

Development and evaluation of a vision-based transfer learning approach for indoor fire and smoke detection

Building Serv. Eng. Res. Technol.
2022, Vol. 43(3) 319–332
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01436244221089445
journals.sagepub.com/home/bse



James Pincott, Paige Wenbin Tien* , Shuangyu Wei  and John Kaiser Calautit 

Abstract

Fire poses a significant risk across industrial and domestic settings, especially to firefighters who must tackle the blaze. Current technology for detection in indoor environments are smoke detectors and flame detectors. However, these detectors have several limitations during 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 recognise objects and activities in indoor spaces. This study aims to develop a vision-based fire and smoke detection system. A deep learning technique that incorporates convolutional neural networks (CNN) was utilised to develop the real-time detection approach that can potentially provide necessary information for fire services, including identifying the position and size of the fire and how the fire spreads. A transfer learning approach using a pre-trained model was used to train the detector. Based on the detection and recognition tests using indoor fire and smoke videos, results indicated that the fire detection achieved up to 92.37% correct detections while the smoke detection did not perform as well. Hence, further improvement and evaluation of the detection approach will be conducted in future work, focusing on the impact of different parameters such as the detection model, building type, indoor space size and positioning of the detection camera. The present study provides an insight into the capabilities and potential applications of the concept.

Keywords

artificial intelligence, vision-based approach, fire detection, smoke detection, indoor environment

Received 15 December 2021; Revised 8 March 2022; Accepted 8 March 2022

Practical application

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 UK, baseline figures of

Department of Architecture and Built Environment, University of Nottingham, Nottingham, NG, UK

Corresponding author:

Paige Wenbin Tien, Department of Architecture and Built Environment, University of Nottingham, University Park, Nottingham, NG NG7 2RD, UK.

Email: paige.tien@gmail.com

missed detection were found to be between 21–45%. This study proposes a vision-based deep learning approach for fire and smoke detection and recognition for buildings, particularly indoor spaces. This will be tested for both domestic and commercial (offices) settings. There could be several 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. The initial results presented here show the practicality of such an approach that could potentially be integrated with firefighting systems for various building spaces and environments.

Introduction and literature review

Indoor fires can threaten lives and damage property, causing significant economic losses.

There can be many fire hazards in the indoor environment, from common fire hazards such as electrical appliances to special fire hazards such as chemicals, combustible dust and flammable liquids.¹ The UK government reported a total of 555,759 incidents that were attended by the fire service, with 28% being a fire-related event.¹ However, most incidents attended false fire alarms, accounting for 41% of the total incidents attended,¹ which cost the fire services money and resources. In addition, it can disrupt businesses and lead to loss of productivity. Hence, fire alarm systems are being integrated with more intelligent technologies to minimise the number of false alarms. In addition, although the figures by the UK government show a decreasing number of cases, incidents are still high, thus demonstrating that fires are still a significant risk. Current technology for detection in indoor environments are smoke detectors and flame detectors.² However, these detectors have several limitations during the ignition phase of a fire and propagation. Furthermore, current systems cannot detect an exact position of the fire or how it is spreading or its size, all of which is necessary information for fire and rescue services when dealing with these incidents.

There has been an increase in published studies each year about artificial intelligence,³ with a significant push for the development and application

towards the improvement of building systems in terms of security, along with heating and cooling management.⁴ A major area of development for artificial intelligence is computer vision-based systems, as they can be used for multiple purposes.^{5,6} This was immediately taken advantage of with multiple studies utilising such developments for fire detection.⁷ However, the detection was initially completed using hand-designed feature extraction techniques,⁸ which may not be robust enough to be accurate, especially in various scenarios.⁹

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 extracting features and classifying them in one step. This saves time by replacing the need to create hand-designed feature extraction programmes and decreasing training time. Several studies have focused on utilising this ability for fire detection and smoke detection,^{9,10} demonstrating that CNN can yield better performance than some relevant conventional video fire detection methods.¹¹ However, such studies focus on the outdoor environment, particularly forest fires.¹⁰ Very few studies have focussed on indoor fire detection, especially in office and residential spaces. Indoor spaces such as offices and houses provide several challenges when using vision-based systems, such as obstacles blocking the view to the desired detection area and reflections that could interfere in fire detection. Early detection of fire and smoke is important for residential spaces such as bedrooms where there is typically only one exit, and the person may be asleep or on medication. This study will build on the previous works by Tien et al.^{12,13} and Wei et al.,^{14,15} where a vision-based artificial intelligence (AI) approach was used to detect and recognise the usage of indoor spaces for aiding demand-driven control systems.

The present study aims to develop a real-time fire and smoke detection and recognition approach for buildings, based on a similar approach to achieve a faster and more localised detection. 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 crosschecking ability for systems such as fire detection, which will minimise false alarms. This could improve overall safety in workplaces and at home and save emergency services time and money by reducing the number of false alarms.

Method

The following section outlines the proposed framework approach and the methods to develop, assess and evaluate a vision-based deep learning fire and smoke detection and recognition approach.

Framework and approach

The proposed approach and framework is presented in Figure 1. It can be divided into detection model development and detection performance evaluation. First, a suitable detection model was selected and trained to perform fire and smoke detection and recognition tasks.

