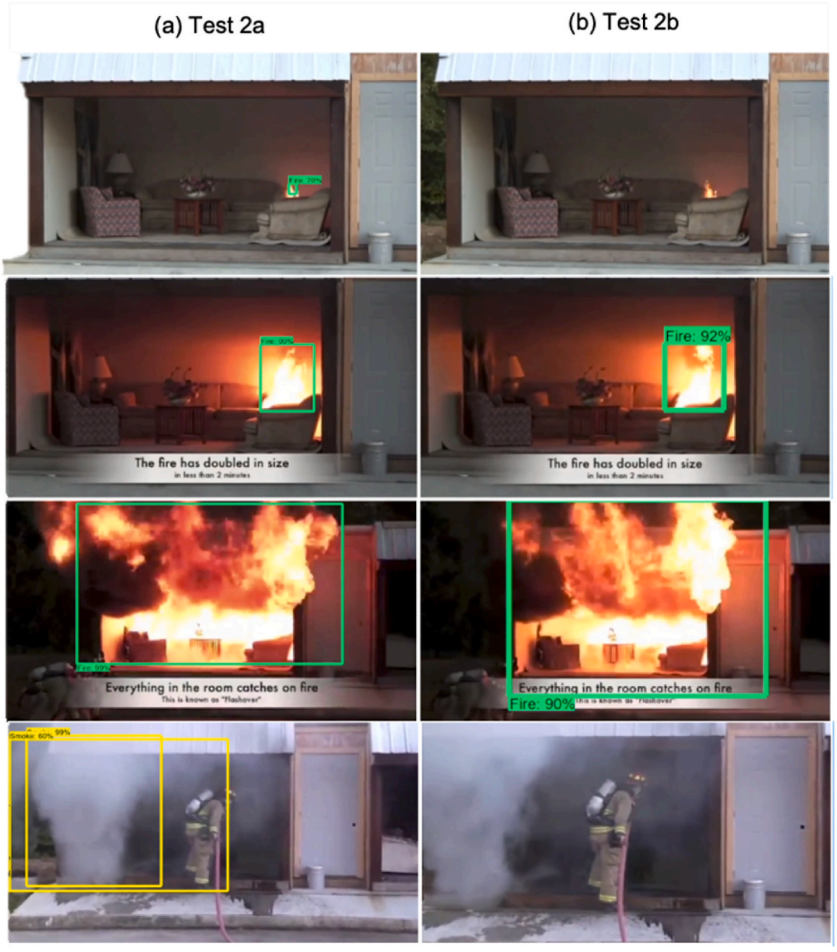


Fig. 14. Test 2 detection and recognition results, **see Video 2**.

of 9.92% and 6.88% were achieved for the applications of Models A and B. Furthermore, as shown in the confusion matrices for all the tests, it indicated the possibility of achieving a high percentage of false negatives for the cases where smoke was not identified, as Test 1a achieved a prediction up to 100% and an average of up to 80.20% using Model A and 86.34% for Model B were identified. However, the results present adequate performance results for an initial model approach but suggest there is a need for further improvements towards the model to achieve a higher percentage of true positive values to become applicable for real-time applications within buildings.

Based on the confusion matrices given in Fig. 20, the following results (Table 5) in terms of the common evaluation metrics were generated. Since more training images of fire were used compared to smoke, it led to the achievement of higher accuracies. Results indicated that 7 out of the 10 tests (Tests 1a, 1b, 2a, 2b, 3a, 4a and 5a) fire achieved higher accuracies in comparison to smoke. Furthermore, with the evaluation metrics of the F_1 Score accounting for the impact of false positive and false negative achieved, Model A achieved an average F_1 Score of 0.61312 for fire and 0.14742 for smoke. This was higher compared to Model B, with an F_1 Score of 0.0358 and 0.143925. Hence, this verifies that the Faster R-CNN with Inception V2 used in Model A does perform better than the SSD MobileNet V2.

Based on the proposed research method for developing the indoor fire and smoke detector using the computer-vision approach (Fig. 1), real-time detection is designed to generate information to assist the fire and safety system for buildings. To enable the detection and recognition made, data were generated in form of profiles. Fig. 21 presents an example of the generated data made during Test 3a for fire and smoke. In this figure, the results were plotted against an actual observation profile indicating the actual fire and smoke conditions across time. Results present errors in detection, suggesting the need for improvements. However, it indicates the demand for such real-time approach that could provide responsive detection through informing the fire and safety systems of such building with propagation tracking, informing fire fighters, and combining detection with an HVAC system.

4.4. Summary and comparison

Unfortunately, due to some papers not presenting all figures mentioned in this work, and some papers only investigating fire or only smoke, a small comparison can be made to other works. This comparison averaged values derived from Zhang et al. [14], Zhang et al.

