

CitizenHelper-training: AI-infused System for Multimodal Analytics to assist Training Exercise Debriefs at Emergency Services

Rahul Pandey

George Mason University
rpandey4@gmu.edu

Brenda Bannan

George Mason University
bbannan@gmu.edu

Hemant Purohit

George Mason University
hpurohit@gmu.edu

ABSTRACT

The adoption of Artificial Intelligence (AI) technologies across various real-world applications for human performance augmentation demonstrates an unprecedented opportunity for emergency management. However, the current exploration of AI technologies such as computer vision and natural language processing is highly focused on emergency response and less investigated for the preparedness and mitigation phases. The training exercises for emergency services are critical to preparing responders to perform effectively in the real-world, providing a venue to leverage AI technologies. In this paper, we demonstrate an application of AI to address the challenges in augmenting the performance of instructors or trainers in such training exercises in real-time, with the explicit aim of reducing cognitive overload in extracting relevant knowledge from the voluminous multimodal data including video recordings and IoT sensor streams. We present an AI-infused system design for multimodal stream analytics and lessons from its use during a regional training exercise for active violence events.

Keywords

Training Exercise, Emergency Preparedness, AI system, Learning Analytics, Responder Training

INTRODUCTION

The recent years have shown the rapid and successful adoption of Artificial Intelligence (AI) techniques across different processes and services of the private and public sectors (Wirtz, Weyerer, and Geyer, 2019). Examples of such AI use-cases range from supporting human information processing in the customer services to real-time information and knowledge management in the organizational workflows. A key driver for the recent interest in such AI-infused information systems is the growing technological capabilities to collect a variety of data sources, including physical sensing streams such as IoT sensors and drones, social and web streams such as microblogs and news feeds, and so on. These data sources provide unprecedented access to big data during crises (Castillo, 2016), with redundant information in various modalities of data (e.g., text and videos). The significance of AI technologies is recognized for their ability to help transform such big multimodal data into actionable knowledge that could help diverse emergency management stakeholders.

There has been growing research interest in AI for emergency management, although the current focus on designing AI-infused systems is heavily biased toward machine learning and natural language processing for emergency response phase (e.g., AIDR system (Imran et al., 2014), CitizenHelper-adaptive system (Pandey and Purohit, 2018)). Traditionally, researchers have shown different applications of AI technologies for emergency management domain, such as multiagent systems for simulations (Dugdale et al., 2010), ontologies for intelligent knowledge management (De Nicola, Melchiori, and Villani, 2019), and semantic data processing (Abel et al.,

2012). There is a greater opportunity to exploit the AI technologies for all phases of emergency management, from mitigation, preparedness, response, to recovery.

In this research, we focus on the emergency preparedness and mitigation phases, investigating the potential application of the emerging AI technologies of computer vision and deep learning to address the challenges of training and not response operation, toward the less explored space of learning analytics. This research study complements the traditional focus of AI applications on simulation modeling for preparedness. We note that response training is an essential component of the emergency services, especially those that may involve interactions among different types of stakeholders. The collaborative work is an important part of the emergency management profession (Saoutal, Cahier, and Matta, 2014). An Emergency response operation is a complex environment that requires an incident response with multiple teams (e.g. fire and rescue teams, dispatch) to efficiently and effectively carry out the required activities for the response mission. Simulation training and emergency exercises provide a systematic experiential learning environment for the emergency management and response personnel, to progress toward more efficient task coordination as well as effective communication (Bannan et al., 2019). However, inefficient tools to assist the instructors in training simulation exercises and the trainees may result in a less effective learning experience (Buck, Trainor, and Aguirre, 2006; Feese et al., 2013).

The traditional practice of the training instructors includes direct observation and radio-based audio communication for collecting data on trainee behaviors and interactions and may not be as effective for large-scale simulation exercises, especially with several exercise participant teams. The loss of valuable information due to limited human resources for observations and human errors in the data collection limits the learning feedback to an individual trainee and team during the training exercise debrief due to a variety of data-centric limitations (Dubrow and Bannan, 2019). The incompleteness of the human-observed data can lead to the inefficient extraction of behavioral knowledge from observed data for learning. The recent advancement in the ability to sense and collect information through diverse multimodal data sources during the training exercises could address our challenges of data-centric limitations (Purohit, Dubrow, and Bannan, 2019). Especially, video streams through IP cameras, sensory information from wearable devices, as well as online microblog streams from social web platforms present an opportunity to capture complementary and redundant information streams for instructors during simulations and enhance debriefing sessions (Kranzfelder et al., 2011; Feese et al., 2013; Dubrow et al., 2017). Prior research (Sawyer and Deering, 2013) indicates that the time when most learning occurs during the training simulation exercises is the debriefing or after-action review session after a live simulation training event, where the instructor and trainee responders reflect and discuss on their behavioral activities. Thus, the capability of AI technologies to mine relevant behavioral events from the multimodal data streams and integrate information from such unconventional sources at large-scale is invaluable. It has the potential to enrich the information collected and analyzed for assisting the training instructors during debriefs (Bannan et al., 2019).

