A 360-Degree Video Analytics Service for In-Classroom Firefighter Training

Ayush Sarkar*§, Anh Nguyen†§, Zhisheng Yan†, Klara Nahrstedt*
*University of Illinois Urbana-Champaign, †George Mason University
*{ayushs2, klara}@illinois.edu †{anguy59, zyan4}@gmu.edu

***Abstract*—Demonstrating firefighting operations in search and rescue missions through videos is a common approach to in-classroom firefighter training. Unfortunately, traditional 2D cameras have fundamental weaknesses – they can only capture a narrow field of view and miss a lot of information coming from the surroundings of the firefighter, which may become the matter of life and death in certain situations. In this paper, we propose a system combining the advantage of 360° videos and deep learning to automatically detect important objects in the panoramic scene, assisting firefighting instructors in classroom teaching scenarios. Specifically, we summarize the salient objects and events relevant to firefighting through an interview with an experienced firefighting instructor. Leveraging this knowledge, we investigate the detection of firefighting objects on 360° videos through a transfer learning approach. We report insightful results for object detectors trained on generic objects and 2D videos and discuss the next steps in designing a customized object detector.**

***Index Terms*—360° video, firefighting, object detection**

I. Introduction

Firefighting is a dangerous activity that demands extensive training of firefighters both inside and outside of the classroom. Instructors (often the Incident Commanders) are in need of advanced cyber-tools inside of the classroom to point out, demonstrate, and show to trainees the appropriate behaviors needed to contain fire effectively before the trainees go out to a training ground to learn in a hands-on fashion how to act to save lives. Firefighting institutes such as the Illinois Fire Service Institute [2] provide training capabilities for future firefighters. They use several techniques in the classroom involving animations, slide-lectures, and recordings that are created using 2D video cameras to showcase both proper and faulty behaviors in firefighting while also educating trainees about diverse situations that firefighters can face at the incident scene.

However, existing 2D video-based tools suffer a fundamental limitation – incident commanders (ICs) are only able to view and display a single view of a scene at a time. This makes it infeasible for ICs and firefighter trainees to discuss the entire picture of the emergency scene, as events of interest can occur outside of a camera's coverage. With the advances of new camera hardware such as 360° video cameras and artificial intelligence (AI) technologies utilizing deep learning, new innovative tools and techniques can be investigated to assist firefighting instructors more efficiently during in-classroom teaching.

§Equal contribution

Fig. 1: The main viewport of a 360° firefighting video overlaid with a mini-view showing an event of interest – fire rollover.

360° video cameras offer multiple views, allowing the ICs to leverage broader content from different views to demonstrate firefighting objects and events. Deep learning AI algorithms can allow for the intelligent selection and inference of events and objects on top of these teaching videos. We present the vision of a 360° video analytics service that enables ICs to interact with and showcase all views in the entire emergency scene. This is achieved by allowing the ICs to not only switch the viewport manually within a 360° panoramic scene, but also visualize machine-detected events of interest that are highlighted on the teaching videos. As shown in Fig. 1, if the video can show not only the first-person view of a room under search, but also has "eyes in the back" to identify the fire rollover behind through a mini-view, the IC is better equipped to educate trainees about the best practices in such a chaotic environment.

Achieving this vision requires us to overcome two major challenges. The first challenge is to understand which events are of higher and lower importance in 360° firefighting videos. This is important because showing and detecting too many visuals might cause information overload. While common sense is often utilized in prior firefighting technology design, they may not meet the domain need. One needs domain expertise from firefighting institutes to identify salient events and objects. The second challenge is to develop a 360° service pipeline that would automatically analyze the recorded 360° videos via appropriate machine learning algorithms and infer the desired objects that the IC can then use to achieve teaching objectives for the trainees. Despite prior research in object detection, the second challenge is non-trivial, as it is unknown whether existing detectors for general purposes can detect specialized firefighting-related objects and events. The

accuracy of detecting firefighting objects in 360° videos is also in question, as these videos are stored in an equirectangular format which distorts objects in each frame (e.g., stretched objects at the top and bottom of each video frame).