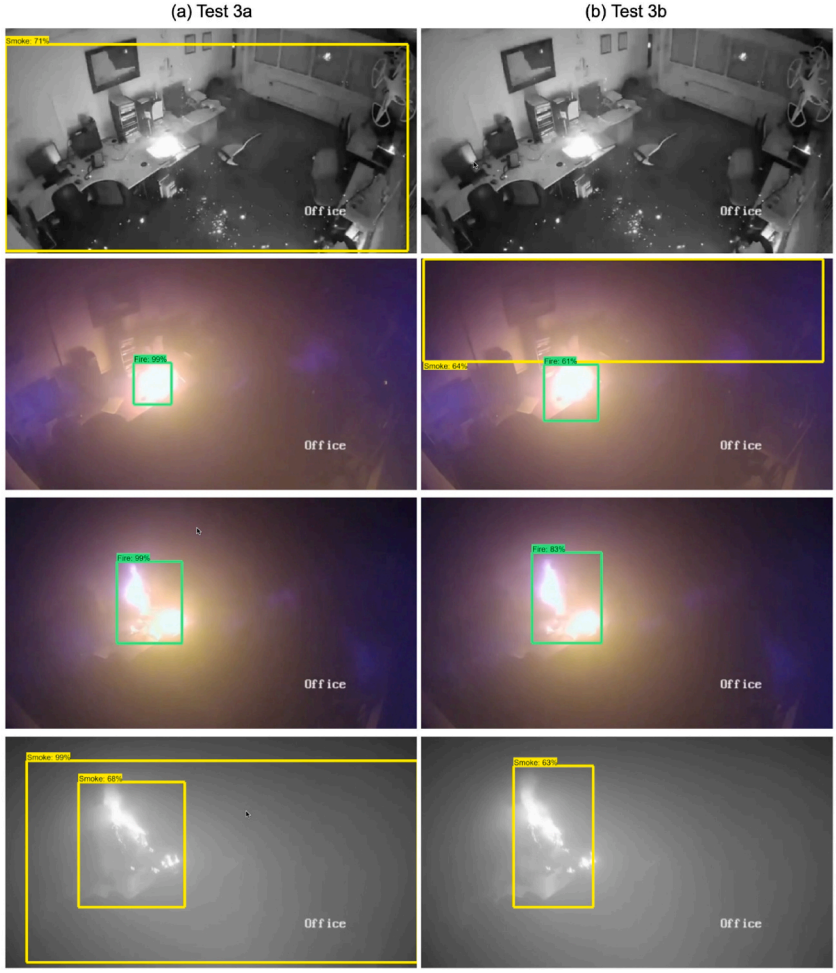


Fig. 15. Test 3 detection and recognition results, **see Video 3**.

[43] and the accuracy value from Kong et al. [26]. The values from the tests found were also averaged. The values are shown in Table 6.

The values not able to be filled in are marked with a dash. The key piece of information to bear in mind when assessing all the data; is that the model was trained on 480 RGB images; this is one of the lowest numbers of training images seen through the literature review. Despite the considerable low number of images, the results are fairly positive, showing that training size may not have significant sway over the accuracy, though the shortcomings seen through the high rate of missed detection is most likely due to the small training dataset.

The values obtained for the SSD model show a failing in almost every aspect with lower averages for average accuracy than that of the comparison papers, the false detection rate is positive; however, the higher averages of missed detection than that of the comparison papers is concerning. As mentioned previously, the reason for testing an SSD based model was for the potential of providing a low-cost, light-weight solution in order to tailor a system that was easier to integrate. The model's in-built accuracy making it computationally lightweight may make it suitable for other tasks however, when missed detection could lead to loss of life it simply is not useable in this scenario.

The results gained from the faster R–CNN are substantially better than that of the SSD, though it is relatively equal to the other works. The fire detection accuracy was similar to the results found by Zhang et al. [14], while the smoke was significantly (36%) less accurate than Kong et al. [26]. This equality shows potential if a greater number of images were used through training, there may be potential to surpass the comparison papers. The false detection rate was similar to that of the SSD and Zhang et al. [14] for fire and 3% higher than the SSD for smoke. The significantly low values show robustness built into the model through the training. The biggest shortcoming however, is the missed detection rate. Both missed detection values for smoke and fire seem very high, though there are two things to consider when assessing them. First, is the statistic stated in the literature review of a current sensor failing to respond in 38% of cases in 2018 [24]. Second is that a missed detection is better than a false detection. Considering a standard of 38%, it is clear that the faster R–CNN overperforms for fire by 17% though smoke underperforms by 10%.

Overall, the Faster R–CNN performs better than the SSD and stands up to other works. The improved detection should be expected due to the difference between using a Faster R–CNN vs an SSD; however, the accuracy of smoke detection showing that a hand-made

