

Detecting Flashover in a Room Fire based on the Sequence of Thermal Infrared Images using Convolutional Neural Networks

M. Hamed Mozaffari^{†,*}, Yuchuan Li[‡], Yoon Ko[◇]

^{†, ‡, ◇} Fire Safety Unit, Construction Research Centre,
National Research Council Canada,
Ottawa, Ontario, Canada, K1A 0R6

Abstract

Flashover phenomena accompanying rapid fire propagation in a room occur when the hot smoke from a fire accumulates in the room's upper part. This phenomenon presents one of the most frightening and challenging situations for firefighters. A typical approach to mitigate and prevent the impact of flashover is to train firefighters to monitor a few common indicators of fire in pre-flashover time, such as moving dark smoke, high heat, and fire rollover. In actual compartment fire events, these pre-flashover indicators are hard to recognize. Furthermore, determination of exact flashover time is difficult by just observing fire activities while there are other vital rescue duties to do by firefighters. Hence, automatic detection and prediction of flashover in real time are of paramount importance to save lives and reduce the cost of damages. Flashover prediction is still an open area of research by fire safety experts. Deep convolutional neural networks are currently dominating the area of computer vision, and these state-of-the-art deep learning models have been successfully used in various applications, including object detection, localization, and segmentation. Unlike previous studies that use RGB images, sensors, and gauges, we utilized the power of deep learning techniques to detect flashover from image sequences captured by thermal infrared (IR) cameras. Our experimental results indicate that not only our proposed approach can detect flashover in IR video data with high precision, but it can detect flashover a few frames before happening. Our technique is a promising approach that can be used in future for flashover prediction in real time.

Keywords: flashover detection, deep learning, flashover prediction, room fire safety

1. Introduction

A fire that started in a room could grow and spread to the building. The increasing number of buildings has boosted this problem to a higher dangerous level each year, and firefighters encounter room fires routinely in emergency response calls. More than 30,000 firefighters get injured each year during firefighting operations [1]. In a room with modern furniture containing many combustible materials [2], such as chemical fibre/foam products, fire tends to grow fast, which would leave a short period of time available for the firefighters to rescue, operate, and escape until the room flashover. When the room flashover occurs, all the combustible materials in the room are suddenly ignited near-simultaneously since the hot smoke layer accumulated near the ceiling rapidly heats the exposed surface of all the items in the room. Figure 1.A to 1.D shows a few images of a test compartment fire. As can be seen in Figure 1, it takes less than 3 mins from ignition to flashover onset.

Failing to recognize signs of the impending flashover could cost the lives of firefighters and occupants in the building. There is no absolute criterion for flashover prediction since flashover occurrence depends on several factors, such as room specifications, fuels in the room, and the sizes of windows and doors. In a typical room fire, the onset of flashover corresponds to the time when the upper smoke layer in the room reaches a temperature around 600 °C [3]. Also, at the time of flashover occurrence, the heat flux (HF) from the

*mhamed.mozaffarimaaref@nrc-cnrc.gc.ca

upper smoke layer reaches approximately 20 kW/m^2 when measured on the floor [3]. In addition, as another criterion for flashover detection, the onset of flashover can be estimated by monitoring the growth of the Heat Release Rate (HRR). While varied with the size of the room, doors, and windows, the average HRR value needed for flashover is more than 1000 kW and less than 2000 kW for a small room (the standard compartment mentioned in ISO 9705 standard [4]) with one open doorway.

Various investigations have been conducted to develop a new method for flashover detection and prediction, employing HRR, HF, and temperature data [3–5]. Simulation data by CFD (Computational Fluid Dynamics) modelling and data analysis by machine learning techniques have been used to study flashover occurrences and to predict flashover onset time [6–8]. Many of these studies used data such as temperature, HF, and HRR trends to detect flashover time, the acquisition of which requires various sensors and gauges. The main drawback of their methods is that several sensors such as thermocouples and heat flux gauges with sufficient high-temperature resistance should be fixed readily in specific places in a fire compartment. Therefore, their flashover prediction approaches cannot be easily utilized by firefighters in actual fire situations. Rather, they could be used for limited analyses of a fire in controlled environments and simulations.

