Wang, Z., Zhang, T. & Huang, X. Explainable deep learning for image-driven fire calorimetry. Appl Intell 54, 1047–1062 (2024).
https://doi.org/10.1007/s10489-023-05231-x

Explainable Deep Learning for Image-Driven Fire Calorimetry

Zilong Wang, Tianhang Zhang, Xinyan Huang*

Research Centre for Fire Safety Engineering, Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

**Corresponding to xy.huang@polyu.edu.hk (X. Huang)*

Abstract: The rapid advancement of deep learning and computer vision has driven the intelligent evolution of fire detection, quantification and fighting, although most AI models remain opaque black boxes. This work applies explainable deep learning methods to quantify fire power by flame images and aims to elucidate the underlying mechanism of computer-vision fire calorimetry. The process begins with the utilization of a pre-trained fire segmentation model to create a flame image database in various formats: (1) original RGB, (2) background-free RGB, (3) background-free grey, and (4) background-free binary. This diverse database accounts for factors such as background, color, and brightness. Then, the synthetic database is employed to train and test the fire-calorimetry AI model. Results highlight the dominant role of determining fire dimensions in fire calorimetry, while other factors displaying minimal influence. Enhancing the accuracy of flame segmentation significantly reduces error in computer-vision fire calorimetry to less than 20%. Finally, the study incorporates the Gradient-weighted Class Activation Mapping (Grad-CAM) method to visualize the pixel-level contribution to fire image identification. This research deepens the understanding of vision-based fire calorimetry and providing scientific support for future AI applications in fire monitoring, digital twin, and smart firefighting.

Keywords: *Flame images; Fire Engineering; Smart firefighting; Semantic segmentation; Image analysis*

Abbreviation

ANN	Artificial neural network	IoU	Intersection over union
CNN	Convolution neural network	*MSE*	Mean square error
FCN	Fully convolutional network	*ReLU*	Rectified linear units
Grad-CAM	Gradient-weighted Class Activation Mapping	*VGG*	Visual Geometry Group
HRR	Heat release rate		

1. Introduction

Fire is still one of the most common hazards in modern society, posing significant threats to people and causing substantial economic losses (Fig. 1). In 2021, the United States fire service department responded to about 486,500 structure fires and twice as many outdoor fires [1]. These fires tragically led to about 3,800 fatalities, including 135 on-duty firefighters and incurred substantial economic losses amounting to $15.9 billion. In a firefighting operation, the intensity and intensity of a fire, often quantified by the heat release rate (HRR), plays a pivotal role in assessing fire danger and its progression [2]. Accurate knowledge of the HRR is instrumental in estimating the minimum fire-hose water flow rate for fire suppression, which is critical in firefighting emergency response [3–6].

However, the absence of comprehensive and real-time fire scene information places firefighters in a situation, where they must rely on their judgment and experience to rapidly assess the fire scenario and forecast the fire development. Firefighters typically estimate the intensity of fire by visually inspecting factors, such as the size and color of fire and smoke, drawing from their memory and past experiences. This reliance on visual observation and past knowledge can occasionally lead to misjudgment of critical fire events, such as flashover and backdraft, resulting in delayed response times, increased risks to firefighters, and potentially tragic consequences. To ensure firefighters' safety and enhance their decision-making during emergency responses, using fire calorimetry to measure fire power and other critical information becomes essential. This approach aims to provide accurate and real-time data concerning fire intensity, development and associated hazards.

Fig. 1. Typical images of (a) building fires, (b) vehicle fires, and (c) forest fires.

So far, many efforts have been made to quantify the fire HRRs of different burning fuels. Conventionally, two principal approaches have been employed in laboratory experiments. The first approach entails the measurement of the fuel mass burning rate at which the fire consumes the fuel (e.g., wood, paper, or plastic). Subsequently, the fire HRR is calculated based on the fuel mass burning rate, in conjunction with factors including the heat of combustion and the stoichiometry of the combustion reaction. The second method is grounded in the principle that the amount of oxygen consumed by a fire is directly proportional to the amount of heat released [7]. This method involves monitoring the oxygen concentration in the exhaust gas from the fire, which is then leveraged to determine the HRR. Additional fire-calorimetry techniques, including those relying on temperature rise [8] and carbon dioxide generation [9], have also been developed. However, these methods are primarily suited for controlled laboratory settings as they necessitate specialized equipment like scales and gas sensors to be pre-installed in proximity to the fire before measurement. Consequently, these conventional fire-calorimetry methods are inapplicable in real firefighting scenarios.

Today, fire images and videos are often available from CCTV cameras or mobile phone cameras that can record flame and smoke behaviors [10–14]. However, their accuracy and reliability are sensitive to camera settings, background light, and experimental interferences. Long and complex image post-processing is also needed, so it cannot measure the real-time fire power. With the development of artificial intelligence (AI) technology [15, 16], especially deep learning methods, real-time analysis of fire images can be realized. So far, most AI computer-vision methods can only identify the existence of fire and smoke from a photo [17–19], but they cannot calculate the fire HRR or determine the fire hazard level. Nevertheless, flame images include massive information on fire behaviors and characteristics, such as flame size, height, color, brightness, and oscillation frequency, as well as their spatial temporal evolution. In-depth analysis of flame images by deep learning can deliver valuable information about fire development.

Previously, we proposed an ideal of fire calorimetry based on numerical fire smoke images and video flame images [20, 21]. Preliminary results show that the deep learning model trained by synthetic fire smoke images or real fire images can well identify the transient fire HRR for various burning fuel and fire cases. However, the influence of the image background and camera settings is unclear. In other words, if we use different cameras to shoot the fire images at a random scene, whether the deep learning mode can continue determining the fire HRR is unexplained. More importantly, the governing information (e.g., flame size, brightness, or color) for this AI-driven fire calorimetry based on fire images is still unknown, posing a big knowledge gap.

To solve these issues, semantic segmentation [22] is needed to achieve the separation of the flame region and the background of the fire image. Recent progress in semantic segmentation are driven by advancements in deep learning. Traditional methods, such as pixel-based classification and graph cuts [23–25], laid the foundation for subsequent approaches. The introduction of fully convolutional networks (FCNs) [26] marked a significant milestone, enabling end-to-end learning and pixel-wise predictions. However, FCNs struggled to capture fine-grained details and handle complex scenes due to their limited receptive field. To address these limitations, U-Net [27] introduced skip connections, which allowed the model to capture both local and global contextual information effectively. By combining the encoder and decoder paths with skip connections, U-Net improved the accuracy of segmentation by preserving spatial information and leveraging multi-scale features. Recently, the Swin-Unet model [28, 29] has emerged as a further advancement in semantic segmentation. Swin-Unet combines the strengths of the Swin Transformer architecture, originally proposed for image classification, with the U-Net structure. The Swin Transformer introduces a hierarchical self-attention mechanism that captures long-range dependencies, enabling the model to have a more comprehensive understanding of the image context. By integrating the Swin Transformer into the U-Net architecture, Swin-Unet achieves state-of-the-art performance in semantic segmentation tasks.

In the domain of fire research, although considerable progress has been made in the development of semantic segmentation models, several challenges continue to persist. Notably, the presence of class imbalance, characterized by a small fire area compared to a large background, poses difficulties in accurately detecting and segmenting fire regions. Additionally, the presence of reflective surfaces in fire scenes presents challenges in precisely delineating boundaries between fire and non-fire regions.

Moreover, real-time performance remains a critical concern in fire-related applications, demanding the development of efficient and fast segmentation algorithms. Consequently, further research must address these challenges and advance the state-of-the-art in fire segmentation methodologies.

This work further explores fast fire segmentation methods and image-based fire calorimetry for burning different fuels. The separation of background, color, and brightness are achieved through semantic segmentation, image greyscale, and binarization, as a way to explore the key factors of fire calorimetry by deep learning models. Finally, Gradient-weighted Class Activation Mapping (Grad-CAM) [30] is adopted to visualize the weight distribution map that highlights the important regions in the image for fire calorimetry. In this way, the framework of explainable image-driven fire calorimetry is proposed to evaluate intensity of fire and hazard under different environmental conditions and with different cameras.

2. Methodology

The objectives of this study are (1) to investigate how the deep learning algorithm correlates the fire HRR to the fire image information, and (2) to determine what is the key information (e.g., image color, brightness, and contrast) that governs the image-driven fire calorimetry. Then, we can seek a generalized model of this new fire calorimetry under any environmental condition and camera setting. The workflow of the proposed flame-image based fire calorimetry framework is shown as Fig. 2.