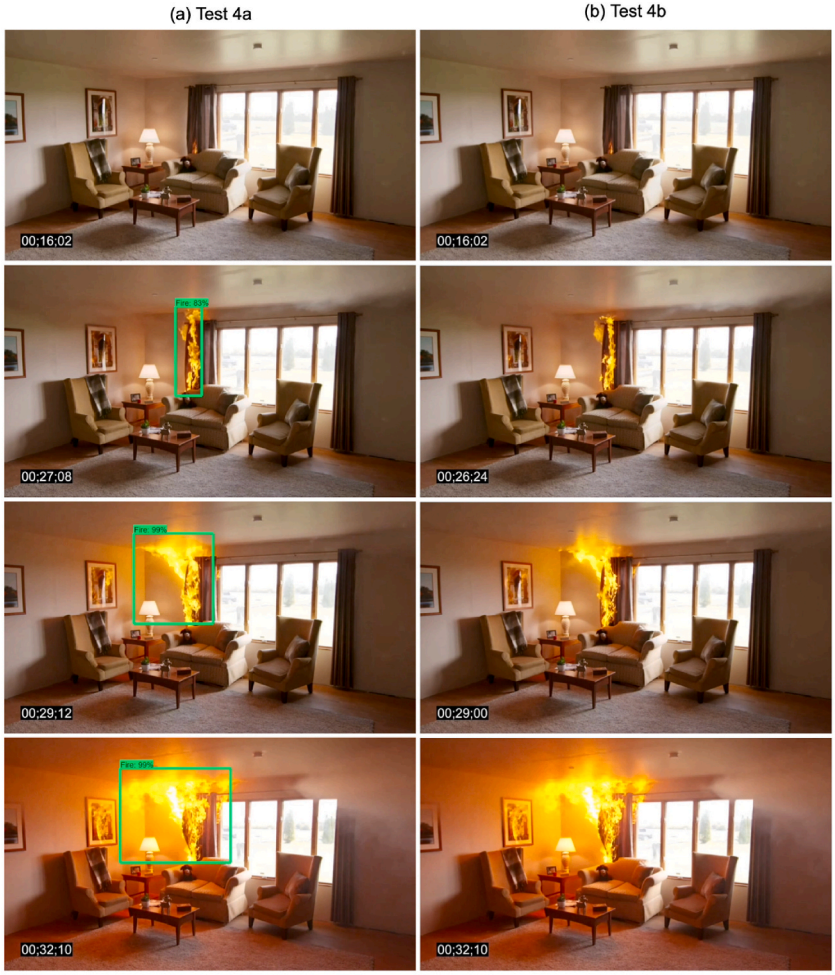


Fig. 16. Test 4 detection and recognition results, **see Video 4.**

algorithm outperformed a deep learning neural network is surprising. This may have been down to the low number of training data, especially with a bias towards fire rather than smoke. Despite this, once the fire and/or smoke is detected through the given camera positioned in the building space, for most cases, a high IoU accuracy rate is achieved (approximately above 73%), unless it was not detected. The false detection rate being greater than that of other works also may be down to not enough training data being used; however, it should be noted that the regions of false detection often were low in accuracy and also short lived. This means that if this model were to be implemented, an algorithm could be written to prevent the alarm from sounding if a detected region were lower than the average accuracy. To further ensure false alarms did not sound, the detection could also have to stay above average accuracy for a set period, i.e., 2 s.

The detection speed of both the SSD and Faster R–CNN models was very variable and heavily situationally dependent. However, SSD has a slow response speed while the Faster R–CNN can compete with current sensors. Further tests regarding response time should be carried out to collect a more suitable dataset size to draw an average response speed.

Although the initial results seem promising, especially for the Faster R–CNN–based detection model, there is more work to be done before it can be implemented in buildings and different indoor spaces. The technology is still at an early stage of development, and it would be difficult to compare it against mature technologies such as thermal and smoke sensors in terms of performance and economical cost/viability. Furthermore, the number of units needed for suitable coverage of a space needed for detection could vary significantly due to complexity in internal design, meaning that the viability of this technology may be limited in some instances. Future developments include enhancing training data and models, using more advanced detection models, and integrating the proposed approach with the fire fighting and HVAC control systems. Furthermore, it should be noted that we are not suggesting that the proposed detection approach will replace existing thermal and smoke sensors; instead, it could complement them, enhancing the firefighting response and saving money/resources.