With recent advancements in digital technology and computer science, thermal IR (Infrared) and digital RGB cameras have become more affordable and powerful than before. Thus, firefighters have been equipped with different types of these cameras (see Figure 1.E). Consequently, using these cameras currently available on the fireground could bring straightforward solutions for detecting and predicting flashover. In addition, the new vision-based solutions could be powered by the state-of-the-art machine learning (ML) techniques. For instance, a recent study [9, 10] employed RGB and IR cameras to predict flashover in room fires [9], and the authors in [11] proposed a preliminary method for prediction of flashover time based on the dynamic change of measured smoke area in RGB video data.

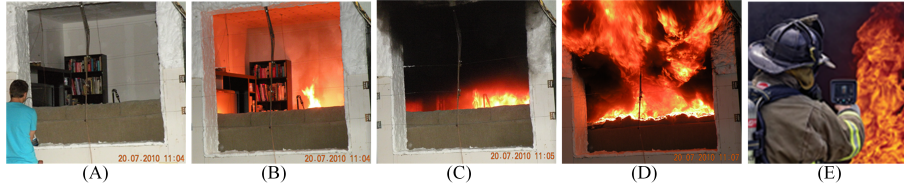


Figure 1. (A-D) Quick growth of fire in a test room fire. Flashover happens in less than three minutes in a room with furniture and book shelves. (E) A firefighter uses one thermal IR camera before going into a building with fire.

To maximize the data analytics of both RGB and thermal IR data, which is essential in developing a powerful vision-based technique for flashover detection and prediction, it is necessary to extract features such as the temperature of smoke from the different RGB and thermal IR frames by aligning them from the same field of view and in the same scale. To mitigate this problem, [9, 10, 12] modified Generative Adversarial Networks (GANs), which is a new deep learning method for image transformation, followed by image segmentation techniques are used to find superimposed RGB and thermal IR data. The drawback of this idea is that transformed data could not be completely accurate and reliable, and the whole process is computationally expensive. The authors in [13] studied flashover prediction in image data generated by CFD-based simulations. For this reason, an approach relying only on IR data or RGB video data entices fully automatic solutions saving some loops of image processing.

2. Method

In this study, we proposed a new classifier using deep learning to detect flashover occurrences in room fires captured by a thermal IR camera. Our approach is fully automatic and end-to-end since it relies only on the IR data without using RGB video data, which enhances the accuracy and speed of the solution. Convolutional Neural Networks (CNN) are feedforward neural networks with several consecutive layers designed, inspired by the visual cortex region in the human brain, to extract hierarchies of digitalized data features such as images from low-level to high-level patterns [14]. A typical CNN model consists of an input layer, many convolutional blocks, a few fully connected layers, and an output layer. Each convolutional block consisting of convolutional layers might be followed by pooling, drop-outs, and batch-normalization layers [14, 15]. Mathematically, CNN is a simple linear and non-linear weight multiplication from input to output. By fine-tuning thousands of weights in different layers of CNN, it can memorize or be trained to extract features of input data. Depending on the application, the output layer acts as a classifier and classifies input data based on extracted features. Training of a CNN network is an iterative optimization process in which the gradient of error (the difference between the output images and ground-truth images) is calculated layer by layer using the chain rule and back-propagated from the output layer to the input layer in each iteration. After several iterations, network weights are tuned to extract features from even novel input images. Deep details of CNN architectures and functionality of each layer can be found in deep learning literature [14–16].

In spite of the achievements of CNN models in various computer vision tasks, the number of convolutional layers in original CNN models was limited to a few layers. The reason is that the back-propagated gradient, calculated from the output layer toward the input layer, decreases after each backward layer. For a CNN architecture with hundreds of layers, the gradient reaches zero ahead of the input layer. This problem is called the vanishing gradient. One of the recent enhancements in CNN models to mitigate this problem is convolutional blocks named Residual block [17]. Figure 2 shows a sample of the residual block where gradient could also skip between CNN layers to avoid vanishing. This new idea helps researchers to propose and train various deeper convolutional neural networks applicable to more complicated computer vision problems than the previous CNN models. In the present study, we selected the Residual Network model [17] with 18 layers (ResNet18). We trained ResNet18 from scratch for detecting flashover in thermal IR image sequences. We also trained ResNet18 using transfer learning [18] by freezing all layers of the CNN network while training only the last fully connected layer. ResNet18 was initialized in both training approaches by pre-trained weights using ImageNet [19] dataset. We modified the last layer of ResNet18 to have two neurons. Details of the ResNet18 network used in this study are presented in Figure 3.

