IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 October 2024, accepted 9 January 2025, date of publication 14 January 2025, date of current version 29 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3529484

APPLIED RESEARCH

Bringing Intelligence to SAR Missions: A Comprehensive Dataset and Evaluation of YOLO for Human Detection in TIR Images

MOSTAFA RIZK[1,2] **AND ISRAA BAYAD**[2]
[1]School of Engineering, Lebanese International University, Beirut 146404, Lebanon
[2]Faculty of Sciences, Lebanese University, Beirut 90656, Lebanon

Corresponding author: Mostafa Rizk (mostafa.rizk@liu.edu.lb)

ABSTRACT Effective search and rescue (SAR) missions are critical for locating and assisting injured or missing individuals while optimizing resource allocation and minimizing costs. This work aims to enhance the efficiency of these missions by exploring advanced deep learning techniques for precise and efficient human detection in thermal images. The primary focus of this work is on YOLOv8, the latest version of the You Only Look Once (YOLO) object detection method. The paper also evaluates YOLOv7-Tiny, which is the most streamlined variant derived from YOLOv7. To support the investigation, a novel dataset comprising 17,148 thermal images with 90,882 instances of human subjects representing various conditions and scenarios has been carefully curated. This dataset is used for training and evaluating different variants of YOLOv8 and YOLOv7. The evaluation of the trained models reveals the efficacy of YOLOv8 in detecting humans in thermal images, achieving an average precision rate of 95% with the largest model, YOLOv8x, and an average precision rate of 91% with the smallest model, YOLOv8n. The evaluation of YOLOv7-Tiny shows that it achieves an average precision similar to YOLOv8n, which is 48% lighter in size and more practical choice for real-world deployment. Also, the trained models are deployed on graphical processing units. The tiniest trained model, YOLOv8n, achieves an inference rate of 273.6 frames per second (FPS) while the largest model, YOLOv8, achieves an inference rate of 100.29 FPS. The achieved inference rates along with the achieved detection performances meet with the requirement of fast detection of humans in SAR missions.

INDEX TERMS Thermal images, object detection, deep learning, convolutional neural networks, YOLO, human detection, search and rescue missions.

I. INTRODUCTION

Urgent search and rescue (SAR) missions exert substantial pressure on resources and efforts, thereby impacting the overall operational efficiency and effectiveness. The speed at which wounded or lost individuals are located is a critical factor in determining the success of these missions. Rapidly pinpointing their whereabouts and gathering essential information is vital for guiding rescue teams and medical professionals in providing timely aid [2].

An intriguing avenue to significantly enhance the chances of saving lives while also reducing operational costs lies in the application of deep learning-based object detection models. This cutting-edge approach has the potential to improve the efficiency and effectiveness of search and rescue missions. Convolutional neural networks (CNNs) have been instrumental in the field of object detection. Several CNN-based architectures have been developed to perform the task of object detection such as region-based CNN (R-CNN), single shot detector (SSD), Mask R-CNN, and YOLO [3].

On the other hand, infrared thermal imaging plays a pivotal role in search and rescue missions. Thermal cameras capture heat emissions from subjects under surveillance,

The associate editor coordinating the review of this manuscript and approving it for publication was Ravibabu Mulaveesala.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

creating images based on infrared (IR) radiation, known as thermograms [4]. This imaging technology aids in pinpointing the locations of missing or injured individuals, even in challenging terrains or adverse conditions. In the past, search and rescue teams had to rely solely on visual cues to locate individuals, a task made extremely difficult, especially at night or in adverse weather conditions [5]. In contrast, the utilization of infrared thermal imaging allows rescuers to discern the unique thermal signature of missing persons from a distance, significantly expediting their detection. Thermal cameras possess the unique ability to penetrate smoke and darkness, proving invaluable in scenarios involving collapsed structures, fiery environments, or hazardous areas. When integrated with drones and complementary detection technologies, their effectiveness is further amplified. Noteworthy instances include their capacity to discern human presence in smoke-filled environments during fires [5], locating flood victims as exemplified in North Carolina in 2018 in the aftermath of Hurricane Florence, and rescuing earthquake survivors situated beyond the immediate line of sight of rescuers. A recent compelling case occurred after Turkey's tragic earthquake in 2023, where adverse weather conditions necessitated round-the-clock efforts, highlighting the critical role played by thermal cameras in expediting the rescue mission.

Recent advancements in thermal imaging technology have revolutionized thermal camera sensors, eliminating the need for gas-filled lenses and refrigeration to maintain optimal performance. This innovation has made thermal cameras more portable and cost-effective. Portable thermal cameras now seamlessly integrate thermal imaging capabilities. Furthermore, the advent of dual-camera systems, which combine thermal and visual cameras results in clearer and more informative images [6].

The integration of thermal imaging and deep learning has emerged as a potent tool for human detection, as increased sensor resolution allows for the capture of finely detailed thermal signatures, enhancing the ability to detect individuals under various conditions. Moreover, this integration capitalizes on recent breakthroughs in deep learning. The rapid evolution of deep neural networks has introduced several object-detection approaches. Emergent approaches elevate the detection performance to a level that they are considered as par with human performance. However, the high detection performance of deep learning-based approaches comes at the cost of hardware resources and power consumption particularly for real-time applications. In recent years, You Only Look Once approach has emerged as an efficient solution for object detection in real-time with high precession and accuracy. Several architectures of YOLO have been proposed. YOLO's architecture, renowned for its real-time object detection capabilities, has the potential to overcome some of the limitations of earlier methods, such as slow processing times and reduced accuracy in challenging environments [7].

In the realm of computer vision and deep learning, the effective detection of humans in thermal images holds significant implications, particularly in applications such as search and rescue missions, surveillance, and public safety. CNN-based object detection has proven to be a powerful tool in this context. However, the success of such systems is profoundly influenced by the quality and diversity of the dataset used for training, as well as the choice of the neural network architecture.

The fundamental premise of this research is rooted in understanding that the dataset is the lifeblood of deep learning models. The number and diversity of images within the dataset play a pivotal role in shaping the accuracy of the neural network's output. Furthermore, the selection of the neural network architecture significantly impacts the detection capabilities of the model. Numerous endeavors have been made to improve human detection in thermal images, focusing either on dataset augmentation and refinement or on the development of novel CNN models for precise detection. However, despite these efforts, existing datasets present certain limitations. Many suffer from a lack of diversity in terms of conditions under which the images are captured, often featuring uniform locations, distances, and environmental factors. Some datasets exhibit inaccuracies in human annotations, hindering model training. Additionally, certain datasets possess a high degree of image similarity due to closely spaced frame capture rates, and some are plagued by a paucity of images. These limitations collectively present a challenge that must be addressed: the creation of a dataset that overcomes these constraints, encompassing a larger and more diverse collection of images, scenarios, conditions, and accurate human annotations.

The second facet of this research work pertains to the selection of an object detection model capable of facilitating precise human detection in search and rescue missions, where rapid and accurate identification is paramount. The latest iterations of the YOLO family of models have made significant strides in this domain. YOLOv8, the latest offering in this lineage, offers a range of architectures, from Nano to Extra-large, harnessing cutting-edge advancements in deep learning and computer vision. YOLOv8 has demonstrated outstanding performance in terms of both speed and accuracy, surpassing its predecessors when tested on the Common Objects in Context (COCO) dataset [8], which is a widely used collection of images with detailed annotations, encompassing a diverse range of object categories in complex scenes.

Although the existing research on human detection in thermal imaging has shown notable advancements, the following gaps can be highlighted:

- **Detection Approach:** the emergent object detection approaches based on deep neural networks are not examined for real-time human detection in thermal images.

- **Dataset Limitations:** current works make use of limited datasets that lack diversity in terms of scale, perspective, number of humans in the frame and specifications of surrounding environment.
- **Limited Scope of Testing Conditions:** many works assess their approaches in specific scenarios and environments, overlooking the challenges imposed by real-life situations.
- **Inference Rates:** published works do not present their achieved inference rates, which are curricular for real-time applications.