In this paper, we present a user-centered design of an AI-infused system for multimodal stream analytics to augment the learning experience of trainees and assist instructors during the training debriefs through visual presentation of relevant knowledge for learning feedback. We define training exercise as an event in which the emergency service personnel simulate a real-life emergency response incident. The data coming from multiple sensors deployed to capture exercise activities has varied modalities like videos, numbers, text, etc. that is referred as multimodal data streams. The relevant information extracted from different data streams can assist instructors in facilitating key pieces of information to provide enhanced trainee feedback through replaying targeted episodes that are visualized through redundant data sources (e.g. video segments from IP cameras and signals from IoT occupancy sensors, both presenting information of victims to be rescued in an area). Knowledge of multiple behavioral events (e.g., time of inflection points of the meeting of responders and wounded victims/patients) can be extracted from such data streams. The improved, integrated knowledge visualizations of the multimodal data streams collected during a training exercise provide an improved awareness for reflection and experiential learning for emergency service trainees during post-exercise debriefs.

In summary, the main contributions of this research are the following:

- We design and develop an AI-infused system for scalable, real-time processing of multimodal data streams during training exercises of emergency responders, to augment the learning experience of trainees during training debriefs immediately after the exercises.
- We demonstrate an application of the state-of-the-art AI techniques of computer vision for training exercises, through a case study of the developed system for training debriefs during a full-scale training exercise for an active violence event on a university campus.

In the remainder of this paper, we first present the related works in the literature, followed by a theoretical foundation to collect, process, and analyze multimodal data streams for an enhanced learning experience and instructor feedback capabilities during training debriefs. We then present the system architecture for multimodal stream processing, followed by a case study of employing the system during a training exercise and lessons learned.

RELATED WORK

We discuss three areas of research relevant to this study on designing the AI-infused system for improving training debriefs during emergency simulation exercises.

Emergency Services and Training Exercises.

Simulation training exercises are heavily utilized in the emergency management that ranges from tabletop discussions to high-fidelity in-situ simulations (Buck, Trainor, and Aguirre, 2006). These different forms of training are centered around specific types of team learning, coordination, and performance. Tabletop exercises (TTX) are based on meetings without any operational activity and focus on helping the emergency managers plan future response techniques such as damage assessment. Drill exercises focus on training for a single function while functional exercises are more focused on the interactions between individuals, teams, and agencies when responding to an event. Thus, for such training exercises, multimodal data collection of video and audio streams can be valuable to study that can provide greater knowledge of the state of interactions at any point during training. Lastly, the full-scale in-situ exercise type is a high-fidelity simulation that approximates the response to a real-life emergency incident (California Hospital Association). While the full-scale exercises are the most realistic, they also challenge training instructors to collect the most information from the full-scale exercise runs, due to a variety of environmental factors such as noise, vision, and heavy equipment. Therefore, redundant information streams captured by multimodal data collection methods are significant to collect, process, and analyze in real-time, however, a major challenge is to process such multimodal data streams to extract knowledge for relevant behaviors useful for the debrief sessions.

Multimodal Stream Analytics with Simulation and Training.

Prior research has demonstrated several applications of AI for stream analytics in different domains. For instance, a variety of sensing methods, video surveillance, and audio communication technologies for data collection and processing using cyber-physical systems are used in practice for the operations and training within the safety-critical applications in industries (Rajkumar et al., 2010). Similarly, the agent-based modeling, live-simulation sensing, and virtual systems are practiced in the military training exercises, while leveraging the video and sensor data streams for operational training purposes that require streaming data analytics (Buller et al., 2010). Likewise, in the aviation domain, extensive use of simulation and training leverage audio communication and sensor data streams to study team behavior and reflect on training processes. Similarly, open-source intelligence methods used for modern surveillance and monitoring systems have extensively collected, mined, and integrated information from open multimodal streams (Glassman and Kang, 2012). The above-mentioned illustrations of streaming data analytics across different domains motivate our research to design a system for assisting training instructors of emergency services during full-scale, live-simulation exercises, with an ability to collect, process and analyze multiple data streams. Extracting and integrating relevant information from these multimodal streams for behavioral events during a training exercise requires AI techniques for large-scale automated processing in real-time.

AI Systems for Disaster Management.

There exist several efforts in the last two decades to develop information systems to assist emergency management for real-time analytics, although heavily biased toward emergency response. While such systems for AI-infused stream analytics in real-time incorporate multiple data sources, they often focus on different modalities to study a specific type of disaster event for response (e.g., Twitcident (Abel et al., 2012), CrisisTracker (Rogstadius et al., 2013), AIDR (Imran et al., 2014), Twitris (Sheth et al., 2014), CitizenHelper-Adaptive (Pandey and Purohit, 2018)). Most of these systems rely on the prebuilt machine learning models for data analytics that are based on the datasets from prior disaster events to train and adapt for the current event with the incoming data, to provide a real-time analytical system. To address the scarcity challenge of labeled data for training and updating the current model, researchers have used transfer learning or domain adaptation methods, in addition to active learning methods that enable expert feedback to revise the prebuilt models (e.g., CitizenHelper-Adaptive (Pandey and Purohit, 2018)). However, majority of these systems have two limitations. First, these are primarily focused on

processing social and web data streams, and second, the focus on extracting relevant behaviors from the data streams is biased toward extracting intelligence to assist response operations, such as informative images with damage infrastructure information (Alam et al., 2018). In contrast, this research focuses on designing AI-infused system with machine learning models focused on real-time multimodal analytics for extracting relevant knowledge, which helps augment the learning experience and the debrief discussions of first responder training exercises.