In this paper, we present a novel 360° video analytics service framework that will be used in a training tool for ICs. We make two important contributions. First, we conduct extensive interviews with our collaborator Richard Kesler, an expert in the IFSI (Illinois Fire Service Institute), to understand the needs for detecting objects and events in firefighter training. We summarized the list of important firefighting objects and events that can be used by the community for the future design of firefighting tools. We show the interview results in Section III. Second, we conducted a study to understand whether or not and how accurately existing object detectors developed for generic 2D videos can perform on 360° firefighting videos. This is the first step towards utilizing a transfer learning mechanism to design 360° firefighting object and event detectors, which can potentially deal with the lack of the ground-truth-annotated 360° firefighting videos. The results and analysis are shown in Section IV. In Section V, we discuss next steps regarding our cyber-tool that can allow ICs to showcase events and objects in real-time to the trainees via pop-up mini-views. We conclude the paper in Section VI.

II. Related Work

Media Technologies in Firefighting. Research has been conducted to understand how audio and video can improve the training and operation for firefighters. In [6], the authors specified a list of requirements in terms of services and security for wireless communications among first responders. Later, the same authors proposed a framework to evaluate the performance of media-assisted firefighting systems. Their evaluation approach relied on the movement patterns of first responders [7], [8]. However, previous first responder systems focused on 2D videos were limited by the field-of-view of 2D cameras. The situational awareness of these media systems can be enhanced to further assist training and operation.

Recently, immersive media technology has begun to attract attention in firefighting studies. Bellemans *et al.* [3] simulated various hazard training scenarios happening on Navy decks using virtual reality. It provides a realistic, flexible, and cheap way to train firefighters while reducing the associated risks. However, their main focus is training in virtual environments. This is different from our work since we are using 360° videos and identifying events of interest happening in the real world in order to assist training.

Object Detection. Object detection is one of the fundamental problems in computer vision. In early days, handcrafted features such as Kalman filtering were widely used to identify objects, but they were less robust as they could not cope with the variety of textures in images. To overcome this limitation, later works applied deep learning to train a model directly on image pixels without relying on a fixed and limited set of handcrafted heuristics. Modern detectors, such as SSD [11] and algorithms in the YOLO family, rely on a one-shot architecture, allowing them to perform detection in real time while maintaining competitive detection accuracy. Despite this fact, modern object detectors can only detect a small number of generic object categories [5]. These categories do not necessarily cover objects/events considered important by firefighters. Furthermore, these detectors are mostly trained on 2D videos/images, which are free from the projection distortions inherent in 360° videos. In this paper, we take the first step to understand how modern object detection architectures perform on firefighting objects and 360° videos. We aim to shed some light on the design of a full-scale object and event detector customized to 360° firefighting videos.

III. Identifying Significant Objects and Events

Typical firefighting scenarios involve a myriad of objects and events that can potentially be detected and identified; however, many of these instances are nonessential and contribute little towards firefighter in-classroom training. To achieve visual clarity for trainees in the classroom setting, the 360° analytics framework must only highlight salient objects and events that demand high priority while also emphasizing swift scene comprehension. This begs the fundamental question that has not been addressed by previous studies - what critical objects and events should be explicitly highlighted for firefighter training? Answering this question is not straightforward and requires specialized knowledge from domain experts. To this end, we conducted an interview with a domain expert.

A. Identifying Subject to Interview

We conducted an interview with Richard Kesler, the Deputy Director of Research Programs at the IFSI. His work focuses on examining the physiological demands of firefighting activity and the impact of firefighting on the firefighter. He serves on the technical committee for the National Fire Protection Association. He teaches in numerous IFSI classes and is a physical training instructor for the IFSI Basic Operations Firefighter Academy. Richard is currently pursuing his PhD in Kinesiology and serves as a volunteer Assistant Chief with the Savoy Fire Department.