Thus, this research work is driven by the imperative to evaluate the effectiveness of YOLOv8, particularly in the context of human detection using thermal images. A crucial aspect of this assessment revolves around determining whether reducing the model's size for faster detection compromises precision, or if it maintains robust performance. This work aims to bridge the gap by developing an enriched dataset that addresses existing limitations and by conducting a comprehensive evaluation of YOLOv8 and its variants in the context of human detection in thermal imagery. The detection performance of the YOLO models is evaluated using a wide set of videos and images captured by thermal images showing humans in diverse scenarios and various environments. The requirement of real-time inference is assessed by deployment of the models and recording the rate of actual processed frames. The overarching goal is to contribute to the advancement of human detection systems, enabling faster and more accurate responses in critical applications such as search and rescue missions.

The research work presented in this paper focuses on YOLOv8 for human detection, with a supplementary exploration of YOLOv7-Tiny, each chosen for their distinct advancements in object detection. YOLOv8 models are designed based on its predecessors' strengths and offer enhanced accuracy and efficiency, which is important for the complex task of detecting humans in TIR images within varied environments. The availability of several models, from light networks to dense ones, supports flexible deployment across different devices, prioritizing high-speed inference for real-time applications.

YOLOv7 adds value to this study through its architectural advancements, offering a comparative perspective on model effectiveness. The selection of YOLOv7 and YOLOv8 is based on their state-of-the-art design and performance, making them ideal for accurately detecting humans under challenging conditions. A detailed description of these models' architectures and specifications, including their deployment flexibility, is provided in Section III-B.

The main contributions of this work are:

1) create an innovative huge dataset of thermal images showcasing humans in diverse conditions and scenarios with the hugest number of accurate annotations compared to existing datasets,
2) train all available YOLOv8 models of different sizes and architectures using the prepared dataset,
3) train a previous model of YOLO, so called YOLOv7-Tiny, using same dataset to make further comparison with YOLOv8,
4) evaluate the detection performance of trained models in terms of well-known performance metrics,
5) deploy the trained models on high-end graphical cards and examine their corresponding inference speeds.

The rest of the paper is organized as following. Section II presents the literature review, delves into the concepts of thermal images, object detection and human detection and provides a survey about the available datasets for human detection in thermal imagery. Section III demonstrates the adopted methodology. Section IV presents the obtained results and a thorough analysis of YOLOv8 model performance and conducts a comparative assessment between YOLOv8 and YOLOv7. Section V shows the evaluation of the models targeting diverse real-life scenarios. Section VII describes the deployment of the models on high-end GPU devices and shows the obtained results. Finally, Section VIII concludes the paper and suggests potential avenues for future work.

II. BACKGROUND AND RELATED WORK

This section provides an exploration of thermal imaging and human detection. It starts with a short introduction of thermal imaging, encompassing its diverse capturing techniques and pivotal image characteristics. Also, it delves into the intricate interplay between image processing and artificial intelligence (AI) techniques that underpin the object detection and recognition. Subsequently, its pivot the focus toward human detection as a sub-field of object detection. A thorough examination of the current state-of-the-art in human detection through thermal imaging is provided. In addition, a comprehensive survey of available datasets containing thermal images of humans is presented.

A. THERMAL IMAGES

1) DEFINITION AND CAPTURING TECHNIQUES

Infrared radiation has longer wavelengths than visible light and is not visible to the human eye. The IR spectrum, which is illustrated in FIGURE 1, is split into various bands based on wavelength: the Near-Infrared (NIR) band ranging from $0.7\mu m$ to $1\mu m$, the Short-Wave Infrared (SWIR) band ranging from $1\mu m$ to $3\mu m$, the Mid-Wave Infrared (MWIR) band ranging from $3\mu m$ to $5\mu m$, the Long-Wave Infrared (LWIR) band ranging from $8\mu m$ to $14\mu m$, and the Very Long-Wave Infrared (VLWIR) band, which has wavelengths greater than $14\mu m$. The NIR and SWIR bands are often referred to as reflected infrared radiation, while the MWIR and LWIR bands are known as thermal infrared (TIR) radiation.

Thermal cameras utilize heat rather than visible light to generate images. Thermal imaging involves the conversion

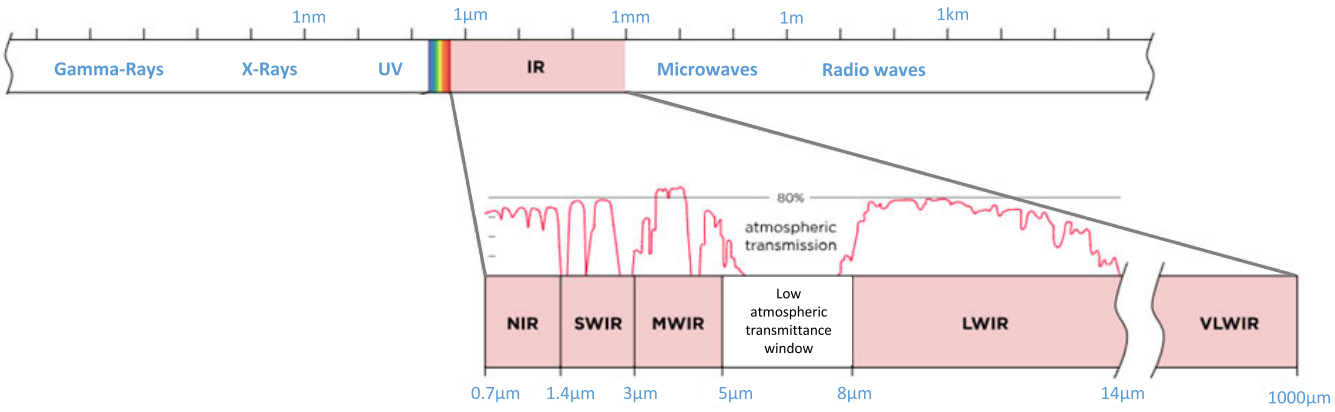

FIGURE 1. A visualization of the electromagnetic spectrum showcasing different sections attributed to IR wavelengths.

of infrared (IR) radiation, or heat, into images that depict the spatial distribution of temperature differences within a viewed scene. All objects at temperatures above absolute zero (−173 °C) emit infrared radiations, which intensities are directly related to their temperature [9]. IR thermal sensors have the ability to capture images of scenes and objects based on either the reflection of IR light or the emission of IR radiation. Thermal cameras operating in the thermal IR radiation bands do not need additional light or heat sources since TIR cameras can detect and form an image solely based on the thermal energy emitted from the objects under observation. This fact makes TIR cameras immune to varying light weather conditions and capable of operating in full darkness [10].

TIR cameras utilize different specialized sensors, yet the most used one is known as a microbolometer; which consists of an array of small pixels. Each pixel is sensitive to infrared radiation and its intensity reflects the temperature of the region it corresponds to. The microbolometer is made up of a thin temperature-sensitive material that undergoes changes in temperature when exposed to infrared radiation. The temperature changes result in alterations in the material's electrical resistance, which are then measured by the camera's electronics. This information is used to determine the temperature of the object being captured [11]. The camera converts these temperature readings into a visual image through a process called thermography. Thermography involves assigning different colors or shades of gray to different temperature readings, resulting in the creation of a heat map that represents the object being imaged. Additionally, thermal cameras can provide temperature measurements for specific areas within the image, allowing for the identification of hot spots or regions with unusual high temperature [10].