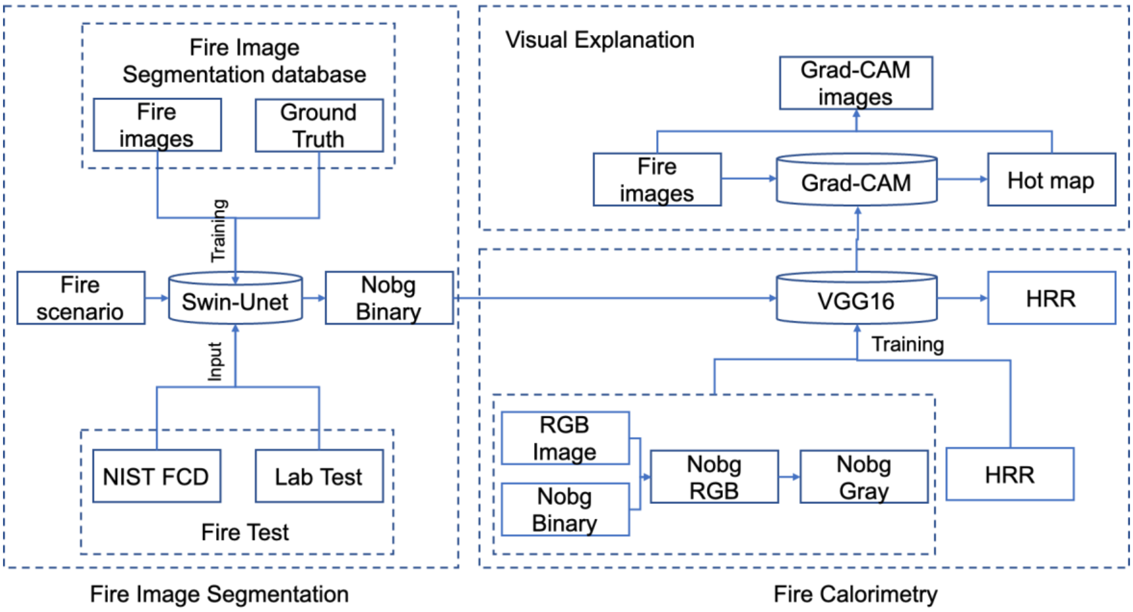

Fig. 2. Workflow of the proposed flame-image based fire calorimetry framework.

To eliminate the effect of image background on fire calorimetry, semantic segmentation is adopted to split the flame and background. Then, four types of fire images are obtained, (1) the original image, (2) the image of flame without the background, (3) the greyscale flame image without color information and background, and (4) the binary flame image without color and brightness information as well as background. These image variations are employed individually to train the fire calorimetry model. The performance of these models is then compared using identical fire test images to discern the factors that contribute most significantly to the model's accuracy. In addition, an explainable algorithm of Grad-CAM is leveraged to highlight the crucial regions within fire images, providing interpretability to the proposed AI fire calorimetry approach.

2.1 Fire-image database

The semantic segmentation model and fire calorimetry model have different requirements (e.g., data structure and label) for the training and validation database [22, 31]. Thus, two different image databases are established to train the flame segmentation model and fire calorimetry model for different functions respectively. The semantic segmentation model for fire images is to remove the background items and only keep the bright flame area (i.e., the filtered flame image). Then, the training database should contain both the original image and the filtered flame image. We choose three databases of fire images, (1) Corsican Fire Database [32], (2) FiSmo Database [33], and (3) videos from NIST Fire Calorimetry Database [34], to train the semantic segmentation model of fire images (Fig. 3).

The Corsican Fire Database contains 2911 multi-modal wildfire images and videos, including original fire images, near-infrared images and background-free binary images. They are segmented manually according to background and flame by an expert. For the FiSmo Database, there are four sub-databases. We choose the BowFire sub-database, which consists of 226 images taken in the urban area (with and without fire). Most fire images in Corsican and FiSmo databases are collected outdoors.

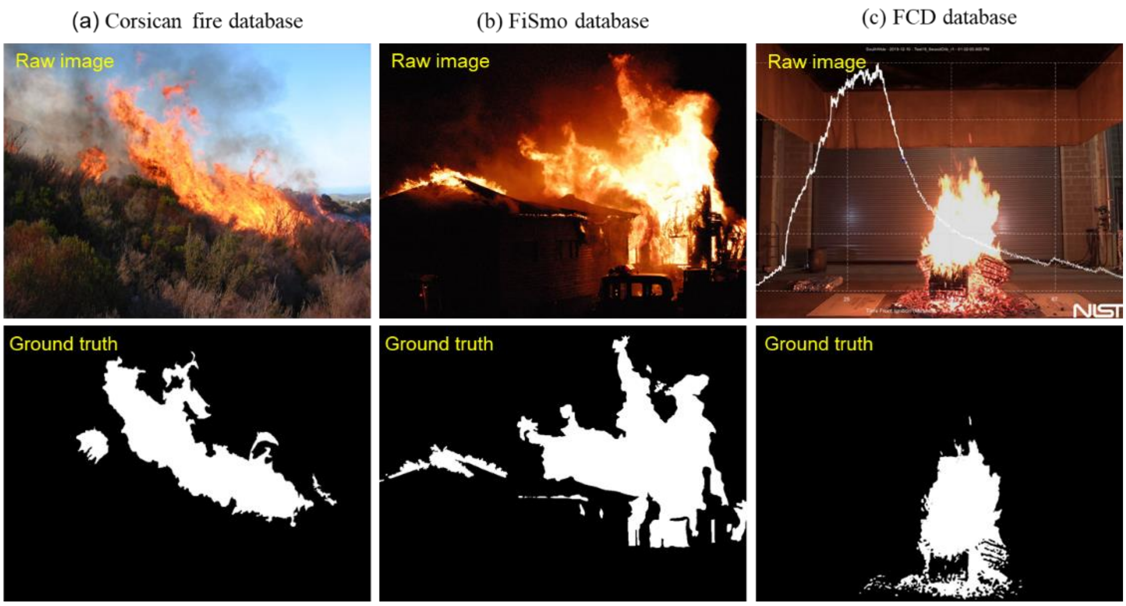

Fig. 3. Fire images in (a) Corsican fire database, (b) FiSmo database, and (c) NIST Fire Calorimetry database.

Comparatively, the fire images of the NIST database were taken in laboratory environments, containing various fuel types like single-burning items, furnished rooms, and controlled burners. The real-time fire HRR of these fires are measured by oxygen calorimetry [34], and its principle is to measure the rate of oxygen consumption in the fire to evaluate its rate of heat release. Each image corresponds to its HRR at that time, and fire tests with a peak HRR of less than 50 kW were ignored because no significant flame behavior was observed. A total of 347 fire images randomly selected from the NIST Fire Calorimetry Database are segmented manually (Fig. 3c). Finally, a total of 1,060 fire images, together with their ground truth images, are used to train the fire segmentation model, and 112 fire tests, including 69,662 fire images, are adopted for fire calorimetry model training.

2.2 Fire segmentation using Swin-Unet

For the original flame image, there is much information irrelevant to the intensity of fire in the background. To eliminate the influence of these extraneous factors, especially the background, we first perform semantic segmentation of the fire image to keep only the bright flame area for the training of AI model. However, conventional image extraction methods require manual processes, neither automatic nor real-time. Thus, a fire segmentation model is applied to separate flame and background automatically. Before training the flame segmentation model, all fire images with their ground truth are segmented into small blocks with the size of 224×224 (Fig. 4). These small image blocks can reduce computational complexity and cost while keeping input and output of the same size.

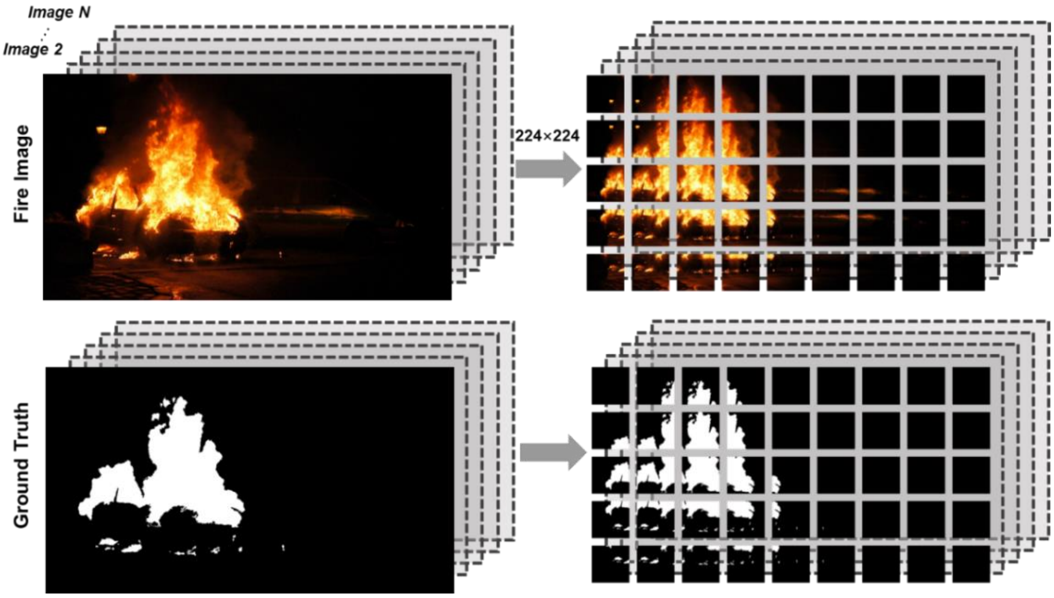

Fig. 4. Dataset pre-processing for fire segmentation.