B. Data Collection

We prepared a series of questions to ask Richard during a qualitative virtual interview:

1) What objects and events require an alert during firefighter operations?
2) What are your requirements for remote incident command systems? What types of information signals would you prefer, and what would be the preferred frequencies and durations of these signals?
3) What is the priority of objects/events that you would like to highlight for firefighter trainees?
4) Should environmental events be flagged and differentiated from events involving people?

Based on the answers to these questions, we were able to compile a list of important objects and events during firefighting scenarios. Each object or event is classified based on whether

TABLE I: Events and objects of importance in firefighting

Type	Context	Name	Priority
Object	Training	Helmet off	Low
Object	Training	Unfastened SCBA strap	High
Object	Training	SCBA facepiece not secure	High
Action	General	Changes in fire condition	Very High
Action	General	Increases in smoke density	Very High
Object	General	Person (Non-firefighter)	Very High
Object	General	Person (Firefighter)	Low
Action	General	Human entering/exiting building	High
Action	General	Integrity of building material	High
Action	General	Water coming out of hose	Low

they can be detected through either object detection or action detection. We further grouped each object or event based on the context of the situation – training-specific objects/events and general objects/events important in both training and emergency scenarios. Table 1 lists the results with the priority value signifying relative importance.

General Objects and Events. We found that events of the highest priority always involved occupants within a building or structure. It was important to explicitly signal when an individual entered and exited a structure, and the most important alerts involved movement inside of a structure that did not belong to a member of the fire crew. An example of an important event would include an occupant hanging outside of a window. It was also considered a priority to differentiate environmental events from events involving people. Environmental events to consider involved changes in smoke and fire condition, changes imperative to highlight to let the IC know whether the frontline firefighters are able to contain the fire effectively.

Training-Specific Objects and Events. Training-specific objects and events are closely related to the errors that trainees often commit during their lessons but are not highly pertinent to emergency incident response scenarios. These mistakes are important to highlight during the in-classroom session, particularly the errors involving the self-contained breathing apparatus (SCBA), a respiratory device that delivers breathable compressed air to firefighters to help them in lethal environments involving toxic particulates. Detecting the mistakes involving the SCBA are of high priority, as they are not as salient and the instructor may possibly miss them during the live session. The training specific instances involving the SCBA starkly contrast with the training-specific case of a firefighter's helmet falling off. This is because the latter case is very salient in the training videos without explicit highlighting and the error itself is detected quite easily by the trainees.

C. Analysis of Interview Data

The collected expert insight provided us with a better understanding of what objects to detect for specifically training purposes. We conclude that many events (e.g., water failing to flow out of the hose), while considered critical and of high priority during incidence response, are of low priority to highlight because they are immediately noticeable by trainees in the classroom. This is somewhat unexpected from a non-expert point of view. However, it provides significant insight for future design of firefighting technologies. To achieve teaching objectives for trainees, ICs must prioritize visual clarity. This fundamental idea also applies to the priority difference between firefighters and non-firefighters. During incident response, the safety of individuals not part of the fire crew, such as occupants within a burning structure, intrinsically carries more importance than the location of firefighters within the crew. Firefighters typically take up large portions of the video, but highlighting each and every instance of a firefighter throughout the video leads to more clutter and information overload.

IV. Detecting Objects for Firefighting

Detecting firefighting-related objects in 360° videos is non-trivial. First, specialized objects, e.g., fire, present radically different textures, shape, and color from everyday objects. There is additional complexity even for the same typical objects - among humans, a firefighter is considered to be in a different class from a non-firefighter, such as an occupant within a structure. Moreover, distorted objects in 360° videos may further confuse the object detectors.

Modern object detectors have been applied in general 2D video content. We ask ourselves a question – can a deep learning model, trained on large public non-firefighting datasets, be applicable to firefighting object detection tasks? If yes, then how effective will the trained classifiers and model parameters be, and what needs to be done to further extend the object detector to be applicable to 360° firefighting videos? Successfully addressing these questions is important in 360° firefighting object detection as it could save a tremendous amount of effort in data collection, annotation and modeling. However, they have never been answered before. Previous object detection frameworks primarily focused on finding humans and everyday objects in 2D videos where smoke and other fire hazards were not presented [15], [17]. In this paper, we focus on understanding whether or not and how a specialized detector can be designed.