2) CHARACTERISTICS OF THERMAL IMAGES

a: DIFFERENTIATION BETWEEN OBJECTS

Using thermal images we can distinguish between humans, animals and other objects, making it particularly useful when targeting specific elements in the image. A thermal image can distinguish humans for example since humans have a distinct thermal signature due to their unique body temperature and heat distribution. The human body typically has a higher temperature than the surrounding environment, and certain areas such as the face, torso, and limbs emit more heat than others. Thermal imaging can detect these temperature variations and identify the human body as a warmer object compared to the surrounding objects or background [12].

b: BLURRING IN THERMAL IMAGES

Thermal images offer valuable insights, but blurring in thermal images can hinder accurate object detection. The key factors that cause blurring are:

- **Environmental conditions and temperature fluctuations**: In challenging environments with frequent temperature fluctuations, such as desert areas, thermal imaging encounters reduced performance in human detection [12]. When the body temperature closely aligns with the ambient temperature, accurate identification becomes increasingly difficult. Factors like heat-induced from the surroundings, steam generated nearby, and heat reflections can further contribute to blurring and affect the overall image quality.
- **Interference from external heat sources and thermal light sources**: Heat-generating objects or other heat-emitting sources may interfere with distinguishing human targets from non-human heat signatures, potentially leading to false positives or mis-identifications. Additionally, thermal light sources like open flames or hot surfaces can create bright spots and overexposure, obscuring critical image details, especially in low-light or nighttime conditions. To ensure accurate detection and clarity, it's essential to mitigate interference, optimize differentiation between heat sources, and adapt imaging techniques to handle thermal light sources effectively [7].
- **Thermal camera limitations**: Thermal camera itself can contribute to image blurring. Factors such as hot air during summer days, especially at noon, can distort

thermal images. Additionally, continuous operation and prolonged use of thermal cameras can lead to an increase in internal temperature, compromising image quality. Maintaining optimal camera temperature and implementing suitable cooling mechanisms are vital to mitigate blurring effects and improve the overall performance of thermal imaging systems [7].

c: THERMAL REFLECTIONS

Thermal images capture the heat emitted by an object, which is then reflected onto the surrounding surfaces like the floor or the wall. These reflections in thermal images share similar characteristics with the objects themselves, such as brightness, shape, and pattern. An example of thermal reflections is shown in FIGURE 2 [7]. In the figure, the thermal reflections are traced in red. However, telling apart thermal reflections from actual objects can be challenging due to their close connection. To accurately identify the specific area of an object in a thermal image, it is necessary to detect and differentiate the thermal reflection.

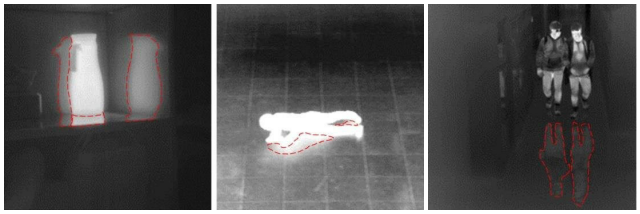

FIGURE 2. Thermal images with thermal reflections.

d: LESS DETAILED THAN VISUAL IMAGES

While thermal images provide valuable insights into temperature distributions and variations, they lack details and visual clarity provided by traditional visual images. The differences between thermal and visual images are significant, with thermal images relying on color-coded temperature representations rather than capturing visual appearance. This makes thermal images generally have lower spatial resolution compared to visual images. The lower resolution can make it challenging to identify and discern details of distant human bodies or small objects in the thermal image. However, the advancements in thermal imaging technology have led to improved resolution in recent years.

e: DAY AND NIGHT VISION

Thermal imaging is effective in both daylight and low-light conditions. Unlike visual images that rely on reflected light, thermal cameras detect the emitted heat energy from objects. Therefore, they can capture thermal signatures even in complete darkness or when objects are obscured by smoke, fog, or other visual barriers [12].

3) APPLICATION DOMAINS

Thermal imaging cameras are indispensable tools in search and rescue operations. They aid firefighters by improving visibility and mobility in smoke-filled rooms, detecting the origin of fires, and identifying hazardous elements. TIR cameras are valuable in locating unconscious victims in various environments, responding to road accidents, and aiding search and rescue efforts during natural disasters and collapses. They are also used in maritime, mountain, forest, and desert rescues, making rescues more efficient and effective.

In addition, to search and rescue missions, the usage of thermal imaging spans among several vital domains. In medical and healthcare applications, thermal imaging is used for thermography, disease detection, and injury assessment. Infrared thermography (IRT) enables rapid non-contact monitoring of body temperature and has been successful in diagnosing various medical conditions [13]. Thermal imaging is wildly used in industrial inspection for non-destructive testing and quality control, where it offers a non-contact control method, making it valuable in hazardous industrial settings. In addition, thermal imaging is used for monitoring the temperature of electric components, which helps in detecting damaged parts emitting excess heat. It is also used for assessing energy efficiency and insulation in buildings, reducing heating and cooling costs. Furthermore, thermal imaging is utilized to study wildlife behavior, monitor habitats, track migration patterns, detect poaching activities, assess temperature variations, water leaks, and energy consumption in urban areas. Thermal imaging is employed in agriculture for crop monitoring, irrigation management, and pest control. It helps identify crop stress areas, detect irrigation issues, and track pest movement, leading to more efficient and sustainable agricultural practices [14]. For security and law enforcement sector, Thermal imaging cameras provide law enforcement with an edge in detecting criminals in pitch-black conditions without requiring additional lighting, saving crucial time before their arrival. Thermal imaging cameras are now utilized in means of transportation. The cameras, which act as passive safety systems, provide the drivers with real-time accurate information about the tracks in poor visibility conditions such as fog, rain, or snow allowing to alert drivers about obstacles within their field of view [15].

Moreover, the use of thermal images combined with RGB images enhances the performance of computer vision applications such as detection [16], segmentation [17] and tracking [18], [19]. The fusion of images captured in diverse spectral ranges can improve the performance at a low computation cost [20]. As technology continues to advance, the applications of thermal imaging are expected to expand further, enhancing various fields and improving human safety, efficiency, and understanding of the world around us.

B. OBJECT DETECTION

Object detection is a specialized task within the field of computer vision. Object detection involves determining whether an image contains specific objects from predefined

classes and accurately pinpointing their locations within the image. Several vital application fields rely currently on the advancements in object detection techniques. The use of advanced object detection methods has been wiledly demonstrated in the field of remote sensing [21], intelligent transportation [22], and search and srescue missions [2], [23].

Object detection differs from image classification, which focuses on recognizing the objects present without specifying their locations. To achieve object detection, a bounding box, which is a rectangular region defined by the pixel coordinates of its four corners. The bounding box encapsulates the object of interest, and its design aims for minimalism, striving to be as compact as possible while still containing the entirety of the object [24]. This approach contrasts with semantic segmentation, which seeks to categorize individual pixels as part of an object category or not.

Several techniques have been proposed over the past two decades to perform the task of object detection. These techniques are commonly divided into two distinct categories: (1) traditional object detectors and deep learning-based detectors [25]. The following subsections introduce a brief overview about most significant techniques.

1) TRADITIONAL OBJECT DETECTORS

Before the emergence of today's revolutionary deep learning-driven object detection techniques, the roots of object detection stretch back to the 1990s, highlighting the innovative designs within early computer vision. These methods predominantly relied on handcrafted features due to limited image representation capabilities at that time. To overcome these limitations, experts ingeniously designed intricate feature representations and employed various acceleration techniques. The popular traditional object detectors and their corresponding techniques are briefly introduced in the following paragraphs.

The detecting approach proposed by Viola and Jones in [26] has been considered a significant advancement in computer vision, enabling real-time human face detection without being restricted by factors like skin color segmentation. It uses the approach of sliding windows and incorporated the concepts of ''integral image'', ''feature selection'', and ''detection cascades'' for efficient detection.

Building on the work of Viola-Jones, Dalal and Triggs have introduced in [27] the Histogram of Oriented Gradients (HOG) feature descriptor. It balances feature invariance and non-linearity, leading to improved detection accuracy. HOG involves computing gradients across a grid of cells and applying local contrast normalization.