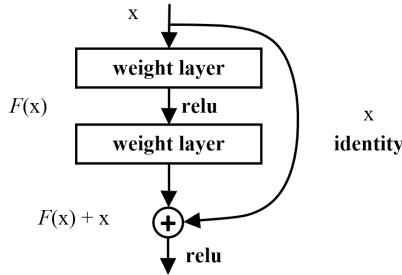


Figure 2. A simple residual block. The gradient of error also passes through a skip connection to avoid the gradient vanishing problems.

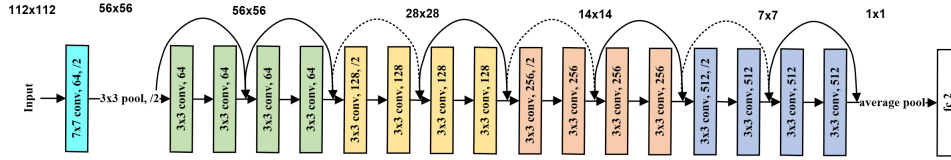


Figure 3. Details of ResNet18 architecture used in our study.

3. Experiments and Discussions

3.1. Dataset Preparation

In order to train, validate, and test our ResNet18 model, we gathered data consisting of thermal IR image frames captured from six room fires, which were conducted by the Fire Safety Unit, the National Research Council of Canada (NRC), Canada [20] for the Characterization of Fires in Multi-Suite Residential Dwellings (CFMRD) project. Thermal IR video and sensor data were acquired from the six tests using different room specifications and materials. By synchronizing thermal IR videos with the data from sensors, we could find the flashover onset and determine the corresponding IR frame. In general, we considered four videos for training and validation with the ratio of 90% and 10%, respectively. Table 1 provides metadata of the videos used in our experiments. To test our trained ResNet18 models, we considered two novel full thermal IR videos: 23-SI-76 comprising flashover (Figure 4.A) and 22-SI-22 (Figure 4.B) containing no flashover. As it can be seen from Figure 4, both videos have similar scene and fire conditions; however, analysis of the data recorded by sensors and thermal IR video verified that flashover did not occur in the test 22-SI-22. It is noteworthy to mention that thermal IR videos contain thermal information where the intensity of pixels is not in the range of RGB video data (i.e., [0-255]) used in computer vision applications. However, normalizing thermal IR video frames into the range of [0-255] made them not only useable as similar to RGB format but also preserved the original thermal information in the form of intensity levels. It means that the higher the intensity, the higher the temperature in those pixels.

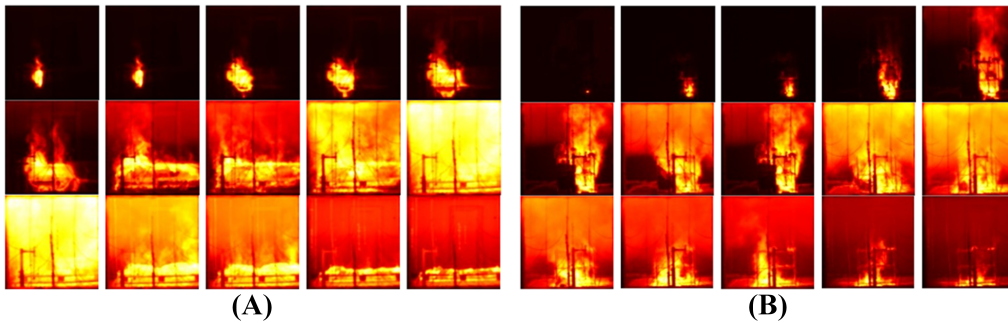


Figure 4. Few randomly selected frames from the thermal IR video (A) 23-SI-76, (B) 22-SI-22. Note the similarity between the two videos in terms of fire growth and scene characteristics.

Table 1 also itemized the number of thermal IR frames used in our experiments as well as the frame identified as the flashover onset, which was determined using the common flashover criteria (Temperature, HF, and HRR). The detail of the test room specifications, the number of sensors, the types of burning materials in each room and the test data can be

IR Video name	Utilization	Burning items	The number of IR frames from ignition to end of flashover	The frame number of the flashover onset
PRF-07	Train/Validation	Combination	113	92
08-SI-04	Train/Validation	Mattress	64	48
14-SI-06	Train/Validation	Mattress	34	22
21-SI-10	Train/Validation	Bed and Mattress	104	60
22-SI-22	Test no Flashover	Toys	364	-
23-SI-76	Test Flashover	Bed assembly	228	163

Table 1. Metadata related to thermal IR videos used in our experiments.

found in the original CFMRD report [20]. Our train/validation dataset contains 222 frames of no-flashover scenes and 93 flashover scenes, in total 315 thermal IR frames. For this reason, the dataset is not balanced between the two classes of no-flashover and flashover. To mitigate the problem of small dataset size, we used a simple data augmentation consisting of normalization, random and center cropping and resizing, and horizontal flipping.