A. Object Detection for Firefighting Operations

To answer the aforementioned questions, we employ YOLOv3 [16], a state of the art one-stage object detector known for its speed and accuracy. YOLOv3 will be trained on the COCO dataset and evaluated on 2D videos provided by the IFSI as well as sample 360° videos collected online. We chose COCO due to its size and versatility of annotated data (1.5 million 2D images). On the other hand, the dataset from IFSI [2] includes 28 2D videos. These videos range from instruction and handling equipment to firefighting in a practical environment.

At this moment, we focus on investigating the potential of domain adaptation with transfer learning, testing the efficacy of object detectors trained with COCO on the IFSI-provided videos and our self-collected videos. We investigate two levels of knowledge transfer – transfer learning from everyday objects to firefighting objects and transfer learning from 2D videos to 360° videos. Details regarding the potential of transfer learning are discussed in section V-A1.

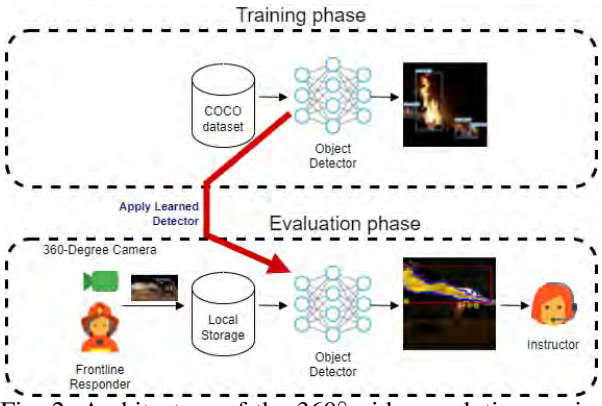

Fig. 2: Architecture of the 360° video analytics service.

Details of our approach are described in Fig. 2. There are two main steps. In the training phase, YOLOv3 was trained on the COCO dataset. In the evaluation phase, we used the trained YOLOv3 framework to perform detection on test videos, as indicated by the red arrow. All videos were stored locally and then later delivered to the back end. Here, videos containing objects defined in Section III-C were selected for evaluation. After the YOLOv3 framework performs detection on the videos, relevant objects can be highlighted to assist the commander when he/she is reviewing the videos.

B. Detection Experiments and Results

1) Results on 2D Videos: We utilized the YOLOv3 architecture to perform detection frame-by-frame on our 2D videos. We discovered that firefighting-relevant objects such as the firefighters and their SCBA breathing air cylinders were detected accurately, demonstrating the potential of a transfer learning approach. However, YOLOv3's performance degrades significantly when detecting objects in low-light and smoke-filled conditions. Fig. 3 demonstrates the degradation in performance of the detection model for 3 IFSI videos. The first video (the first row) captures a scene of a firefighter team crawling into a burning building. The second video (the second row) is of a scene during which two firefighters are dragging a downed, unconscious firefighter out of a smoke-filled environment. The third video (the third row) consists of a firefighter assist and search team searching the interior of a flaming building for other firefighters in distress. For the first video, the detection confidence score for humans dropped from 0.95 (left figure) down to 0.53 when the firefighter entered the smoke-filled environment (right figure). In this environment, YOLOv3 failed to track the breathing air cylinders and the firefighter in front of the one detected, as their textures were blurred and occluded. For the second video, none of the firefighters were detected in the smoke-filled environment (right figure). They became detectable once on the verge of exiting the building (left figure). The third video also demonstrates a stark contrast – almost all of the individuals in the video are perfectly detected once they have exited the smoke-filled building (left figure). However, the firefighter inside the building was not detected (right figure).