In [28], the authors have introduced Deformable Part-Based Model (DPM), which extends the HOG framework by decomposing objects into deformable parts during training. In the inference phase, detections are assembled based on the arrangement of these parts. DPM has introduced concepts like mixture models, hard negative mining (HNM), bounding box regression, and context priming [25].

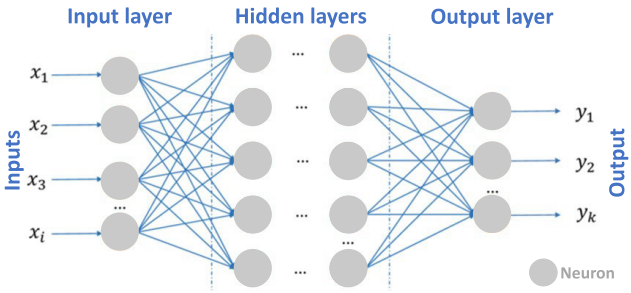

FIGURE 3. A schematic of a neural network.

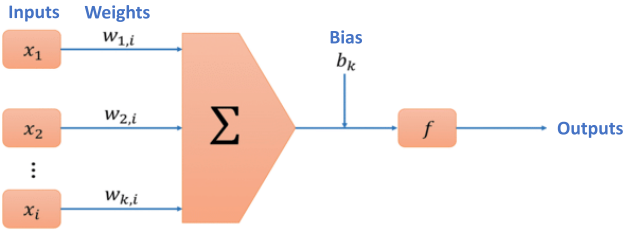

FIGURE 4. A schematic of each neuron in the network.

2) DEEP LEARNING-BASED DETECTORS

Traditional methods for object detection, though innovative for their time, have faced challenges in handling the increasing complexity of object detection tasks. They rely on manually crafted features and rule-based systems, which show limitations in dealing with diverse real-world scenarios, variations in object attributes, and complex backgrounds. In response to these limitations, AI-based techniques play a pivotal role. As AI matured, incorporating more sophisticated machine learning algorithms and optimization methods, novel detectors become more refined. However, the momentum of object detection technology has accelerated in the recent years because of the rapid progress in deep learning based methods and tools. The computational capacity delivered by graphical processing units (GPUs) has risen the performances of object detection and tracking tasks [12].

Deep learning, a branch of AI that emerges from the exploration of artificial neural networks, integrates structures like the multi-layer perceptron with multiple hidden layers. CNNs play dominant role in classifying and locating multiple objects in images and video sequences resulting in high accurate detection [29]. A typical CNN comprises two primary stages: (1) feature extraction and (2) classification. Feature extraction identifies distinctive constituents of objects while classification gauges the confidence of the object's presence in a specific location. Higher confidence indicates the presence of the object. The underlying principle of feature extraction advances from lower-level elements like edges and colors to higher-level attributes encompassing recognizable objects [12].

To enhance the network's capability, layers are sequentially interconnected in a feed-forward neural network architecture. This architecture transforms inputs through layers step by

step. FIGURE 3 presents a general schematic of a neural network. The first layer of a neural netrwork is the input layer. Hidden layers are allocated between input layer and the final output layer.

Artificial neurons compose the fundamental building blocks of the layers. FIGURE 4 illustrates a schematic of a neuron. These neurons process individual input values to create a vector $\mathbf{x}$ and produce scalar outputs $\mathbf{y}$. The inputs connect to weights $\mathbf{w}$, which are adjusted during learning. The computation involves two steps: calculating a linear combination of inputs and weights, followed by an activation function application, often a sigmoid-like logistic function. For tasks needing multiple output dimensions, like multi-class classification, multiple neurons work together. The output of each neuron corresponds to a unique dimension in the output space. To achieve this, a layer with numerous neurons, equal to the desired output dimensions, is used. Neurons process inputs independently through vector-matrix multiplication and activation functions [24].

Training artificial neurons involves iteratively adjusting weights using training examples. The neuron's output error E is measured by comparing its output y with the desired output d. This involves minimizing the error through gradient descent, which is a nonlinear optimization technique. Back-propagation, which is widely used for training neural networks, has two phases. In the feed-forward phase, a training example moves through the network, and its error is computed using functions like mean squared error or cross-entropy. The back-propagation phase adjusts weights to minimize the output error. This requires computing gradients of the error with respect to weights and updating them in the opposite gradient direction. However, calculating gradients for all weights is challenging due to inter-layer dependencies. Back-propagation starts by computing the gradient of the last layer's error and iteratively works backward, using the chain rule to propagate the error from the last layer to the first layer [24].

CNNs excel in computer vision tasks as they include convolutional and pooling layers as shown in FIGURE 5 [30]. The convolutional layer processes local inputs, differing from fully connected layers. Neurons' receptive fields cover overlapping windows, capturing spatial correlation and promoting weight sharing. Shared weight filters convolve across input images. This layer tackles translation variance. By applying filters across positions, it captures patterns regardless of location. Learned filters convolve with inputs during training, generating feature maps. Complementing this is the Pooling Layer, which reduces dimensionality. Max pooling, common in CNNs, subsamples by selecting maximum values within convolutional windows. It fosters scale invariance, enhancing feature learning and detecting larger objects. However, it may obscure small objects. CNNs excel in image processing, using convolutional and pooling layers for intricate feature extraction, maintaining translation invariance, and managing dimensionality [24]. In image analysis, a challenge is extracting meaningful features for reliable classification. Deep neural networks with many hidden layers address this by autonomously extracting features from sufficient training data. Challenges in training deep networks due to vanishing gradients have been overcome by hardware advances and innovations.

Deep learning-based object detection models can be divided into two distinct categories: (1) two-stage detectors and (2) one-stage detectors. Each category approaches object detection differently. Generally, as mentioned before, deep learning object detectors extract features from input images or frames, grasping the visual context and leading to a dual challenge: first, detecting a variable number of objects (possibly even zero) and then classifying each object and precisely defining its dimensions using a bounding box. Thus, the process is by either breaking down these tasks into two separate stages or by unifying both tasks into a single step.

a: TWO-STAGE DETECTORS

The process of two-stage detectors comprises: (Stage 1) proposing regions with potential objects using computer vision or deep neural networks and (Stage 2) object classification and precise dimension estimation using bounding box regression. While these detectors offer high accuracy, they operate at a slower speed due to multiple inference steps per image, which impacts the processing speed.

The following paragraph explores briefly the most popular two-stage detectors.

Region-based convolutional neural network (R-CNN) [31] has been introduced as the first two-stage object detector based on deep learning. R-CNN uses selective search for object proposals. Each proposal passes through a pre-trained CNN model for feature extraction and employes linear classifiers for object presence prediction and category recognition. While it significantly improves results, it is slow due to repeated feature calculations for numerous overlapping proposals. Spatial Pyramid Pooling Networks (SPPNet) has been later introduced where it addressed speed issues by introducing a spatial pyramid pooling (SPP) layer that generates fixed-length region representations, eliminating the need for repeated convolutional feature computations. It achieves remarkable speed improvements while maintaining accuracy. Fast R-CNN [32] refines R-CNN and SPPNet by simultaneously training a detector and bounding box regressor within the same network configuration, which leads to a significant speed boost and improved accuracy. Faster R-CNN [33] has been introduced shortly after introducing Fast R-CNN as the first near-real-time deep learning detector. It incorporates a region proposal network (RPN) that enabled almost cost-free region proposals, achieving a good balance between speed and accuracy [25]. Feature Pyramid Networks (FPNs) approach has introduced a top-down architecture with lateral connections to establish high-level semantics across scales, effectively utilizing the inherent feature pyramid structure of CNNs.