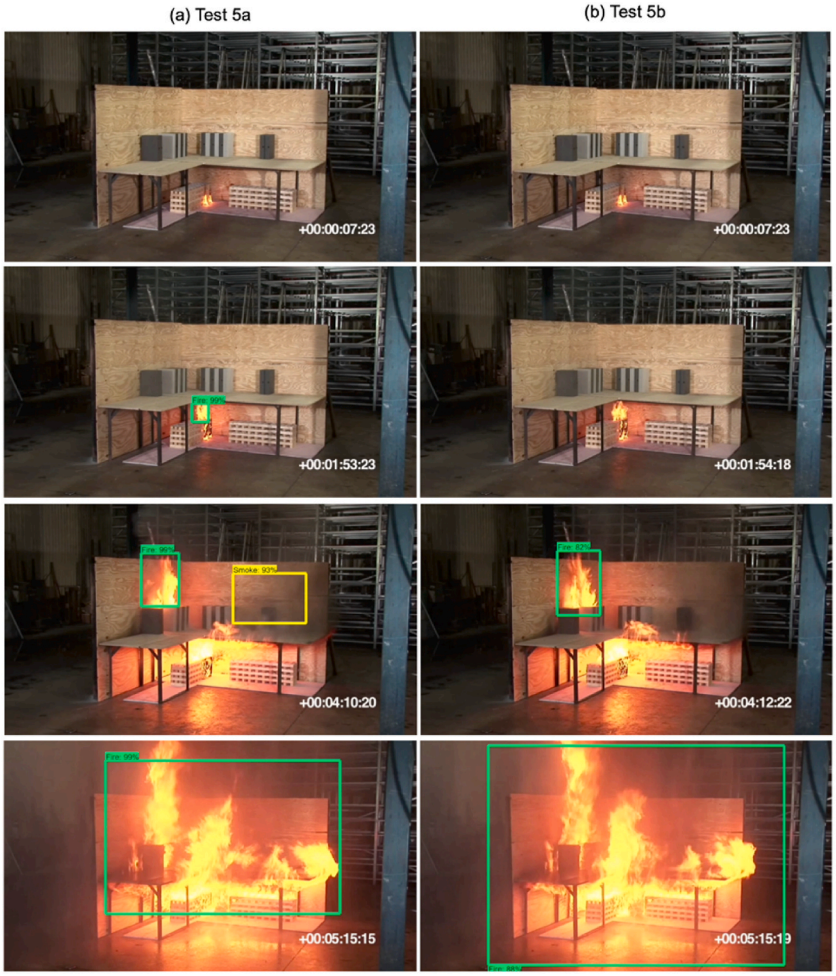


Fig. 17. Test 5 detection and recognition results, **see Video 5**.

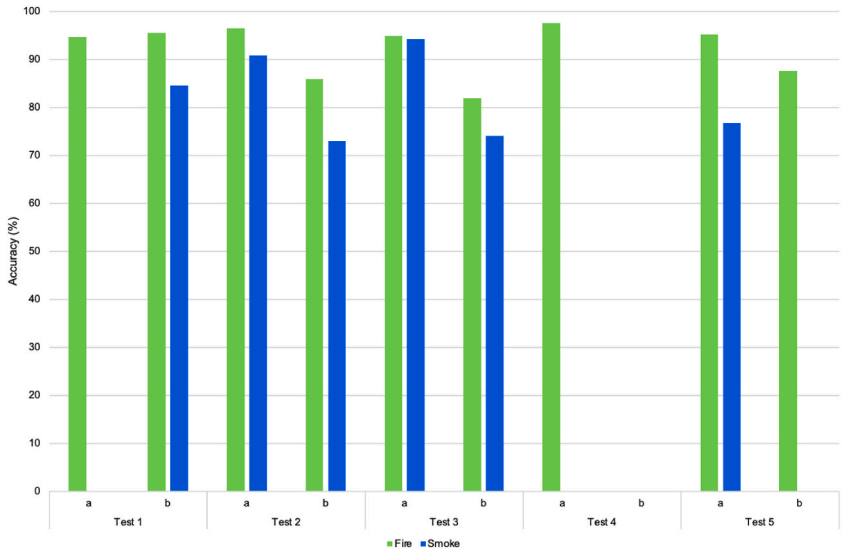


Fig. 18. Average IoU detection accuracy for both fire and smoke during the video feed tests.

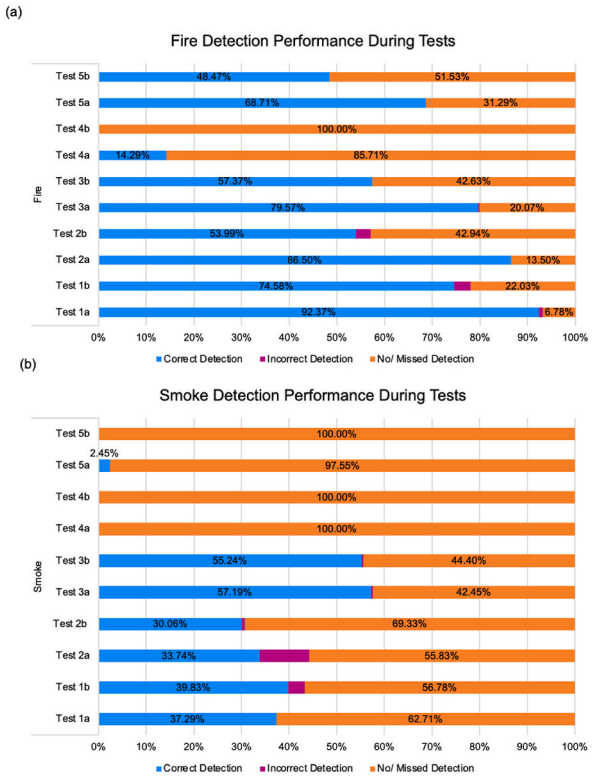


Fig. 19. Detection performance of fire and smoke during all four video feed tests with a). Model A and b). Model B. Identification of the percentage of time achieving correct, incorrect, and no/missed detections.

5. Conclusions and future works

This work showed that a Faster R–CNN Inception V2 vision-based system could be integrated with current indoor fire safety systems to improve the detection of fire in buildings, through the initial data collected. If implemented in both domestic and commercial settings, benefits could be seen especially if integrated along with the HVAC control based on the same detection approach. The promising results were achieved from using only 480 training images. Despite the success achieved by the Faster R–CNN Inception V2, the SSD MobileNet V2 model showed poor accuracy and missed detection results. This is mostly due to the model being trained to a loss of 0.85439 compared to the Faster R–CNN of 0.00498. A low number of training images could be part of this; however, the loss had reached convergence, implying that the model had been effectively trained. SSD models are generally considered less accurate, allowing them to be computationally lightweight (18% GPU usage vs 26%). The initial data achieved for the current SSD model means that it is not suited for current fire detection tasks.