Fig. 3: Difference in performance when the model detects objects in normal conditions (left) versus smoke-filled and low-light conditions (right) for 3 videos (3 rows).

In addition, we quantitatively evaluated the model performance on 2D firefighting videos. We evaluated the model performance as well as the change of model score when the video content transitions from a normal training condition to a fire hazard environment. In these testing videos, half of the content included the inside of the building, which was dark and covered in smoke.

We use mAP scores to benchmark the detection accuracy [10]. The mAP metric is commonly used to reflect the reliability of predicted bounding boxes, requiring two types of input, the ground-truth bounding boxes and predicted bounding boxes. The predicted bounding boxes are produced directly from YOLOv3. On the other hand, since firefighting objects have never been annotated in previous video datasets including the IFSI dataset, the ground truth bounding boxes must be manually annotated by us. We manually annotated the positions of bounding boxes of objects appearing in the video frames, and to facilitate the annotation process, we developed a Python program plotting positions of ground truth bounding boxes on frames given the input positions in numerical format. The plotted images provide visual cues, allowing the annotator to shift the bounding boxes until they fit the target objects in the images. The adjusted bounding boxes are accepted as ground truth if the annotator cannot find gaps between an object and its bounding box within three seconds.

Table II reports the average mAP results. We observe that the mAP score when YOLOv3 performs detections for normal conditions ranges from 13 to 44 points higher than that of hazard conditions. This big drop further confirms the significant impact of factors such as light condition and fire/smoke occlusion on model performance. We also notice that the YOLOv3 performance in normal conditions is not far from the typical performance of YOLOv3 on the COCO dataset, proving the potential of transfer learning from regular

TABLE II: The mAP scores for all training videos

Average duration (seconds)	mAP normal	mAP hazard
16.5	44.2	15.6

Fig. 4: The distorted detection results on a 360° video.

objects to firefighting objects.

We conclude that current off-the-shelf object detectors could detect some important objects closely related to the context of a firefighting operation. However, the performance of these detectors degrades in the face of smoke-filled and fire environments, demanding that our training data incorporate images of objects within these scenarios. Further improvement directions are discussed in detail in the Discussion Section.

2) Preliminary Results on 360° Videos: We looked to evaluate how the performance of the object detector trained on 2D videos degrades for firefighting-specific 360° content. We had YOLOv3 perform detection tasks on a publicly available 360° fire safety video stored as equirectangular format [1]. Many objects in this video were close to the equator of the projection, allowing for many successful detections as the distortions near the equator are small. Despite this, many objects were still misclassified or not detected at all due to the small distortions of the equirectangular 360° format. We point out one particularly egregious misclassification caused from the distortion of the video in Fig. 4 – the stretching of the building structure coupled with the elongated curvature of a firefighter's shadow led the YOLOv3 object detector to classify a large area of the video as a boot. In another instance, a firefighter crawling near the camera was misclassified as a motorbike due to the distortions from the projection elongating his arms and legs substantially more than his back.

V. Discussion

A. Improving Event-Object Detectors

Event-Object detector is the core component in our system. The following discuss several approaches to improve its performance.

1) Transfer Learning: Transfer learning is a popular technique to reduce the limitation of the small data size. The idea is to have the model trained on a large dataset originating from a closely related domain before continuing the training on the target dataset. By doing this, the model remembers some universal basic patterns across domains such as color and texture, which helps increase the model accuracy. Transfer learning has been successfully applied to transfer knowledge in some computer vision tasks e.g. transferring 2D saliency to panoramic saliency [13].

We have completed initial steps towards using transfer learning to solve our problem. In particular, we showed that YOLOv3 can leverage some knowledge of the COCO dataset to identify important firefighting objects we previously identified on both 2D videos and 360° videos. To fully utilize transfer learning techniques, our next step is to design a training procedure to force YOLOv3 to learn features for hazardous environments and distorted objects in 360° videos using a customized dataset while still retaining previous knowledge from COCO.