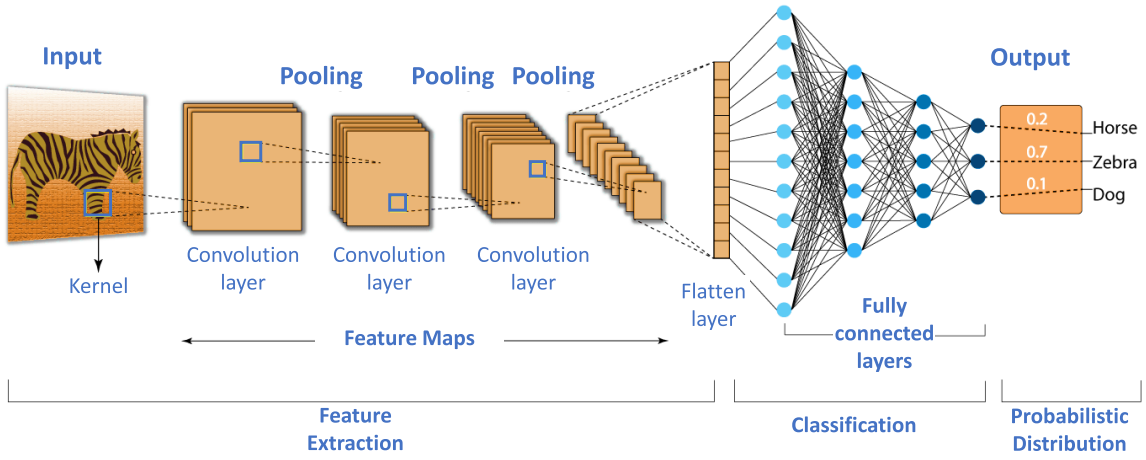

FIGURE 5. Architecture of a CNN.

b: ONE-STAGE DETECTORS

One-stage detectors, in contrary with two-stage detectors, predict bounding boxes over images without the requirement for an initial region proposal step. This streamlined process translates to faster processing of frames, making one-stage detectors more ideal for use in real-time applications. In the following paragraphs the most recent one-stage detectors are described.

In [34], the authors have introduced YOLO as a pioneer one-stage detector, which adopts a unique approach that utilizes one neural network to detect multiple objects in the entire image. In YOLO workflow, the input image is partitioned into non-overlapping cells consisting a grid. Each cell has the responsibility of making predictions for several bounding boxes and assessing the confidence of an object's presence. Multiple forecasts are generated relative to the centers of grid blocks. These forecasts are then organized into categories, and low-confidence detections are removed to create the ultimate output. This approach allows YOLO to capture object context effectively by examining the entire image at once, reducing false-positive detections compared to methods like R-CNN [34].

The architecture of YOLO comprises three main parts: the backbone, neck, and head. The backbone extracts the features from the input image, while the neck and head layers predict bounding boxes, object classes, and confidences. The loss function combines squared differences for bounding box coordinates, square-rooted width and height differences, and class predictions, with coefficients to prioritize localization and object detection accuracy. YOLO achieves exceptional speed, when compared to available detectors, while attaining a PASCAL VOC07 mean average precession (*mAP*) of 52.7%. The first version of YOLO [34] achieves a VOC07 *mAP* of 63.4%. Building on this success, subsequent versions of YOLO have been developed, each refining and expanding upon the original concept. FIGURE 6 shows the evolution of YOLO versions.

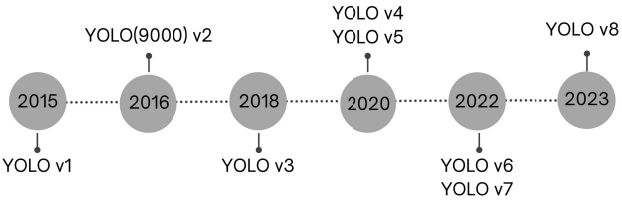

FIGURE 6. The evolution of YOLO versions.

YOLOv2 (also known as YOLO9000) [35], has addressed precision and recall concerns while leveraging anchor boxes for more efficient object localization. By adopting Darknet-19, which is a CNN architecture with 19 layers for feature extraction, YOLOv2 attains impressive accuracy of 91.2% and maintained a high processing speed of 67 FPS [35]. In [36], YOLOv3 has been proposed. YOLOv3 improves the performance through the utilization of Darknet-53 and multiscale prediction. Its logistic regression-based objectness scoring and binary cross-entropy loss for class predictions makes it especially adept at detecting small objects. In [37], YOLOv4 has been introduced. It compromises innovative features like data augmentation, an anchor-free detection head, and a new loss function. Based on YOLOv4, YOLOv5 has been launched by streamlining workflows with hyperparameter optimization, integrated tracking, and automatic export capabilities. YOLOv6 [38] has found practical application in autonomous delivery robot; whereas,YOLOv7 [39] expands the model's capabilities by tackling additional tasks like pose estimation. The latest iteration, YOLOv8 [40] stands as a state-of-the-art model, advancing performance, adaptability, and efficiency across diverse vision AI tasks such as detection, segmentation, pose estimation, tracking, and classification.

In [41], Liu et al. have introduced SSD, which innovatively incorporates multi-reference and multi-resolution detection techniques. SSD operates across various network layers to

detect objects of different scales. This enables a combination of speed and accuracy achieving COCO *mAP*@0.5 of 46.5%. Lin et al. have proposed in [42] RetinaNet to tackle the accuracy gap faced by one-stage detectors compared to two-stage counterparts. The focal loss function, a core component of RetinaNet, focuses on challenging, misclassified examples during training, rectifying the foreground-background class imbalance. This leads to attaining a high detection speed (COCO $mAP@0.5 = 59.1\%$) while achieving acceptable detection performance in terms of accuracy when compared to available two-stage detectors. In [43], EfficientDet has been proposed by the Brain team at Google Research as a scalable and efficient object detector. EfficientDet achieves COCO *mAP*@0.5 of 52.2%. CornerNet and CenterNet, which both embrace keypoint-based detection approache, have been introduced respectively in [44] and [45]. CornerNet predicts keypoints and subsequently forms bounding boxes through smart grouping; whereas, CenterNet eliminates post-processes, streamlining the detection process. Both models exhibit impressive performance, with CornerNet achieving COCO *mAP*@0.5 of 57.8% and CenterNet achieving 61.1%. In [46], the authors have introduced in Data-Efficient Image Transformer (DETR), which signaled a shift by employing Transformers, replacing traditional convolutions with attention-based calculations. This novel perspective treats the task of object detection as a set prediction problem, resulting in transformative end-to-end detection. Later, Deformable DETR has emerged to overcome convergence challenges and limitations with small object detection, setting a new benchmark with a COCO *mAP*@0.5 of 71.9% [25].

C. HUMAN DETECTION

Human detection is a specialized field within computer vision that involves localizing instances of humans in images or videos using advanced algorithms [47]. It is a sub-field of object detection, focusing specifically on identifying and locating human beings within a given scene. Hence, the progress in object detection models shapes the evolution and refinement of human detection techniques, resulting in wider and more efficient applications across diverse domains.

Human detection has various important applications, including: video-based surveillance systems [48], biometrics and identity verification systems [49], autonomous vehicles [48], and human-computer interaction [49].

Particularly noteworthy, human detection is valuable in disaster response scenarios. Equipped with a human detector, a device can assist rescue teams in locating survivors after a disaster [11]. By accurately detecting and reporting the specific locations of trapped individuals, the efficiency of rescue operations can be greatly enhanced [2], [50].

D. HUMAN DETECTION IN THERMAL IMAGING

The techniques employed for object detection have demonstrated their adaptability to the nuanced task of human detection. Using TIR images as an input has emerged as a particularly promising avenue, enhancing detection performance in challenging conditions such as low visibility or adverse weather. This technique finds its roots in the broader context of object detection methodologies, which have been widely applied across a spectrum of applications. The realm of human detection using thermal images has seen continuous exploration, dating back to the inception of image processing techniques. From the early days of shape-based and geometric approaches to the transformative advent of deep learning, researchers have striven to elevate the accuracy, adaptability, and real-time applicability of human detection systems.