3.2. Experiments and Results

As mentioned in the previous section, we developed two versions of ResNet18 by training the whole network and training the last layer by transfer learning method using pre-trained weights from the ImageNet dataset [19]. The input size of both networks is 224×224 , and we selected Cross-Entropy loss function, and Stochastic Gradient Descent (SGD) optimization algorithms [15] for training and validation. The learning rate of SGD gradually decreases from 0.001 in every seven iterations using parameter values of 0.9 for momentum and 0.1 for gamma. In general, we used the mini-batch training method by randomly selecting mini-batches of size 5 for 25 epochs for both ResNet18 models. The models with the best loss and accuracy values in the validation stage were saved for the subsequent testing stage. The whole experiments were conducted by a Tower PC equipped with an NVidia Titan RTX GPU in Python language and using PyTorch deep learning library.

Our flashover detection experimental results revealed that both ResNet18 models performed similarly for the test video 23-SI-76. In contrast, for the test video 22-SI-22, the fine-tuned ResNet18 by transfer learning performed better with fewer false flashover detections. Figure 5 to 8 present flashover prediction of our method, compared with the sensor data from the room fire tests. In the figures, the flashover period is marked from its onset estimated based on the sensor data. The temperature of the near ceiling region, HRR and heat flux on the floor are plotted and compared with our model results of flashover probability since fire-safety experts consider all flashover criteria together (i.e., the temperature of near ceiling region, HRR, door, and window specifications, and heat flux on floor level). In Figure 5 and 6, it can be seen that ResNet18 can detect the impending flashover quite earlier (specifically, on average 41 seconds (or 21 frames) in advance with an average probability of 94.4% flashover detection) than the actual flashover onset verified by the test sensor data. One reason for this early detection is that in visual terms dissected in a second, flashover onset is not a sharp phenomenon, and pre-flashover frames have a high visual correlation with the frames of the flashover region, which is well captured by our model.

Our experimental results also show that the sharp increase in ResNet18 predictions before the actual flashover (such as in Figure 5) can be generalized as a promising solution for the prediction of flashover. To evaluate the performance of ResNet18 models for situations where the fire doesn't grow to the flashover level, we conducted the same comparison study of our model for the test video 22-SI-22. The sensor data from the test verified that no flashover

happened in the 22-SI-22 experiment. As compared in Figure 7 and 8, the ResNet18 model trained by the transfer learning method performed better than the ResNet18 model trained from scratch. This was not an unexpected conclusion since many studies have proved that training a deep learning model by the idea of transfer learning and domain adaptation increases the performance of the model in most cases.

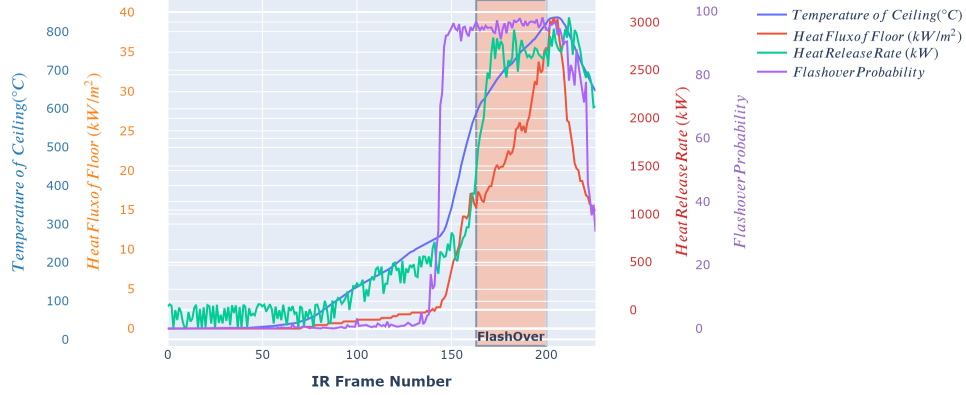


Figure 5. Results of flashover detection by ResNet18 model trained from scratch, which is compared with the flashover period verified by the sensor data from the room fire test of 23-SI-76.