The Faster R–CNN model achieved acceptable values; however, the still relatively high missed detection could be improved. If this model were to be implemented into a real-world detection system, it is recommended that a further algorithm should be developed and integrated into the model, which will evaluate the detected fire and smoke and ensure no or minimal false alarms are triggered. This system could be integrated with current sensors to crosscheck detection to further minimize false alarms.

Further research could be done on the overall training of models to see how low to high-resolution images in the dataset impact the final results alongside the number of images and training time. These can have significant effects on how well the models performed; however, there seems to be little data in order to draw comparisons on these points. Furthermore, it is recommended that the better performing faster R–CNN model should be carried forward to real-world testing. Other CNN models such as the YOLO series should be explored. Using real-time camera detection would allow a proper assessment of the performance of the model. It is noted that generating fires in indoor environments are near impossible without it being the same as the videos used through the testing section of this study. Despite this, testing in real-time would demonstrate clear areas where training improvement would need to be made to ensure minimal false detection and false alarms.

The integration with the fire fighting and HVAC control systems should be investigated. For example, the information can be used to control the ventilation system according to the fire, and smoke detected. This is done as controlled ventilation in fire scenarios can lead to the fire decaying faster, and so the system could also act on the scenario. Future works can focus on integrating this with our previously developed detection approaches. An all-in-one (AIO) system will be developed in the future that could have multiple functions such as control of heating, ventilation, air conditioning systems (HVAC), and fire safety. Finally, a comparison against current technologies such as thermal and smoke sensors should be carried out in terms of performance and economical cost/viability.

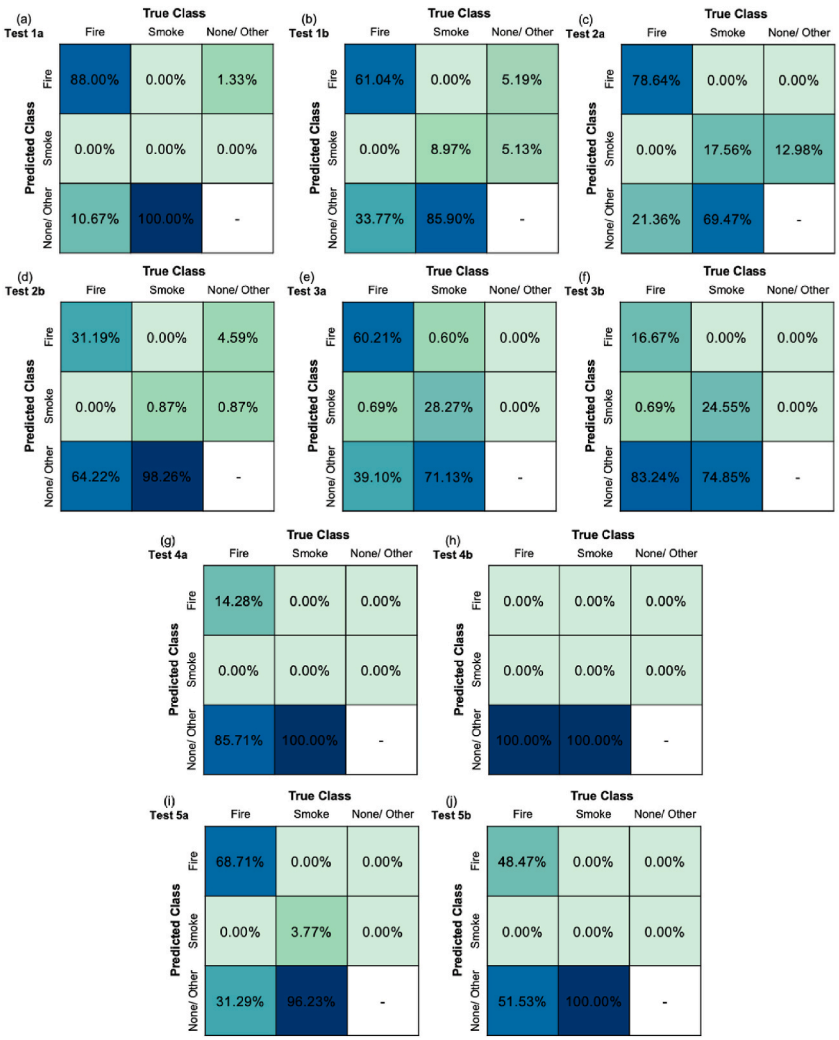


Fig. 20. Video feed test detection performance evaluated in the form of the confusion matrices based on the percentage of labels identified.

Table 5
Evaluation of the video feed test detection performance based on the common evaluation metrics.