After the data pre-processing, Swin-Unet [28] is used to achieve the separation of flame and background. The Swin Transformer block [29] is adopted to build an encoder, bottleneck and decoder, which has good performance on medical images. Compared with the Convolution Neural Network (CNN) [35, 36], the Swin Transformer can learn global and remote semantic information interactions well without the limitations of convolutional operations. In addition, Swin-Unet can achieve local global semantic feature learning by feeding the tokenized image blocks to the Transformer-based U-shaped En-Decoder architecture via a jump connection.

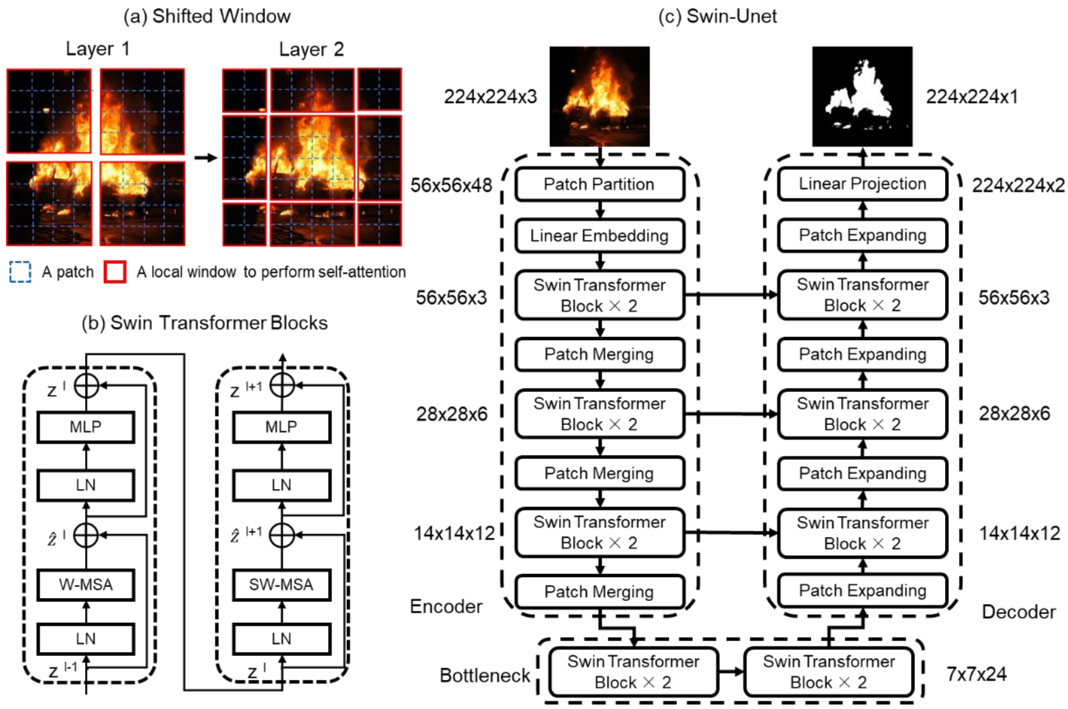

Fig. 5. The architecture of Swin-Unet in fire segmentation.

Fig. 5 shows the detailed architecture of Swin-Unet used for fire image segmentation, consisting of 14 Swin Transformer blocks, three patch merging blocks, four patch expanding blocks and one linear embedding layer with 5,644,477 parameters. The introduction of the shifted window self-attention mechanism (Fig. 5a) allows for better information exchange with other windows and a great reduction in computational complexity. The small block of fire image, as input, is segmented into non-overlapping image patches using patch partition first. Each patch, considered a token, is fed into a transformer-based encoder (Fig. 5b) to extract the main feature. The extracted features are then up-sampled by the decoder with a patch expanding layer and concatenated with multiscale features of the encoder. This is done via a jump connection to recover the spatial resolution of the feature map for further segmentation.

2.3 Fire identification using VGG16

Previously, we tested the effectiveness of VGG16 [21] in measuring fire HRR in the NIST database and got a promising accuracy for most fire images. However, it is unclear what is the key image information that the deep-learning algorithm focuses on to achieve a good or bad measurement. It is assumed that the AI model can recognize both the flame and background images through continuous learning. However, the trained model is easily affected by the no-flame background, especially when the background looks like a flame. What's more, if the same fire scene is shot with different camera apertures, shutter, and ISO (the sensitivity of the camera's image sensor to light), the AI model will give very different predictions. In other words, the influence of the image background and camera setting has not been explored.

Analysis of the original images (see Fig. 6a) reveals raw fire images contain a large amount of information, such as flame area, background, color, and brightness. The color, as well as the brightness of the flame, varies with the camera settings, and the background changes with different fire scenarios. To obtain more accurate and robust fire calorimetry, it is necessary to separate these factors and explore which ones are dominant in image-driven fire calorimetry. By inputting the raw RGB image to the fire segmentation model, we can generate the background-free binary image directly (Fig. 6d). We can also process the image step by step, first only removing the background of the raw image and keeping the RGB images of flame (Fig. 6b). Then, remove the image color (Fig. 6c), and finally remove its brightness into the binary image (Fig. 6d).

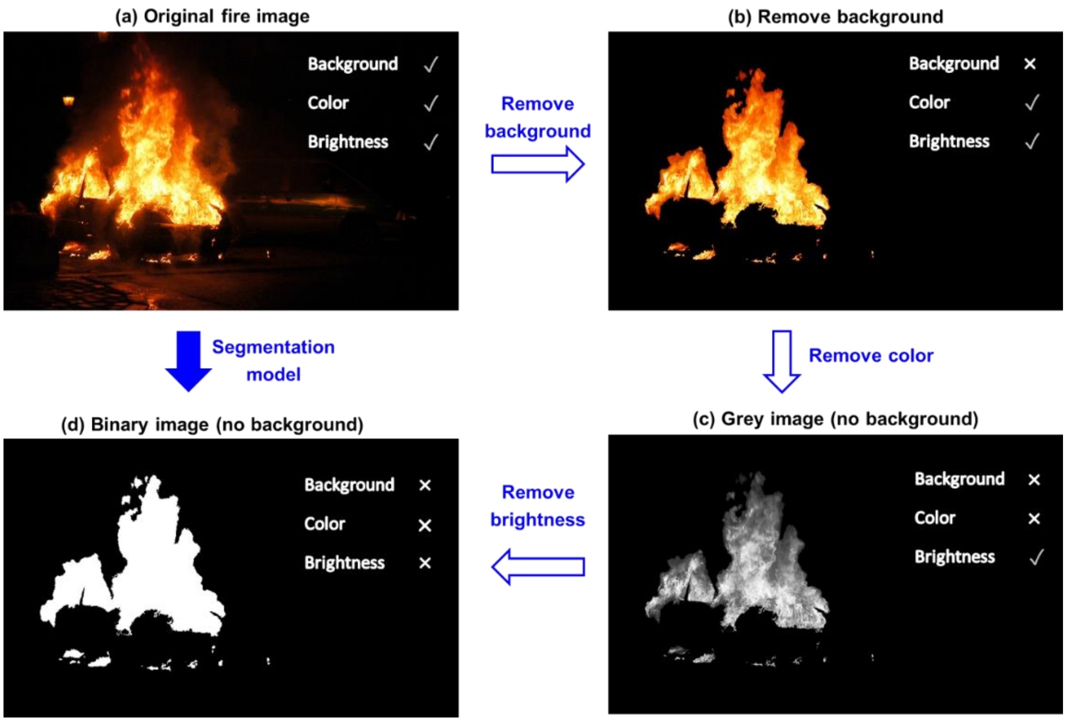

Fig. 6. Different forms of fire images: (a) original RGB fire image, (b) background-free RGB image; (c) background-free grey image, and (d) background-free binary image.

Subsequently, these different forms of fire images are paired with their instantaneous HRRs, resulting in 51,013 sets of paired data. Fig. 7 shows the data extraction process and the distribution of HRR corresponding to all fire images. The fire image dataset consists of raw fire images, background-free flame images, background-free grey images, and background-free binary images, each corresponding to their respective HRR (Fig. 7a). The distribution of HRR values within the database is depicted in Fig. 7b. It is observed that as the HRR increases, the corresponding fire image decreases, leading to an inherent data imbalance. To establish the relationship between these unbalanced data, these distinct datasets are individually inputted into the VGG16 model (Fig. 7c). The VGG16 model used in this work consists of 16 layers, including 13 convolutional layers and 3 fully connected layers. The final layer output is activated using a "linear" activation function to generate the predicted HRR.

Finally, four different fire calorimetry models are generated and the key factors in fire calorimetry can be determined by comparing the performance of different models.

