HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 3 (2022)

Finding the Signal from the Smoke: A Real-Time, Unattended Fire Prevention System Using 3D CNNs

Michael Ngai[1], Eugene Fu[1], Andy Tam[1], Amber Yang[#] and Grace Ngai[#]

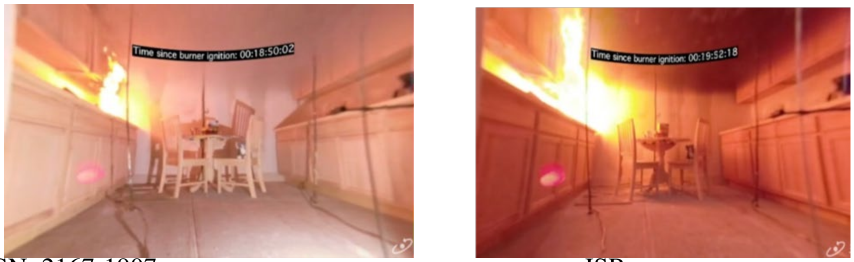
[1] Phillips Exeter Academy, Exeter, NH, USA
[#]Advisor

ABSTRACT

Cooking fires are dangerous. Every year, they are responsible for taking away more than 500 lives in the U.S. alone. Existing approaches using sensors usually require expensive retrofitting and are not feasible in real-life situations. This research presents Finding-Signals-from-Smoke (FiSS), a robust fire machine learning prediction model that aims to prevent cooking fires from starting using videos captured with a normal camera. FiSS is based on a 3-dimensional Convolutional Neural Network, which analyzes the video signals and models the complex relationships of the spatial-temporal features of smoke signals with fire ignition. It uses a segment-based video sampling and modeling framework that is able to generalize to a variety of kitchen/stove settings and achieve promising prediction performance. FiSS is trained and evaluated with video data from 30 full-scale kitchen fire experiments and can predict potential fire ignitions as early as 60 seconds before the moment of ignition. As a result, FiSS can be used in an early warning system to prevent fire ignitions and help to reduce casualties and injuries from cooking fires.

Introduction

Cooking fires are the primary cause of home fires in the U.S. A study conducted by the National Fire Protection Association (NFPA) *(Ahrens, 2020)* shows that cooking fires were responsible for an annual average of about 172,000 home fires, 550 deaths, 4,820 injuries, and more than $1.2 billion in direct property damage between 2014 to 2018. The same study found that the leading cause of cooking fires is unattended cooking in which food is left on hot cooking equipment without any supervision. The National Fire Incident Reporting System *(USFA 2022)* further observes that cooking oil, fat grease, and related substances were responsible for approximately 52% of first-ignited food. Fig. 1, which shows two screenshots from a full-scale cooking fire experiment in a mock-up kitchen *(Hamins et al., 2018a),* illustrates how dangerous cooking fires can be. When cooking oil is ignited, nearby combustible materials, such as wood cabinets above the cooktop, are likely to be ignited. Fig. 1b shows that the fire grows significantly in less than a minute, and such fires are notoriously difficult to suppress. A study from NPFA *(Ahrens, 2020)* reveals that more than 50% of cooking fire injuries were due to improper fire suppression. As cooking fires remain the primary cause of home fire injuries and the second leading cause of home fire deaths, new approaches are needed to address this problem.


Time since burner ignition: 00:18:50:02
Time since burner ignition: 00:19:52:18

a) b)

Figure 1. A screenshot of the video record for a) the moment right after cooking oil ignition and b) the fire condition after 1-min ignition.

Efforts have been made to reduce unattended cooking fires, but these approaches have limitations. The standard for the safety of household electric ranges, namely, UL 858 *(Underwriters Laboratory, 2014)*, was recently revised and approved. It requires stovetops manufactured after June 2018 in the United States to pass an oil pass/fail ignition test. The test criterion is that the average pan temperature cannot exceed 385 °C with the stovetop on its highest power setting for at least 30 minutes. The revised standard can effectively prevent cooktop fires from unattended cooking. However, the latest UL 858 is only applicable to new stovetops with electric heating elements, and there is a lack of standards for older and other types of stovetops. Since the UL 858 is the only safety standard for cooktops in the U.S., alternative fire prevention systems are needed in order to prevent unattended cooking fires from using the older and other types of stovetops. Furthermore, this criterion only serves to *prevent* cooktop fires from igniting when left unattended for the specified duration and does not serve to *detect* fires that have already been ignited in real-life settings.