2) Data Collection: To conduct the transfer learning process, an annotated dataset for 360° firefighting videos is needed. To our best knowledge, such a dataset does not exist. We plan to collect and manually annotate our own dataset. Regarding the data sources, gathering data from firefighting training sites of the IFSI is certainly not enough. We also need to collect relevant videos from other digital platforms such as YouTube and Vimeo. To ensure the success of the model training, collected videos must include the previously defined important objects/events. Furthermore, the content should be diverse enough to reduce the overfitting effect.

3) Working with 360° Video: We see that YOLOv3 detects objects reasonably well in the regions near the equator but performs poorly near the pole areas due to the distortion from the projection in 360° videos. To overcome this challenge, we plan to employ a multi-directional projection (MDP) technique in our algorithm [14]. This technique generates different versions of a 360° frame, each with a different projection orientation. The fundamental idea is to alter the original projection of the panoramic frame to force distorted objects near the pole areas towards the equator in new versions of the frame with different projection orientations. This method could allow a YOLO-based object detection framework to perform well for 360° formats even if it is largely being trained on 2D video datasets.

4) Dealing with smoke-filled conditions: As previously discussed in Section IV-B, the detection performance dropped significantly due to the interference of smoke. To deal with these challenges, we plan to investigate two potential approaches. Firstly, frames could be transformed from the RGB color space to YUV, or the event-object detector could utilize a thermal imaging component to highlight features in dim or smoke-filled environments, where RGB images fail. Secondly, we will try augmenting frames with simulated smoke and a low light level to adapt the model to fire hazard conditions.

5) Event Detection: Our current system can detect objects of various types. We aim to improve the model to detect both objects and events. To accomplish this goal, our model should be able to accommodate spatiotemporal data from multiple consecutive frames to detect time-dependent events such as increases in smoke density, changes in fire condition, and the structural collapse of building materials. There are multiple techniques to solve these problems, involving long short term memory networks [12], transformers [9], or 3D convolutional networks [4]. Finding the suitable method for our problem will be left for future work. The developed event detection

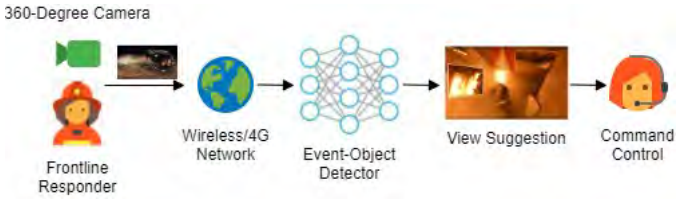

Fig. 5: Future system architecture.

model and the object detector will constitute an Event-Object Detector that can be used in more emergency management scenarios outside of firefighter in-classroom training.

B. Future System Architecture

While the service framework discussed in this paper focuses on an offline instruction system in classrooms, our vision in the long term is to bring the 360° video analytics service into real world situations, e.g., deploy the system online for active firefighting operations. We plan to create a real-time system transferring omnidirectional information captured from frontline responders to the IC, allowing him/her to quickly grasp the situation and make the most-informed decision.

The details of this future system are described in Fig. 5. In the system, 360° cameras are attached to each dispatched frontline responder. As the frontline responders proceed to engage and interact with the environment, the panoramic cameras capture the scene. The 360° video data for each frontline responder is encoded and transferred over the network to a server at the back end. The main viewport corresponding to the current orientation of the firefighter carrying the 360° camera is rendered and displayed to the IC. At the same time, a deep learning model performs directly on the video frames in the playback buffer to detect critical objects and events. Relevant information not shown in the current viewport will be displayed in a mini-view. The commander will then achieve an enhanced situational awareness by viewing the mini-view or quickly switching to a new viewport using the mini-view.