1) IMAGE PROCESSING-BASED DETECTORS

In [51], an innovative approach has been proposed, harnessing the discriminatory capabilities of the Shape Context (SC) descriptor and Linear Discriminant Analysis (LDA) for robust human detection in thermal images. The proposed method integrates SC and boosting techniques to extract features that effectively represent human information. The conducted comparisons against a rectangle feature-based classifier have demonstrated its significant advantage in performance. Achieving a detection rate surpassing 70% at a false positive rate of 5%, it clearly outperforms the rectangle feature-based detector, which achieves a 30% rate. While showcasing superior detection, the authors also have recognized the need to address the computational cost of SC descriptor calculation for enhanced real-time applicability, underlining the practical considerations essential for successful implementation. Building on this foundation, the authors in [52] have introduced a pragmatic method for pedestrian detection in thermal imagery. This approach seamlessly merges HOG features with geometric characteristics, showcasing promising detection capabilities. By capitalizing on the robustness of HOG features and the adaptability of geometric attributes, the proposed method has demonstrated versatility in recognizing pedestrians and other diverse targets present in the OSU Thermal Pedestrian Database [53].

In [54], the authors have delved into pedestrian detection in thermal images, exploring the application of the CENsus TRansform hISTogram (CENTRIST) [55] visual descriptor and its variants. The well known Linear Support Vector Machine (SVM) classifier is employed. The comparison of the proposed approach against the Histogram of Oriented Gradients (HOG) feature descriptor shows the superior accuracy of CENTRIST and its variants, accompanied by significantly reduced training and testing times. This marked achievement emphasizes the practical benefits of CENTRIST-based methods, highlighting their computational efficiency and detection accuracy. In [56], the authors have introduced a novel perspective by ingeniously combining intensity-based region of interest (ROI) extraction with feature-based validation for the application of pedestrian detection in thermal images. Addressing the challenge of body part splitting in ROI extraction, the proposed approach tailors the process to the pedestrian's scale, supported by

empirical analysis highlighting pedestrian statistics' scale variations. An automated solution for scale identification, along with an adaptive threshold for pedestrian region growth, leads to enhancing the detection accuracy. The incorporation of the Curvelet Energy Entropy feature further reduces false positives in the validation step, showcasing meticulous attention to methodological detail.

2) DEEP LEARNING-BASED DETECTORS

In [3], the authors have utilized YOLO detector for identifying individuals in thermal images. The work meticulously has examined challenging conditions such as winter nights, varying weather, and distances from the camera, ranging from $30m$ to $215m$. Although thermal images possess unique visual characteristics, the authors have hypothesized that features learned by YOLO on RGB images could serve as a baseline for thermal analysis. The initial YOLO model has exhibited limited success, with a modest average precision (AP) of 7% for person detection, highlighting the substantial dissimilarity between visual and thermal data. However, fine-tuning the model with a custom thermal dataset achieves a significant leap, elevating the AP to approximately 30%, thereby underscoring the potential of thermal data augmentation to enhance detection performance.

In [57], the authors have utilized YOLOv3 to detect humans in thermal images. Their proposed method achieves an average precision of 95% and the inferense rate of 58 *FPS* and 0.125 *FPS* when runing the trained model on a GPU and CPU respectively. The specifications of the used GPU and CPU are not mentioned. Also, the Intersection over Union (IoU) used while computing the precession is not specified.

In [58], the authors have utilized YOLOv4 model for human detection in low-visibility scenarios. The work focuses on scenarios featuring heavy smoke and impaired visibility, such as fire scenes. Employing a thermal imaging camera in compliance with NFPA1801 standards, the proposed work achieves remarkable accuracy, with the YOLOv4 model detecting individuals across different postures with an IoU of 50%. Notably, the model's convergence within 4000 epochs on a single Nvidia GeForce 2070 GPU facilitated real-time detection, operating at 30.1 *FPS*.

The authors in [59] have explored human detection from an aerial perspective using two prominent object detection algorithms: Faster R-CNN and SSD. The work presents a comprehensive performance analysis of these algorithms in the context of aerial thermal images. Notably, Faster R-CNN ResNet50 emerges as the frontrunner in terms of mean average precision, while SSD Mobilenet-v1 excels in detection speed. The achieved results highlights the suitability of these algorithms for distinct scenarios, with Faster R-CNN favoring accuracy-centric applications and SSD better aligned with real-time contexts. Adjusting anchor parameters yielded performance improvements, further reinforcing the potential of algorithmic fine-tuning.

In [60], the authors have exploited deep learning techniques targeting the application of tracking people in thermal images for search and rescue missions in mountains. YOLOv5 is used to detect people in thermal videos captured by small drones equipped with thermal imaging camera. The developed tracking method exploits the centroid of the bounding box generated by YOLOv5 model.

Multitude AI-based techniques have been recently introduced as solutions to perform efficiently the complex task of object detection. These techniques attain different levels of accuracy and precision and require different computational resources to run in real-time. Also, they differ in the number of the objects they are able to detect simultaneously in images and videos. YOLO technique has emerged as a superior solution for real-time object detection [34], [37]. Published works targeting object detection in several application domains [29], [50], [61], [62], [63], show that YOLO outperforms other available one-stage methods such as SSD [41] and RetinaNet [42] and other two-stage detection techniques like R-CNN [31] and its variants such as Fast R-CNN [32] and Faster R-CNN [33]. Also, the conducted experiments targeting COCO dataset illustrate that YOLO-based models attain higher inference rates and are and more accurate than the real-time neural network EfficientDet provided by Google [43].

In the continuum of research efforts, the up presented works collectively illuminate the unwavering commitment to elevate human detection methodologies across diverse contexts. By undertaking a comprehensive comparison with other YOLO models, this work contributes to the ongoing evolution of object detection technologies, unveiling fresh insights and innovative methodologies that continue to propel the field forward. In line with this trajectory, our current work sets out to explore the performance of the YOLOv8 model in human detection using thermal images, with a supplementary exploration of YOLOv7-Tiny.

YOLOv8 provides improved accuracy and efficiency, making it particularly suitable for the complex task of detecting humans in thermal images with varying sizes, scales, perspectives and environmental contexts. Its design focuses on fast inference, crucial for real-time monitoring. Having variants from Nano to Extra-Large, YOLOv8 supports flexible deployment across various devices. The design of the models prioritizes high-speed inference, which is essential for real-time applications and immediate decision-making in SAR missions. YOLOv8's versatility is showcased through its range of model sizes, each finely balanced in terms of parameters and computational complexity. This adaptability enables customized deployments to meet specific computational resources and requirements, whether on edge devices with limited power or more robust systems. The anchor-free design of YOLOv8 further enhances its flexibility, reducing the chances of suboptimal results due to manual specification of anchor boxes.

The use of YOLOv7's in this work highlights its strength and architectural advancements, enriching the research with a comparison of model effectiveness in real-life scenarios. The use of YOLOv7 provides supplementary perspectives and

contributes to a broader understanding of the advancements in object detection models. The selection of YOLOv7 and YOLOv8 is based on their cutting-edge design and performance, making them leading choices for object detection.

E. SURVEY ON THE AVAILABLE DATASETS OF HUMANS IN THERMAL IMAGES

This section provides a systematic survey of the available datasets specifically designed for human detection in thermal images, which are essential for data-driven deep learning approaches. Significant efforts have focused on curating datasets of thermal images for human detection. These datasets, spanning diverse scenarios and applications, play a pivotal role in training and evaluating deep learning models. In this context, this section explores a range of notable datasets designed to enhance person detection algorithms, offering insights into real-world challenges and advancements in thermal imagery analysis.