THEORETICAL REASONING FOR ENABLING ENHANCED LEARNING AND TRAINING EXPERIENCE WITH MULTIMODAL ANALYTICS

This section describes our three theoretical motivations to create a foundation of the design of AI-infused system for multimodal analytics, which could help training instructors in providing enhanced learning experience to emergency service trainees.

First, prior research in learning sciences and cognitive psychology has shown that many professionals have difficulty reasoning about complex systems due to the unpredictability of multiple interacting elements that tax working memory (Funke, Fischer and Holt, 2018). Further, events at one level of the system (e.g., individual actions) can have unexpected consequences at another level (e.g., team and multiteam systems) (Wilkerson-Jerde and Wilensky, 2015). In the complex socio-technical system of emergency management training, therefore, our first design goal is to facilitate a systematic approach to *seamlessly collect* data of all states of training activity through multimodal sources and *efficiently process* data for relevant patterns (AI technologies have an important role to play in identifying patterns from large-scale streams) to assist the debriefing session.

Second, another important theoretical consideration is to understand why to conduct multimodal data analytics, at which stages, and for whom. Research has shown that learning through reflection on action in a simulated environment or live experience can enrich the source of objective information, which contributes to effective feedback from the situation (Jenvald and Morin, 2004). Debriefing sessions after a training exercise provides the opportunity for reflection on action. Although, discussion alone in the debrief may not be enough to enhance learning, but rather an ability to visualize targeted replay information can enrich reflection in providing relevant details immediately following the exercise. Furthermore, prior research (Ayers and Cierniak, 2012) has shown that integrated information is more effective for learning especially if the individual components of information are found to be less intelligible in isolation, or hard to interpret for any meaningful analysis. Instead, a strategic information integration approach that avoids the split attention effect can lead to a more intuitive processing of relevant information for enhanced learning (Chen, Woolcott, and Sweller, 2017). Thus, our system design goal is to save time in providing relevant knowledge for learning feedback to instructors for trainee debriefs, with appropriate redundancy from multiple data streams that can be captured in the integrated knowledge and visualized for enhanced feedback.

Lastly, prior research on the design expectations of the first responders suggests that the responders wish to gain an overall picture or mental model of the incident, which can be quickly inferred through information processing of various available data sources. For instance, in a fire incident exercise, it is valuable for a commander to gain information about the structure of the building, location, number of victims, and any hazardous materials on-site immediately (Xiaodong, et al., 2004). This research study from the field studies notes the key factors essential to address the technology design for relevant knowledge extraction for emergency managers as follows: a.) accountability of resources and personnel that are critical to the management and coordination of response activity, b.) assessment of the situation through multiple sources of information while avoiding information overload, and c.) resource allocation for incident commanders that is essential for situational awareness.

In summary, the desired system should aim to save time for the efficient knowledge extraction from multiple, redundant data streams for quick visualization of relevant behavioral events for debriefing feedback. This approach allows trainees and instructors to browse content from multiple streams on the same visual dashboard and learn. The systems should also reduce effort, by seamlessly collecting, analyzing, and frequently integrating knowledge from the common information streams and eliminating the need for instructors to check several dashboards for providing different feedback. Finally, the system should facilitate scalable monitoring of trainee learning by allowing more instructor roles to access the same dashboard views.

CITIZENHELPER-TRAINING SYSTEM DESIGN

We utilized a user-centered design approach (Wallach and Scholz, 2012) to ideate, refine, and execute the design of a multimodal stream analytics system, to support the training of emergency services. In the following, we first go through the multimodal data collection pipeline, followed by machine learning (AI) models for data processing

and finally, visualizing the results to gain meaningful knowledge out of the multimodal data.