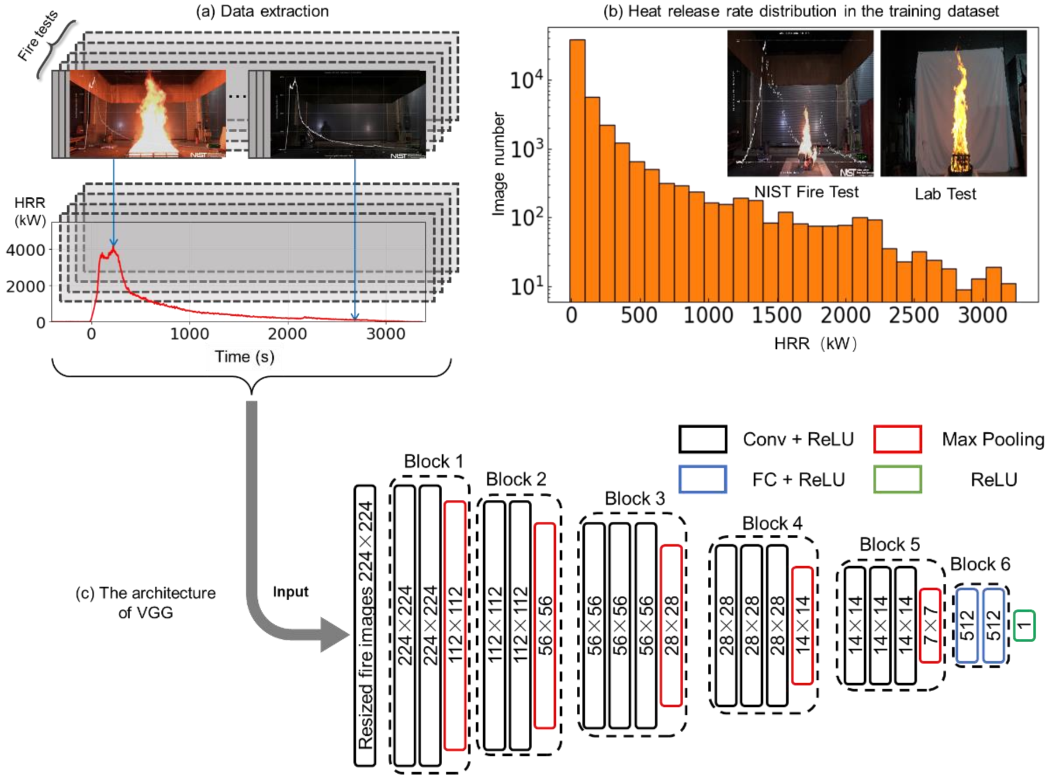

Fig. 7. (a) data extraction, (b) HRR distribution in the dataset, (c) the architecture of VGG for fire identification, updated from [21].

2.4 *Visual explanation using Grad-CAM*

The analysis of image parameters in the vision-driven fire calorimetry starts to interpret the black box of the deep learning model. However, it is still unclear why it is so predictive and where it focuses its attention. Grad-CAM is applied to analyze all fire images to explain further which image regions or pixels that the network focuses on for a given fire scene. Then, it is possible to get a visual explanation for the deep learning model.

Fig. 8 shows the detailed architecture of Grad-CAM. First, the input image is forward propagated to get the feature layer (the output of the last convolutional layer) and the output HRR. To clarify the focused region of the input image for the fire calorimetry, the output HRR is backpropagated to get the gradient information of the feature layer. Such a gradient represents the contribution of each element in the feature layer to the output HRR. The larger the gradient (contribution), the more important the pixel in the image is considered. Finally, the gradient information of each channel in the feature layer is weighted and summed to obtain the Grad-CAM showing the contribution of each pixel in the original image to the identified result.

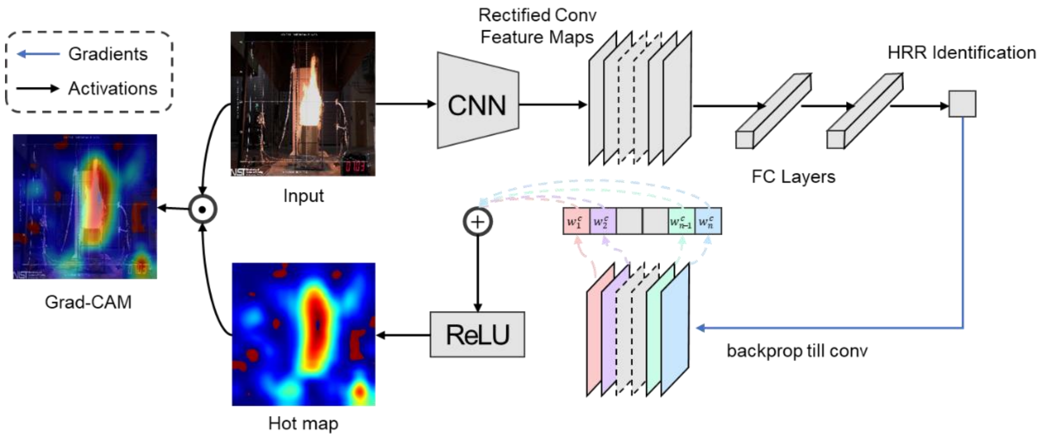

Fig. 8. Grad-CAM for fire calorimetry model

3. Results and discussion

The explainable fire calorimetry models consist of three parts in total: fire segmentation used for separation of influencing factors, fire calorimetry model training and test, as well as visual explanations of fire calorimetry models, which are discussed separately in the following sections.

3.1 Fire segmentation

During the training process for the fire segmentation model, Sigmoid is selected as the activation function of the final output to convert the raw output values into probabilities. The focal loss [37] is used as the loss function to guide the direction of gradient descent. It can up-weight flame pixels in the loss value during classification, thus achieving accurate classification in unbalanced data (Flame pixels are much smaller than background pixels). In addition, the accuracy and Intersection over Union (IoU) metrics are employed to assess the effectiveness of the segmentation model in accurately identifying and classifying flame pixels present in an image. A dropout rate of 0.2 is applied after each multi-layer perceptron (MLP) layer, Swin Transformer block, and skip connection to avoid overfitting.

All the images selected from the Corsican fire database, FiSmo database, and FCD database are further divided into two parts, the training set (90%) for training and validating the model and the test set (10%) for the demonstration purpose. It takes about 17 h in total for 50-epoch training on a server with 32 CPU cores and a Tesla P100 GPU card. After the model training, the deep learning model can achieve real-time flame image segmentation to facilitate fire calorimetry. The model performance during the training is shown in Fig. 9. After 50 epochs of training, the deep learning model reaches convergence with a maximum accuracy of 0.982 and a maximum IoU of 0.936, so the fire segmentation model can accurately identify flame pixels and background pixel in the training dataset.

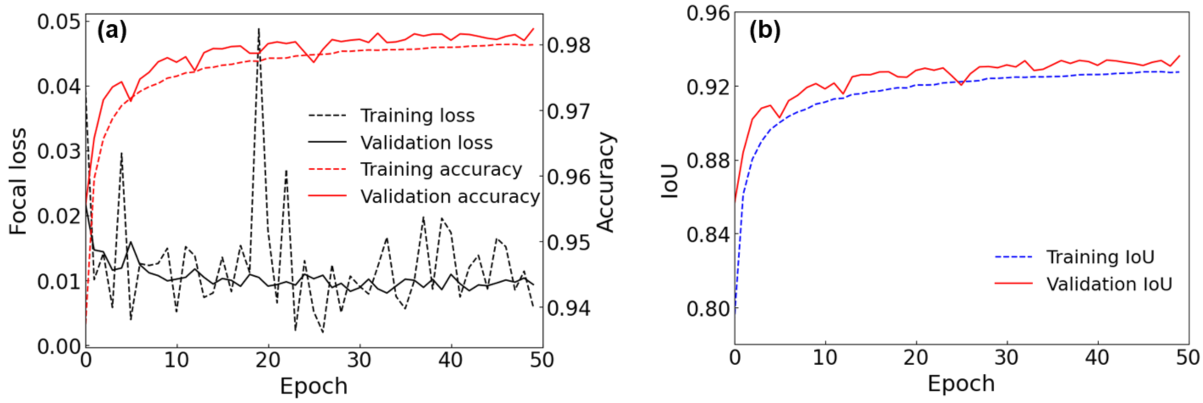

Fig. 9. (a) Focal loss and accuracy and (b) IoU of training and validation during the training process.

In order to further validate the efficacy of the proposed model, a comparative analysis is conducted with other semantic segmentation methods utilizing the fire image dataset. In this comparative experiment, two widely used semantic segmentation methods, namely FCN8 and U-Net, are selected and contrasted against the proposed approach. The evaluation of model performance is based on accuracy and IoU, with consideration given to the segmentation speed relative to the total number of parameters. It is worth noting that a smaller parameter count corresponds to a faster segmentation speed.

Table 1. Comparison of segmentation results using different segmentation models

Method	Size	Accuracy	IoU
FCN8	14,758,030	0.957	0.903
U-Net	7,760,648	0.845	0.733
Swin-Unet	5,644,477	0.982	0.936

As compared in Table 1, the Swin-Unet model outperforms FCN8 and U-Net models in fire segmentation. With a parameter size of merely 5,644,477, it exhibits remarkable accuracy and segmentation speed, achieving values of 0.982 and 0.936, respectively. These findings serve as evidence of the superiority of the Swin-Unet model in accurately segmenting fire images while maintaining efficient computational performance.

After completing the training of the model, the test set is used to verify the segmentation performance of the model, which has never been used in model training. Three images from the three databases mentioned are selected for the model demonstration, representing the segmentation performance of the model in urban areas, forested areas, and laboratory environments, respectively. As shown in Fig. 10, the proposed model can effectively segment fire images in different scenes and environments, with accuracies greater than 0.99 and IoUs more than 0.83. Overall, the Swin-Unet demonstrates the ability to perform real-time and accurate image segmentation, serving as a foundational component for reliable HRR identification independent of camera settings.