1) DATASET 1: BORDER SECURITY THERMAL DATASET

The Border Security Thermal Dataset [64] is designed for training machine learning and deep learning models to detect illegal movements in border and protected areas. The dataset includes thermal videos and images captured under various weather conditions, including clear weather, rain, and fog, during nighttime. The data simulates real-world scenarios by recording in forest areas. It consists of 7,412 manually labeled images extracted from LWIR video frames. The dataset evaluates the impact of standard and telephoto lenses on thermal image quality and person detection. It comprises annotated frames from different weather scenarios, processed from about 20 minutes of clear weather, 13 minutes of fog, and 15 minutes of rainy weather footage. Annotations include object centroid positions and bounding box dimensions using the YOLO annotation format. FIGURE 7 shows sample images from Dataset 1.

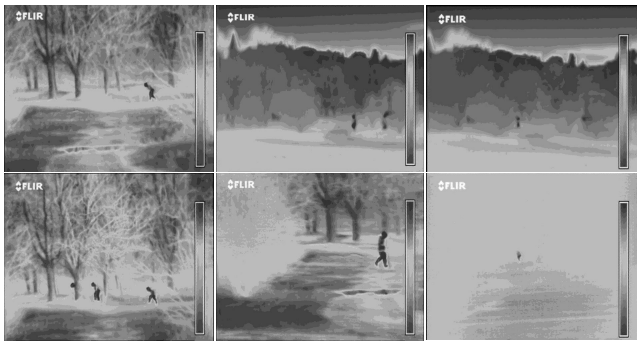

FIGURE 7. Sample images from dataset 1.

2) DATASET 2: THERMAL PERSON DETECTION DATASET - OHIO STATE UNIVERSITY CAMPUS

The Thermal Person Detection Dataset from Ohio State University [53] is intended for developing and evaluating person detection algorithms in thermal imagery, especially for surveillance, security, and pedestrian analysis applications. The dataset is captured using a Raytheon 300D thermal sensor core with a 75mm lens. It features 10 sequences with 284 8-bit grayscale bitmap images of 360 $\times$ 240 pixels resolution. The dataset includes environmental information and ground truth bounding box annotations, encompassing people with similar aspect ratios and at least 50% visibility to ensure high quality. The camera is mounted on an 8-story building on the university campus, providing an elevated viewpoint of a pedestrian intersection. FIGURE 8 shows sample images from Dataset 2.

FIGURE 8. Sample images from dataset 2.

3) DATASET 3: VEHICLE-MOUNTED THERMAL AND VISIBLE CAMERA DATASET

The Vehicle-mounted Thermal and Visible Camera Dataset [65] is curated for object detection, tracking, and scene understanding in real-world driving scenarios. It includes data captured using a Teledyne FLIR Tau 2 thermal camera with a 13mm $f/1.0$ lens, as well as a Teledyne FLIR BlackFly S visible camera with a 52.8-degree HFOV lens. The dataset ensures accurate alignment of thermal and visible data through time-syncing. It provides validation videos for tracking metric computation, featuring diverse and relevant frames while excluding redundant footage, ensuring a focus on meaningful visual scenarios encountered during vehicle-mounted thermal and visible camera capture. FIGURE 9 shows sample images from Dataset 3.

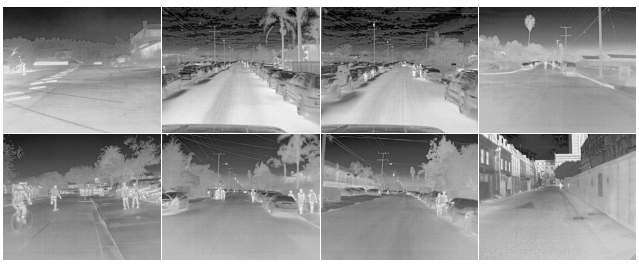

FIGURE 9. Sample images from dataset 3.

4) DATASET 4: HIGH-ALTITUDE INFRARED THERMAL FOR UNMANNED AERIAL VEHICLE (HIT-UAV) DATASET

The HIT-UAV dataset [66] comprises 2898 infrared thermal images extracted from 43470 frames captured by an Unmanned Aerial Vehicle (UAV) across diverse scenes like schools, parking lots, roads, playgrounds, etc. The dataset

encompasses multifaceted aspects such as object categories (person, bicycle, car, other vehicle), flight altitudes (ranging from 60 to 130 meters), camera perspectives (ranging from 30 to 90 degrees), and daylight intensity (day and night). It's particularly pertinent for advancing object detection research with UAV-based infrared thermal imagery. FIGURE 10 shows sample images from Dataset 4.

FIGURE 10. Sample images from dataset 4.

5) DATASET 5: THE PTB-TIR: THERMAL INFRARED PEDESTRIAN TRACKING BENCHMARK

The PTB-TIR dataset [67] is tailored to impartially assess thermal infrared trackers. It emphasizes the significance of TIR pedestrian tracking, especially for low-light scenarios. The dataset includes 60 TIR sequences with manual annotations. It offers sequences with diverse attributes to comprehensively evaluate TIR tracking, encompassing factors like thermal crossover, intensity variation, occlusion, scale variation, background clutter, low resolution, fast motion, motion blur, and out-of-view targets. Nine attribute labels accompany each sequence, facilitating attribute-based evaluation. The shooting videos in the dataset are obtained from existing commonly used thermal sequences and video websites, including the OSU Color-Thermal dataset, Terravic Motion IR, BU-TIV, LITIV2012, CVC-09, CVC-14, VOT-TIR2016, INO dataset, and video sequences posted on YouTube. Additionally, this dataset covers a wide array of factors, enabling the assessment of TIR pedestrian trackers under varying challenges. FIGURE 11 shows sample images from Dataset 5.

FIGURE 11. Sample images from dataset 5.

6) DATASET 6: THE PERSON AND DOG THERMAL IMAGES DATASET

This dataset [68] encompasses 489 annotated thermal infrared images in YOLO format, with a specific focus on people and dogs. The images in this dataset depict subjects captured in park and home settings at varying distances. Captured via the Seek Compact XR Extra Range Thermal Imaging Camera for iPhone, the images employ the Spectra color palette. The dataset includes both portrait and landscape orientations, with annotations designed to align correctly regardless of image orientation using Roboflow's auto-orientation feature. Additionally, the dataset features augmented versions of each source image through transformations including horizontal flips, random cropping, and brightness adjustments. FIGURE 12 shows sample images from Dataset 6.

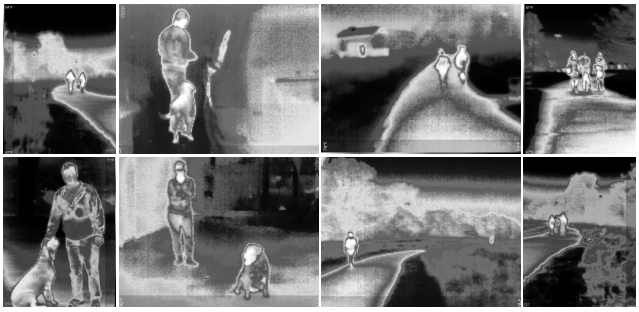

FIGURE 12. Sample images from dataset 6.

III. METHODOLOGY

This section demonstrates the adopted the methodology and practical steps of dataset creation, training and evaluation of the models in order to realize the research work contributions. An overview of the methodology employed in our work is summarized in FIGURE 13.

A. PREPARATION OF THE DATASET

In order to train different YOLO targeted models, a huge dataset that fuels the training process is meticulously constructed. The collected dataset comprises a variety of thermal images featuring varying numbers of humans, ranging from single individuals to crowded places. FIGURE 14 presents a selection of sample images collected to form the dataset. These images are sourced from diverse available resources [53], [64], [65], [66], [67], [68] and are captured using various devices, including surveillance cameras, hand-held cameras, vehicle-mounted cameras, and drone cameras. The scenes encompass a wide spectrum, both indoors and outdoors, including homes, schools, university campuses, sidewalks, roads, playgrounds, parking lots, parks, and forests. The images are taken at different times of the day, during both daylight and nighttime, and under various weather conditions, such as clear weather, rain, and fog. The collected dataset also encompasses variations in camera motion, camera viewpoints, human distances (far, middle, near), human sizes (big, middle, small), and human positions (standing, walking, sitting, crawling, hiding).