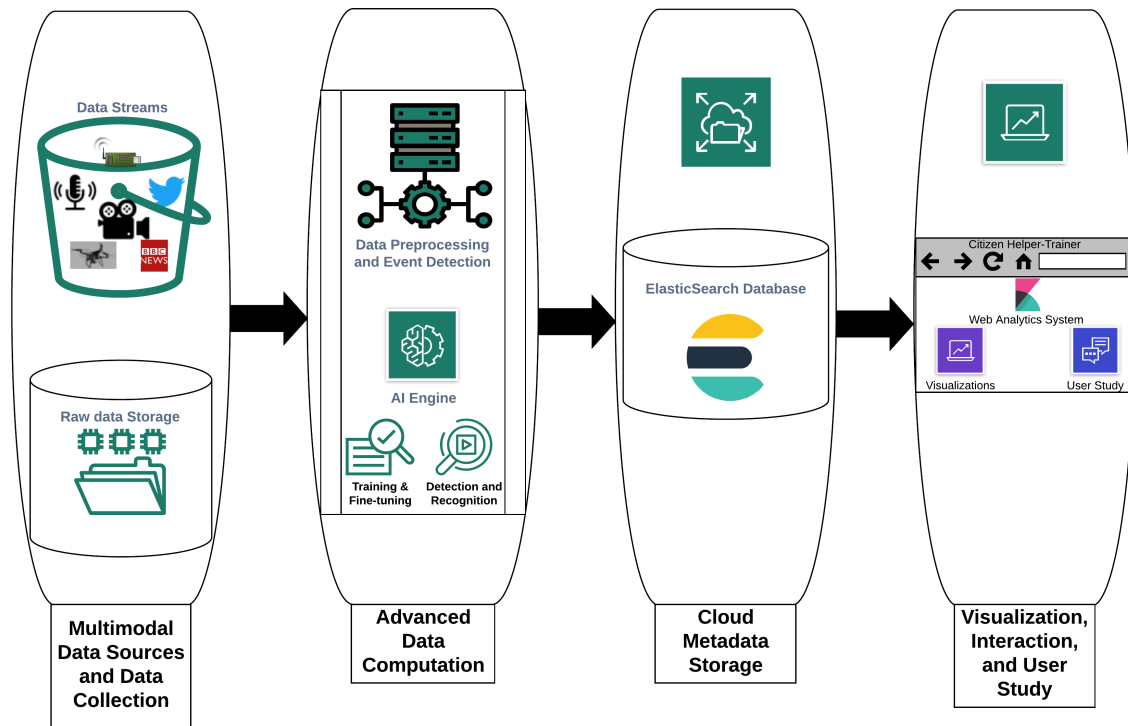


Figure 1: System architecture for multimodal stream analytics.

Our overall system has 4 major components as shown in figure 1:

- a.) Data Collection from multimodal data sources,
- b.) Advanced Data Computation,
- c.) Cloud Metadata Storage for storing the resulting pattern metadata from the previous component, and finally,
- d.) Visualization Dashboard, which helps to interact with the knowledge embedded in the metadata of the computed results from the multimodal data streams.

We describe each component in the following.

Multimodal Data Collection

This module of the CitizenHelper-training facilitates the collection and storage of multimodal data streams. During a training exercise, one can collect data from different sensing sources. These sensing technologies including wearables can be either static at certain locations such as IP cameras, occupancy sensors on walls, etc., or dynamic like drone devices, IoT sensors, etc. tied to a first responder or other exercise participants, such as patients or victims. During an exercise, we collect data from all the sensing devices at the same time and store it in a database for further processing in the Advanced Data Computation component.

Advanced Data Computation

In this component, all the collected semi-structured data from multimodal data sources are filtered to extract the knowledge and output semantically structured metadata. Different steps are taken to obtain knowledge from the raw data source. We explain each component that is present in this module with an example of a video data stream input for better understanding.

Data Preprocessing and Event Detection

The first step of the Advanced Data Computation includes data preprocessing. In this approach, we try to identify the events of our interests in the raw data that have been collected from different data sources. An event can be defined as the change of signals from the stable state of the collected sensory data. During an activity, there can be multiple possibilities of occurrence of events and each event may contain information that can be interpreted as knowledge by the targeted user (training instructor).

For instance, in the case of video streams coming from a fixed IP camera during a training exercise, a sudden increase in the intensity of the person objects present in the background could indicate the possibility of the beginning of an event. Although we would have data streams for the entire duration of the exercise since the time the IP camera was active, the real information can be present only when there is a fairly large disturbance in the normal influx of data (in this case, the sudden change in the content of video frames). Hence, our data preprocessing step includes the detection process for the times when an event has occurred and conveys that information to the next component for information filtering.

AI Engine

The next component of our Advanced Data Computation module is the machine learning algorithm that drives the knowledge from the preprocessed raw sensory data. We call this an AI Engine. There are two sub-components of the AI Engine: *Training & Fine-tuning* and *Detection & Recognition*. We will go through each sub-component to better explain the working of an AI Engine.

1. Training and Fine-tuning

To extract knowledge and insights from the event data, we need a robust machine learning model trained for the relevant knowledge patterns. But since we only have a limited amount of data from an ongoing or past training scenario, it is hard to create a robust machine learning model, as discussed earlier in the related work. Hence, we use the transfer learning and domain adaptation techniques to leverage the model already trained with a huge number of generic datasets (e.g., ImageNet (Deng et al., 2009) knowledge base with person objects). Finally, we fine-tune the pre-trained generic robust model on a small but relevant dataset (e.g., person objects of type firefighter for images/videos) of prior training scenarios to make the final trained model.

2. Detection and Recognition

Once we have the robust model from transfer learning with acceptable performance on the small dataset relevant to a new exercise scenario, we use that model to filter targeted information by detecting and recognizing semantically relevant objects from the input event data. Furthermore, once the first pass of information filtering is done, we further analyze the data to extract more in-depth information that can be useful to generate knowledge for the training exercise.

Cloud-based Metadata Storage

Once we have identified interesting patterns of relevant target objects out of the large event-centric data windows of sensory data streams, we store it in the cloud database, such as Elasticsearch database. The reason to store this information on the cloud is for faster processing for interactive visualization. Databases like Elasticsearch uses the indexing technique, which is very efficient in getting search results in real-time.

Visualization

Once the filtered information of relevant pattern metadata is stored in the cloud in the Elasticsearch database, we use the Kibana interface to visualize different time series patterns to gain knowledge. The reason to use Kibana was its easier integration with the Elasticsearch database (also, Kibana is provided by the creators of ElasticSearch), which gives benefits in handling different frontend interface changes and filters efficiently as the user browses the dashboard interface in real-time. We used a hierarchical visualization approach for the dashboard interface, in which as we go down on the interface, each visualization widget unpacks very specific information to share knowledge. In this way, a trainee or instructor can follow a guided explanation of the relevant knowledge one step at a time for discussion during debrief (example provided in the case study with Figure 3).

ILLUSTRATIVE CASE STUDY

We first describe the scenario of the simulation training and then, the system implementation and deployment.