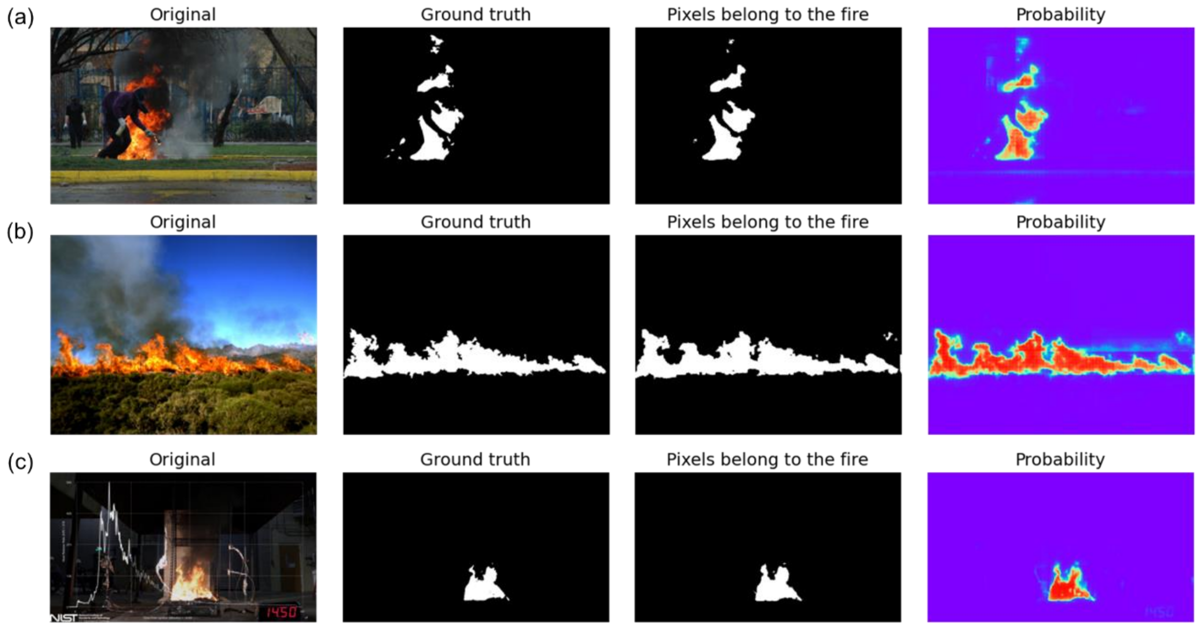

Fig. 10. Segmentation of different database: (a) FiSmo database (Acc=0.99, IoU=0.83), (b) Corsican fire database (Acc=0.99, IoU=0.87), (c) FCD database (Acc=1.00, IoU=0.90).

3.2 Fire Calorimetry Model training

In general, fire images taken by cameras contain much information, including fire pixels, backgrounds, color, and brightness. However, not all information is useful for fire identification, for example, background information in the image does not influence the size of the fire, and the color of the flames and background can vary with different camera settings.

To find out which factors have the greatest influence on fire identification and to obtain the optimal fire identification model, the above information is separated one by one and then used for the training of VGG16 network, and by comparing the fire identification results of each model in the test set, the optimal fire identification model and the key factor that affect the fire calorimetry performance can be obtained. Four different forms of fire images, including raw images (fire pixel, background, color, brightness), background-free images (fire pixel, color, brightness), background-free grey images (fire pixel, brightness) and background-free binary images (fire pixel) are used to train the AI model.

Fig. 11 shows the AI model performance and R^2 of the training set and validation set. During the training process, each model using a specific form of fire images reaches convergence with optimal R^2 from 0.979 to 0.991, which indicates that the AI model trained from either the raw fire images or the processed fire images can identify the fire development in the validation set very well. The AI model trained by raw fire images has the highest R^2 of 0.991 and the minimum MSE of 885, while those trained by background-free RGB images, background-free grey images, and background-free binary images have the same R^2 of 0.979. The decrease of R^2 shows that partial loss of information in post-processed images can result in the trained model becoming less effective in the validation set compared with the model trained by raw RGB images, but in general, the identification results are still within acceptable limits. Meanwhile, the same R^2 of 0.979 for these three models trained with background-free images also shows that the color and brightness of the flame have little effect on the identification of the fire in the validation set, background-free binary images containing only flame area information can also perform well in the validation set.

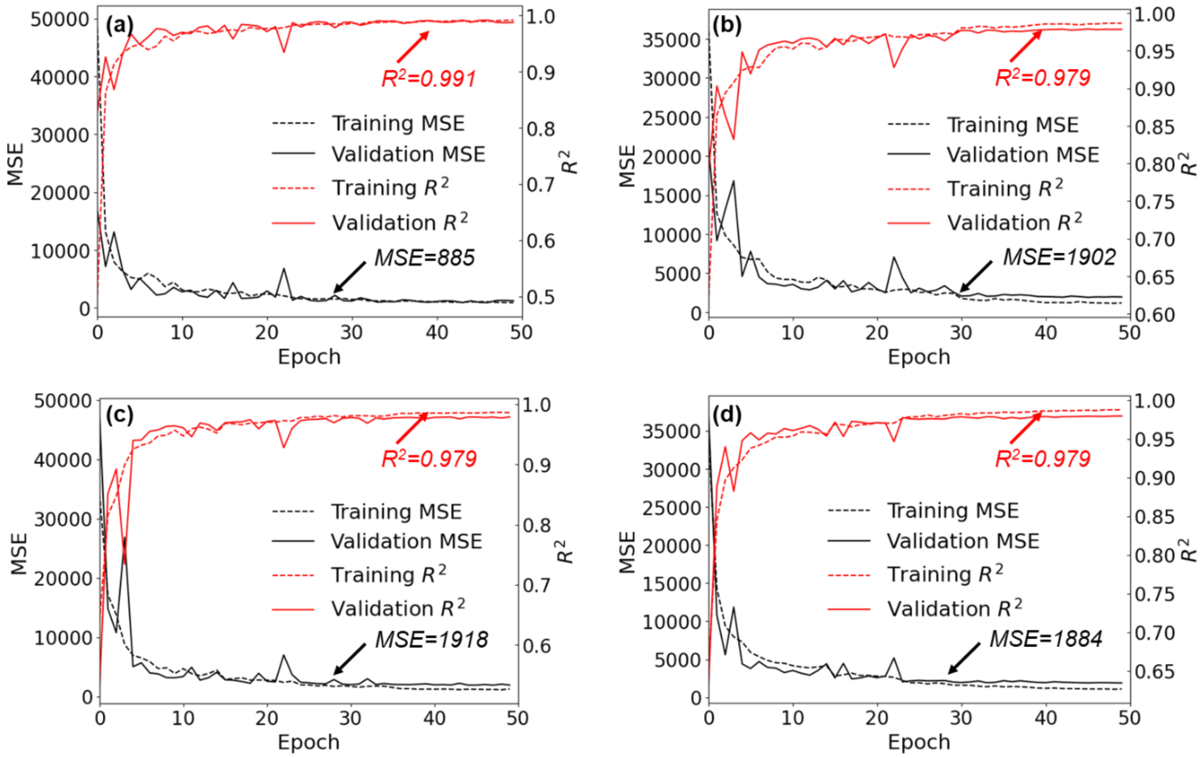

Fig. 11. The loss and R^2 curve during the training process (a) raw RGB images, (b) background-free RGB images, (c) background-free grey images, (d) background-free binary images.

3.3 Model test using the NIST database

After completing the training of these models, the same test set as in [21] was used to validate the identification performance of various models. The plastic bag fire (small HRR), trash can fire (medium HRR), wood pallets fire (large HRR) identified by various models are taken as examples to show the real-time identified fire HRR by AI model. Fig. 12 summarizes the comparison between the true HRR and the predicted HRR using the AI model trained by different image forms, where a savgol filter is used to smooth results. The detailed identification process for the transient fire HRR of a trash can fire using different image forms is demonstrated in Video S1. In general, the AI model performs well in fire calorimetry throughout the fire development process.

Overall, all four models can accurately determine the time-dependent fire HRRs, and among these, the model trained by raw RGB images shows the best performance, whether in small, medium, or large fires. For post-processed images such as background-free RGB images, background-free grey images and background-free binary images, due to the inevitable loss of some flame information during image processing, for example, inaccurate flame segmentation, the model trained from post-processed images is less effective in recognizing smaller fire scenarios, as shown in Fig. 12a.