VI. Conclusion

This research addresses the lack of situational awareness of traditional 2D video tools when they are used to monitor search and rescue operations in firefighter training. Our main contributions are two fold. First, we perform a thorough interview process to identify objects/events relevant to search/rescue in firefighter training. We discover that objects of high priority in real-world scenarios may not need to be highlighted in the training because they may be easily noticeable. Second, we propose a framework to combine the advantages of 360° cameras and deep learning to help the IC in teaching firefighting operations. Our preliminary result shows that the detector trained on generic 2D video datasets can predict the existence of several important objects relevant to the firefighting operations on both 2D and 360° videos. We also point out our next steps for the system to reliably and accurately perform in real time. The insight from this research will help the community to better understand the need for technology in firefighting as well as how existing technologies may be adapted to the emergency response domain.

VII. Acknowledgements

We would like to acknowledge Richard Kesler for his instrumental support, providing us with necessary domain expertise to develop our architecture. This work was funded by National Science Foundation (NSF) grants NSF IIS 2140645 and NSF IIS 2140620. All opinions and statements in the above publication are of the authors and do not represent NSF positions.

References

[1] 360 msa fire safety virtual reality. https://www.youtube.com/watch?v=VsC45ORxp8U, 2022. [Online; accessed 19-Feb-2022].

[2] Ifsi training files. https://www.fsi.illinois.edu/research/videos.cfm, 2022. [Online; accessed 29-Jan-2022].

[3] Michel Bellemans, D Lamrnens, J De Sloover, Tom De Vleeschauwer, Evarest Schoofs, W Jordens, B Van Steenhuyse, J Mangelschots, Shivam Selleri, Charles Hamesse, et al. Training firefighters in virtual reality. In *2020 International Conference on 3D Immersion (IC3D)*, pages 01–06. IEEE, 2020.

[4] Ali Diba, Mohsen Fayyaz, Vivek Sharma, Amir Hossein Karami, Mohammad Mahdi Arzani, Rahman Yousefzadeh, and Luc Van Gool. Temporal 3d convnets: New architecture and transfer learning for video classification. *arXiv preprint arXiv:1711.08200*, 2017.

[5] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[6] Ying Huang, Wenbo He, Klara Nahrstedt, and Whay C Lee. Requirements and system architecture design consideration for first responder systems. In *2007 IEEE Conference on Technologies for Homeland Security*, pages 39–44. IEEE, 2007.

[7] Ying Huang, Wenbo He, Klara Nahrstedt, and Whay C Lee. Corps: Event-driven mobility model for first responders in incident scene. In *MILCOM 2008-2008 IEEE Military Communications Conference*, pages 1–7. IEEE, 2008.

[8] Ying Huang, Wenbo He, Klara Nahrstedt, and Whay C Lee. Incident scene mobility analysis. In *2008 IEEE Conference on Technologies for Homeland Security*, pages 257–262. IEEE, 2008.

[9] Yanghao Li, Saining Xie, Xinlei Chen, Piotr Dollar, Kaiming He, and Ross Girshick. Benchmarking detection transfer learning with vision transformers. *arXiv preprint arXiv:2111.11429*, 2021.

[10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[12] Jun Ma, Jack CP Cheng, Feifeng Jiang, Weiwei Chen, Mingzhu Wang, and Chong Zhai. A bi-directional missing data imputation scheme based on lstm and transfer learning for building energy data. *Energy and Buildings*, 216:109941, 2020.

[13] Anh Nguyen, Zhisheng Yan, and Klara Nahrstedt. Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction. In *Multimedia Conference*, pages 1190–1198. ACM, 2018.

[14] Jounsup Park. Real-time object detection in 360-degree videos. In *Real-Time Image Processing and Deep Learning 2021*, volume 11736, page 117360C. International Society for Optics and Photonics, 2021.

[15] Andres Quan, Charles Herrmann, and Hamdy Soliman. Project vulture: A prototype for using drones in search and rescue operations. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 619–624. IEEE, 2019.

[16] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[17] Ruidong Zheng, Ruigang Yang, Keqiao Lu, and Shesheng Zhang. A search and rescue system for maritime personnel in disaster carried on unmanned aerial vehicle. In *2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, pages 43–47. IEEE, 2019.