Training Scenario

We staged an active violence incident (AVI) mass casualty live simulation exercise occurring in a large arena at a university campus in the mid-Atlantic region of the U.S. Local fire, EMS and law enforcement first responders participated the simulation-based multiteam training activity exercise to help better understand how these technology systems can be used to enhance public safety and response effectiveness in emergency situations. This scenario is part of continued research into smart building and smart first responder technology systems that can

help first responders save lives and improve public safety. This user experience design effort brought together leading innovators and researchers in artificial intelligence, learning science in the smart building/Internet of things/smart first responder space to address mission-critical issues identified by emergency services and law enforcement agencies.

Approximately 70 first responders, law enforcement personnel and volunteers participated in the simulated active shooter scenario. Technologies tested included sensors and displays designed to improve the operational and energy efficiency of the arena. Should an emergency occur these same videos and sensors can be available to help responders more rapidly determine the location and type of emergency, help find victims more quickly and, ultimately, save lives. The goals of the training simulation debrief after different scenario runs included the questions related to specific events of interest, such as what was the time between the response dispatch to the time of the neutralization of the shooter, time of interaction between the responders and victims or patients, etc. We collected a set of such relevant questions for debriefing discussion and in-depth analysis via our prior engagement with the regional first responder training exercise planners.

Implementing CitizenHelper-training Prototype

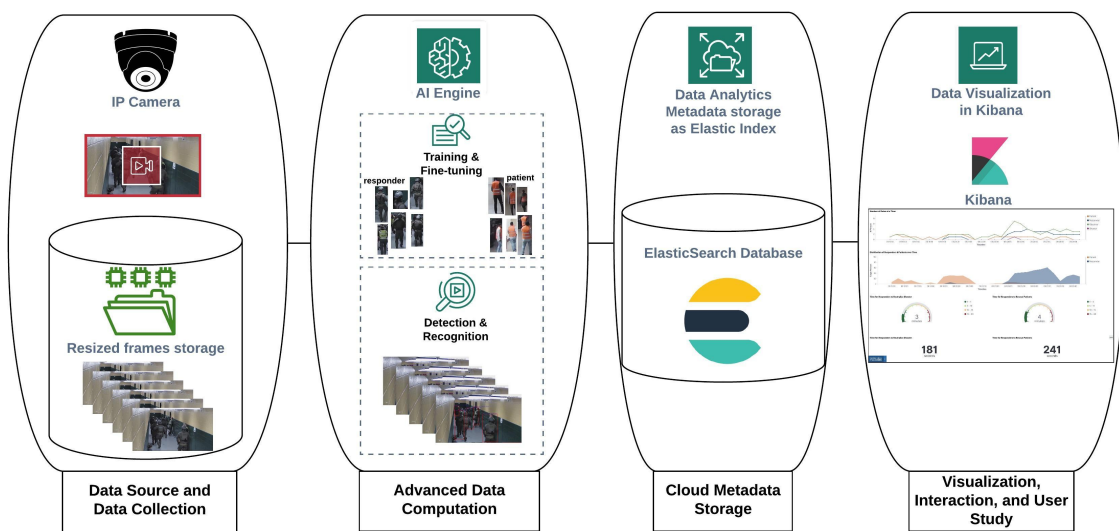


Figure 2: CitizenHelper training system prototype implemented for video data during a simulation of an active shooting emergency response exercise.

For this case study and demonstrating the role of AI technologies for multimodal analytics, we have specifically focused on IP camera sensors, which provided video streams to understand the behavior patterns of different category of person roles (actors) who participated in the training scenario. We describe the implementation of each component of the CitizenHelper-training system in the following. The specific workflow of our system implementation can be seen in Figure 2 and the final visualization dashboard can be seen in Figure 3.

Data Source

We set-up static IP cameras at different locations of the training site in the arena, which collected the video streams throughout the training simulation exercise runs. For each IP camera that has been set up, we analyze its captured video stream based on the proposed system architecture.

Data Collection and storage

Since the data from the IP camera are being generated continuously, our first task is to identify the start of an event. Since this training exercise was a planned scenario, we fixed the start and end time of the exercise and collected the video streams for only that duration. The frame rate of our installed camera was around 15 frames per second, but for simplicity, we only took one frame per second to capture and extract the knowledge with minimal information loss. We also resize the frame images to 224x224, since that is the required size for our AI engine (machine learning model) and in this way, we optimize the space complexity. We store those images on the server for further processing.

Advanced Data Computation

Our primary task of analytics for this exercise was to identify different actors who participated in the emergency response scenario and annotate them to one of the categories based on their role: Patient, Responder, Observer, or Others. This automated annotation produced semantic metadata of relevant training actors over time, which enables the visualization for relevant knowledge for discussion during debriefs. To do this task, we have performed people detection using a pre-trained model followed by training and fine-tuning recognition model as explained below.

1. People (Actor) Detection

Our biggest challenge with training a people detection module was the lack of data. Hence, as discussed in the system design, we used a transfer learning technique to conduct model training and fine-tuning. We utilized pretrained FasterRCNN (Region-based Convolutional Neural Network) (Ren et al., 2015) to detect people or actors in the training exercise. The adapted FasterRCNN model has a deep learning-based architecture of an Inception V2 model (Szegedy et al., 2016) and trained with the MS Coco dataset (Lin et al., 2014). Coco is an extensive general image corpus that has over 200k images with 1.5 million object instances over 80 image categories. "Person" is also one of the categories of the Coco dataset. This is the key reason for choosing the model, as it can represent a generic structure of the people's images very effectively. Hence, with the trained FasterRCNN model, we detect different actors from the video streams during the time of an event. We store this information to recognize and tag different actors next.

2. Actor Recognition and Tagging

This step classifies different actors into different categories based on their role. This was a hard problem as in video frames (images), there are very few differentiable cues in the detected actors to generate a robust recognition model. Hence, we decided to fine-tune a pre-trained image recognition model. For this implementation, we have fine-tuned a pretrained MobileNet (Howard et al., 2017) to detect different types of actors from patients to victims based on their poses. MobileNet is the simpler deep neural network models that are much faster than the traditional big deep models, but without losing much efficiency. The reason we have used a smaller and much faster model to classify different actors was to make the overall AI engine near real-time. Also, since we were testing our implementation in a controlled environment, we believe that even a smaller model can achieve better results. Also, there is less chance for the presence of an actor that can be an anomaly in the data. Hence, to fine-tune the data, first, we have annotated the actors detected by the FasterRCNN model on video data from the past training exercises to any of the categories of different actors. We use this weakly-labeled dataset to fine-tune the MobileNet model. We first make all layers except the last layers untrainable so that it can retain the trained weights to capture the generic information of the images (actors in this case). We make the last layer before SoftMax trainable, which will learn the differentiable patterns for different actors in the training set. Once we have the fine-tuned model ready, we use that model to recognize every actor in the current exercise. The actor recognition in a video stream gives more specific information in the exercise for more in-depth knowledge of an event, and thus, contribute significantly to the debrief discussions, as we learned from the first responders during the training exercise.

Visualization, Interaction, and User Study

We used the combination of Elasticsearch (ES) and Kibana for storing the metadata as well as creating the dashboard interface for visualization of relevant patterns, with an interactive interface for the targeted users. Hence, after storing the information provided by the AI engine in the ES database, we created different visualizations to extract knowledge from the data for training. Each visualization is in the order of providing generic to in-depth information for learning and debriefs discussions, as shown in Figure 3. We describe each component of our visualization below.

1. *Number of roles at a time*: This line chart describes the intensity of actors of different roles at each interval of time who were present near the specific location of a given IP camera. This gives us the idea about the time in which there is a burst in actor occupancy that may signal a possible event of interest for the responder teams. This type of visualization helps the trainers when and which camera to look at during an exercise.
2. *Distribution of Responders and Patients over time*: This area chart informs us about the number of responders versus patients that were present in the vicinity area of the camera. This helps the trainer to compare at what place there was a necessity of responders when the patients were more. Moreover, it also warns about the places where there was an unusually large number of responders present than the required based on the number of patients to treat.

3. *Time for Responder to Neutralize Shooter and Time for Responders to Rescue Patients:* This visualization conveys more specific information based on the analysis of the video streams of the entire event. It is very important information for the trainees and instructor, as time is a key factor to consider during this kind of active shooting response event. Also, it is actually very hard to infer this information by just looking at the video or any other multimodal data stream. Hence, we have used two different types of visualization: Gauge Meter, which is easy to visualize and total seconds, which can help in more in-depth analysis.

Based on this type of visualization interface, trainees can better learn the insights explained by the training instructor that would not be impossible in a short amount of time if they try to interpret everything just by looking at the video or other multimodal data streams during the debrief session. The visual interface can also help the trainees to not get distracted and overloaded with other information that is contained in the video streams but is not necessary during an event for learning.

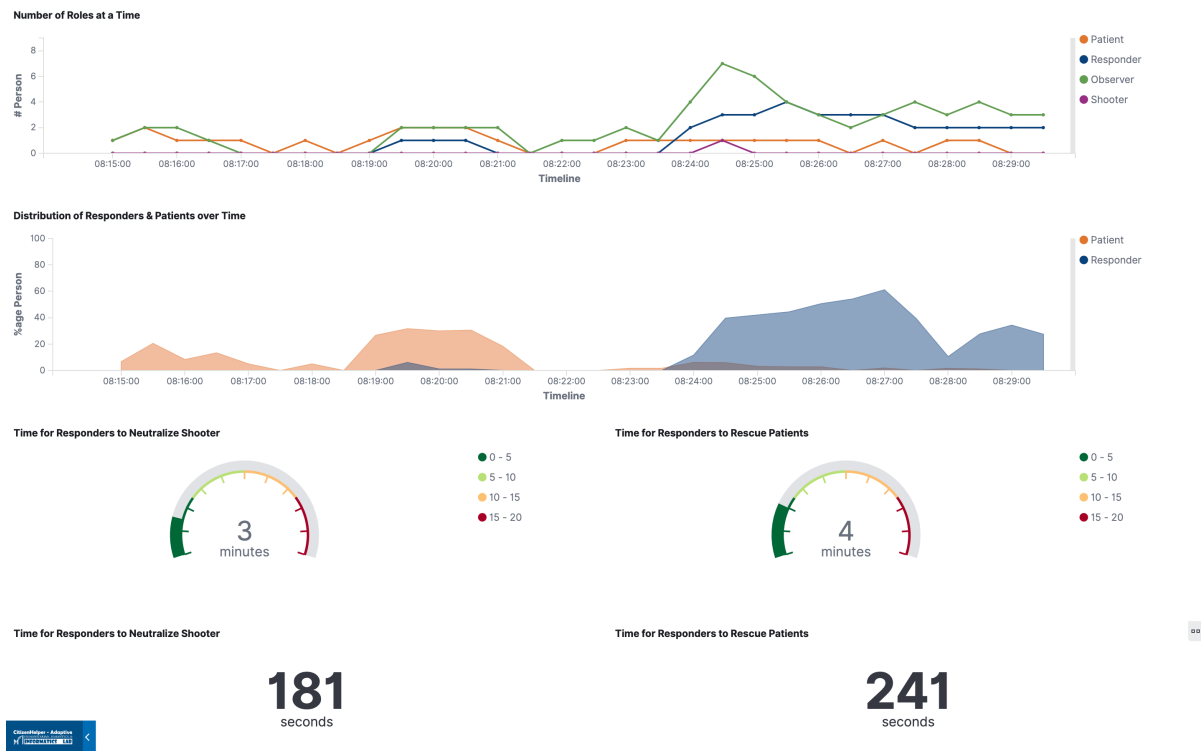


Figure 3. System dashboard from a simulation exercise run. From top, the first widget shows the time series data of the occurrence of all identified person roles from the AI engine, followed by the second widget with temporal distribution of specific roles of interest for debriefs discussion in the vicinity of a camera location, and similarly, the widgets below showing the metrics of interest for debriefing. All the widgets are interactive and allow an instructor to filter/focus a specific period on the timeline.

DISCUSSION - LESSONS LEARNED AND LIMITATIONS

This research has demonstrated a feasibility study on the role of AI technologies (particularly, computer vision and deep learning) in enhancing the real-time data analytics capability of instructors to support post-training debriefs of emergency response training. While the application of the proposed AI-infused system was demonstrated for specific AI technologies, the system design is generalizable for data processing with other types of AI techniques, such as natural language processing to process audio communication of the exercise participants to provide more insights for enriching the debrief discussions.

Similarly, the application of these AI technologies could be explored using our AI-infused system for different types of emergency scenarios, such as fire evacuation, or search and rescue during natural hazard events. Additionally, while we focused on emergency response training, the proposed design of the AI-infused system can be adapted and enhanced toward other types of training and learning environments, such as sports activities and performing arts on university campuses.

We also note the factors for performance limitations from the implementation of our system. First, we limited ourselves for manual error analysis-based evaluation of the AI techniques in the advanced data computation stage

of our case study and aim to address in an extended future study. Also, the performance of AI techniques can be severely limited by the scarce labeled data in the new exercise scenario, for fine-tuning the models for relevant pattern recognition to extract knowledge from raw data. Thus, future work could explore ways to identify and incorporate different related sources of data to enable better learning of relevant patterns through efficient domain adaptation. Second, the external environment of the training exercise could also cause false-positive errors for the prediction model in the pattern detection and recognition stage. For instance, in the initial testing of the object recognition module, we noticed that the algorithm picked the person in the image on the wall in the background at the scene of the simulation exercise. Thus, better awareness and remedy of such errors from the external environment should also be considered in the future work direction. Lastly, the simulation training exercise, especially the full-scale version, is a complex socio-technical environment with multiple actors, resources, and user interactions. Thus, the resolution of patterns for the relevant behaviors for visual analysis might vary from an instructor to an incident commander to a trainee of a specific team, and so, future work should consider personalization of interfaces for different stakeholders of the emergency response training.

CONCLUSION

This paper presented a theoretically motivated design of AI-infused system to augment learning experience of emergency responders in the simulation training exercises, by supporting instructors with multimodal learning analytics for enhanced training debriefs. We demonstrated an application of AI methods of computer vision and machine learning to process multimodal data streams in real-time, collected by diverse sensing technologies during training exercises. We presented a case study of employing the developed system during a regional emergency response exercise. The lessons learned from this initial prototyping inform the potential of leveraging AI technologies for emergency preparedness and improving the experiential learning of trainees during training debriefs, with an ability to conduct instant replays of relevant behavioral events extracted with the help AI methods.

ACKNOWLEDGMENTS

Authors thank Gaurav Bahl, Yasas Senarath, and Ganesh Nalluru for helping in the system implementation and U.S. National Science Foundation award IIS-1815459 for partial research support.

REFERENCES

- Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., and Tao, K. (2012) Twitcident. In: Proceedings of the 21st International conference companion on World Wide Web - WWW '12 Companion.
- Alam, F., Ofli, F., and Imran, M. (2018). Crisismmd: Multimodal twitter datasets from natural disasters. In: Twelfth International AAAI Conference on Web and Social Media.
- Ayres, P. and Cieffarniak, G. Split-Attention Effect., In: N. M. Seel (ed.), Encyclopedia of the Sciences of Learning, Springer US, Boston, MA, pp. 3172-3175 (2012).
- Bannan, B., Dubrow, S., Dobbins, C., Zaccaro, S., Purohit, H., and Rana, M. (2019). Toward wearable devices for multiteam systems learning. In: Perspectives on Wearable Enhanced Learning (WELL) (pp. 79-95). Springer, Cham.
- Buck, D. A., Trainor, J. E., and Aguirre, B. E. (2006). A critical evaluation of the incident command system and NIMS. In: Journal of Homeland Security and Emergency Management 3.3.
- Buller, M., Welles, A., Jenkins, O. C., and Hoyt, R. (2010). Extreme health sensing: The challenges, technologies, and strategies for active health sustainment of military personnel during training and combat missions. In: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX. Vol. 7666. International Society for Optics and Photonics, p. 766610.
- California Hospital Association. Emergency Preparedness: Types of Exercises [Website]. Retrieved Dec 14, 2019, from <https://www.calhospitalprepare.org/post/types-exercises>
- Castillo, C. (2016) Big crisis data - social media in disasters and time-critical situations, In: Cambridge University Press.
- Chen, O., Woolcott, G., and Sweller, J. (2017). Using cognitive load theory to structure computer-based learning including MOOCs. In: Journal of Computer Assisted Learning 33.4, pp. 293–305.
- De Nicola, A., Melchiori, M., and Villani, M. L. (2019). Creative design of emergency management scenarios driven by semantics: An application to smart cities. In: Information Systems, 81, 21-48.