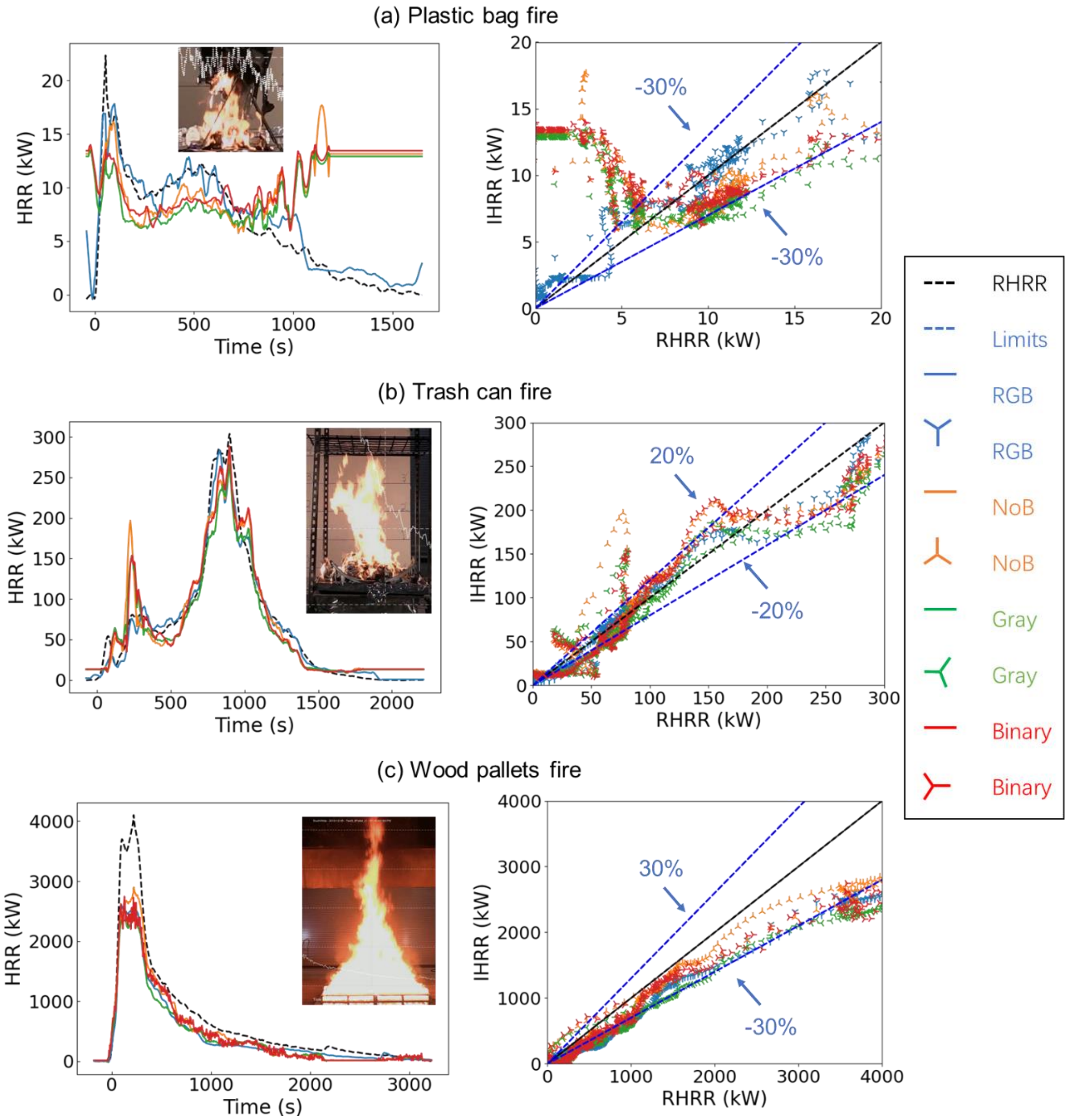

Fig. 12. Evolution of fire HRR by model prediction and the oxygen calorimeter by NIST, (a) plastic bag fire, (b) trash can fire (see Video S1), and (c) wood pallets fire.

As the fire HRR increases, the flame region will gradually dominate the image. At this point, the loss of flame region information caused by flame image segmentation only accounts for a small proportion, and the post-processed image can still provide effective identification of the fire HRR, as shown in Fig. 12b-c. In particular, in larger fire scenarios, as shown in Fig. 12c, the flame image after post-processing, i.e., image segmentation, has better identification results than the raw RGB image at the peak of the fire HRR, which indicates that flame image segmentation can improve the effectiveness of fire calorimetry, but is influenced by the effect of image segmentation.

When the fire HRR is small, the flame region occupies a small portion of the image, and at this time, any influence on the information of the flame region will cause false identification of the fire HRR. While the flame area in the image becomes larger, i.e., the fire heat release rate becomes higher, the loss of a small amount of information in the flame area will have minimal effect on the fire calorimetry but will facilitate the identification of large HRR fire scenarios.

Besides, the comparison of the identification results of different models shows that the size of the flame area dominates the process of fire calorimetry. From raw RGB images to background-free binary images, the information contained in the image is gradually reduced, where the binary image with the background removed contains only the regional information of the flame. Nevertheless, the binary image containing only the flame area information can still identify the transient development of fire HRR well, and adding the flame color and brightness information on this basis, i.e., RGB image with background removed and greyscale image with background removed does not effectively improve the fire calorimetry performance, which indicates that the background information, flame color and brightness information do not contribute much to the fire calorimetry, and the trained model mainly achieves the identification by the flame area.

3.4 Model test under lab environment

Note that NIST's database has similar experimental conditions, i.e., camera settings, shooting distance, and background. Therefore, using only the fire test in NIST database to test the fire identification model is not sufficient to illustrate the generalizability of the model. To further demonstrate the feasibility of the AI model for fire calorimetry, a wood crib fire ignited by 100 mL of propanol [21] was carried out to validate the deep learning model. The burning process and the mass loss rate were recorded by the camera and balance, respectively. Since the distance of our experimental shots in the lab is different from the NIST fire test, the images need to be scaled to match the training set before testing, which can avoid scale effects on fire identification.

Different formats of fire images were used to figure out which parameters of the image dominate the identification of HRR. Fig. 13 summarizes the comparison between the measured HRR in experiment and the identified HRR using the AI model trained by different image forms. The detailed process of fire calorimetry using different image forms is demonstrated in Video S2. In general, the AI model performed well in lab environment with an error <30%. In particularly, the AI model, trained by the raw RGB image model, shows the best performance in the small HRR stage of fire. However, due to the difficulty of flame segmentation in small HRR scenarios to fully identify the flame region, the model trained by the post-processed image is not as effective as the model trained by the original RGB image in low HRR scenarios, as shown in Fig. 13b-d.

For the fully developed fire scenario, the model trained by the binary images with the background removed outperforms the other three models, which indicates that the background color and brightness are important factors that interfere with the identification of the model. After removing the effect of these disturbing factors, the model's identification error of HRR is basically within 20%.

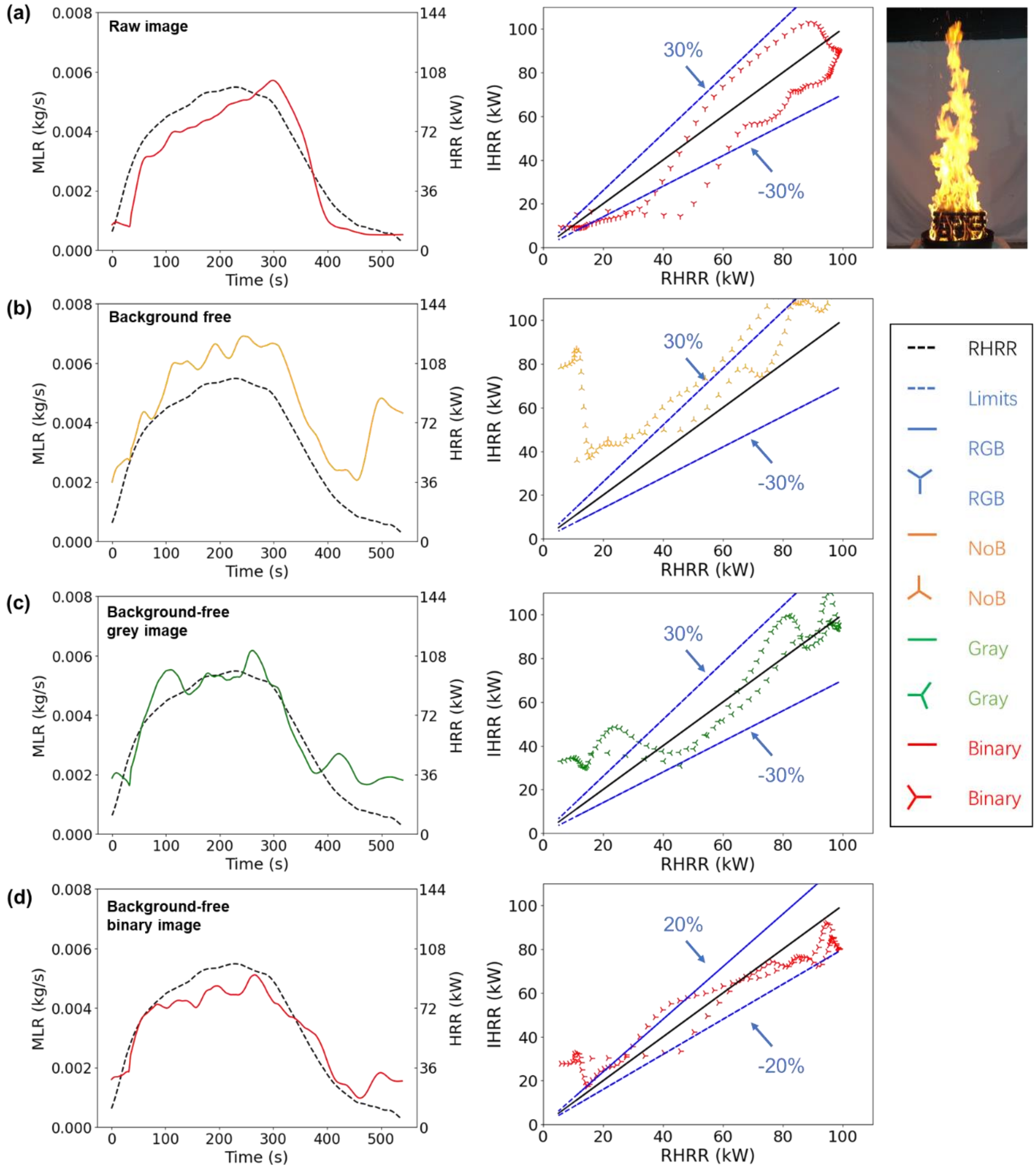

Fig. 13. Evolution of small wood crib fire by model prediction and the mass-balance measurement, (a) raw RGB image, (b) Background-free RGB image, (c) Background-free grey image and (d) binary image (see Video S2).

