HIVE: A Novel Algorithmic Framework for Standoff Concealed Threat Detection

Funded by the Department of Homeland Security Science and Technology Directorate (DHS-S&T)

Developed by MIT Lincoln Laboratory

This material is based upon work supported by the Department of Homeland Security under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Homeland Security.

Executive Summary

The real-time detection of concealed threats is critical for protecting public transportation, sports arenas, and other open, difficult-to-secure environments. New sensing technologies, such as standoff active-RF imaging, can help security personnel screen customers and bags quickly without affecting the flow of traffic. However, accurately detecting threats in the complex environment of crowds carrying everyday items remains a challenge.

The HIVE algorithmic framework enables new approaches to protecting people and infrastructure in areas where traditional security checkpoints are not feasible. HIVE (Hierarchical Inference for Volumetric Estimation) is a custom deep convolutional neural network architecture that interprets volumetric video generated by a standoff, active-RF imagers. The architecture performs multi-resolution detection, classification, and segmentation of objects in the scene at various scales in order to produce automated threat detections, alerts, and visualization products – all without requiring the person to stop, pose, or remove their belongings.

Enhanced Detection: While state-of-the-art computer vision algorithms have shown outstanding performance in object detection in photographic images, these techniques cannot be directly applied to RF imagery since the underlying characteristics, appearance, and statistics are very different. HIVE was specifically engineered to use all the information contained in a complex-valued, three-dimensional RF image volume to better discriminate objects based on their unique characteristics in this modality. Detection with HIVE, using the full 3D volume, is significantly more accurate than processing 2D imagery with COTS algorithms.

Enhanced Intelligence: HIVE can be trained to detect specific threat items (such as explosives or large amounts of certain materials), while ignoring common benign items (such as laptop computers). Additionally, multiple detectors can be run as an ensemble, to allow the user to build custom rules for a specific deployment scenario – since security needs in a sports arena may be different than a transportation setting. In contrast, existing commercial systems (such as metal detectors or millimeter-wave portal scanners) generally output a binary "threat/no threat" alert and do not have much flexibility to adapt to new deployment scenarios.

Privacy Preserving: A key benefit of algorithmic screening using HIVE is privacy protection for people imaged by the sensor, as a human security operator does not have to view the RF imagery. For instance, a data product may show a localized alert overlaid on top of color video, never revealing RF imagery of the person's body.

Finally, HIVE is agnostic to the source sensor and can be paired with other active-RF systems in the future to add new capabilities and improved detection performance as commercial imaging technology advances.

Figures

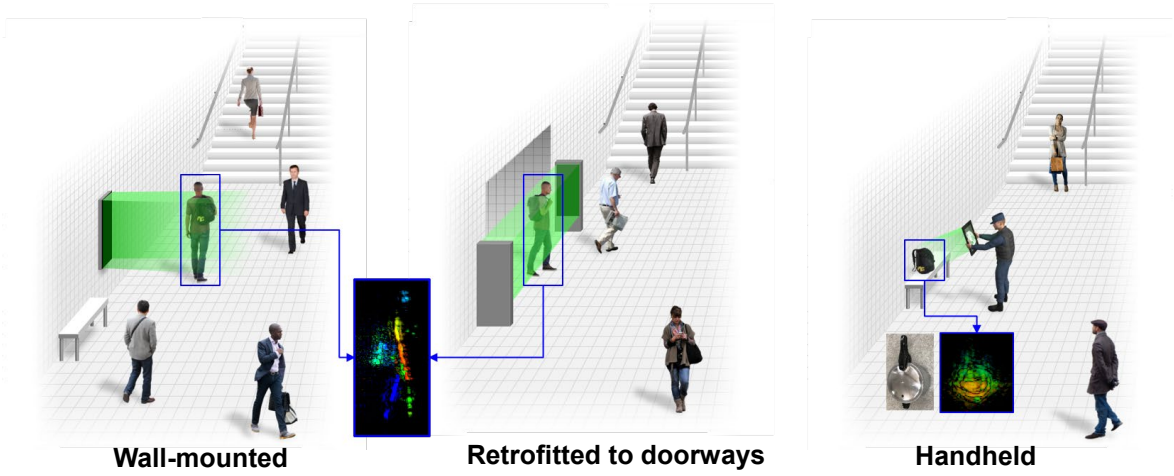

Figure 1. Notional configurations of standoff microwave imaging sensors with automated threat detection in a transit station.