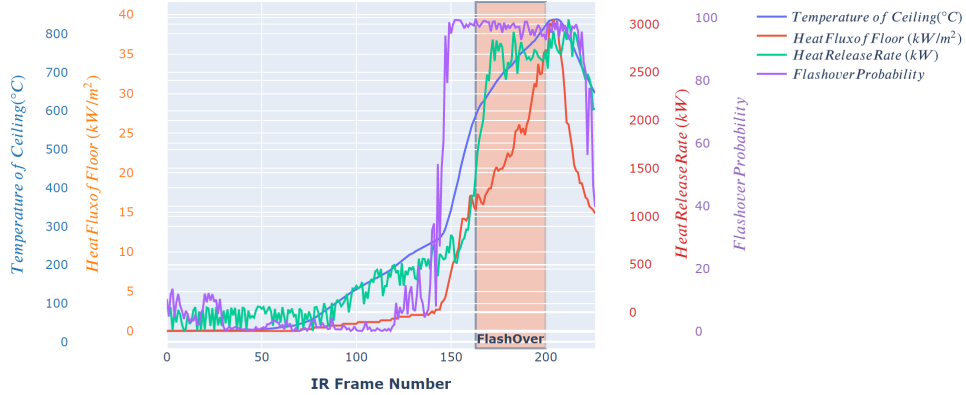


Figure 6. Results of flashover detection by ResNet18 model which only its last layer trained by transfer learning technique, which is compared with the flashover period verified by the sensor data from the room fire test of 23-SI-76.

There are plenty of predetermined classification and detection metrics (see Table 2) that allow us to assess the performance of our deep learning models. To calculate those metrics, the first step is to determine the confusion matrix for each experiment that presents the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) predictions. Binary classification is a particular problem where there are just two classes of data, in our case: flashover (positive) and non-flashover (negative). Since the output of the last dense layer of deep learning models is the probability of flashover occurrence in every frame of test videos, a threshold value should be applied to convert the probabilities into binary values. For this reason, the values of classification metrics are dependent on the

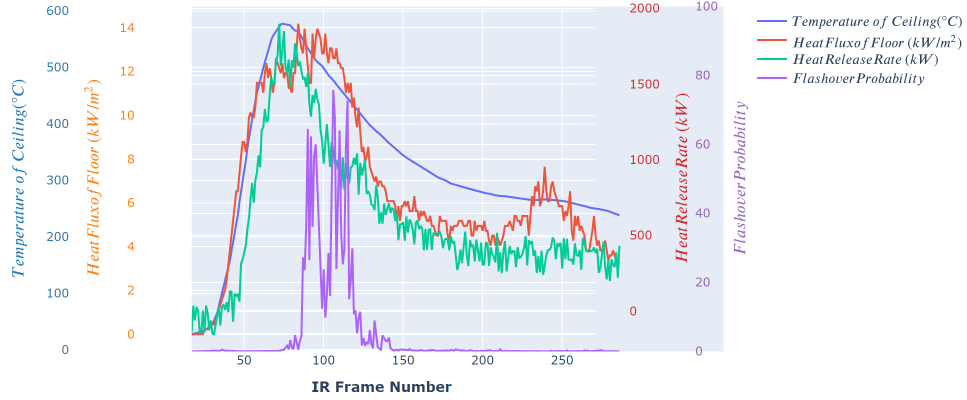


Figure 7. Results of flashover detection by ResNet18 model trained from scratch, which is compared with the flashover period verified by the sensor data from the room fire test of 22-SI-22.

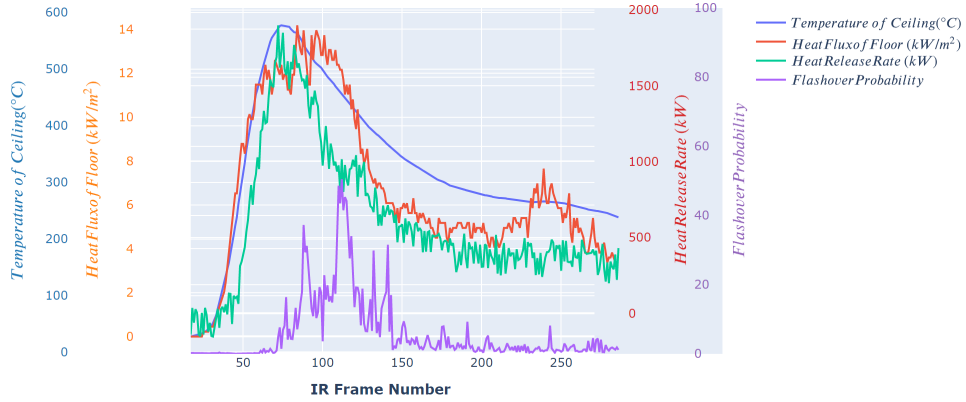


Figure 8. Results of flashover detection by ResNet18 model which only its last layer trained by transfer learning technique, which is compared with the flashover period verified by the sensor data from the room fire test of 22-SI-22.

selected threshold value. We evaluated our ResNet18 models on the two test thermal video data. Figure 9 present classification metrics rates for each experiment.