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database". In: 2009 IEEE conference on computer vision and pattern recognition. Ieee, pp. 248–255.
- Dubrow, S., and Bannan, B. (2019, July). Toward Improving Situation Awareness and Team Coordination in Emergency Response with Sensor and Video Data Streams. In: International Conference on Human-Computer Interaction (pp. 257-269). Springer, Cham.
- Dubrow, S., Dobbins, C., Bannan, B., Zaccaro, S., Peixoto, N., Purohit, H., Rana, M., and Au, M. (2017). Using IoT sensors to enhance simulation and training in multiteam systems. In: Proceedings of the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Published, pp. 1–10.
- Dugdale, J., Bellamine-Ben Saoud, N., Pavard, B., and Pallamin, N. (2010). In: Simulation and emergency management (Vol. 10). Chapter.
- Feese, S., Arnrich, B., Troster, G., Burtscher, M., Meyer, B., and Jonas, K. (2013). CoenoFire: monitoring performance indicators of firefighters in real-world missions using smartphones. In: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. ACM, pp. 83–92.
- Funke, J., Fischer, A., and Holt, D. V. (2018). Competencies for complexity: problem solving in the twenty-first century. In: *Assessment and teaching of 21st century skills* (pp. 41-53). Springer, Cham.
- Glassman, M. and Kang, M. J. (2012). "Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT)". In: Computers in Human Behavior 28.2, pp. 673–682.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). "Mobilenets: Efficient convolutional neural networks for mobile vision applications". In: arXiv preprint arXiv:1704.04861.
- Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S. (2014) AIDR: Artificial intelligence for disaster response. In: Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion.
- Jenvald, J. and Morin, M. (2004). Simulation-supported live training for emergency response in hazardous environments. In: Simulation and Gaming 35.3, pp. 363–377.
- Jiang, X., Hong, J. I., Takayama, L. A., and Landay, J. A. (2004). Ubiquitous computing for firefighters: Field studies and prototypes of large displays for incident command. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp. 679–686.
- Kranzfelder, M., Schneider, A., Gillen, S., and Feussner, H. (2011). New technologies for information retrieval to achieve situational awareness and higher patient safety in the surgical operating room: the MRI institutional approach and review of the literature. In: Surgical endoscopy 25.3, pp. 696–705.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, pp. 740–755.
- Pandey, R., and Purohit, H. (2018, August). CitizenHelper-Adaptive: Expert-Augmented Streaming Analytics System for Emergency Services and Humanitarian Organizations. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 630-633). IEEE.
- Purohit, H., Dubrow, S., and Bannan, B. (2019, July). Designing a Multimodal Analytics System to Improve Emergency Response Training. In: International Conference on Human-Computer Interaction (pp. 89-100). Springer, Cham.
- Rajkumar, R., Lee, I., Sha, L., and Stankovic, J. (2010). Cyber-physical systems: the next computing revolution. In: Design Automation Conference. IEEE, pp. 731–736.
- Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp. 91–99.
- Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., and Laredo, J. A. (2013) CrisisTracker: Crowdsourced social media curation for disaster awareness. In: IBM Journal of Research and Development, 57, 5.
- Saoutal, A., Cahier, J. P., and Matta, N. (2014, May). Modelling the communication between emergency actors in crisis management. In: 2014 International Conference on Collaboration Technologies and Systems (CTS) (pp. 545-552). IEEE.
- Sawyer, T. L., and Deering, S. (2013). Adaptation of the US Army’s after-action review for simulation debriefing in healthcare. In: Simulation in Healthcare, 8(6), 388-397.

- Sheth, A., Jadhav, A., Kapanipathi, P., Lu, C., Purohit, H., Smith, G. A., and Wang, W. (2014) Twitris: A system for collective social intelligence. In: Encyclopedia of social network analysis and mining, 2240–2253.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.
- Wallach, D. and Scholz, S. C. (2012). User-centered design: why and how to put users first in software development. In: Software for people. Springer, pp. 11–38.
- Wilkerson-Jerde, M. H., and Wilensky, U. J. (2015). Patterns, probabilities, and people: making sense of quantitative change in complex systems. In: Journal of the Learning Sciences, 24(2), 204-251.
- Wirtz, B. W., Weyerer, J. C., and Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. In: International Journal of Public Administration, 42(7), 596-615.
- Yoon, S. A. and Hmelo-Silver, C. (2017). Introduction to special issue: Models and tools for systems learning and instruction. In: Instructional Science, 45(1), 1-4