Test	Class	Category	Accuracy	Precision	Recall	F_1 Score
1a	1	Fire	94.00%	0.9851	0.8919	0.9362
	2	Smoke	50.00%	–	0.0000	0.0000
1b	1	Fire	80.52%	0.9216	0.6438	0.7581
	2	Smoke	54.48%	0.6362	0.0460	0.1646
2a	1	Fire	89.32%	1.0000	0.7864	0.8804
	2	Smoke	58.78%	0.5750	0.2018	0.2987
2b	1	Fire	65.60%	0.8717	0.3269	0.4755
	2	Smoke	50.44%	0.5000	0.0880	0.0172
3a	1	Fire	79.81%	0.9901	0.6021	0.7488
	2	Smoke	63.79%	0.9762	0.2827	0.4384
3b	1	Fire	58.04%	1.0000	0.1657	0.2843
	2	Smoke	74.11%	0.9727	0.2470	0.3939
4a	1	Fire	57.15%	1.0000	0.1249	0.2501
	2	Smoke	9.09%	–	0.0000	0.0000
4b	1	Fire	9.09%	–	0.0000	0.0000
	2	Smoke	9.09%	–	0.0000	0.0000
5a	1	Fire	82.37%	1.0000	0.6474	0.7860
	2	Smoke	52.11%	1.0000	0.0421	0.0808
5b	1	Fire	74.23%	1.0000	0.4847	0.6529
	2	Smoke	50.00%	–	0.0000	0.0000

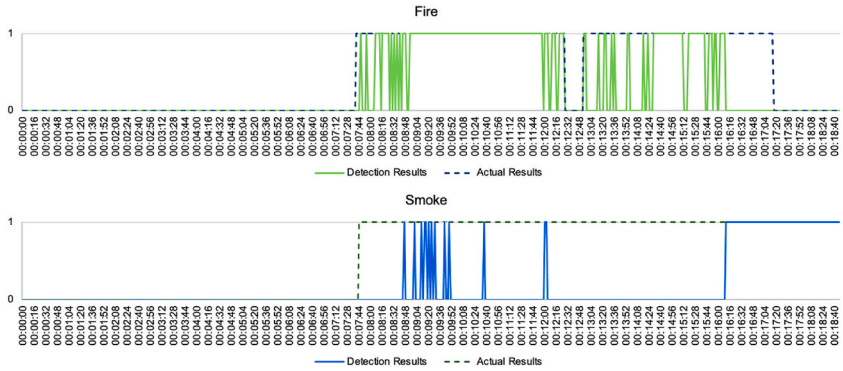


Fig. 21. Comparison between the actual observation (ground truth) and the detected fire and smoke in Test 3.

Table 6
Summary and comparison of data.

	Average Accuracy of Detected Area		False Detection Rate		Missed Detection	
	Fire	Smoke	Fire	Smoke	Fire	Smoke
Faster R–CNN (Model A)	95%	62%	0.4%	4%	14%	54%
SSD (Model B)	88%	77%	2%	1%	39%	56%
Zhang et al. [14] CNN-Pool 5	94%	–	2%	–	38%	–
Zhang et al. [43] Faster R–CNN	–	–	–	–	15%	
Kong et al. [26] hand made	–	98%	–	–	–	–

CRediT author statement

James Pincott: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft. Paige Wenbin Tien: Methodology, Software, Validation, Formal analysis, Investigation, Data visualization, Writing - review & editing, Visualization. Shuangyu Wei: Writing - editing. John Kaiser Calautit: Conceptualization, Resources, Methodology, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Department of Architecture and Built Environment, University of Nottingham, and the PhD studentship from EPSRC, Project References: 2100822 (EP/R513283/1).

References

[1] Gov, Fire & rescue incident statistics, Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/884271/fire-and-rescue-incident-dec19-hosb1120.pdf, 2019, 9th November 2021.
[2] Gov, Fire alarms-property management, Available: https://www.london-fire.gov.uk/safety/property-management/fire-alarms/, 2020, 9th November 2021.
[3] P.W. Tien, S. Wei, J.K. Calautit, A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand, Energies 14 (2020) 156.
[4] P.W. Tien, S. Wei, T. Lui, J.K. Calautit, J. Darkwa, C. Wood, A deep learning approach towards the detection and recognition of opening of windows for effective management of building ventilation heat losses and reducing space heating demand, Renew. Energy 177 (2021) 603–625.
[5] S. Wei, J. Calautit, Development of deep learning-based equipment heat load detection for energy demand estimation and investigation of the impact of illumination, Int. J. Energy Res. 45 (5) (2021) p7204–7221.
[6] J. Huang, et al., Speed/accuracy trade-offs for modern convolutional object detectors, Available: https://arxiv.org/abs/1611.10012, 2017, 9th November 2021.
[7] E. Soltanaghaei, K. Whitehouse, Practical occupancy detection for programmable and smart thermostats, Appl. Energy 220 (-) (2021) p842–855.
[8] A. Brunetti, Computer vision and deep learning techniques for pedestrian detection and tracking: a survey, Neurocomputing 300 (-) (2018) p17–33.
[9] J. Laufs, et al., Security and the smart city: a systematic review, Sustain. Cities Soc. 55 (-) (2020) p10–23.
[10] G. Healey, D. Slater, T. Lin, B. Drda, A.D. Goedeke, A system for real-time fire detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 1993, June, pp. 605–606.
[11] M. Shah, N. Lobo, Flame recognition in video, in: Workshop on Applications of Computer Vision, 2000.
[12] A. Jadon, A. Varshney, M.S. Ansari, Low-complexity high-performance deep learning model for real-time low-cost embedded fire detection systems, Procedia Comput. Sci. 171 (2020) 418–426.
[13] K. Avazov, M. Mukhriddin, M. Fazliddin, Y.I. Cho, Fire detection method in smart city environments using a deep-learning-based approach, Electronics 11 (no. 1) (2021), 73.
[14] Q. Zhang, J. Xu, L. Xu, H. Guo, Deep convolutional neural networks for forest fire detection, in: 2016 International Forum on Management, Education and Information Technology Application, Atlantis Press, 2016.