Then the model was deployed to a camera to monitor an indoor space. If the emerging situation is happening within the monitored space, the detection

output will provide the fire and safety system information to respond and operate rapidly and automatically to save emergency services time. To evaluate the feasibility of this approach, videos taken in different spaces were used to test its detection and recognition performance. Based on the testing results and the comparison with other solutions, the potential fire and safety system responses and operations were designed for different scenarios to provide a faster, more accurate, and effective solution.

Deep learning method

Recently, vision-based object detection and recognition^{16, 17} developed using computer vision, and deep learning methods have increased in popularity. Among the deep learning methods used to generate different detectors, a suitable method for the vision-based fire and smoke detection application would be CNN, due to its accuracy and speed. It is the most widely used method and provides excellent performance in various computer vision tasks as it can extract features from images or videos directly.¹⁸

In order to develop the fire and smoke detector based on the CNN, the following steps should be carried out. Firstly, a set of model preparation tasks will be conducted including suitable CNN model

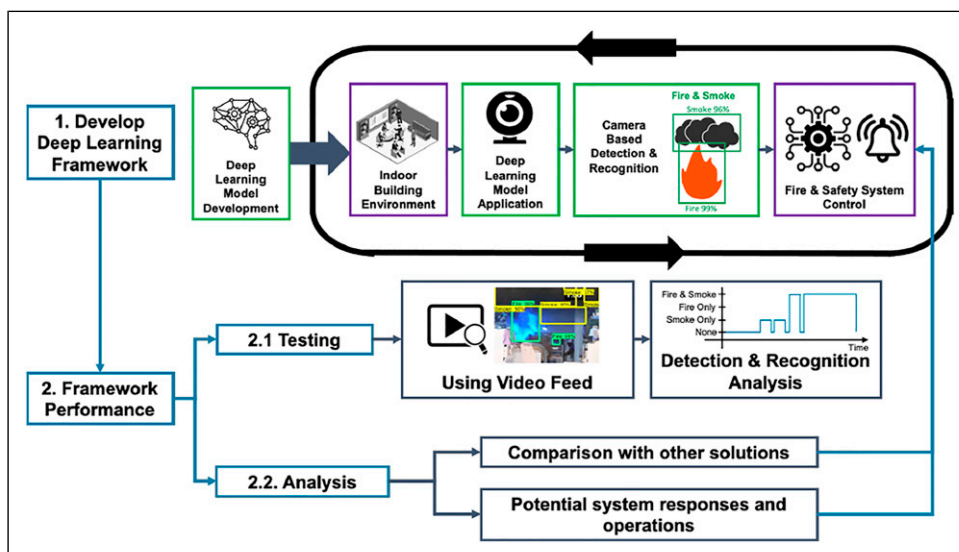


Figure 1. Overview of proposed framework and approach.

selection, input data (images) processing and model training. After the training process, the trained model will be deployed to an AI-enabled camera. Then, the camera could be employed to perform real-time fire and smoke detection tasks in an indoor space. Further details are presented in the following section.

Model development. In this study, as the CNN model was selected, the TensorFlow Object Detection API was used as the framework platform as it provides various pre-trained models and allows the configuration of the desired model. This allows the use of the transfer learning method to learn a new object. The transfer learning stores and extracts the knowledge gained from solving one problem and then uses the obtained knowledge to deal with a different but relevant problem. This can effectively achieve high detection performance while requiring less network training time and a smaller input dataset size.

The commonly used CNNs for object detection are Region-based Convolutional Neural Network (R-CNN), Fast R-CNN and Faster R-CNN. R-CNN, which combines regional proposals and CNN, initially scans the input image to search for the possible objects within the image. During this process, thousands of proposed regions are created and fed into a CNN to extract the features.¹⁹ The outcomes of the CNNs for all the generated regional proposals are then inputted into a support vector machine (SVM) layer to classify the regions containing possible objects. These regions are illustrated through bounding boxes.²⁰ This process reduces the number of the regions, which will be fed into the classification layer, as the feature extraction is implemented on the input image before the selection of the proposed regions. This leads to a significant increase in the speed of completing the detection for each image.¹⁹

However, as R-CNN still takes a large amount of time, a SoftMax classifier was selected to replace the SVM to increase the speed because multiple SVMs are not required to be trained.²⁰ The performances of SoftMax and SVM are often considered comparable,²¹ and therefore, the replacement is not considered to cause a remarkable change when performing detection on one or two classes. After applying these improvements to the R-CNN, the Fast R-CNN was

formed.²² With further evaluation, the Faster R-CNN, the most up to date model, was developed based on the Fast R-CNN by changing the selective search algorithm to a region proposal network (RPN) to predict the region proposals.²³ This once again enhanced the implementation speed. Compared to the R-CNN and Fast R-CNN, the Faster R-CNN significantly reduced the run speed, benefiting real-time detection and recognition activity.

Additionally, the Inception network was an important milestone in the development of CNN classifiers, which improved accuracy and reduced the required computational time. It improves the utilisation of the computing resources inside the network to provide higher classification accuracy. The Inception network is presented in many forms. This includes Inception V1–V4 and the Inception ResNet, whereby each of these versions is an iterative improvement of the architecture of the previous one.

Following the analysis of the Inception networks in Ref. 24, along with the availability of the pre-trained models in the TensorFlow Detection Model Zoo,²⁵ the ‘Inception V2’ provides a good middle ground between accuracy, CPU and GPU memory usage, and the implementation speed. Thus, the Faster R-CNN with Inception V2 was employed to train this study’s proposed model. The general outline of this model is presented in Figure 2.

To enable the model to carry out fire and smoke detection and recognition in an indoor space, input data in the form of RGB images with various pixel densities were clustered to extract and learn the related features of fire and smoke. Some images only presented either fire or smoke, and the other images consisted of both fire and smoke within different indoor environments, such as office spaces and dwellings. These images were randomly divided into training and testing image datasets.

Examples of images consisting of fire and smoke are presented in Figure 3, along with examples of how images were labelled using LabelImg.²⁶ The regions, including fire or smoke specific regions of interest, were labelled manually in bounding boxes in each of the collected images. In some cases, these specific regions of interest were labelled by multiple bounding boxes to ensure that all parts were covered. Moreover, for some images, overlaps occurred

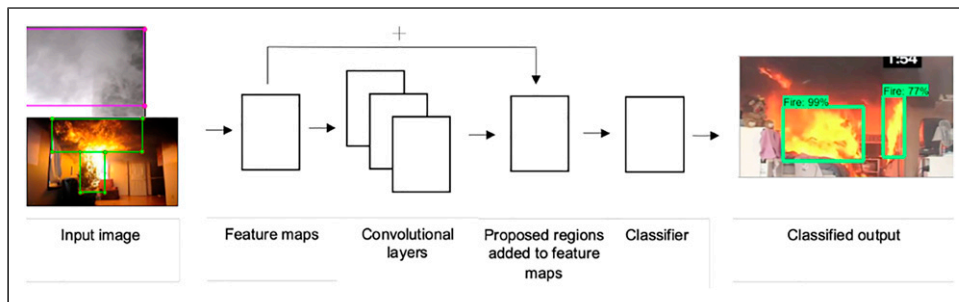


Figure 2. General outline of the architecture of the pre-existing Faster CNN with Inception V2 model used to train the fire and smoke detection model.



Figure 3. Example dataset images of fire and smoke with the highlighted region of interest (ROI) of labelled images. Images are obtained via Google.



between the assigned bounding boxes as both fire and smoke were included and at a close distance in these images.

Initial model application and testing. After the training process, the performance of the trained model was tested using videos that included both indoor fire and smoke. For the initial analysis, two scenarios were selected, which provided an insight into how the detection approach can be applied in buildings. Clearly, this is limited, and the tests may be open to many uncontrolled aspects, and hence, more

scenarios should be evaluated in future works to provide a more rigorous assessment of interrogation of the method and model. [Table 1](#) provides a detailed description of the two different indoor fire and smoke scenarios.

Two methods were employed to assess the model's performance after carrying out the fire and smoke detection. First the average value of Intersection over Union (IoU) and the percentage of the time, which achieved correct, incorrect and no or missed detections were obtained to evaluate the ability of the model to detect and recognise indoor

Table 1. Summary of the videos used for testing the detection approach.

Test	Video	Description	Image example from video
1	Bedroom fire test – Lancashire Fire ²⁷	The video shows a mock-up of a child's bedroom with a lot of materials and basic fittings in the way of lights. It also contains an old television screen which allows for the training of reflections. This scene then develops into a full flashover event	
2	Kitchen-Fire accident caught on camera – Anjan Amirishetty ²⁸	A CCTV video for a domestic setting. Consisting of a toaster that fails and catches fire. An incident with dense smoke rather than a flashover features high contrasting zones from a lighting perspective. The video also contained moving people through the space, which will help see how motion could potentially lead to false detections	

fire and smoke. Then a confusion matrix was used to further evaluate the results of the tests. The classification evaluation metrics, including accuracy, precision, recall and F1 score, were used to assess the detection and recognition performance from the confusion matrix. In this study, the values presented in the confusion matrices were the percentage of the correct and incorrect predictions on each class instead of the number of predictions because of each class's unequal number of labels.

Results and discussion

This section presents the developed deep learning-based fire and smoke detector results and discussion. It includes the analysis of its performance on a series of video feed tests and the analysis in terms of its detection performance and its potential as a solution to enhance the fire safety of buildings.

Detection model training results

Based on a transfer learning approach where the pre-trained model of the Faster-RCNN with InceptionV2 was used to train the model, the model for this present study was trained for a total of 90,973 steps for 5 h, 45 min and 54 s. An average loss of 0.10,534 was achieved with a minimum loss of 0.00,498. Correspondingly, the training results are presented in Figure 4, indicating that the convergence of the loss

function implies that the model has been effectively trained.

The trained model for the vision-based detector was validated by the performance on the feed of testing image data. The 120 images within the test dataset were used to assess the initial recognition performance. 156 labels would be assigned to the images as some achieved more than one response label assigned. Results based on the percentage of the total response labels assigned to the application of the images from the dataset were presented in terms of a confusion matrix in Table 2. The model provided a correct classification of fire up to 74.73 and 93.85% for smoke. The only drawback is that the response for fire is sometimes predicted as smoke (10.99%), and for some instances, predictions were not made. Furthermore, the associated results in terms of the evaluation using the common evaluation metrics were generated. As given by the F₁ Scores, it suggested the models provide sufficient ability towards the detections of both fire and smoke.

Framework test results

This section discusses the application of the vision-based detection model on video tests. Continuous detection and recognition of fire and smoke were performed during both video feed tests. For a given instance where fire and/or smoke was detected, it would display the prediction bounding boxes along

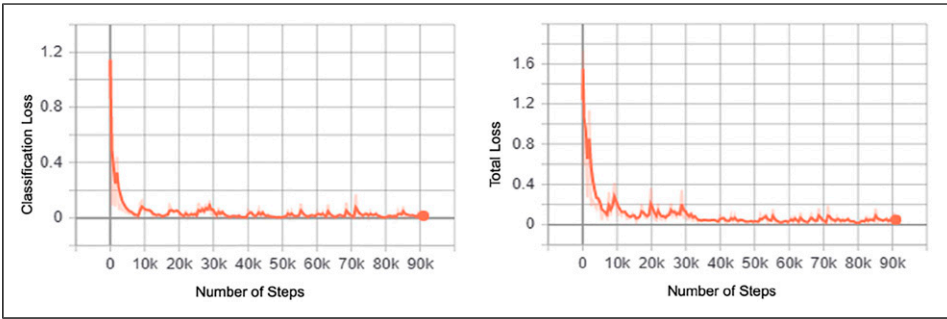


Figure 4. Deep learning model training results in terms of the classification loss and the total loss against the number of training steps.

Table 2. Confusion matrix results and the model performance results based on the common evaluation metrics to assess the trained detection model.

Confusion matrix				Category	Accuracy, %	Precision	Recall	F ₁ Score
Predicted Class	True Class			Fire	87.36	1.0000	0.7472	0.8553
	Fire	Smoke	None/ Other	Smoke	91.43	0.8220	0.9531	0.9163
	Fire	Smoke	None/ Other					
	Fire	Smoke	None/ Other					



Figure 5. Example snapshot at various key points of the stages of the application of the detection model on video 1, test one.

with the IoU values that were achieved over time. The following sections of The test one and The test two evaluate the detection during these tests.

Test one. The key stages of Test one are presented in Figure 5. In terms of detecting fire, the model was slow to react to the ignition. The first fire detection and recognition were made 21 s after the fire had started.

As shown by the snapshot images, throughout the test, when the fire was recognised, the bounding box would appear to be closely assigned to the areas where the fire is at its strongest, where regions that were presented in the colours of bright yellow and orange. Furthermore, this test, also suggests that accurate fire detection was achieved as the detections

made were not dependent on factors such as the given shape or size of the fire.

Overall, good performance was achieved based on the detection rate and the accuracy. However, it did not achieve any correct smoke detection over the collection period. Despite these shortcomings, the reflection caused by the television screen opposite the bed did not result in any false detection, which shows the model's significant robustness. Therefore, this indicates that the model enabled accurate fire detections within an indoor environment such as a bedroom.

Test two. Figure 6 presents the application of a vision-based fire and smoke detector with video 2 of

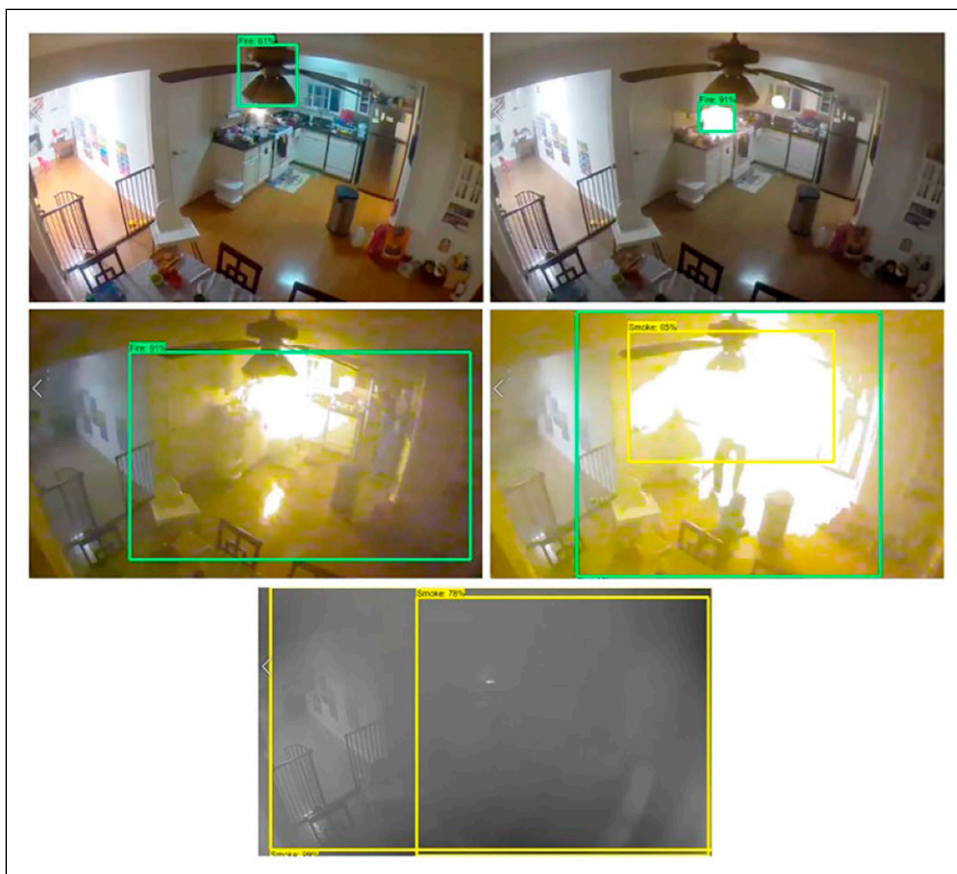


Figure 6. Example snapshot at various key points of the stages of the application of the detection model on video 2, test two.

a domestic-based setting. The initial stage of the video shows light smoke coming from the toaster on the kitchen side. Through this, there were no smoke detection made from the application of the detector and only false fire detection on a foreground element. Through the rest of the video, fire detection was only momentarily seen at flashes with correct fire detection and some amounts of false detection. Towards the end of the video, the scene is filled with smoke, this is where most of the smoke detection occurs.

Analysis of detection performance

Test one achieved an average Intersection over Union (IoU) accuracy of 94.64% for fire, and Test two achieved 79.75% for fire and 97.83% for smoke. Overall, consistent prediction accuracy was achieved. Since the content given in video 1 and 2 were very different, the initial results indicate that the detection accuracy can be dependent on the content of the video. 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, and also the distance from the fire/smoke with the camera. Therefore, this reflected upon the variation in accuracy between all tests. Overall, this indicates the ability of the model to enable viable detections across a range of different environments and its practicality for an effective detection approach.

Table 3 presents the detection performance results in terms of the detection labelled assigned at every second during each of the video feed tests with the

classification as a correct detection, no detection or incorrect detection.

Both fire and smoke video tests indicated that fire achieved a higher percentage of correct detections than smoke, with the combined average percentage of time with correct detections was 85.97% for fire and 47.24% for smoke. Comparing the performance on test 1 (video 1) and test 2 (video 2) indicated that test two achieved better performance with 68.38% of the time achieving correct detections, compared to 64.83% of the time in test 1. This further indicated that the colour (black and white and not colour (RGB)) in test 2, did not influence the detection performance. Moreover, the recognition ability becomes more dependent on identifying the shape of the selected response. Hence, comparing the results based on both tests shows that fire and smoke detection performance has many contributing factors that were individually based on each of the test videos.

For detecting and recognising fire, Test one achieved the highest percentage of correct detections with 92.37%. This indicates that settings such as the one shown in video 1 would be best to access the performance in terms of fire detection, and video 2 would be more suitable for assessing smoke detection. Therefore, the evaluation of these results in such form enables the provision of a greater understanding of the effectiveness of the detection performances under the different scenarios.

The following provides a further evaluation of the detection performance during each video feed test using both models. For both tests, the confusion matrix (Figure 7) enabled the identification of results

Table 3. Detection performance of fire and smoke during all tests in terms of identifying the percentage of time achieving correct, incorrect and no/missed detections.

Category	Video test	Percentage of time achieving		
		Correct detections	Incorrect detections	Missed detections
Fire	1 (bedroom)	92.37%	0.85%	6.78%
	2 (kitchen)	45.21%	10.96%	43.32%
	Average	68.79%	5.91%	25.05%
Smoke	1 (bedroom)	37.29%	0.00%	62.71%
	2 (kitchen)	61.64%	6.85%	31.51%
	Average	49.47%	3.43%	47.11%

in terms of the true positives, true negatives, false positives, and false negatives for both fire and smoke. Overall, a high variation in the results was achieved as the performance differs between the tests. For example, the percentage of true positive values ranged from 9.08% in Test 2–88% in Test 1. This gave an average of 48.54% of true positive values for fire. No true positive was achieved for smoke in test one and only 45.10% in Test 2. However, as shown in the confusion matrices, it indicated the possibility of achieving a high percentage of false negatives for the cases where smoke was not identified, as Test one achieved a prediction up to 100% and 45.10% in Test 2. Despite this, the results present adequate performance results for an initial model approach but suggest a need for further improvements towards the

model to achieve a higher percentage of true positive values for real-time applications within buildings.

Results in Table 4 indicated that fire detection achieved higher accuracy than smoke for both tests. With the evaluation metrics of the F_1 Score accounting for the false positive and false negative achieved, an overall F_1 Score of 0.5471 was achieved for fire and 0.3088 for smoke.

In addition to discussing how the model performances can be solely based on each of the individual tests, the results in Table 4 also reflect this. Hence, solely based on applications of the developed models and the two video feed tests, it suggests that a stronger performance in the detection of fire was achieved under the two different situations than smoke.

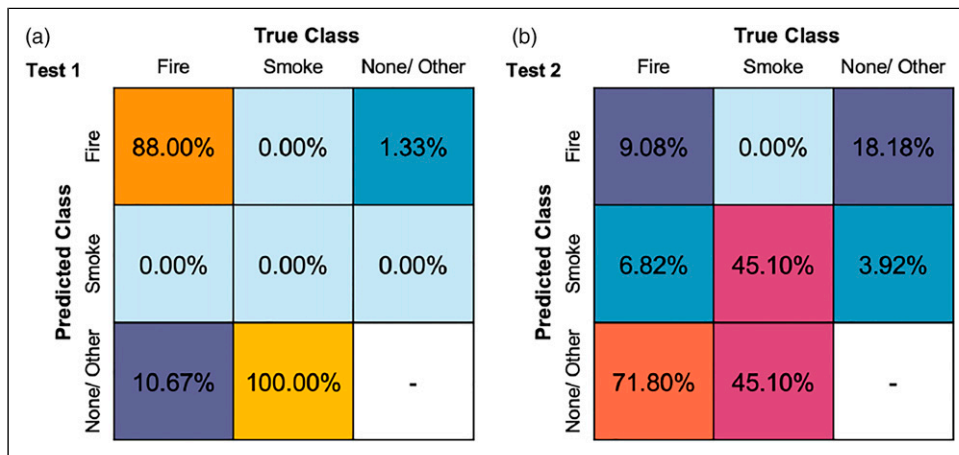


Figure 7. Video feed test detection performance evaluated in the form of the confusion matrices based on the percentage of labels identified.

Table 4. Evaluation of the video feed test detection performance based on the common evaluation metrics.

Test	Class	Category	Accuracy, %	Precision	Recall	F_1 Score
1	1	Fire	94.00	0.9851	0.8919	0.9362
1	2	Smoke	50.00	—	0.0000	0.0000
2	1	Fire	79.81	0.9901	0.6021	0.7488
2	2	Smoke	63.79	0.9762	0.2827	0.4384
Average	1	Fire	72.80	0.6591	0.4977	0.5471
	2	Smoke	61.04	0.8077	0.2500	0.3088

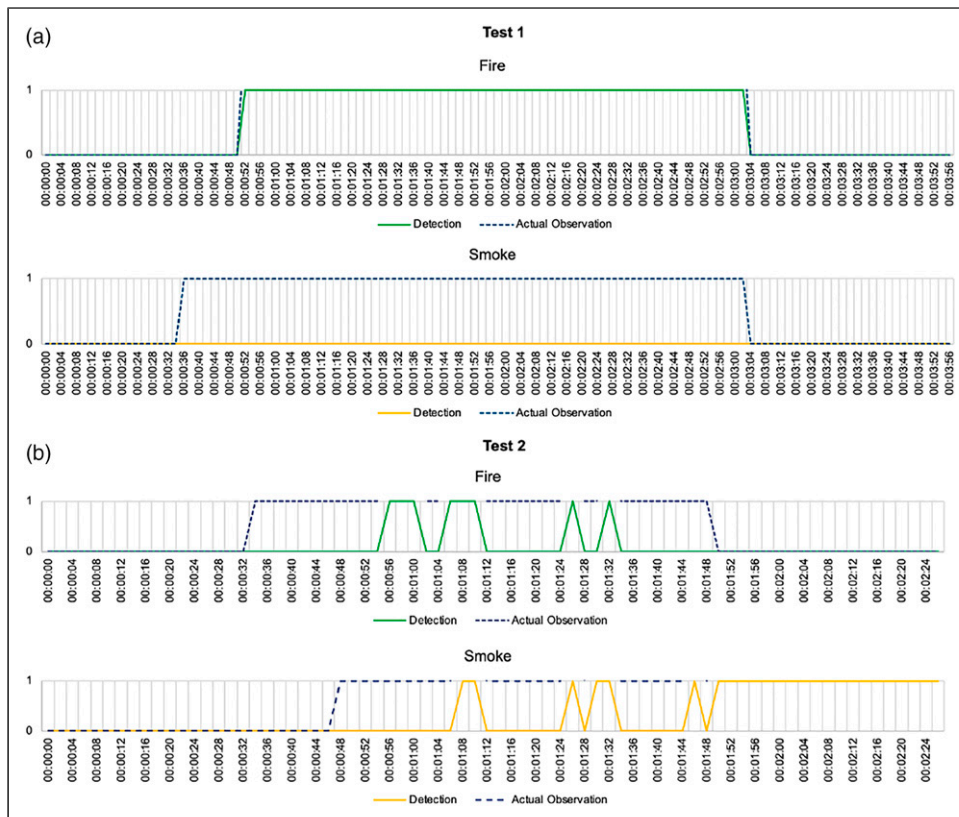


Figure 8. Comparison between the actual observation (ground truth) and the detected fire and smoke during the two tests, (a): Test one and (b) Test two.

Figure 8 compares the actual observation (ground truth) results with the real-time detection of fire and smoke during the two tests. The following results given in Figure 8 are in line with the findings in the previous section. Overall, it identifies the need for further improvements towards the detection ability to provide an effective framework approach. Sufficient improvements in the detection approach will be required by reducing the errors between the detection results and the actual observation to provide a fast, responsive solution that correctly identifies the times when fire and/or smoke appears within an indoor building environment.

This section presents an overview of the developed model's performance on detecting and recognising fire within indoor environments during the two tests performed. A comparison is made with

existing strategies. This includes the fire detection approach developed by Zhang et al.,¹⁰ the fire and smoke detection by Zhang et al.,²⁹ and also fire flame detection by Kong et al.³⁰ The developed vision-based approach in this study consists of using the lowest numbers of training images. Despite this, the results are positive, though the shortcomings seen through the high rate of missed detection is most likely due to the small training dataset.

Overall, the fire detection accuracy was 7% lower than Zhang et al.,¹⁴ while the smoke was significantly less accurate than Kong et al.¹⁵ The results suggest the performance is variable and heavily situationally dependent, and the environmental setting can impact the detection performance. It should be acknowledged that the existing solutions mentioned in Table 5 were designed to enable the detection of both

Table 5. Summary and comparison of data.

	Average accuracy of detected area	False detection rate	Missed detection			
	Fire	Smoke	Fire	Smoke	Fire	Smoke
Proposed detection model	87%	49%	6%	3%	25%	47%
Zhang et al. ¹⁰ CNN-Pool 5	94%	—	2%	—	38%	—
Zhang et al. ²⁹ faster R-CNN	—	—	—	—	15%	—
Kong et al. ³⁰ hand made	—	98%	—	—	—	—

fire and smoke within outdoor settings. Hence, results cannot be directly comparable as the concept of the developed framework in this present study is designed to enable detection and recognition within indoor environments to enhance the fire safety of buildings. However, the results achieved identifies the potential of using such a developed vision-based approach to recognise instances when both fire and smoke are generally present within an indoor space to help enhance the safety of buildings. Effectively, this is a preliminary study that would require further validation and assessments under different scenarios that consist of many uncontrolled aspects to provide a better insight into the design and performance of the developed framework approach.

Conclusion and future work

This present study shows that a Faster R-CNN Inception V2 vision-based detection system could be a suitable replacement for current fire detection systems. If implemented in domestic or commercial settings, benefits could be seen especially if integrated along with the HVAC control from previous work. The promising results were achieved by using only 480 training images. The Faster R-CNN model achieved acceptable values; however, there were still relatively high missed detection, suggesting further improvements required. Suppose this model were to be implemented into a real-world detection system. In that case, it is recommended that a further

algorithm be added to the model that ignored detections below the average model accuracy and high accuracy detection regions to remain detected for at least 2 s should ensure no false alarms are triggered. This system could also be integrated with current sensors to crosscheck detection.

Future works could look at the feasibility of both a physical sensor (i.e. smoke sensor) and a vision-based detector being combined into one system. Furthermore, the detector presented through this study could be trained further to detect the difference between controlled smoke (i.e. cigarettes and toast burning) and uncontrolled (i.e. smoke from a fire). This would be especially useful in domestic settings. It is only possible to do this through vision systems as object detection can be utilised to detect the source of smoke (i.e. if smoke from cigarettes is detected, no alarm). The combination of both physical and vision detectors would most likely result in close to no false alarms.

Further research could be done towards the overall training of models to see how low to high-resolution images in the dataset impact the final results alongside several images and training time. These can significantly affect how well the models performed; however, there seems to be little data to draw comparisons on these points. Further evaluation of the detection approach will be conducted in future works, focussing on the impact of different aspects such as the building type, indoor space size and positioning of the detection camera. More scenarios should be evaluated in future works to provide

a more rigorous assessment of interrogation of the method and model. Furthermore, it is recommended that the Faster R-CNN model be carried forward to real-world testing. Using real-time camera detection and recognition would allow a proper assessment of the model's performance. It is noted that generating fires in indoor environments is near impossible without it being the same as the videos used through the testing section of this report. Although 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. Finally, 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 and ventilation³¹, air conditioning systems (HVAC)³² and fire safety.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the University of Nottingham and the PhD studentship from EPSRC (2100822 (EP/R513283/1)).

ORCID iDs

Paige Wenbin Tien  <https://orcid.org/0000-0003-0123-248X>

Shuangyu Wei  <https://orcid.org/0000-0002-1053-1652>

John Kaiser Calautit  <https://orcid.org/0000-0001-7046-3308>

References

- Home Gov.UK. *Office. Fire and Rescue Incident Statistics*. England, 2019. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/884271/fire-and-rescue-incident-dec19-hosb1120.pdf, (accessed April 2021).
- London Fire Commissioner. *Property management: Fire alarms.*, 2021. Available at: <https://www.london-fire.gov.uk/safety/property-management/fire-alarms/>, (accessed April 2021).
- Wang Z and Srinivasan RS. A review of artificial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models. *Renew Sustain Energy Rev* 2017; 75: 796–808.
- Soltanaghaei E and Whitehouse K. Practical occupancy detection for programmable and smart thermostats. *Appl Energy* 2018; 220: 842–855.
- Brunetti A, Buongiorno D, Trotta GF, et al. Computer vision and deep learning techniques for pedestrian detection and tracking: a survey. *Neurocomputing* 2018; 300: 17–33.
- Laufs J, Borrión H and Bradford B. Security and the smart city: A systematic review. *Sustain Cities Soc* 2020; 55: 102023.
- Healey G, Slater D, Lin T, et al. *A system for real-time fire detection*. Proceedings Fifth IEEE Workshop on Applications of Computer Vision 1993: 605–606.
- Philips W, Shah M and Da Vitoria Lobo N. *Flame recognition in video*. Proceedings Fifth IEEE Workshop on Applications of Computer Vision 2000: 224–229.
- Jadon A, Varshney A and Ansari MS. Low-complexity high-performance deep learning model for real-time low-cost embedded fire detection systems. *Proced Comput Sci* 2020; 171: 418–426.
- Zhang Q, Xu J, Xu L, et al. Deep convolutional neural network for forest fire detection. International Forum on Management, Education and Information Technology Application, 2016, pp. 568–575.
- Chen T, Wu P and Chiou Y. An early fire-detection method based on image processing. *ICIP* 2014; 04: 1707–1710.
- Tien PW, Wei S, Calautit JK, et al. A vision-based deep learning approach for the detection and prediction of occupancy heat emissions for demand-driven control solutions. *Energy and Buildings* 2020; 226: 110386.
- Tien PW, Wei S and Calautit J. A computer vision-based occupancy and equipment usage detection approach for reducing building energy demand. *Energies* 2020; 14: 156.
- Wei S, Tien PW, Calautit JK, et al. Vision-based detection and prediction of equipment heat gains in

- commercial office buildings using a deep learning method. *Appl Energy* 2020; 277.
15. Wei S and Calautit JK. Development of deep learning-based equipment heat load detection for energy demand estimation and investigation of the impact of illumination. *Int J Energy Res* 2020; 45(5): 7204–7221.
 16. O'Mahony N, Campbell S, CarvalhoHarapanahalli AS, et al. Deep learning vs. traditional computer vision. In: Arai K and Kapoor S (eds). *Advances in Computer Vision. CVC*. Cham: Springer; 2019.
 17. Voulodimos A, Doulamis N, Doulamis A, et al. *Deep Learning for Computer Vision: A Brief Review*. Computational Intelligence and Neuroscience, 2018, p. 7068349.
 18. Yamashita R, Nishio M, Do RKG, et al. Convolutional neural networks: an overview and application in radiology. *Insights Into Imaging* 2018; 9: 611–629.
 19. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA; 2014, 24–27.
 20. Zhao Z-Q, Zheng P, Xu S-T, et al. Object detection with deep learning: a review. *IEEE Transactions Neural Networks Learning Systems* 2019; 30: 3212–3232.
 21. Li F, Krishna R and Xu D. *CS231n: Convolutional Neural Networks for Visual Recognition*, 2021. Available from: <https://cs231n.github.io/linear-classify/>, (accessed April 2021).
 22. Gkioxari G, Girshick R and Malik J. Contextual action recognition with R*CNN. *ICCV '15. Proc 2015 IEEE Int Conf Comput Vis (Iccv) December* 2015; 2015: 1080–1088.
 23. Ren S, He K, Girshick R, et al. Faster r-cnn: towards real-time object detection with region proposal networks. In: *Advances in neural information processing systems*; 2015, pp. 91–99.
 24. Raj B. *A Simple Guide to the Versions of the Inception Network*, 2021. Available from: <https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202https://cs231n.github.io/linear-classify/>, (accessed April 2021).
 25. TensorFlow. *Tensorflow Detection Model Zoo*, 2020. Available from: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tfl_detection_zoo.mdhttps://cs231n.github.io/linear-classify/, (accessed April 2021).
 26. Tzutalin. *LabelImg*, 2015. Available from: <https://github.com/tzutalin/labelImg>, (accessed April 2021).
 27. LancashireFire. *Bedroom Fire Test*, 2013. Available from: <https://www.youtube.com/watch?v=ezJ6SorlpJo&t=46s>.
 28. Amirishetty A. *Fire Accident Caught on Camera*, 2018. Available from: <https://www.youtube.com/watch?v=HHWEU6lsgQE>, (accessed April 2021).
 29. Zhang Q-X, Lin G-H, Zhang Y-M, et al. Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images. *Proced Eng* 2018; 211: 441–446.
 30. Kong S. G., Jin D., Li S, et al. Fast fire flame detection in surveillance video using logistic regression and temporal smoothing. *Fire Saf J* 2016; 79: 37–43.
 31. Tien PW, Wei S, Calautit J, et al. A deep learning approach towards the detection and recognition of opening of window for effective management of building ventilation heat losses and reducing space heating demand. *Renewable Energy* 2021; 177: 603–625.
 32. Tien PW, Wei S and Calautit JK. Vision-based human activity recognition for reducing building energy demand. *Building Services Engineering Research and Technology* 2021; 42(6): 691–713.