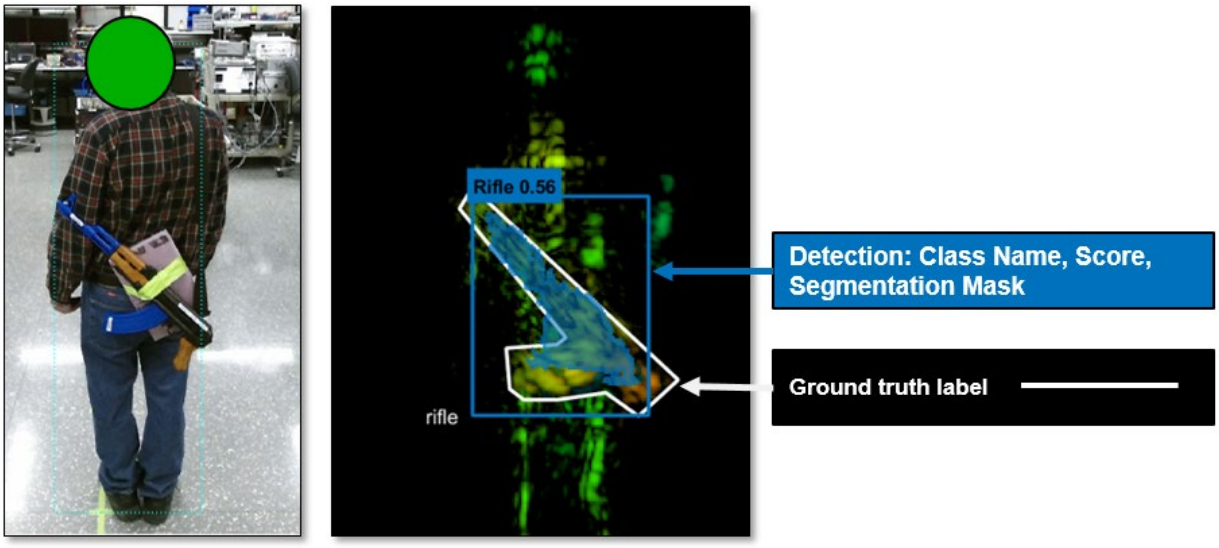

Figure 2. Example output from HIVE analytics that highlights detection of carried threat item (rifle mock-up). Left: corresponding red-green-blue image from co-located color camera. Right: Detected bounding box, segmentation mask, class name and score overlaid over 2D maximum-intensity projection of millimeter-wave image cube. The ground truth label is shown in white, identifying the location of the item in the scene.