False Positive Rate (FPR) (Equation 1) is also known as type I error. This criterion shows the fraction of false alerts that our model predicted. It is important for fire safety research because as the False Positive alarm raised by the system would be a massive waste of fire safety resources. False Negative Rate (FNR) (Equation 2), also known as type II error, is one of the most critical metrics for fire safety research which shows what portion of the flashover frames were predicted by the model as non-flashover cases. True Negative Rate (TNR) (Equation 3) or Specificity measures how many cases out of all the non-flashover cases the model could classify as non-flashover correctly. Similarly, True Positive Rate (TPR) (Equation 4) or Recall, also known as Sensitivity, looks for how many flashover cases correctly have been classified as a flashover situation. These four metrics are barely used alone for evaluations of the classification results, and they are utilized for calculating more reliable metrics. From figure 9, both models show relatively low values for FPR and FNR and high values for TNR and TPR for thresholds between 30 and 70 which can be considered

Type I Error	$FPR = FP / (FP + TN)$	Equation 1
Type II Error	$FNR = FN / (TP + FN)$	Equation 2
Specificity	$TNR = TN / (TN + FP)$	Equation 3
Recall	$TPR = TP / (TP + FN)$	Equation 4
Precision	$PPV = TP / (TP + FP)$	Equation 5
Accuracy	$ACC = (TP + TN) / (TP + FP + TN + FN)$	Equation 6
F1 Score	$F1\ score = 2TP / (2TP + FP + FN)$	Equation 7

Table 2. Definition of classification evaluation metrics used in this study.

as good classifiers. FPR value rarely reached zero, and the reason is that there is no clear, sharp border between flashover and non-flashover region in our test thermal IR video data. It also indicates that our system should work pretty well for fire cases similar to the ones in our dataset. At the same time, the imbalanced nature of our dataset, with fewer positive samples than negative ones, is the reason for getting non-zero FNR for thresholds of more than 30.

We calculated Positive Predictive Value (PPV) (Equation 5) or Precision to measure how many flashover cases are in fact actual flashovers. Our experimental results show that our models can reach the precision of 0.8, which is good for our dataset with a limited number of samples. Usage of Accuracy (ACC) (Equation 6) metric for an imbalanced dataset like ours will provide a better result. However, we determined the accuracy by measuring how many cases, both flashover and non-flashover, were correctly classified. Our best ResNet18 model could reach an average accuracy of 91.3%, which is considerable. From the combination of precision and recall, we could find F1 score criteria (Equation 7). It is one of the most commonly used metrics for classification tasks in the machine learning area. The higher the F1 score is, the classification model performs better. It is also crucial for the dataset with unbalanced samples on performance judgement. Thus, it would tell more than other metrics, such as accuracy, considering the nature of our dataset. Our best ResNet18 model could reach the average F1-score of 84.1%. From the F1 score, we can also determine which threshold value could be selected for our experiments. It can be seen from Figure 9 that the results of our models will not be changed significantly by changing the thresholds, which indicates the robustness of our ResNet18 models.

There is a trade-off between many classification metrics. For instance, TPR and FPR are in contrast. To figure out this trade-off better, we draw TPR and FPR for every threshold value in a graph known as Receiver Operating Characteristic (ROC). For all threshold values, a larger area under the ROC Curve (AUC) means better classification results. Similar to the F1 score, AUC has advantages in revealing the real performance on an unbalanced dataset by considering the impact of both the positive samples and negative samples. Therefore, it is vital to our model. Likewise, we can analyze the trade-off between precision and recall values and determine the AUC of this graph as well. Figure 10 shows the ROC and Precision-Recall graphs related to both ResNet18 models for the test thermal IR video 23-SI-76, where the AUC for the ResNet18 model trained from scratch was 0.88, and the model fine-tuned its last layer by transfer learning reached to AUC of 0.93.