3.5 Visual explanations based on Grad-CAM

Although the conclusion that the key factor affecting the fire calorimetry is the flame region can be obtained from the comparison of model performance, we still do not know how the trained model identifies the size of the fire and which regions of the image contribute to the fire calorimetry. Thus, Grad-CAM is adopted to analyze the learning of AI model on flame images and to present the weighting of the contribution of each region of the flame image to the fire calorimetry, which allows us to visualize the AI model's identification of fire images.

Wang, Z., Zhang, T. & Huang, X. Explainable deep learning for image-driven fire calorimetry. Appl Intell 54, 1047–1062 (2024).
https://doi.org/10.1007/s10489-023-05231-x

Taking the image in trash can fire as an example, generated Grad-CAM images for different models are shown in Fig. 14. Since fire images are resized to the size of 224*224 before inputting the AI model for fire calorimetry, Grad-CAM images are generated based on the resized fire images.

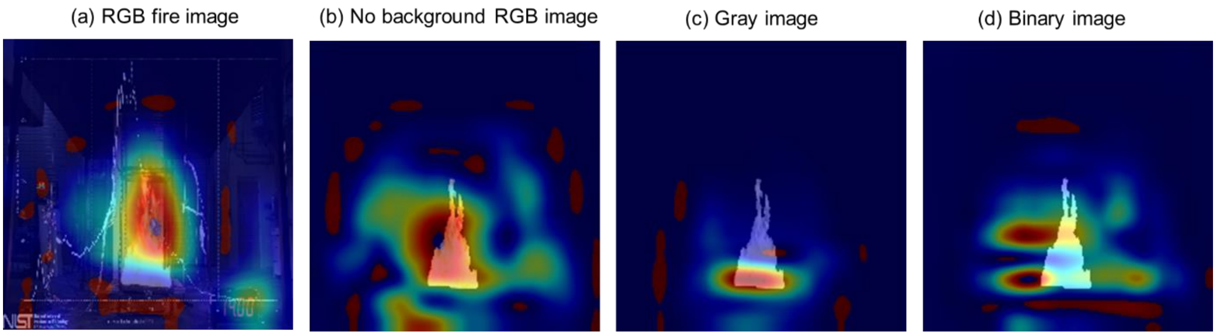

Fig. 14. Generated Grad-CAM image for (a) RGB fire image, (b) No background RGB image, (c) Grey image and (c) Binary image.

For the model trained by raw RGB images, the flame region in the image contributes a larger weight in the identification of the HRR, which indicates that the raw RGB image trained model can effectively identify the flame region. However, since the electronic clock used for notation is red in color and exhibits similar characteristics to the flame, it also has a high weight in the identification of the model. This result shows that although VGG16 can effectively reduce the influence of the background on the identification results, if there are objects with similar features to the flame in the image, it will inevitably have an impact on the model performance.

For the model trained by post-processing fire images, the trained model, the main area of interest for each training model is the flame region, although there are still areas in the background region that have a high weight in fire identification, but since the background has been removed, it does not have an impact on the identification results. In this case, the performance of the model depends on the accuracy of the flame segmentation and the optimization of the model weights. With the gradual reduction of image information, the fire calorimetry can be achieved even by images with only flame area information remaining.

3.6 Discussion

In summary, the application of deep-learning models trained on diverse forms of fire images has proven effective in accurately identifying fire size. Notably, the model trained on raw RGB images exhibits the highest performance in scenarios involving small fires, demonstrating an R^2 value of 0.8196, as illustrated in Table 2. Conversely, the model trained on post-processed images yields enhanced performance in medium to high HRR scenarios, with a minimum R^2 value of 0.7990, owing to the removal of disruptive background information. Through a comparative analysis of the identification outcomes obtained by different models, it becomes evident that the flame area primarily governs the fire HRR identification process conducted by the fire image and AI model. In contrast, the contributions of image background, flame color, and brightness information toward fire calorimetry are found to be limited.

Table 2. Performance of deep-learning models trained by different forms of fire images

AI model trained by	Model's predictive accuracy (R^2)					
	Training	Validation	Plastic bag	Trash can	Wood pallet	Wood crib
raw RGB	0.9917	0.9876	0.8196	0.9634	0.8453	0.8040
background-free RGB	0.9867	0.9786	-1.5233	0.8709	0.8937	-0.010
background-free grey	0.9867	0.9772	-1.3236	0.8633	0.7990	0.7024
background-free binary	0.9875	0.9794	-1.4373	0.8638	0.8455	0.7998

According to the lab test using wood crib fire, it has been observed that even after the removal of the background, the binary image exhibits a commendable level of accuracy, with only a marginal decrease in the R^2 value from 0.8040 to 0.7998 when compared to the AI model trained using raw RGB images. This finding indicates that the application of image segmentation techniques can effectively mitigate the influence of camera settings on HRR identification without compromising the overall accuracy of the recognition process. Nevertheless, it is important to acknowledge that the exclusion of background information, achieved through the process of fire segmentation, can result in the partial loss of fire-related information. Consequently, there is a possibility that the predicted results may deviate from the true HRR due to the inherent limitations of the segmentation approach. Hence, enhancing the accuracy of flame segmentation serves as a pivotal step in mitigating the identification errors in vision-based fire calorimetry. To cater to the intricate flame segmentation scenarios, there is a demand for additional fire images in the fire segmentation database and more effective deep learning model to realize more accurate fire segmentation.

Another challenge of this work is the impact of the distance. Note that the present work is demonstrated by a fire image database shot at a fixed distance, i.e., the fire images in NIST database and our lab tests. However, the situation in real fire scenarios could be more complicated. The location of the camera and the fire source may change as the fire develops, making it impossible for us to scale the image based on distance. Thus, how to obtain the distance of the fire image from the shooting position automatically will also be major challenges in the practical application of this technology that deserves more research.

4. Conclusions

In this work, the fire calorimetry using fire images and deep learning method is presented and further studied. A series of fire tests from NIST fire database are selected to form a big fire image database and different formats of fire images corresponding to the HRR are used for AI model's training. Results show that the AI model, trained by different forms of fire images, can achieve accurate fire calorimetry and the size of the fire area dominates the fire identification process.

Comparison shows that the main factor affecting the fire calorimetry is the size of the fire area in the image, while other factors like as background, color, and brightness have little effect on the results. Therefore, improving the accuracy in flame segmentation plays the most important role in reducing the error in vision-based fire calorimetry. As long as the fire segmentation model can capture the area of flame precisely, their identification error of HRR is no more than 20%. Comparison shows that the Swin-Unet model outperforms FCN8 and U-Net models in fire segmentation.

Overall, this work deepens the understanding of this new vision-based fire calorimetry. This AI method can further be programmed into an APP and coupled with cameras of smart phones, UAVs, and firefighter hamlet, which can guide the future AI applications in fire monitoring, digital twin, and smart firefighting. In the future, we will continue to study the automatic measurement of fire images and explore the effect of scaling on HRR identification.

Statements and Declarations

Acknowledgments

This work is funded by the Hong Kong Research Grants Council Theme-based Research Scheme (T22-505/19-N). ZW thanks the support from the SFPE Foundation Student Research Grant. TZ thanks the support from the Hong Kong PhD Fellowship Scheme.

Authors contribution statement

Zilong Wang: Investigation, Methodology, Writing - original draft, Formal analysis. **Tianhang Zhang**: Investigation, Resources, Writing - review & editing. **Xinyan Huang**: Conceptualization, Supervision, Writing - review & editing, Funding acquisition.

Ethical and informed consent for data used

The authors declare no conflict of interest.

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.