The fire research community has made attempts to develop early detection systems for ignited cooktop fires, but the existing methods are unreliable and difficult to deploy in regular kitchen settings. For example, Mensch et al. *(2019)* presented a retrofitted device for cooktop ignition prevention. The device was constructed using 16 different sensors (i.e., aerosols, hydrocarbon, smoke, carbon monoxide, carbon dioxide, volatile organic compounds, temperature, and humidity sensors). It was approximately 25 cm long, 7 cm wide, and 3 cm tall, and was required to be installed in the exhaust duct above the ceiling. Additional wiring was needed to transmit the data. Utilizing a threshold-based algorithm, the device can determine if there is a fire based on the streaming sensor signals. However, the prediction performance was significantly diminished when the detection system was applied to a different cooking setting (i.e., a turned-off range hood or a gas cooktop instead of an electric coil cooktop). In fact, due to the size, the installation requirements, and the use of 16 different sensors, the device has not been successfully implemented outside of laboratory environments. Therefore, a practical early fire detection system needs to be low-cost, simple to use, reliable, and robust in terms of model performance.

This research presents the Finding-Signals-from-Smoke (FiSS) detection model for unattended cooking fires. FiSS uses 3D Convolutional Neural Networks (CNNs) to analyze video data streamed from a typical camera. The following sections will describe our data preprocessing, model development, and discuss our results. A conclusion is presented at the end. FiSS can provide potentially life-saving early detection of unattended cooking fires that reduces civilian casualties.

Fire Data and Fire Dynamics

FiSS is developed based on data from *(Mensch et al. 2021)*. Fig 2 shows the data collection setup, which is a mock-up kitchen equipped with a 30-inch-wide cooktop with four different heating elements and a range hood located about 35.5 inches away from the cooktop. A pan was placed onto a heating element turned to its maximum power. Video records were made using a Nikon 500D single-lens reflex digital camera positioned on a tripod about 2 ft away from the cooking pan. Video capture commenced when the cooking was started and ended when the cooking oil was ignited. Baking soda was manually poured to suppress the fire when the oil ignited.

The dataset contains multiple scenarios with kitchen setups that mimic different cooking practices, which can be used to test the robustness of the detection model. Both electric coil cooktops and gas cooktops were considered, with two heating settings: 1.1 kW and 1.8 kW for the electric coil cooktops and 3.4 kW and 4.0 kW for the gas cooktops. The experiments mainly use a single pan, with four of them using two pans. A variety of different cooking oils were used based on the testing conditions found in UL 858. Table 1 summarizes the 30 different tests. We refer to this as the "*Cooking-Practice*" dataset.

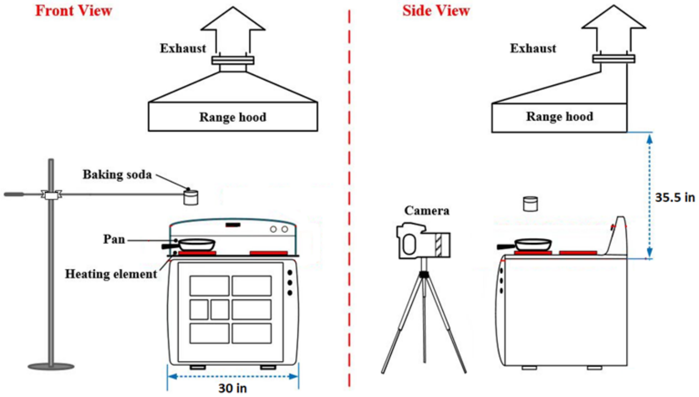


Figure 2. The schematic of the mock-up kitchen with a) front view and b) side view.

Table 1. The 30 different cooking experiments in the dataset.

Set of Exp.	Oil Type	Amount	Heating Type	Heating Size	Pan Type
8	Canola Oil	50 mL	Electric	Small	Cast Iron
1	Canola Oil	100 mL	Electric	Small	Cast Iron
1	Canola Oil	50 mL	Electric	Small	Aluminum
1	Canola Oil	50 mL	Electric	Small	Cephalon
1	Canola Oil	50 mL	Electric	Small	Stainless Steel
1	Canola Oil	200 mL	Electric	Small	Cast Iron
1	Canola Oil	50 mL	Electric	Big	Cast Iron
1	Canola Oil	100 mL	Electric	Big	Cast Iron
1	Corn Oil	50 mL	Electric	Small	Aluminum
1	Corn Oil	50 mL	Electric	Small	Cast Iron
1	Corn Oil	100 mL	Electric	Big	Cast Iron
1	Soy Oil	50 mL	Electric	Small	Cast Iron
1	Soy Oil	100 mL	Electric	Big	Cast Iron
1	Olive Oil	50 mL	Electric	Small	Cast Iron
1	Olive Oil	100 mL	Electric	Big	Cast Iron
2	Sunflower Oil	100 mL	Electric	Big	Cast Iron
1	Butter	45.68 g	Electric	Small	Cast Iron
1	Canola Oil	50 mL and 2L water	Electric	Small	Cast Iron
2	Canola Oil	50 mL and 100 mL	Electric	Small & Big	Cast Iron
1	Olive Oil	50 mL and 100 mL	Electric	Small & Big	Cast Iron
1	Canola Oil	100 mL	Gas	Big	Cast Iron

Figure 3 contains screenshots from an experiment involving a cast iron pan with 50 ml of canola oil and a 1.8 kW electric heating element, shows the evolving fire at four different moments: 90 s before ignition, 60 s before ignition, 30 s before ignition, and the moment of ignition. It can be easily understood that when the heating element is turned on, the cooking oil temperature begins to rise. As described in *(Hamins et al., 2018a and 2018b)*, oil temperatures above 200 °C will generate vaporized gases/smoke at a rate correlated to the temperature. At approximately 410 °C, the cooking oil will be ignited. These evolving cooking fire processes are captured in Fig. 3. Since the flame after ignition provides additional heat to the cooking oil, the fire's size and height will grow. If manual suppression is not done in time or not done correctly, a localized fire is likely to escalate into a room or house fire that can no longer be extinguished by manual suppression, in a process similar to that shown in Fig. 1. In principle, warnings to residents and an interruption to the heating element will prevent a fire ignition. This requires a robust fire detection model that can predict if there is potential for the ignition of a cooking fire. This paper proposes a model that relies solely on streamed video data that can be obtained with easily-accessible equipment

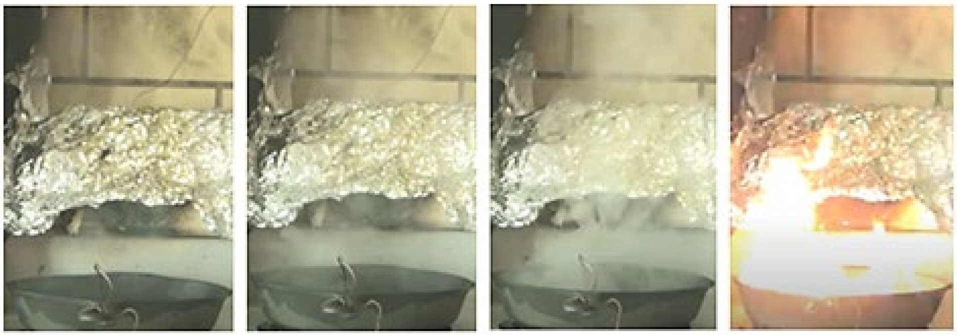


Figure 3. Evolving fire for 50 ml canola oil on a cast iron pan using a 1.8 kW electric heating element at four different moments: (from left to right) 90 s before ignition, 60 s before ignition, 30 s before ignition, and the ignition moment.

Research Questions and Hypotheses

As seen in Fig. 3, smoke behaviors, such as the amount of smoke at a particular moment and the rate of generation of smoke over a period of time, are indicative of the potential for fire ignitions. It is intuitive that a growing amount of smoke from a cooking pan is dangerous. However, it is difficult to determine whether and when there will be an ignition. With visual computing and machine learning techniques, algorithms can effectively capture and model subtle signals to provide reliable predictions. This research aims to develop a FiSS model to encode the smoke signals in videos to predict unattended cooking fires.

Challenges and Research Questions

The major challenge of this task is that smoke has both temporal (i.e., change in time) and spatial (i.e., change in position) features and our model needs to model both aspects captured from the videos and correlate them for ignition detections. The data being considered in this study is by nature very complex as it is influenced by a multitude of environmental factors (e.g., air movement caused by temperature difference). Although an intricate model, which can process and learn as much of the temporal information as possible, is expected to provide better performance, the drawbacks are that training such a model requires large amounts of data, which is expensive for this kind of extreme and dangerous fire events (every experiment requires intentionally setting a kitchen fire). The amount of data available to us is limited,. Thus, we must keep the model less complex to avoid over-fitting, in which the model only memorizes the data without learning any useful features. The balance between training a powerful model and one that is robust (i.e., applicable to other scenarios) is a practical challenge for us.

Given these research challenges, we propose the development of the FiSS model by using a 3D Convolutional Neural Network (3D CNN) with a segment-based sampling technique. Our research will address the following questions:

Q1. Can the 3D CNN predict cooking fires from streaming video data?

Q2. To that end, what is the most effective segment-based sampling framework?

Hypotheses

The following are the hypotheses being made in this study:

H1. A 3D CNN can encode spatial-temporal smoke signals and predict potential cooking fires effectively.

H2. The use of segment-based sampling technique and the proposed modeling framework can contribute to provide an accurate and numerically efficient machine learning-based fire detection model.

Methods

3D Convolutional Neural Networks

Spatial information (e.g., size and shape) of smoke signals is an important indicator of potential ignition. However, in real applications, a camera can be placed anywhere (e.g., varying angle and distance to the cooktop) in a multitude of kitchen settings (e.g., the color of the pan and the decoration of the kitchen), which adds difficulties for extraction and modeling of generalized smoke spatial information. Our first challenge is to yield effective location and background invariant detection. Convolutional Neural Networks (CNNs) have demonstrated their capabilities to learn representative features from complex spatial information in images/videos. We thus propose to apply a CNN-based model for FiSS development.

Our second challenge is to model both spatial and temporal information together, since the smoke's temporal information (e.g., moving speed and generation rate) is also crucial to the prediction. 2D CNNs can effectively handle spatial information in images, but they are less compelling in modelling temporal information in videos. Recent works *(Kamnitsas et al., 2017; Lin et al., 2019; Tran et al., 2015; Yuan et al., 2018)* have successfully applied 3D CNNs to learn spatial-temporal features from videos, achieving promising performance in many applications such as action recognition. In this research, we propose to apply a 3D CNN to model the spatial-temporal features of smoke signals to predict unattended cooking fires.

We design and develop our 3D CNN model based on the C3D model proposed in *(Tran et al., 2015)*, which has demonstrated its effectiveness in various domains *(Lin et al., 2019; Yuan et al., 2018)*. Given the relatively small dataset, we modify the original model to obtain a lightweight version of C3D to avoid over-fitting.

Fig. 4 shows our model architecture. Our model comprises five 3D convolutional layers and each convolution layer is followed by a max pooling layer. We make use of the optimal kernel setup suggested in *(Tran et al., 2015)*: setting the convolution kernels as $3 \times 3 \times 3$ for all the convolutional layers and adopting $2 \times 2 \times 2$ kernels for all the max pooling layers except "Pool1", for which the kernel size is set as $2 \times 2 \times 1$. After the convolutional and pooling layers, our model consists of two fully connected layers to further encode the feature representations. Finally, a softmax layer is applied for model outputs and final predictions.

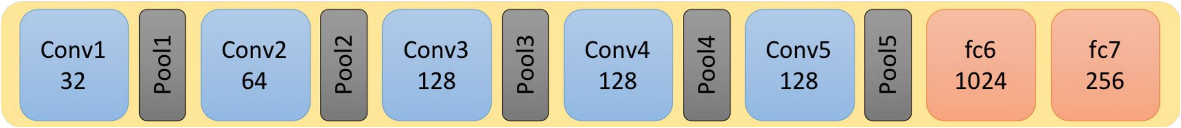


Figure 4. The architecture of the proposed 3D CNN with five 3D convolutional layers. The number of kernels in layer 1 to 5 is 32, 64, 128, 128, and 128, respectively.

Video modeling often adopts sparse frame sampling, which under-samples the video frames. For example, using a sampling rate of three frames per second, a 5-second video is represented as a 15-frame stream. The idea is to preserve sufficient valid information from the original video while reducing the model capacity and complexity by learning the spatial-temporal features from a much smaller number of frames. Our method follows the proposed setup in (Tran et al., 2015), in which a given video instance V is sampled into a *16-frame stream* ($V' = \{f'^1, \dots, f'^{16}\}$). All the inputted frames will be resized and cropped to the size of 112×112. Moreover, our model needs to handle images with RGB channels (3 channels: red, green, and blue). For that, the input dimensions of the model are $3 \times 16 \times 112 \times 112$ (V'). Given such an input, the output of our 3D CNN model is formulated as follows:

$$p = softmax\big(\mathcal{D}(V';\, W)\big) \qquad (1)$$

$$\hat{y} = argmax(p) \qquad (2)$$

where function $\mathcal{D}$ denotes the 3D CNN model with parameters W. The final prediction ($\hat{y}$, such as "*Danger*" or "*Non-Danger*") is made based on the class scores ($p = (p_{Danger},\ p_{Non-Danger})$) output from the model.

Segment-Based Modeling Framework

The second challenge is that smoke generation changes significantly depending on different cooking settings (e.g., the amount of oil and the heating power). For example, cooking a large amount of oil with a low heating power may result in a smaller amount of generated smokes. This setting also leads to slower smoke movement, and it requires more time for the oil to reach ignition conditions. In cases like these, long-range temporal information is needed. For long-range video modeling, to ensure a low model complexity, a smaller sample rate (e.g., 1 frame per second for a 10 s video) is desired to avoid over-fitting the model. However, the reduction of sample rate will result in loss of information in temporal direction. Inspired by previous work *(Wang et al., 2018)*, we propose to incorporate a segment-based sampling and modeling framework for the development of our FiSS (i.e., sampling and modeling on the subsegments of the given video individually and aggregating their learned information together for a final prediction). This framework can help us maintain a low complexity model that minimizes information loss.

Fig. 5 (a) illustrates our framework. Specifically, an input video instance V is segmented into K segments: $\{S_1, \dots, S_K\}$. Each segment (S_i) is further down sampled to 16 frames ($S'_i = \{f'^1_i, \dots,\ f'^{16}_i\}$), and they are fed into the 3D CNN. The final prediction is obtained by aggregating the CNN outputs of all K segments. Mathematically, the FiSS model is formulated as follows:

$$p = FiSS(V) = softmax(\mathcal{H}(g_1, \dots, g_K)) \qquad (3)$$

$$g_i = \mathcal{D}(S'_i;\, W) \qquad (4)$$

where $\mathcal{D}$ is a 3D CNN model with trainable parameters W and $\mathcal{H}$ is the aggregation function. All the K segments are fed into the same 3D CNN with the same parameters W.

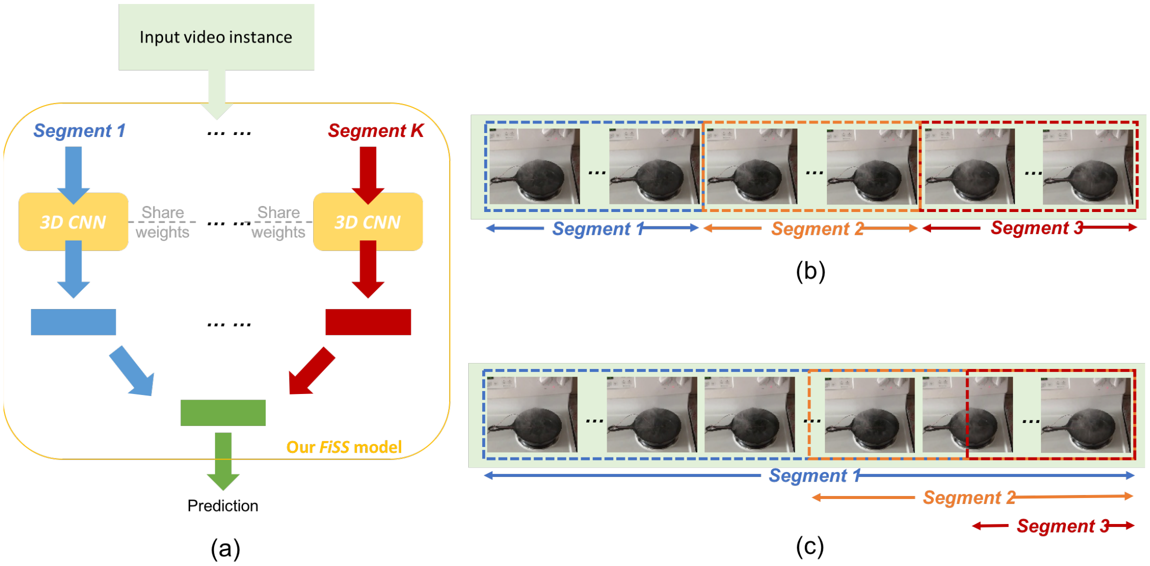


Figure 5. The proposed segment-based sampling and modeling framework (a). The two segmentation methods used in our evaluations (using $K = 3$ as an example): (b) equal-length segmentation and (c) scaling-based segmentation.

Fine-tuning is needed to obtain the optimal settings for the 3D CNN. In this study, we evaluate three different components: 1) the aggregation approach, 2) the number of segmentation *K*, and 3) the segmentation method. For the aggregation approach, we evaluate model performance with two different pooling strategies: max pooling and average pooling. For the segmentation method, we propose two segmentation methods. Inspired by (Wang et al., 2018), our first method divides a video into *K* segments with equal length of duration. For example, for a 9-second video where K is equal to 3, the first segment has video data from 0s to 3s, the second segment has video data from 3s to 6s, and the third segment has video data from 6s to 9s. There are no overlapping temporal segments. We refer to this method as *equal-length* segmentation. In our second method, the video is divided according to a scaling approach, which we denote as scaling-based segmentation. Assuming that the spatial-temporal information closer to the moment of interest (i.e., towards the end of the video when the oil is ignited) contains more salient information, overlapping data near the ignition moment is used. The length of each segment varies to capture temporal behaviors of smoke in different time scales and the length is reduced by a factor of 2. Fig. 6 (b) and (c) illustrate the two segmentation methods.

Dataset and Model Configuration

Dataset

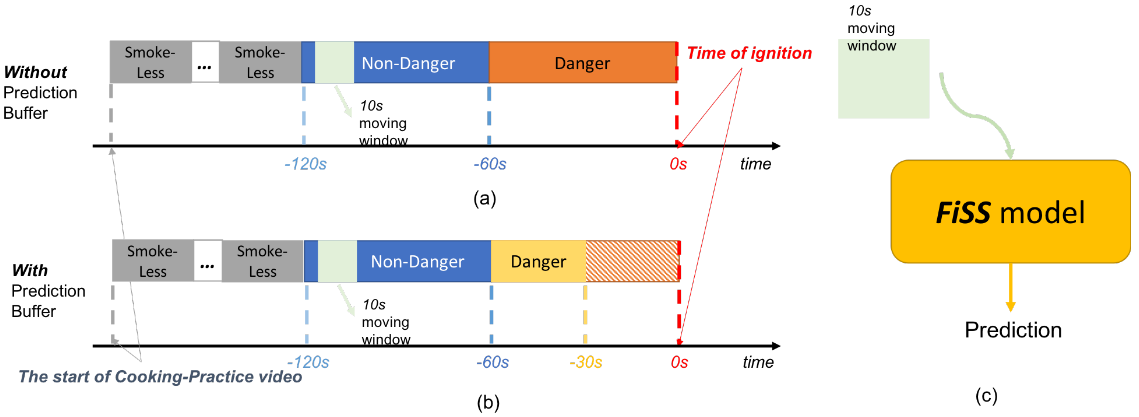


Figure 6. Segmenting each Cooking-Practice video into "*Smokeless*", "*Non-Danger*" and "*Danger*" clips, with (a), and without (b), the prediction buffer. FiSS is trained and evaluated with the video instances (10s moving windows) generated by the presented methods (c).

Given an video instance, our task is to predict whether oil ignition is imminent in the near future (e.g., 60s). In our experiment, we pre-process the *Cooking-Practice* dataset to generate video instances for our model training and evaluation. As shown in Fig. 3, smoke can hardly be seen 120s before ignition. When the oil temperature becomes higher, smoke tends to be generated and as the ignition moment is approaching, we can see in the figures that smoke is generated at a faster rate. This kind of smoke behavior usually occurs in between 120s before ignition to the moment of ignition, and immediate action is needed if the oil is going to ignite within the next 60s. With this domain knowledge, we split each video in the *Cooking-Practice* dataset into three sub video clips at 120s and 60s prior to the ignition moment as shown in Fig. 6 (a and b). We annotate the sub-clips as "Smokeless", "Non-Danger", and "Danger". As mentioned above, since our study targets to model the smoke signals to predict potentially dangerous situations, we discard the "Smokeless" data and our model construction and evaluation will only focus on the "*Non-Danger*" and "*Danger*" clips.

To generate our video instances, a 10-second moving window with a stride of 1 second to the "*Non-Danger*" and "*Danger*" video sub-clips is applied (Fig. 6, a). Thus, each sub-clip contributes 60 instances of "Non-*Danger*" and "*Danger*", respectively. Since the instances are likely to be very similar, 30 instances from each category are then selected to avoid overfitting. Each video in *Cooking-Practice* dataset thus contributes 60 instances with two balanced classes.

Another important aspect of this problem is that we need to develop a model such that it can send warning information and, if possible, automatically turn off the stove when danger is detected in real-life contexts. However, in real applications, it takes time for the pan temperature to return to a safe range after the stove is turned off. During this period, the food and oil are still being heated and there is still a risk of ignition, and once a cooking fire is ignited, it is very difficult to suppress. Taking these practical concerns into consideration, a model that can make a prediction in advance, such that there is an adequate time buffer to allow the pan temperature to reduce to a safe range, is more practical in preventing potential fires. We account for this scenario by introducing a "prediction buffer" (Fig. 6 b). Specifically, we further split the "Danger" clip at 30s before ignition moment (as suggested by domain experts), and discard all instances after 30s before ignition (after which the temperature of the pan is too hot, and ignition will occur even if the stove is cut off). As this reduces the length of the "Danger" clip to 30 seconds, only 30 instances can be extracted, which are all included in our evaluations.

Approximately 1800 instances are obtained for each fire test. For evaluation purposes, we randomly split our cooking practices into three subsets: training, validation, and testing sets with 70 %, 10 %, and 20 %

of the events assigned to training, validation, and testing, respectively. This ensures that instances from the same cooking practice will not be split into two different sets (i.e. none of the instances in the *test* set will have been "seen" by the model during training or validation). This yields 1260, 180, and 360 training, validation, and testing instances, respectively. We refer to this as "*Ignition-Prediction*" dataset. These instances are then split into *K* segments for segment-based modeling as described in "Methods: Segment-based Modeling Framework".

Model

FiSS is trained from scratch on the *Ignition-Prediction* dataset. The "Adam" optimizer (Kingma & Ba, 2014) is adopted in model training with the initial learning rate of 10^{-5}, which is reduced to 10^{-6} and 10^{-7} after 10 and 20 training epochs, respectively. The model is trained for at most 50 epochs with a batch size of 10. Early stopping strategy is employed to prevent over-fitting: if the model has not improved the validation set in 5 epochs, the training process will be terminated regardless of whether the maximum epoch has been reached. A dropout rate of 0.5 (Srivastava et al., 2014), which randomly ignores half of the nodes in each training step, leading to a similar effect of model capacity reduction in the training process, is adopted to prevent overfitting further.

Results and Discussion

Parametric studies are conducted to evaluate the effectiveness of our FiSS models. To systematically explore the effect of segment-based sampling framework, we evaluate the framework with different setups, including different segment number (*K*), segment approaches (equal-length segment (*ES*), scaling-based segment (*SS*)), and prediction aggregation methods (mean-pooling (*Mean*), max-pooling (*Max*)), and compare with a plain 3D CNN model ($K = 1$).

Table 2 shows the model performance without the prediction buffer. Our problem is formulated as a binary classification problem (i.e., classifying an instance as *"Danger"* or *"Non-Danger"*). We first evaluate the classification accuracy. As the "*Danger*" cases are evidently more critical, we further investigate the precision, recall, and f1-score of the "*Danger*" classification.

Our results show that promising performance can be achieved even with just the plain 3D CNN (~79% accuracy), implying that the proposed 3D CNN model is effective in learning representative spatial-temporal smoke features to predict potential cooking fires. It is also encouraging to note that segment-based sampling framework contributes to performance improvement, especially when $K = 2$ is adopted. This indicates that the segment-based sampling framework is able to capture and model rich temporal information in a more efficient way. We also see that the performance does not further improve with more segments, i.e. $K = 3$. One possible reason is that the length of each segment in a 3-segment setting (or the last segment in scaling-based segmentation) is too short to contain sufficient information for modeling, which may be confusing for the model.

Table 2. Model performance without prediction buffer.

Segment method		Accuracy	Precision	Recall	F1-Score
K = 1		78.89%	78.89%	78.89%	78.89%
	ES-Mean	80.83%	81.71%	79.44%	80.56%

K = 2	*ES-Max*	**82.22%**	**84.12%**	79.44%	81.71%
	SS-Mean	79.72%	74.88%	**89.44%**	81.52%
	SS-Max	80.83%	93.02%	66.67%	77.67%
K = 3	*ES-Mean*	79.72%	80.57%	78.33%	79.44%
	ES-Max	72.78%	71.13%	76.67%	73.80%
	SS-Mean	78.06%	76.17%	81.67%	78.82%
	SS-Max	71.11%	68.45%	78.33%	73.06%

The best accuracy (~82%) is achieved by 2-segment setup with equal-length segmentation and max-pooling aggregation ($K = 2$, *ES-Max*), which also yields the highest precision, i.e., having the smallest false alarm rate. The best recall, on the other hand, is obtained by scaling-based segmentation with mean-pooling aggregation ($K = 2$, *SS-Mean*). The model trained with this framework is more likely to predict fire, but also has a higher chance to make false alarms.