4. Conclusion

In this study, we showed that deep convolutional neural networks, specifically the ResNet18 model, can be trained by a small number of IR thermal images using a domain adaptation approach to detect flashover onset in thermal IR videos. Unlike previous attempts of using deep learning methods for flashover prediction, our approach is end-to-end and fully automatic, without using any data processing, video transformations, or extensive data

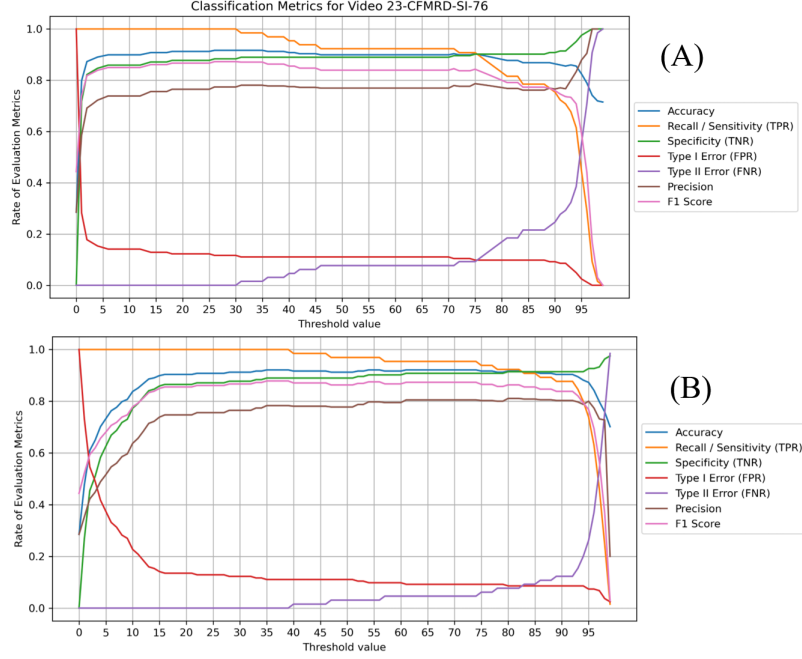


Figure 9. Comparison between classification metrics to evaluate our deep learning models for test thermal IR video 23-SI-76 (A) ResNet18 trained from scratch, (B) ResNet18 trained by fine-tuning the last layer by transfer learning approach.

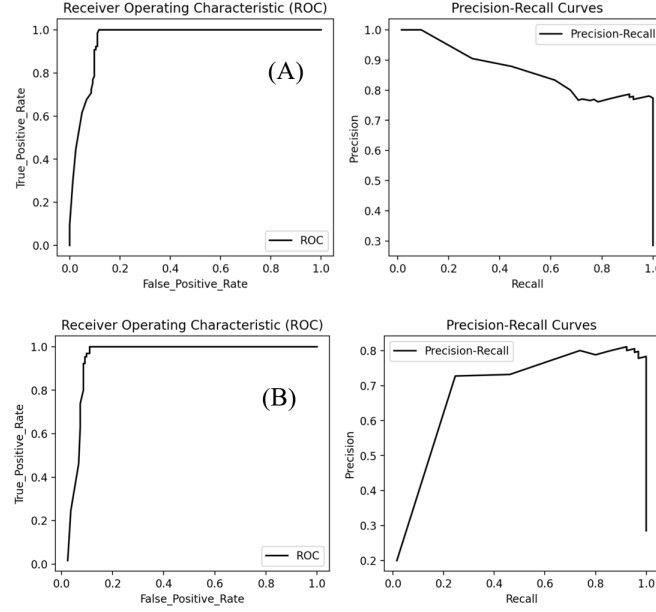


Figure 10. Comparison between classification metrics to evaluate our deep learning models for test thermal IR video 23-SI-76 (A) ResNet18 trained from scratch, (B) ResNet18 trained by fine-tuning the last layer by transfer learning approach.

augmentation. The temperature of the smoke layer near the ceiling and heat flux of floor gathered by sensors and gauges as well as calculated heat release rate from room fire tests were also used to determine the accurate occurrence of flashover in terms of time and the corresponding IR frame in our dataset. Two thermal IR videos were selected, one captured from a room fire test with no-flashover and another room fire test with flashover to evaluate our proposed deep learning methods. The comparison study between the network predictions for the two test thermal IR videos and the sensor data from the room fire tests indicates that the ResNet18 model can detect the impending flashover earlier than the actual flashover onset. We also calculated a comprehensive list of classification performance metrics to evaluate our methods. ResNet18 models could reach an average accuracy of 91% with the ROC of 0.93, which is a considerable result relative to our small dataset. As a future path, our new vision-based technique, benefiting only from thermal IR video data, is a promising solution that can be used in predicting flashover way before it actually occurs.