Data Availability Statements

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Hall S, Evarts B (2022) Fire Loss in the United States During 2021. National Fire Protection Association (NFPA)
2. Drysdale D (2011) An Introduction to Fire Dynamics, 3rd ed. John Wiley & Sons, Ltd, Chichester, UK

3. Johansson N, Svensson S (2019) Review of the Use of Fire Dynamics Theory in Fire Service Activities. Fire Technology 55:81–103. https://doi.org/10.1007/s10694-018-0774-3
4. Sudheer S, Kumar L, Manjunath BS, et al (2013) Fire safety distances for open pool fires. Infrared Physics & Technology 61:265–273. https://doi.org/10.1016/J.INFRARED.2013.09.006
5. Qiu X, Wei Y, Li N, et al (2019) Development of an early warning fire detection system based on a laser spectroscopic carbon monoxide sensor using a 32-bit system-on-chip. Infrared Physics & Technology 96:44–51. https://doi.org/10.1016/J.INFRARED.2018.11.013
6. Dang J, Yu H, Song F, et al (2018) An early fire gas sensor based on 2.33 μm DFB laser. Infrared Physics & Technology 92:84–89. https://doi.org/10.1016/J.INFRARED.2018.05.007
7. Quintiere JG (2016) Principles of Fire Behavior, 3rd ed. CRC Press
8. Smith EE (1996) Heat release rate calorimetry. Fire Technology 1996 32:4 32:333–347. https://doi.org/10.1007/BF01037743
9. Brohez S, Delvosalle C (2009) Carbon dioxide generation calorimetry—Errors induced by the simplifying assumptions in the standard test methods. Fire and Materials 33:89–97. https://doi.org/10.1002/FAM.988
10. Hosokawa Y, Nakashima E, Ueno T (2021) Image analysis of flame behavior for polyolefins and polystyrene in vertical flame test. Journal of Applied Polymer Science 138:49845. https://doi.org/10.1002/APP.49845
11. Toulouse T, Rossi L, Akhloufi MA, et al (2018) A multimodal 3D framework for fire characteristics estimation. Measurement Science and Technology 29:025404. https://doi.org/10.1088/1361-6501/AA9CF3
12. Morandini F, Toulouse T, Silvani X, et al (2019) Image-Based Diagnostic System for the Measurement of Flame Properties and Radiation. Fire Technology 55:2443–2463. https://doi.org/10.1007/S10694-019-00873-1/TABLES/3
13. Xiong C, Fan H, Huang X, Fernandez-Pello C (2022) Evaluation of burning rate in microgravity based on the fuel regression, flame area, and spread rate. Combustion and Flame 237:111846. https://doi.org/10.1016/j.combustflame.2021.111846
14. Bonner M, Fire T, Rein G (2022) Visual Fire Power : An Algorithm for Measuring Heat Release Rate of Visible Flames in Camera Footage , with Applications in Facade Fire. Fire Technology. https://doi.org/10.1007/s10694-022-01341-z
15. Li N, Lee EWM, Cheung SCP, Tu J (2020) Multi-fidelity surrogate algorithm for fire origin determination in compartment fires. Engineering with Computers 36:897–914. https://doi.org/10.1007/s00366-019-00738-9
16. Naser MZ, Seitllari A (2020) Concrete under fire: an assessment through intelligent pattern recognition. Engineering with computers 36:1915–1928
17. Ahmad T, Mishra A, Ghosh S, Casari CS (2022) Identifying Efficient Cooling Approach of Cylindrical Lithium-Ion Batteries. Energy Technology 10:1–10. https://doi.org/10.1002/ente.202100888
18. Huang L, Liu G, Wang Y, et al (2022) Fire detection in video surveillances using convolutional neural networks and wavelet transform. Engineering Applications of Artificial Intelligence 110:104737. https://doi.org/10.1016/j.engappai.2022.104737
19. Li Y, Zhang W, Liu Y, Jin Y (2022) A visualized fire detection method based on convolutional neural network beyond anchor. Applied Intelligence 52:13280–13295. https://doi.org/10.1007/S10489-022-03243-7/TABLES/3

20. Wang Z, Zhang T, Wu X, Huang X (2022) Predicting transient building fire based on external smoke images and deep learning. Journal of Building Engineering 47:103823. https://doi.org/10.1016/j.jobe.2021.103823
21. Wang Z, Zhang T, Huang X (2023) Predicting real-time fire heat release rate by flame images and deep learning. Proceedings of the Combustion Institute 39:4115–4123. https://doi.org/10.1016/j.proci.2022.07.062
22. Guo Y, Liu Y, Georgiou T, Lew MS (2018) A review of semantic segmentation using deep neural networks. International Journal of Multimedia Information Retrieval 7:87–93. https://doi.org/10.1007/s13735-017-0141-z
23. Wang Y, Ren J (2019) Application of KNN Algorithm Based on Particle Swarm Optimization in Fire Image Segmentation. Journal of Electrical Engineering and Technology 14:1707–1715. https://doi.org/10.1007/s42835-019-00194-x
24. Yang X, Chen R, Zhang F, et al (2021) Pixel-level automatic annotation for forest fire image. Engineering Applications of Artificial Intelligence 104:104353. https://doi.org/10.1016/J.ENGAPPAI.2021.104353
25. Madani K, Kachurka V, Sabourin C, et al (2018) A human-like visual-attention-based artificial vision system for wildland firefighting assistance. Applied Intelligence 48:2157–2179. https://doi.org/10.1007/S10489-017-1053-6/FIGURES/17
26. Choi HS, Jeon M, Song K, Kang M (2021) Semantic Fire Segmentation Model Based on Convolutional Neural Network for Outdoor Image. Fire Technology 57:3005–3019. https://doi.org/10.1007/s10694-020-01080-z
27. Frizzi S, Bouchouicha M, Ginoux JM, et al (2021) Convolutional neural network for smoke and fire semantic segmentation. IET Image Processing 15:634–647. https://doi.org/10.1049/IPR2.12046
28. Gomes R, Pham T, He N, et al (2023) Analysis of Swin-UNet vision transformer for Inferior Vena Cava filter segmentation from CT scans. Artificial Intelligence in the Life Sciences 4:100084. https://doi.org/10.1016/J.AILSCI.2023.100084
29. Guo Q, Wang C, Xiao D, Huang Q (2023) A novel multi-label pest image classifier using the modified Swin Transformer and soft binary cross entropy loss. Engineering Applications of Artificial Intelligence 126:107060. https://doi.org/10.1016/J.ENGAPPAI.2023.107060
30. Selvaraju RR, Cogswell M, Das A, et al (2016) Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. International Journal of Computer Vision 128:336–359. https://doi.org/10.1007/s11263-019-01228-7
31. Arif M, Wang G, Chen S (2018) Deep Learning with Non-Parametric Regression Model for Traffic Flow Prediction. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00120
32. Toulouse T, Rossi L, Campana A, et al (2017) Computer vision for wildfire research: An evolving image dataset for processing and analysis. Fire Safety Journal 92:188–194. https://doi.org/10.1016/j.firesaf.2017.06.012
33. Cazzolato M, Avalhais LPS, Chino DYT, et al (2017) FiSmo: A Compilation of Datasets from Emergency Situations for Fire and Smoke Analysis. Brazilian Symposium on Databases
34. Bryant R, Bundy M (2019) The NIST 20 MW Calorimetry Measurement System for Large-Fire Research. NIST Technical Note 2077. https://doi.org/10.6028/NIST.TN.2077
35. Liu H, Zhang C, Deng Y, et al (2023) TransIFC: Invariant Cues-aware Feature Concentration

Learning for Efficient Fine-grained Bird Image Classification. IEEE Transactions on Multimedia. https://doi.org/10.1109/TMM.2023.3238548
36. Liu T, Liu H, Yang B, Zhang Z (2023) LDCNet: Limb Direction Cues-aware Network for Flexible Human Pose Estimation in Industrial Behavioral Biometrics Systems. IEEE Transactions on Industrial Informatics. https://doi.org/10.1109/TII.2023.3266366
37. Lin TY, Goyal P, Girshick R, et al (2020) Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 42:318–327. https://doi.org/10.1109/TPAMI.2018.2858826

Zilong Wang is currently a Postdoc Fellow at The Hong Kong Polytechnic University (PolyU). He received his PhD (2023) from PolyU, MSc (2022) from University of Science and Technology of China, and BEng (2017) from China University of Mining and Technology. His current research fields include Smart Firefighting and AI-driven Fire Modelling.

Tianhang Zhang received his PhD (2023) at the Hong Kong Polytechnic University and his MSc (2020) and BEng (2017) from Zhejiang University. His research interests include building fire digital twin, smart firefighting and tunnel fire.

Dr Xinyan Huang is an Associate Professor at The Hong Kong Polytechnic University. He received his PhD from Imperial College London. He has co-authored over 170 journal papers in Combustion, Fire Science and Smart Firefighting. He is an Associate Editor of *Fire Technology* and *International Journal of Wildland Fire*, an editorial member of *Fire Safety Journal* and *Fire and Materials*. He is also the winner of Bernard Lewis Fellowship from Combustion Institute, Proulx Early Career Award from International Associate for Fire Safety Science, Ricardo Award from Institute of Physics, "5 under 35" Award from Society of Fire Protection Engineers.

Affiliations

Zilong Wang, Tianhang Zhang, Xinyan Huang*

Zilong Wang

zilong.wang@connect.polyu.hk

Tianhang Zhang

tianhang.zhang@connect.polyu.hk

Xinyan Huang

xy.huang@polyu.edu.hk

Research Centre for Fire Safety Engineering, Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong