AI-based Human Detection and Localization in Heavy Smoke using Radar and IR Camera

Hovannes Kulhandjian†, Alexander Davis†, Lancelot Leong†, Michael Bendot†, and Michel Kulhandjian‡
†Department of Electrical and Computer Engineering, California State University, Fresno, Fresno, CA 93740, U.S.A.
E-mail: {hkulhandjian, alexanderdvs, leoxyn, charizardly}@mail.fresnostate.edu
‡Department of Electrical and Computer Engineering, Rice University, Houston, TX, U.S.A.
E-mail: michel.kulhandjian@rice.edu

***Abstract*—One of the main challenges currently firefighters are facing in search and rescue operations is battling the heavy smoke inside a space that needs to be searched for people and animals. In this work, we develop an integrated system composed of two unique sensing mechanisms that are capable of real-time detection and localization of humans and animals in deep smoke to improve the situational awareness of firefighters on the scene. We make use of data from a micro-Doppler sensor and an infrared camera and train a DCNN algorithm to localize a human in dense smoke in real-time. Experimental results reveal that the proposed system can detect a human in heavy smoke with an average of 98% validation accuracy.**

***Index Terms*—Deep learning, human detection in heavy smoke, data fusion, artificial intelligence.**

I. Introduction

Firefighters are tasked with high-stress difficult situations in search and rescue operations where vision can be obstructed by fog and debris. By using intelligent sensing and processing technology, the efficiency and success rate of firefighters in saving victims can be increased.

The goal of this research work is to effectively and efficiently obtain and analyze information gathered from an infrared (IR) camera and a radar sensor, perform spectrogram analysis of the radar data, and relay relevant classification and identification of objects of interest in real time. This system can aid firefighters in real-time by obtaining crucial information from these two sensors for processing. Intelligent inferences from reliable technology during high-stress situations that can involve life or death will allow firefighters to safely navigate through hazardous environments. The proposed system can be integrated into a handheld device, a drone, or a robot that can be sent for missions in search and rescue operations inside a dense smoke environment.

To perform human classification, deep convolutional neural network (DCNN) and data fusion algorithms will be utilized, programmed in Python scripting language on a Raspberry Pi 4. By using DCNN, data fusion algorithms, and a combined micro-Doppler sensor and IR camera system, the detection of objects of interest, e.g., a human or an animal, in low visibility dense smoke environments can be greatly improved.

While there have been several research works, such as [1] and [2], that have studied the use of infrared cameras to assist firefighters in decision-making in high-stress situations, very little work have been done to combine multiple sensors for detecting a human in a heavy smoke environment. With the ability to gather different kinds of information from the micro-Doppler sensor and IR camera in various situations, there is a chance that artificial intelligence (AI) based tools can become a norm for firefighters, and save more lives in the near future.

Related Work: Li *et al.* [3] address various radar technologies such as Doppler, frequency-modulated continuous wave (FMCW), interferometry, and ultra-wideband (UWB) radar about human activity recognition (HAR) and highlight a pertinent issue - actions recorded with radar in "controllable environments" with "little interference". This issue is partially addressed in [4] with a proposed method to reduce noise and remove "non-target micro-motion interference".

Sakai and Aoki [5] discuss the utility of millimeter-wave radar combined with a gyro sensor to address the issue of obstruction of firefighters' views in dense smoke. After that, the data is used to form a signal-reflection dataset that will be utilized to construct a 3D map via 3D image processing techniques.

Authors in [6] and [7] propose convolutional neural networks (CNNs) for human and object detection in thermal images with the explicit goal of aiding firefighting operations. The proposed CNN models are able to achieve a validation accuracy of 96.3%.

Towards wildfire detection and monitoring, [8] provides an overview of current (2016) IR and visible light sensors, continuous monitoring systems such as unmanned aerial vehicles (UAVs), evaluation scenarios (laboratory, controlled burns, field trials), and metrics for human or automated monitoring.

Authors in [9], [10], and [11] all discuss the uses of infrared technology and thermal imaging to obtain data from a distance in various environments. This suggests that the ability to obtain data from afar generates viable data for datasets in identification and can prove useful in practice and various implementations.

Oliveira and Wehrmeister [12] utilize a multi-rotor UAV to perform data gathering and surveillance, altogether lowering the risks of sending actual humans to assess situations that may harm them. Cheaper methods of UAV surveillance in search and rescue missions are discussed, allowing this technology to be more adaptable if it is required.

2023 IEEE Radar Conference (RadarConf23) | 978-1-6654-3669-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/RADARCONF2351548.2023.10149735

Not much work has been done exploring the benefits of both an IR camera and a radar sensor to detect humans in heavy smoke. Therefore, in this work, we develop an integrated system composed of an IR camera and radar sensor that is capable of real-time detection and localization of humans and animals in heavy smoke. We train the two sets of data using a deep convolutional neural network (DCNN) algorithm to localize a human in dense smoke in real time. Experimental results reveal that the proposed system can detect a human in heavy smoke with 98% validation accuracy.

The rest of the paper is organized as follows. In Section II, we describe the system overview, followed by the human detection in heavy smoke using an IR camera in Section III and a radar sensor in Section IV. The deep convolution neural network algorithm is presented in Section V. After illustrating experimentation results in Section VI, we draw the main conclusions in Section VII.

II. System Overview

The system is composed of two parts, a training phase, and a testing phase. During the training phase, different movements in front of the IR camera are captured, cropped, and then organized in predefined folders. Similarly, the data collected by the micro-Doppler radar after amplification and analog to digital conversion are converted to spectrograms, which are then cropped and organized in predefined folders. The organized data set is used for training the DCNN algorithm to produce two pre-trained models, one for the IR camera and the other for the radar spectrograms. Fig. 1 shows the flow diagram of the training and testing phase.

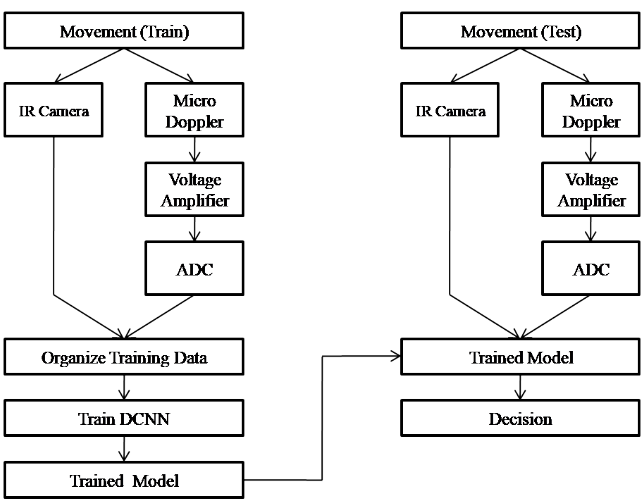

Fig. 1. The data flow for the system training and system testing.

The two pre-trained deep learning models are then ported onto a Raspberry Pi 4 micro-computer for live detection of a human in a dense smoke environment. The two sensors, the IR camera, and the micro-Doppler sensor capture the data in real-time using the Raspberry Pi 4 micro-computer. The gathered IR camera data as well as the newly generated spectrograms obtained from the micro-Doppler radar are fed into the pre-trained deep learning models to detect the presence of a human in heavy smoke. Once the trained model detects a human it informs the user whether he/she is at near, medium, or far range, i.e., $1\,m$, $5\,m$, or $10\,m$ from the sensors, respectively.

III. Human Detection with Infrared Camera

The IR camera captures the intensity of waveforms within the IR spectrum of 300 GHz - 430 THz, as opposed to the visible spectrum of 430 THz - 770 THz. The temperature of an object is correlated with its black body radiation which in turn decreases the wavelength of emitted IR waves. Thus lower wavelengths are generally represented as "bright" while longer wavelengths are generally represented as "dark". We utilize a FLIR Lepton 3.5 IR camera with a breakout board, as shown in Fig. 2. The Lepton 3.5 IR camera has a resolution of 160×120 pixels and a field of view of 71° angle-diagonal.

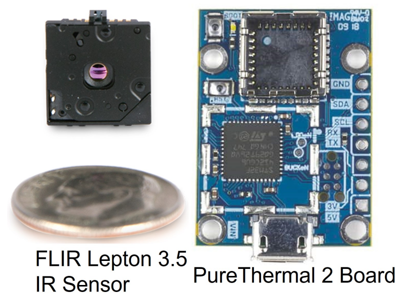

Fig. 2. Flir Lepton 3.5 IR camera with PureThermal 2 Board.

Fig. 3. Human in Heavy Smoke captured using Lepton 3.5 IR camera.

Samples from the FLIR Lepton 3.5 IR Camera with and without heavy smoke presence, are shown in Fig. 3 and Fig. 4, respectively. As we can see, unlike a visible light camera the thermal camera is not affected by heavy smoke, it can still clearly detect a human in a heavy smoke environment.

When gathering images from the IR camera, we observed that the camera would normalize the background, as shown in Fig. 3. Accordingly, we modified the software of the IR camera so that the camera would no longer normalize the images. In Fig. 5 we show several sample images of a human at different distances and one with no human without normalization. We trained our DCNN on the images without normalization as a human might not be seen if normalization happens when an

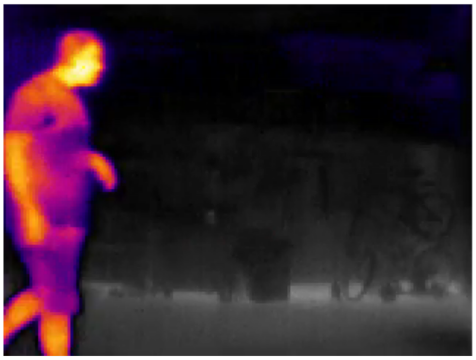

Fig. 4. Human in No Smoke captured using Lepton 3.5 IR camera.

object that is much warmer than a human is within the IR camera view.

Fig. 5. IR images without normalization of a human walking at $1\,m$, $5\,m$, $10\,m$, and No Human.

IV. Human Detection with Radar Sensor

To gather the presence of a human in heavy smoke we utilize the HB100 radar sensor, which is an X-Band Bi-Static Doppler transceiver module. It is a built-in dielectric resonator oscillator and a pair of microstrip patch antenna arrays operating at 10.25 GHz. The HB100 radar sensor as well as the transceiver block diagram are shown in Fig. 6.

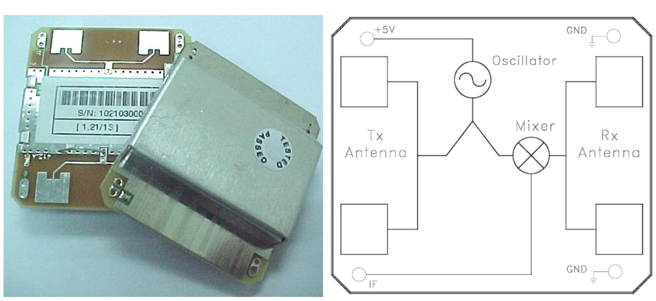

Fig. 6. HB100 Radar Sensor and the Transceiver Block Diagram.

The micro-Doppler radar sensor HB100's output voltage is too low to be analyzed and converted to a spectrogram. Accordingly, it is first amplified with the amplification circuit, shown in Fig. 7, and the output is taken at the output pin of the LM234 Operational Amplifier.

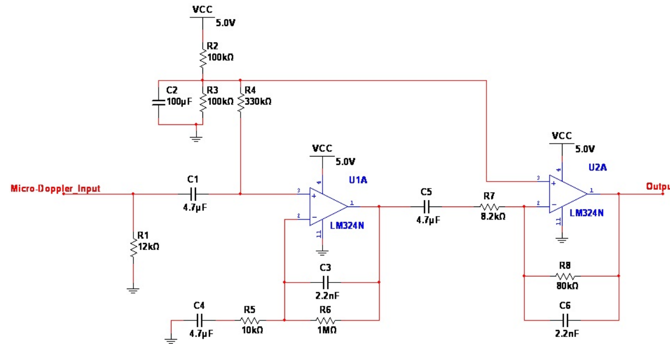

Fig. 7. Amplification circuit for HB100 micro-Doppler sensor.

The soldered amplification circuit for the HB100 micro-Doppler sensor is shown in Fig. 8. It utilizes two op-Amps one acts as an inverting amplifier while the other changes the polarity back to the original positive value.

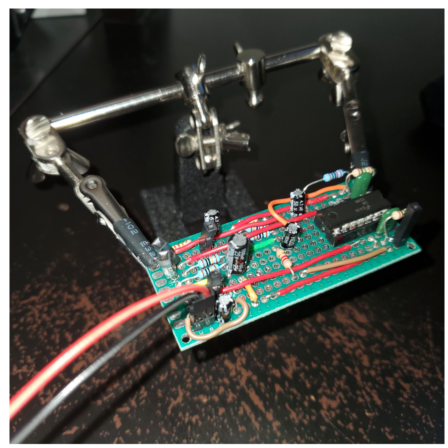

Fig. 8. Soldered amplification circuit for HB100 micro-Doppler sensor.

A spectrogram is used to indicate the magnitude of various frequencies present within a signal over time, thus emphasizing small patterns within a captured waveform that are not easily identified from the waveform itself. The x-axis represents the passage of time, the y-axis indicates the frequency (in the case of the observed micro-Doppler waveform, the side-band frequency), and the color indicates the magnitude in the defined magnitude and frequency "bin".

A fast Fourier transform (FFT) is applied to every A (variable) based on the FFT size, moving forward B (variable) samples based on the FFT time step. The FFT is applied every FFT time step until the end of the waveform is reached by a given FFT operation based on FFT size. The FFT determines the magnitude for every frequency within the FFT size and fills a "bin" on the spectrogram with the width FFT time step.

The "bin" size is determined by the FFT sample step and the sampling frequency.

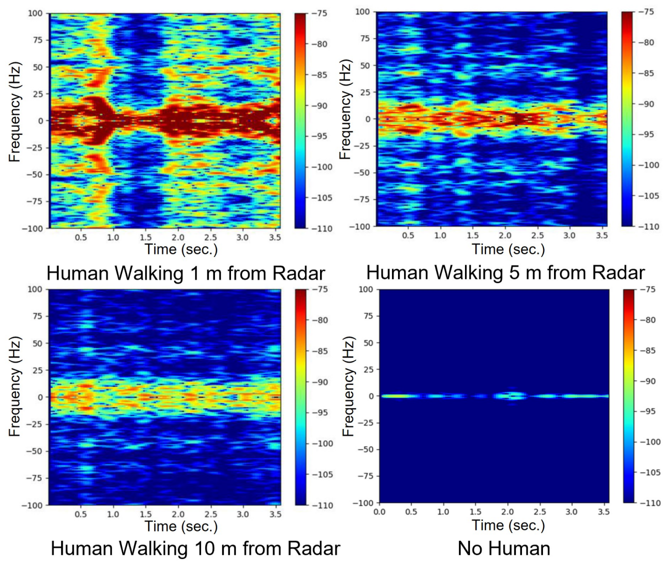

Fig. 9. Human walking from radar at $1\,m$, $5\,m$, $10\,m$, and No Human.

Sample spectrograms generated using the test data are shown in Fig. 9 for a person walking at $1\,m$, $5\,m$ and $10\,m$, and No Human, from the HB100 micro-Doppler sensor, respectively, each taken with an FFT size of 512 samples, FFT time step of 64 samples, and sampling frequency of 1000 Hz. As we can observe from Fig. 9, unique spectral characteristics are apparent in those four spectrogram plots for a person walking at $10\,m$ from the radar sensor versus a person walking at $5\,m$ or $1\,m$ from the sensor and no human present. Using the deep learning algorithms, which are discussed in the next section, we will train it to recognize those unique characteristics for human detection at a different distance from the radar sensor.

V. Deep Learning Model Training

We utilize a DCNN algorithm to determine if a human is present in a heavy smoke environment by using images captured from the IR camera and the generated spectrograms using the micro-Doppler radar sensor.

The architecture of a DCNN algorithm implemented to train the machine learning model for the IR camera image input is shown in Fig. 10. A similar DCNN architecture is used for the spectrogram image machine learning model training. The IR images and spectrograms are first resized to 100×120-pixel images. The input images undergo a feature extraction network by first being processed by the convolution layer consisting of 8 convolution filters of size 20×20. The output from the convolution layer goes through the rectified linear unit (ReLU) function followed by the pooling layer, which employs a max pooling process of 2×2 matrices. This process is repeated several times to create the output and train the machine with the inherent features of the image. The output of the pooling layer is fed into a second convolution layer consisting of 16 convolution filters of size 10×10. Similarly, after passing the output through the ReLU function it undergoes the pooling layer with a max pooling size of 2×2 matrices. Finally, it is passed through the third round of the convolution layer consisting of 32 convolution filters of size 5×5 after, which is processed by the ReLU function, and the pooling layer with max pooling size of 2×2 matrices. The max pooling concept is demonstrated in Fig. 11.

The stride is the sliding window operation, which is used in the convolution layer and in the max pooling operation in which case the stride is 2 [13], [14]. Suppose $n \times n$ convolution is performed, the stride represents the movement by S elements with every step. If the stride is defined as 1 that means the convolution layer will move with a sliding window of 1 pixel and move every third pixel by skipping the second pixel. Max pooling, shown in Fig. 11, is a down-sampling process where it selects the maximum value from each view [15], [16]. Since the IR camera images and spectrograms contain sharp edges max pooling instead of average pooling is used to extract the most important features such as edges. The classifier network consists of a fully connected layer comprised of 100 hidden nodes, which produce a Softmax output that in turn is used for classifying the driver's status. The output layer of the DCNN represents the probability distribution containing the probabilities that each class is assigned in accordance with the input images. Using maximum ratio combining by looking at the two gathered sensor data once the algorithm detects a human in heavy smoke it will send an alert to the user, e.g., a firefighter, if no human is detected it will forward to the algorithm to continue to perform the human detection on the new set of data received by the sensors.

While building the DCNN model, there are a few important inputs that affect the learning and validation accuracies. These inputs are the number of dense layers (DL), layer sizes (LS), and the number of convolution layers (CL).

To train our model for optimum results, we make use of two functions; *Function A* and *Function B* for a simple explanation. Function A is responsible for finding the best input arguments by trying every different combination of the inputs and then training a model on each variation of the inputs. For example, if DL can be 0, 1, or 2 and LS can be 32, 64, or 128, Function A will try to train 9 different combinations and record which specific combination gave the best training accuracy results.

Next, we take the best input values, found in Function A, and feed them into Function B. Function B is responsible for training the final model. Function B will train the model using an $80\% - 20\%$ split for training and validation. We train two DCNN models to account for the two different categories of input images, IR images, and spectrogram images. Overall we have gathered and used 2,000 IR images and 2,000 spectrogram images from the radar, 500 images per classification category per sensor. For each DCNN model, there are four classifications for the input images: close, medium, far, and no human.

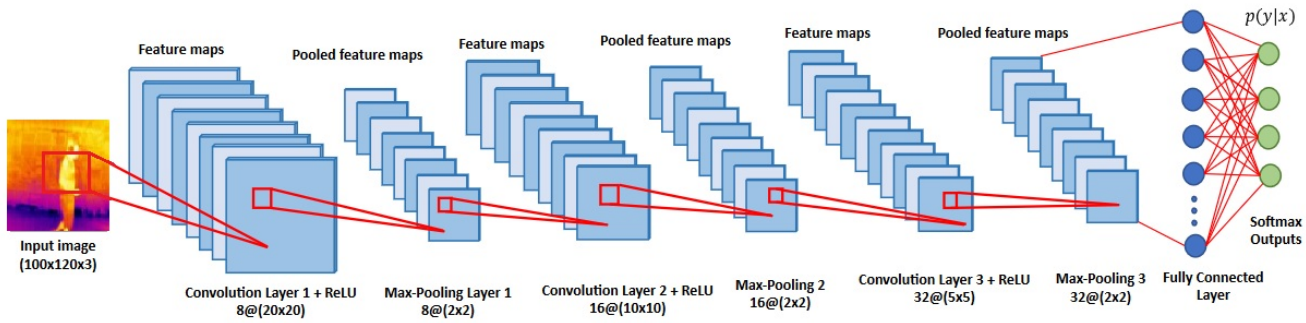

Fig. 10. Architecture of the DCNN algorithm implemented for image classification.

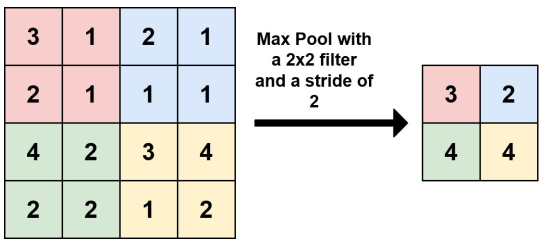

Fig. 11. Max pooling principle.

VI. Experimental Results

Experiments were conducted in a two-car garage in which an artificial smoke was first filled using a VIRHUCK fog machine. The smoke thickness was so high that a human in the smoke was not visible to the naked eye.

The testbed experimental setup is shown in Fig. 12. The right slide shows the IR sensor as well as the radar sensor used on the top shelf along with the fog machine at the bottom self. The top right image shows an instance of how the artificial smoke is filled inside the garage and the bottom right side is when the smoke exits as the garage door is opened.

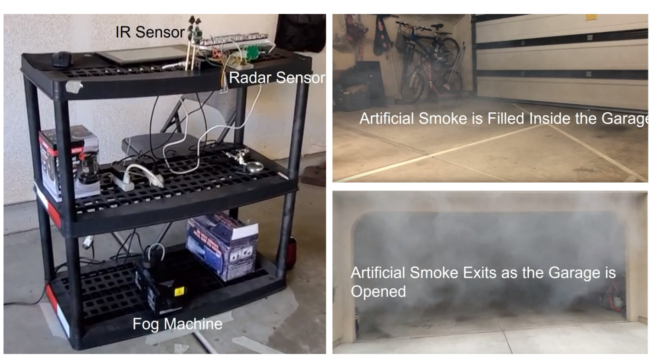

Fig. 12. Testbed experimental setup inside a garage with artificial smoke.

The fully integrated data acquisition system is shown in Fig. 13. It is composed of the IR camera mounted on a 3D printed stand and the HB100 radar sensor along with the radar amplification circuit. All the sensor information is gathered on the Raspberry Pi 4 micro-computer that does all the signal processing for live detection of the human in heavy smoke using the pre-trained deep learning modes discussed in Section V.

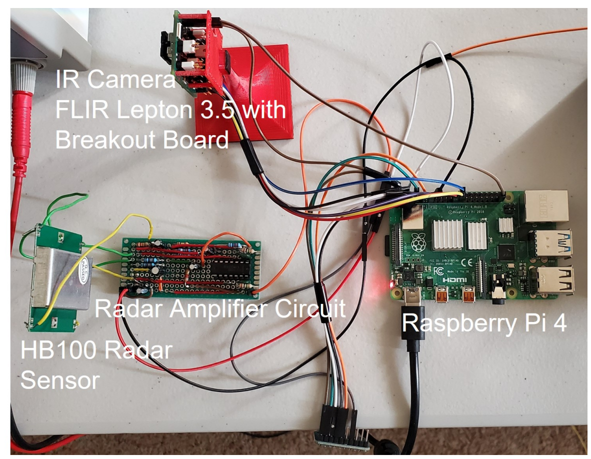

Fig. 13. Fully connected data acquisition system.

Implementing machine learning for image classification greatly reduces the time needed for decision-making in an intense environment where every second counts. Image classification can be more than just a “yes” or “no” decision. The DCNN that was implemented consisted of two models, each with four labels of classification. To enlarge our data set we used data augmentation, sample images of data augmentation for the four categories are shown in Fig. 14. The actual two models were trained using the larger data augmentation data set. The data augmentation enlarged our data set from 2,000 images for each sensor to 14,000. The larger modified data set showed an observed improvement in the validation accuracy of both models.

When training the two DCNN models, overfitting of the dataset was experienced whenever the number of epochs was greater than fifteen. For this reason, the number of epochs was limited to ten when training and 44 iterations per epoch were used. This provided both models the IR sensor and the radar sensor with about 98% validation accuracy with a validation split of $80\% - 20\%$, as shown in Figs. 15 and 16. The benefit of using an IR camera along with the radar sensor is that the IR sensor acts best when the human is not moving fast, e.g., stationary while the radar sensor provides the best results when the human is in motion in a heavy smoke environment.

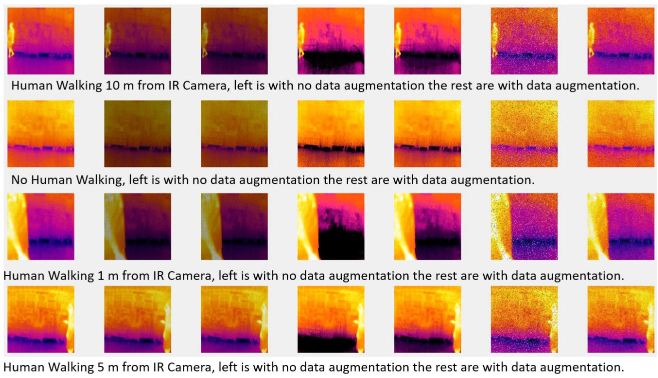

Fig. 14. Data augmentation of our IR camera data set.

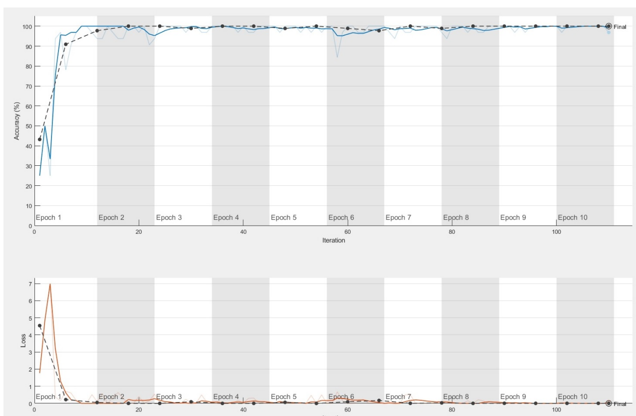

Fig. 15. Validation accuracy and loss plot of the IR model.

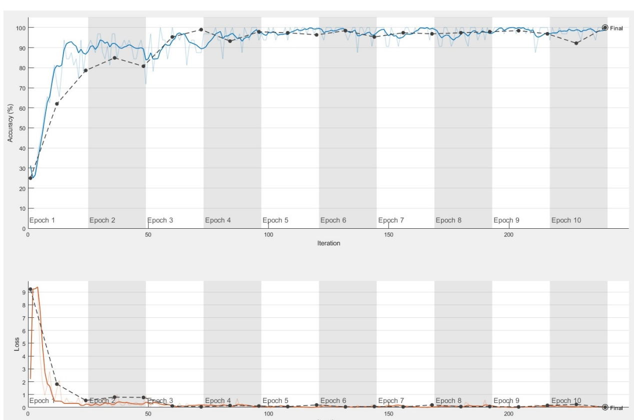

Fig. 16. Validation accuracy and loss plot of the spectrogram model.

VII. Conclusion

We developed an AI-based human detection scheme using two unique sensing mechanisms an IR camera and a radar sensor. The system is capable of real-time detection and localizing humans and animals in deep smoke to improve the situational awareness of firefighters on the scene. We make use of data from a micro-Doppler sensor and an infrared camera and train a deep convolutional neural network algorithm to localize a human in dense smoke at Close, Medium, and Far distances in real time. Experimental results reveal that the proposed system can detect a human in heavy smoke with 98% validation accuracy. A similar AI-based training procedure can be carried out to detect animals in heavy smoke as well.

Acknowledgment

Special thanks to the Claude C. Laval Award for Innovative Technology and Research Grant for supporting aspects of this project.

References

[1] J. W. Starr and B. Y. Lattimer, “Evaluation of navigation sensors in fire smoke environments,” *Fire Technology*, vol. 50, no. 6, pp. 1459–1481, 2014.

[2] J.-H. Kim, J. W. Starr, and B. Y. Lattimer, “Firefighting robot stereo infrared vision and radar sensor fusion for imaging through smoke,” *Fire technology*, vol. 51, no. 4, pp. 823–845, 2014.

[3] X. Li, Y. He, and X. Jing, “A survey of deep learning-based human activity recognition in radar.” *Remote Sensing*, vol. 11, no. 9, p. 1068, 2019.

[4] X. Ma, R. Zhao, X. Liu, H. Kuang, and M. A. A. Al-qaness, “Classification of human motions using micro-doppler radar in the environments with micro-motion interference.” *Sensors (14248220)*, vol. 19, no. 11, p. 2598, 2019.

[5] M. Sakai and Y. Aoki, “Human and object detection in smoke-filled space using millimeter-wave radar based measurement system,” vol. 3. IEEE, 2006, pp. 750–750.

[6] M. Bhattarai and M. MartíNez-Ramón, “A deep learning framework for detection of targets in thermal images to improve firefighting,” *IEEE Access*, vol. 8, pp. 88 308–88 321, 2020.

[7] F. B. Jaradat and D. Valles, “A victims detection approach for burning building sites using convolutional neural networks,” in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 2020, pp. 0280–0286.

[8] R. S. Allison, J. M. Johnston, G. Craig, and S. Jennings, “Airborne optical and thermal remote sensing for wildfire detection and monitoring.” *Sensors (14248220)*, vol. 16, no. 8, pp. 1310 – 1338, 2016.

[9] D. Li, C. C. Menassa, and V. R. Kamat, “Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography,” *Energy and buildings*, vol. 176, pp. 246–261, 2018.

[10] E. Valldor, “Person detection in thermal images using deep learning,” Master’s thesis, Uppsala University, 2018, independent thesis Advanced level (degree of Master (Two Years)).

[11] T. T. Zin, H. Takahashi, and H. Hama, “Robust person detection using far infrared camera for image fusion.” IEEE, 2007, pp. 310–310.

[12] D. C. de Oliveira and M. A. Wehrmeister, “Using deep learning and low-cost rgb and thermal cameras to detect pedestrians in aerial images captured by multirotor uav,” *Sensors (Basel, Switzerland)*, vol. 18, no. 7, p. 2244, 2018.

[13] H. Kulhandjian, P. Sharma, M. Kulhandjian, and C. D’Amours, “Sign Language Gesture Recognition using Doppler Radar and Deep Learning,” in *Proc. of IEEE GLOBECOM Workshop on Machine Learning for Wireless Communications*, Waikoloa, HI, USA, December 2019, pp. 1–9.

[14] H. Kulhandjian, E. Batz, E. Garcia, S. Vega, S. Velma, M. Kulhandjian, C. D’Amours, B. Kantarci, and T. Mukherjee, “AI-based RF-Fingerprinting Framework and Implementation using Software-Defined Radios,” in *Proc. of IEEE International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, HI, USA, Feb. 2023, pp. 1–5.

[15] H. Kulhandjian, N. Ramachandran, M. Kulhandjian, and C. D’Amours, “Human Activity Classification in Underwater using sonar and Deep Learning,” in *Proc. of ACM Intl. Conf. on UnderWater Networks and Systems (WUWNet)*, Atlanta, GA, USA, October 2019.

[16] H. Kulhandjian, J. Barron, M. Tamiyasu, M. Thompson, and M. Kulhandjian, “Pedestrian Detection and Avoidance at Night Using Multiple Sensors and Machine Learning,” in *Proc. of IEEE International Conference on Computing, Networking and Communications (ICNC)*, Honolulu, HI, USA, Feb. 2023, pp. 1–5.