Table 3 shows the performance of our model in the scenario where a prediction buffer is added. This is a more difficult task, and as such, a consistent performance drop is observed across the board for all models. However, it is noteworthy that promising performance is still achieved by the models – specifically, using mean-pooling (*SS-Mean*) and $K = 2$ achieves ~77% accuracy and ~75% precision, albeit with a drop in the recall. This further demonstrates the effectiveness of the proposed 3D CNN and segment-based framework, especially when $K = 2$. Generally, the model adopting $K = 2$ with *ES-Max* sampling and modeling method attains the most robust performance across the board.

Table 3. Model performance with prediction buffer.

Segment method		Accuracy	Precision	Recall	F1-Score
K = 1		73.33%	68.92%	85.00%	76.12%
K = 2	*ES-Mean*	75.00%	68.60%	**92.22%**	78.67%
	ES-Max	76.11%	71.96%	85.56%	78.17%
	SS-Mean	**77.22%**	**75.52%**	80.56%	77.96%
	SS-Max	70.83%	67.12%	81.67%	73.68%
K = 3	*ES-Mean*	73.06%	72.43%	74.44%	73.42%
	ES-Max	73.06%	68.95%	83.89%	75.69%
	SS-Mean	71.11%	67.59%	81.11%	73.74%
	SS-Max	73.33%	70.00%	81.67%	75.38%

Conclusions

This research demonstrates the efficacy of using a 3D CNN model with a segment-based modeling framework to predict unattended kitchen fires. Empirical study is conducted to understand the effectiveness of the proposed model, and explore the optimal setup. The experimental results show that 3D CNNs can effectively learn representative spatial-temporal features which to model smoke signals, yielding promising performance for unattended kitchen fire prediction. Further performance improvement can be achieved by applying segment-based modeling. In the future, this research can be expanded by applying different model structures such as combining a CNN with a Long Short-Term Memory network. Future work with an expanded dataset could also investigate predicting potential other kitchen accidents related to cooking. For example, it is not too difficult to imagine a similar model used for detecting water boil-over, which can also be hazardous in gas cooktops.

Acknowledgments

I would like to thank Eugene Fu and Andy Tam for being my mentors and guiding me through building my model and analyzing the results. I would also like to thank Amber Yang for assisting me in revising my paper. Finally, I would like to thank Grace Ngai for inspiring my love for computer science.

References

Ahrens, M., 2020. *Home Cooking Fires*. National Fire Protection Association, Quincy, Massachusetts.

Hamins, A. , Madrzykowski, D. , Kim, S. and Kent, J. (2018), Investigation of Residential Cooking Fire Suppression Technologies, Technical Note (NIST TN), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.TN.1969 (Accessed February 4, 2022)

Underwriters Laboratories, UL 858, Standard for Safety for Household Electric Ranges, 2014 Edition and 2017 Revision, Underwriters Laboratories, Northbrook, IL.

Kim, S. , Hamins, A. and Bundy, M. (2018), Investigation of Residential Cooktop Ignition Prevention Technologies, Technical Note (NIST TN), National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.TN.1986 (Accessed February 4, 2022)

USFA (U.S. Fire Administration), 2022. https://www.usfa.fema.gov/nfirs/

Mensch, A., Hamins, A., Lu, Z.Q., Kuperschmid, M., Tam, W.C. and You, C., 2019, September. Evaluating sensor algorithms to prevent kitchen cooktop ignition and ignore normal cooking. In *Suppression, Detection and Signaling Research and Applications Conference*.

Mensch, A.E., Hamins, A., Tam, W.C., Lu, Z.Q., Markell, K., You, C. and Kupferschmid, M., 2021. Sensors and machine learning models to prevent cooktop ignition and ignore normal cooking. *Fire Technology*, *57*(6), pp.2981-3004.

Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P. et al., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, 36, pp.61-78.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint* arXiv:1412.6980.

Lin, G., Zhang, Y., Xu, G., Zhang, Q., 2019. Smoke Detection on Video Sequences Using 3D Convolutional Neural Networks. *Fire Technology*, 55(5), pp.1827-1847.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15, 1 (January 2014), 1929–1958.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M., 2015. Learning spatiotemporal features with 3d convolutional networks. *Proceedings of the IEEE international conference on computer vision*, pp.4489-4497.

Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., Van Gool, L., 2018. Temporal segment networks for action recognition in videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2740-2755, 1 Nov. 2019, doi: 10.1109/TPAMI.2018.2868668.

Yuan, Y., Zhao, Y., Wang, Q., 2018. Action recognition using spatial-optical data organization and sequential learning framework. Neurocomputing, 315, pp.221-233.