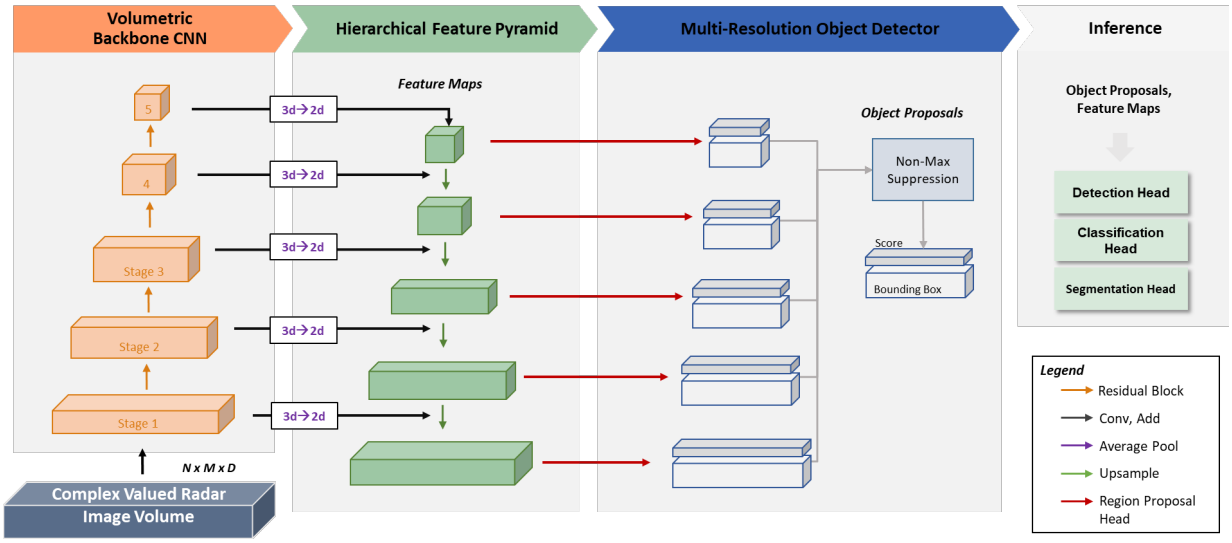

Figure 3. Overview of the HIVE detection engine. The first functional module performs feature extraction on dual-channel/complex-valued, three-dimensional millimeter-wave image data. A novel dimensionality reduction step converts 3D tensors to 2D tensors so that later processing can operate under reduced computational load. Feature maps are enriched using a feature pyramid network before proceeding to object detection, classification, and segmentation steps.

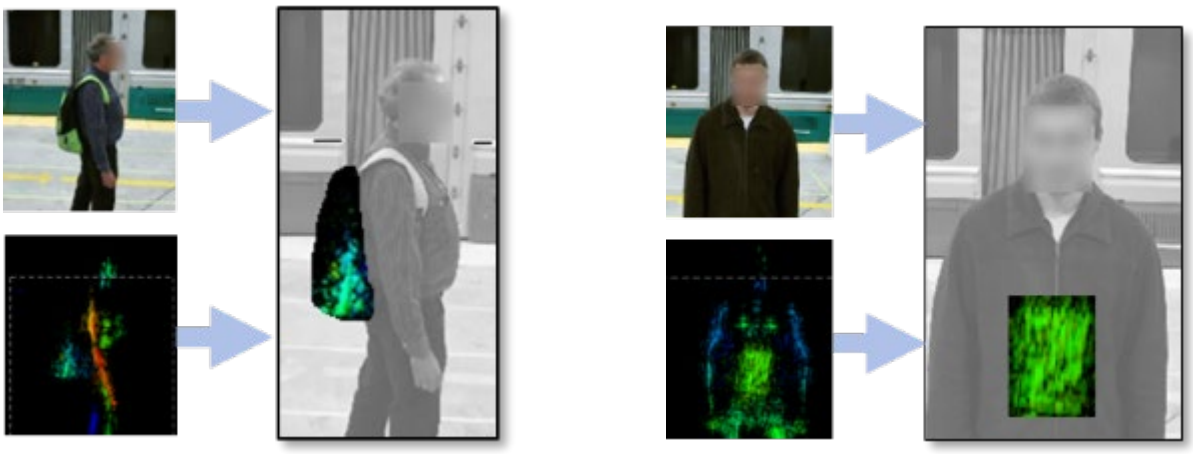

Figure 4: Example uses of fusion between RF and color video for operational setting. Left: information from the color camera and radar are aligned and fused to show the radar imagery only within the boundaries of backpack. Right: only portions of the radar imagery that correspond to a concealed anomaly are shown. The right-hand image is what a security operator might be shown in order to protect passenger privacy and anonymity.

Overview

HIVE detection analytics are designed for automated detection of concealed threats in crowded environments. While many of these settings (transit systems, shopping centers, concert venues) employ security personnel, the volume of traffic makes it impossible to manually screen every person/bag entering. When body scanners or other microwave imagers are used, even if enough staff personnel were available to review imagery, the signatures of threat objects or materials are often not visual in nature, and cannot be detected by visual inspection alone.

The HIVE (Hierarchical Inference for Volumetric Estimation) detection algorithm automates information extraction from video-rate millimeter-wave (MMW), or other RF frequency, imagery that show the reflection of microwave radiation on objects in a scene, and the analytics system uses this intelligence for decision support in the form of automated threat detections, alerts, and visualization products. The architecture leverages the hierarchical nature of deep convolutional neural networks to learn the best representation of the three-dimensional RF imagery and perform multi-resolution detection, classification, and segmentation of objects in the scene at various scales.

An additional key benefit of algorithmic screening is privacy protection for civilians imaged by the sensor, as a human security operator does not have to view the video imagery. For instance, a data product may be created to show the contents of a person's bag or on-body manmade item without revealing imagery of the person's body.

Operation

The HIVE architecture is composed of several functional modules, illustrated in Figure 3. HIVE is inspired by state-of-the-art object detection algorithms such as Mask R-CNN. Unlike Mask R-CNN and other common off-the-shelf deep convolutional neural networks, however, HIVE operates on complex-valued, three-dimensional (3D) input data. The input to a HIVE model can be a **single or dual-channel** RF image volume. Complex valued data can be processed by passing the real and imaginary components as separate channels. Likewise, magnitude and phase components can be ingested in this manner.

HIVE is a hybrid processing architecture in which the early processing is done on 3D imagery, but later processing (and output) is reduced to two dimensions to reduce processing load, leverage common training techniques, and allow for easier visualization. A 3D Convolutional Neural Network operates on the original volume to extract both low- and high-level semantic features maps (e.g., small edges as well as larger structures) at multiple spatial resolutions (i.e., for small and large objects in the scene). Then, an intermediate processing step converts the collection of 3D feature maps to a more common convention of 2D feature maps.

For a given image frame, HIVE outputs the item label, bounding box, and pixel-wise segmentation map for any detected threats in the scene. When operating on video footage (prerecorded or real-time), the developed analytics take advantage of information collected over a short time window in order to improve quality of automated detections while driving down false alarm rates. To accomplish this, the HIVE model is applied to each image frame independently. Those detections are accumulated over a recent time window, and an alert is generated if the number of recent/repeated detections exceeds a certain threshold. This threshold can be set or adjusted by the end user based on system requirements.

The analytic outputs from HIVE may be used to generate additional visualizations and data products to be reviewed by a human operator. By pairing a collocated red-green-blue (RGB) video or RGB-Depth sensor with the microwave imager, it is possible to extract information about people and objects in the scene using computer vision techniques. The color imagery from the RGB source and the volumetric RF data may be combined to produce fused visualizations which provide information to security personnel while preserving privacy.

These fused data products are generated as follows:

1. Segmentation masks of people, bags, and other classes are generated from the RGB color imagery.
2. Faces are redacted or blurred using an off-the-shelf facial detection algorithm.
3. The RGB imagery and RF imagery (and any derived segmentation masks or bounding boxes) are aligned using a registration process.
4. The segmentation masks from the HIVE detection model are used to determine which pixels from the RF image to show, and which pixels to mask using the RGB color image.

Comparison to commercial products

Traditional approaches from signal processing, computer vision, and machine learning have limited success in the security screening application domain. For instance, using hand-crafted features paired with classifiers (versus learning features using a convolutional neural network) to perform object detection provides limited performance, and given the complexity of the data, these traditional methods cannot scale well and provide the performance needed in operational environments. State-of-the-art computer vision algorithms have shown outstanding performance in object detection, classification, and segmentation in photographic images. However, these techniques cannot be directly applied to RF imagery since the underlying characteristics, appearance, and statistics are very different.

The HIVE architecture differs from other state-of-the-art architectures in the following ways:

1. First use of regional convolutional neural networks for the processing of active RF imagery.
2. Use of object detection methods for other purposes, such as material discrimination.
3. Feature extraction is performed using both the real and imaginary components of the complex-valued RF image volume.
4. Hybrid processing regime enables high-fidelity feature extraction within three-dimensional image volumes while reducing overall computational load by converting to two-dimensional processing for final detection, segmentation, and classification.

Furthermore, existing commercial systems that perform concealed threat detection (such as metal detectors or millimeter-wave portal scanners) generally output a binary "threat/no threat" alert and do not have much flexibility to adapt to new deployment scenarios. It may be possible to adjust sensitivity or detection thresholds, but these systems lack the ability to dynamically change – for instance, to ignore certain items or otherwise change the logic of how the system determines risk.

Together, the imager and algorithm system support processing active RF imagery and collocated RGB color imagery to extract a variety of information, such as detection of certain materials, objects, or anomalies in the scene. Because HIVE models can be trained for different purposes, several can be run in parallel to extract complementary information (e.g., object material or size). Depending on the envisioned deployment scenario, the component information derived from the RF and RGB imagery may be combined in unique ways to determine whether a person is carrying a concealed item and whether that item is considered a threat. For instance, security personnel responsible for screening fans entering a sports stadium may have different requirements or prohibited items than that of a transportation hub or museum. Having the ability to configure the screening system for different purposes would enable a dynamic security posture that can adapt to new threats in the future. The unique configuration may be set by the system developer or end-user, or may be learned via machine learning.