References

- [1] M. E. Bowyer, V. Miles, T. N. Baldwin, and T. R. Hales. *Preventing deaths and injuries of fire fighters during training exercises*. Tech. rep. 2016.
- [2] A. Yuen, G. Yeoh, R. Alexander, and M. Cook. “Fire scene reconstruction of a furnished compartment room in a house fire”. In: *Case Studies in Fire Safety* 1 (2014), pp. 29–35.
- [3] R. D. Peacock, P. A. Reneke, R. W. Bukowski, and V. Babrauskas. “Defining flashover for fire hazard calculations”. In: *Fire Safety Journal* 32.4 (1999), pp. 331–345.
- [4] V. Babrauskas, R. D. Peacock, and P. A. Reneke. “Defining flashover for fire hazard calculations: Part II”. In: *Fire Safety Journal* 38.7 (2003), pp. 613–622.
- [5] G. E. Gorbett, R. Hopkins, and P. Kennedy. “The current knowledge & training regarding backdraft, flashover, and other rapid fire progression phenomena”. In: *annual meeting of the National Fire Protection Association, Boston, MA*. 2007.
- [6] J. Wang, W. C. Tam, Y. Jia, R. Peacock, P. Reneke, E. Y. Fu, and T. Cleary. “P-Flash—A machine learning-based model for flashover prediction using recovered temperature data”. In: *Fire Safety Journal* 122 (2021), p. 103341.
- [7] D. Cortés, D. Gil, J. Azorín, F. Vandecasteele, and S. Verstockt. “A review of modelling and simulation methods for flashover prediction in confined space fires”. In: *Applied Sciences* 10.16 (2020), p. 5609.
- [8] S. A. Yusuf, A. A. Alshdadi, M. O. Allassafi, R. AlGhamdi, and A. Samad. “Predicting catastrophic temperature changes based on past events via a CNN-LSTM regression mechanism”. In: *Neural Computing and Applications* (2021), pp. 1–16.
- [9] Y. Li. “Dual-Attention Generative Adversarial Network and Flame and Smoke Analysis”. PhD thesis. Université d’Ottawa/University of Ottawa, 2021.
- [10] Y. Li and Y. J. Ko. *Development of a Hybrid Algorithm to predict room fire flashovers based on Vision data*. Tech. rep. National Research Council of Canada, 2021.
- [11] A. Huyen, K. Yun, S. De Baun, S. Wiggins, J. Bustos, T. Lu, and E. Chow. “Dynamic fire and smoke detection and classification for flashover prediction”. In: *Pattern Recognition and Tracking XXXII*. Vol. 11735. International Society for Optics and Photonics. 2021, p. 1173502.
- [12] K. Yun, J. Bustos, and T. Lu. “Predicting rapid fire growth (flashover) using conditional generative adversarial networks”. In: *Electronic Imaging* 2018.9 (2018), pp. 127–1.
- [13] L.-c. Su, X. Wu, X. Zhang, and X. Huang. “Smart performance-based design for building fire safety: Prediction of smoke motion via AI”. In: *Journal of Building Engineering* 43 (2021), p. 102529.
- [14] M. H. Mozaffari and W.-S. Lee. “Semantic Segmentation with Peripheral Vision”. In: *International Symposium on Visual Computing*. Springer. 2020, pp. 421–429.
- [15] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.
- [16] S. Albawi, T. A. Mohammed, and S. Al-Zawi. “Understanding of a convolutional neural network”. In: *2017 International Conference on Engineering and Technology (ICET)*. Ieee. 2017, pp. 1–6.

- [17] K. He, X. Zhang, S. Ren, and J. Sun. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.
- [18] M Hamed Mozaffari and W.-S. Lee. “Domain adaptation for ultrasound tongue contour extraction using transfer learning: A deep learning approach”. In: *The Journal of the Acoustical Society of America* 146.5 (2019), EL431–EL437.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [20] A. Bwalya, E. Gibbs, G. Loughheed, and A. Kashef. *Characterization of fires in multi-suite residential dwellings: final project report: part 1-A compilation of post-flashover room fire test data*. Tech. rep. National Research Council of Canada, 2014.