[15] T.-H. Chen, P.-H. Wu, Y.-C. Chiou, An early fire-detection method based on image processing, in: 2004 International Conference on Image Processing, 2004, vol. 3, ICIP '04., Singapore, 2004, pp. 1707–1710.
[16] P.W. Tien, S. Wei, J.K. Calautit, J. Darkwa, C. Wood, A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions, Energy Build. 226 (2020), 110386.
[17] P.W. Tien, S. Wei, J.K. Calautit, J. Darkwa, C. Wood, Occupancy heat gain and prediction using deep learning approach for reducing building energy demand, J. Sustain. Develop. Energy Water Environ. Syst. (2020), 1080378.
[18] S. Wei, P.W. Tien, J.K. Calautit, Y. Wu, R. Boukhanouf, Vision-based detection and prediction of equipment heat gains in commercial office buildings using a deep learning method, Appl. Energy 277 (2020).
[19] Ottawa Fire Services, Fire dynamics 3 enclosure fires, Available: https://guides.firedynamicstraining.ca/g/fd203-enclosure-fires-pres/118834, 2020, 9th November 2021.
[20] L. Shi, F. Long, C. Lin, Y. Zhao, Video-based fire detection with saliency detection and convolutional neural networks, in: F. Cong, A. Leung, Q. Wei (Eds.), Advances in Neural Networks - ISNN 2017. ISNN 2017, Lecture Notes in Computer Science, vol. 10262, Springer, Cham, 2017.
[21] Smoke alarms fail in a third of house fires, Available: https://www.bbc.co.uk/news/uk-england-50598387, 2019, 9th November 2021.
[22] T. Celik, H. Demirel, Fire detection in video sequences using a generic color model, Fire Saf. J. 44 (2) (2019) 147–158.
[23] W.-B. Horng, J.-W. Peng, C.-Y. Chen, A new image-based real-time flame detection method using color analysis, in: Proceedings. 2005 IEEE Networking, Sensing and Control, vol. 2005, 2005, pp. 100–105. Tucson, AZ.
[24] Celik, T., Demirel, H., Ozkaramanli, H. and Uyguroglu, M., Fire detection in video sequences using statistical color model. In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (Vol. vol. 2, pp. II-II). IEEE.
[25] T. Celik, Fast and efficient method for fire detection using image processing, ETRI J. 32 (6) (2020) 881–890.
[26] S.G. Kong, D. Jin, S. Li, H. Kim, Fast fire flame detection in surveillance video using logistic regression and temporal smoothing, Fire Saf. J. 79 (2016) 37–43.
[27] Christian Moen, Pirjo-Riitta Salminen, Geir Dahle, Johannes Hjertaas, Ketil Grong, Knut Matre, Is strain by Speckle Tracking Echocardiography dependent on user controlled spatial and temporal smoothing? An experimental porcine study, Cardiovasc. Ultrasound 11 (32) (2013), https://doi.org/10.1186/1476-7120-11-32.
[28] D.C. Wang, X. Cui, E. Park, C. Jin, H. Kim, Adaptive flame detection using randomness testing and robust features, Fire Saf. J. 55 (2013) 116–125.
[29] G. Marbach, M. Loepfe, T. Brupbacher, An image processing technique for fire detection in video images, Fire Saf. J. 41 (4) (2006) 285–289.
[30] B.U. Töreyin, Y. Dedeoğlu, U. Güdükbay, A.E. Cetin, Computer vision based method for real-time fire and flame detection, Pattern Recogn. Lett. 27 (1) (2006) 49–58.
[31] R. Collins, A. Lipton, T. Kanaden, A system for video surveillance and monitor, in: Proceedings of the 8-th International Topical Meeting on Robotics and Remote Systems, American Nuclear Society, 1999.
[32] F. Van der Heijden, Image Based Measurement Systems: Object Recognition and Parameter Estimation, Wiley, 1994.
[33] D.A. Reynolds, R.C. Rose, Robust text-independent speaker identification using Gaussian mixture speaker models, IEEE Trans. Speech Audio Process. 3 (1) (1995) 72–83.
[34] Z.Q. Zhao, P. Zheng, S.T. Xu, X. Wu, Object detection with deep learning: a review, IEEE Transact. Neural Networks Learn. Syst. 30 (11) (2019) 3212–3232.
[35] J. Günther, P.M. Pilarski, G. Helfrich, H. Shen, K. Diepold, First steps towards an intelligent laser welding architecture using deep neural networks and reinforcement learning, Proc. Technol. 15 (2014) 474–483.
[36] H. Wu, D. Wu, J. Zhao, An intelligent fire detection approach through cameras based on computer vision methods, Process Saf. Environ. Protect. 127 (2019) 245–256.
[37] D.Y. Chino, L.P. Avalhais, J.F. Rodrigues, A.J. Traina, Bowfire: detection of fire in still images by integrating pixel color and texture analysis, in: 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, IEEE, 2015, pp. 95–102.
[38] P. Foggia, A. Saggese, M. Vento, Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion, IEEE Trans. Circ. Syst. Video Technol. 25 (9) (2015) 1545–1556.
[39] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, S.W. Baik, Efficient deep cnn-based fire detection and localization in video surveillance applications, IEEE Transac. Syst. Man Cyber.: Systems 1–16 (2018).
[40] K. Muhammad, S. Khan, M. Elhoseny, S.H. Ahmed, S.W. Baik, Efficient fire detection for uncertain surveillance environment, IEEE Trans. Ind. Inf. 15 (5) (2019) 3113–3122.
[41] R.E. Fan, K.W. Chang, C.J. Hsieh, X.R. Wang, C.J. Lin, LIBLINEAR: a library for large linear classification, J. Mach. Learn. Res. 9 (Aug) (2008) 1871–1874.
[42] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Caffe Darrell, Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM International Conference on Multimedia, 2014, pp. 675–678.
[43] Q.X. Zhang, G.H. Lin, Y.M. Zhang, G. Xu, J.J. Wang, Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images, Procedia Eng. 211 (2018) 441–446.
[44] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: European Conference on Computer Vision, Springer, Cham, 2014, pp. 818–833.
[45] O. Maksymiv, T. Rak, D. Peleshko, Real-time fire detection method combining AdaBoost, LBP and convolutional neural network in video sequence, in: 14th International Conference the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), Lviv, 2017, 2017, pp. 351–353.
[46] Z. Wang, Z. Wang, H. Zhang, X. Guo, A novel fire detection approach based on CNN-SVM using tensorflow, in: D.S. Huang, A. Hussain, K. Han, M. Gromiha (Eds.), Intelligent Computing Methodologies. ICIC 2017. Lecture Notes in Computer Science vol. 10363, Springer, Cham, 2017.
[47] A.Z. da Costa, H.E. Figueroa, J.A. Fracarolli, Computer vision based detection of external defects on tomatoes using deep learning, Biosyst. Eng. 190 (2020) 131–144.
[48] R. Yamashita, M. Nishio, R.K.G. Do, et al., Convolutional neural networks: an overview and application in radiology, Insights Imag. 9 (2018) 611–629.
[49] P. Dollar, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: an evaluation of the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2011) 743–761.
[50] S.-H. Tsang, Review: G-RMI - winner in 2016 COCO detection (object detection) [online]. Available from: https://towardsdatascience.com/review-g-rmi-winner-in-2016-coco-detection-object-detection-af3f2eaf87e4, 2020, 9th November 2021.
[51] Tzutalin, LabelImg, Available: https://github.com/tzutalin/labelImg, 2015, 9th November 2021.
[52] TensorFlow, Tensorflow detection model zoo, Available: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md, 2021, 9th November 2021.
[53] J. Xu, Deep learning for object detection: a comprehensive review, Available : https://towardsdatascience.com/deep-learning-for-object-detection-a-comprehensive-review-73930816d8d9, 2017, 9th November 2021.
[54] Linear classification (n.d.), Available: https://cs231n.github.io/linear-classify/#:~:text=The performance difference between the,a bug or a feature, 2020, 9th November 2021.
[55] G. Gkioxari, R. Girshick, J. Malik, Contextual action recognition with r* cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1080–1088.
[56] S. Ren, K. He, R. Girshick, J. Faster Sun, r-cnn: towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, 2015, pp. 91–99.
[57] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Ssd Berg, Single shot multibox detector, in: European Conference on Computer Vision, Springer, Cham, 2016, pp. 21–37.
[58] X. Feng, R. Xie, J. Sheng, S. Zhang, Population statistics algorithm based on MobileNet, in: Journal of Physics: Conference Series, vol. 1237, IOP Publishing, 2016, 022045. No. 2.

[59] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.
[60] Juan Villacrés, Fernando auat cheein, Detection and characterization of cherries: a deep learning usability case study in Chile, Agronomy 10 (2020), https://doi.org/10.3390/agronomy10060835.
[61] LancashireFire, Bedroom fire test, Available: https://www.youtube.com/watch?v=ezJ6SorlpJo&t=46s, 2013, 9th November 2021.
[62] Oak Ridge Fire Department, Flashover demonstration, Available: https://www.youtube.com/watch?v=BtMmymOxdjc&list=WL&index=56&t=8s, 2013, 9th November 2021.
[63] R.M. Videos, Laptop explodes and burns down office building, Available: https://www.youtube.com/watch?v=ehcGWLOH-Js&list=WL&index=58&t=7s, 2018, 9th November 2021.
[64] Home Fire Sprinkler Coalition, Living room fires with and without a fire sprinkler (Timecode), Available: https://www.youtube.com/watch?v=EehF0UHYaYk. Last, 2017, 13th June 2022.
[65] BRE Group, Water mist fire demonstration, Available: https://www.youtube.com/watch?v=kq8N-9TaoZc, 2010, 13th June 2022.