Finally, HIVE is agnostic to the source sensor and can be paired with other active RF systems to add new capabilities and improved detection performance as commercial imaging technology advances.

Table 1: Comparison to related technologies

Product Feature	MIT LL I2S imager + HIVE detection analytics	Body-scanning portals	Passive IR/MMW imagers	Commercial/ off-the-shelf 2D detection algorithms	Competitive Advantage
Unobtrusively screening people while in motion	Yes	No	Yes	N/A	The RF sensor has been developed for the purpose of unobtrusively screening people while in motion, at a distance and without requiring them to stop, pose or remove belongings.
Video-rate image processing	Yes	No	Yes	Yes	Video-rate image reconstruction and processing is critical for screening high foot-traffic environments, and allows multiple frames of a subject to be captured for improved threat detection.
3D image-processing	Yes	Yes	No	No	3D imaging allows for clearer identification of items within images, as well as object size estimation. Thus, the processing of 3D features in the RF data is critical for image exploitation.
Hybrid machine learning architecture for 3D feature extraction and 2D processing	Yes	N/A	N/A	No	Hybrid processing enables high-fidelity feature extraction from 3D imagery while reducing overall computational load by converting to two-dimensional processing for final detection, segmentation, and classification
Screening and object detection in bags	Yes	Maybe	No	Maybe	Given that many threat items are concealed inside luggage, the ability to screen inside bags, backpacks and luggage, and underneath heavy clothing like jackets, is critical. Commercial detection algorithms may be unable to detect concealed items if object and material RF signatures are flattened in 2D.
Automated threat detection	Yes	Yes	Maybe	Maybe	Machine learning automation saves time and resources for staff personnel as they do not need to manually assess every image output, and is more reliable than

					human operators in detecting RF material signatures.
Non-invasive imaging and privacy protection	Yes	No	No	Maybe	Using multiple data streams (RF and RGB imagery) with a flexible detection architecture allows the contents of a person's bag or on-body item to be displayed to a security operator without revealing imagery of the person's body, affording privacy protection to civilians who are being imaged.
Flexible system configuration and interpretable outputs for varied deployment scenarios	Yes	No	No	Maybe	Deep learning approach offers more than a binary "threat/no threat" alert, with the ability to provide security operator with other key information such as the type of object, size and material of the threat detected. Several of these information modalities can be layered together to develop flexible system alert configurations based on specific deployment scenarios (*e.g.*, "only cue on metal objects greater than a certain size").

Legend: Green is "Yes." Red is "No." Purple is "N/A." Orange is "Maybe."

Limitations

The overall system, comprised of an RF imager and HIVE, has performed well during in-lab and field testing. For the HIVE detection algorithm to be supported, the system must run on at least one Graphics Processing Unit (GPU). Labeled data is required to train a model for new use cases, which requires some amount of human effort upfront to collect and annotate radar imagery. Data sufficiency will vary depending on the complexity of the detection problem and the model may need to be retrained or fine-tuned iteratively to accommodate new specifications. The use of multiple (or more powerful) GPUs can expedite and streamline the training process.

Summary

The novel HIVE framework operates on RF imagery to automatically screen for concealed threat items in crowded places. The system of algorithms enables new approaches to protecting people and infrastructure in areas where traditional security checkpoints are not feasible.

The main advantages of this technology include the following:

1. The HIVE architecture is a novel combination of deep neural network components, customized to process complex-valued volumetric data.
2. The algorithms operate on data from a new class of standoff, active RF sensors that have been developed to unobtrusively screening people while in motion and at a distance, without requiring them to stop, pose, or remove belongings.
3. The system goes beyond a binary “threat/no threat” output, providing insight into why the threat was detected.
4. The system may be used on multiple video-rate sensor configurations.
5. Fusing video and radar data streams can preserve privacy of people being screened.
6. The system is compatible with multiple sensor sources and can be paired with other/commercial RF imagers for ease of integration and/or improved detection performance.