The process of compiling the dataset involves integrating existing datasets originally utilized for object detection.

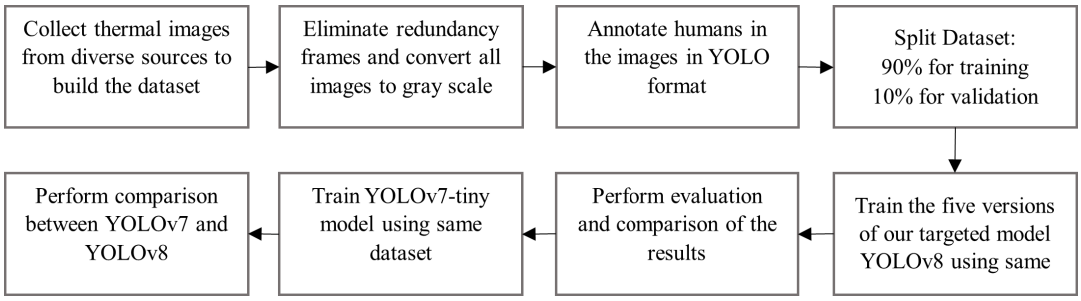

FIGURE 13. An Overview of the adopted methodology.

FIGURE 14. Sample images from the collected dataset showing the diversity of integrated images.

To narrow down the focus solely on human detection, annotations for objects other than humans are removed from these datasets. Additionally, to eliminate redundancy and similarity, similar frames from video-based images are excluded. To ensure uniformity, all images are converted to grayscale, resulting in a standardized representation across the dataset. For the annotated portion of the dataset, existing annotations are reviewed carefully and converted into the YOLO format, where each corresponding label file contains annotations for all humans existing per image in the form of <object-class> <x> <y> <width> <height>. In cases where the annotation data is not previously annotated or only one human is annotated per image, manual annotations are generated using *LabelImg* in YOLO format as well. During the annotation process, special attention is paid to ensure that the bounding boxes only encompassed humans, capturing either complete human figures or visible body parts. Reflections, which do not depict real humans, are excluded from the annotations. Our final Infrared Thermal Images for Human Detection dataset comprises 17,148 grayscale images with annotations for approximately 90,882 humans. Detailed information about the image sources and their distribution is provided in Table 1.

FIGURE 15, presents the labels correlogram, which provides insight into the relationships between different label dimensions (<x> <y> <width> <height>) in the created dataset. It includes a a group of 2 dimension histograms that show each axis of the labels in the created dataset against each

TABLE 1. Dataset statistics.

Dataset	Number of Images	Number of Annotations
[64]	631	957
[53]	284	984
[65]	11,255	67,271
[66]	1,691	12,312
[67]	3,038	9,021
[68]	249	337
Total	17,148	90,882

other axis. These histograms helps in identifying correlations between these dimensions for better understanding of the annotations of humans in the dataset.

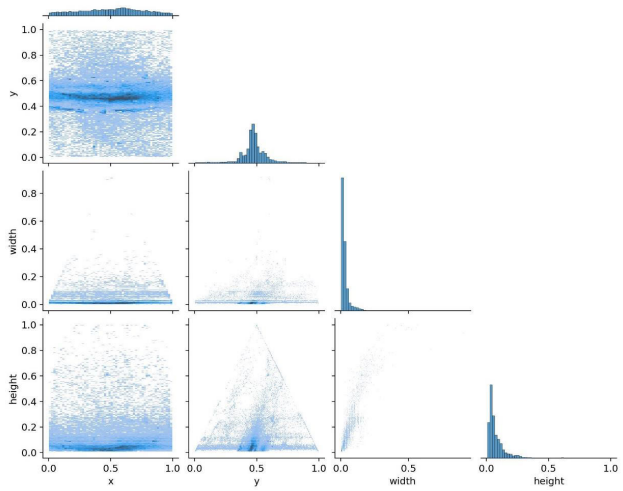

FIGURE 15. Labels correlogram of the created dataset.

Note that the resulting dataset will be released to research community under request to facilitate progress in detection of humans in thermal images using deep learning approaches.

B. TARGET MODELS

1) YOLOv8 MODELS

Our work focuses on YOLOv8, the latest version of the YOLO object detection, classification, and segmentation model developed by Ultralytics [69]. YOLOv8 can be easily implemented on hardware platforms spanning from edge devices to cloud-based APIs. The architectural structure of YOLOv8 is depicted in FIGURE 16 [70]. The structure is based on a modified version of CSPDarknet53, which incorporates 53 convolutional layers and employs cross-stage partial connections. These connections enhance the seamless flow of information between layers, contributing to the model's effectiveness. The architecture of YOLOv8 can be dissected into two main components: the backbone and the head. The backbone forms the core of the model and consists of numerous convolutional layers, which play a pivotal role in extracting essential features from input data. The head of YOLOv8 follows the backbone and comprises multiple convolutional layers followed by a series of fully connected layers. The head layers are in charge of several critical tasks such as predicting bounding boxes, determining objects scores, and estimating class probabilities for objects detected within an image. The architecture of YOLOv8 depicted in FIGURE 16 shows that there are three *Detect* modules which correspond to the three heads responsible for detecting small, medium, and large objects, respectively. By using this multi-scale prediction, YOLOv8 is able to detect objects at different scales, which enhances the detection performance of the model to various sizes of objects.

YOLOv8 series presents five models: YOLOv8 Nano (YOLOv8n), YOLOv8 Small (YOLOv8s), YOLOv8 Medium (YOLOv8m), YOLOv8 Large (YOLOv8l), and YOLOv8 Extra-Large (YOLOv8x) which are designed to meet different application requirements. These models vary in depth, width, and aspect ratio, enabling a spectrum of performance-tradeoff options. Table 2 shows the values of depth, width and aspect ratio corresponding to each variant of YOLOv8 (FIGURE 16). YOLOv8n excels as the fastest and most compact choice, while YOLOv8x stands out as the accuracy-focused counterpart, albeit at a higher computational cost. FIGURE 17 and FIGURE 18 presents the specifications of the targeted variants of YOLOv8 that have been utilized in this work in terms of the number of parameters and FLOPs respectively.

TABLE 2. Values of depth, width, and aspect ratio for YOLOv8 variants.

Model	depth multiple (d)	width multiple (w)	aspect ratio (r)
YOLOv8n	0.33	0.25	2.0
YOLOv8s	0.33	0.50	2.0
YOLOv8m	0.67	0.75	1.5
YOLOv8l	1.0	1.0	1.0
YOLOv8x	1.0	1.25	1.0

The following paragraphs delve deeper into the core components and innovations of YOLOv8.

a: BACKBONE AND C2F MODULE

The backbone of YOLOv8 serves as the foundational feature extractor. Notably, it incorporates the Cross-Stage Partial Bottleneck with Two Convolutions (C2f) module, a critical enhancement over the traditional Cross Stage Partial Layer (CSPLayer). The CSPLayer is essentially a special type of bottleneck layer. In the context of neural networks, a bottleneck layer is a part of the architecture that compresses or reduces the dimensionality of the input data. This helps in learning more compact and abstract representations of the features. The term "bottleneck" refers to the idea that this layer constrains the flow of information, forcing the network to learn more efficient representations. Bottlenecks in the CSPDarknet53 backbone reduce computational complexity while maintaining accuracy. In addition, the use of Spatial Pyramid Pooling Fast (SPPF) layer captures features at multiple scales, which further improves the detection performance. The C2f module plays a pivotal role in enriching detection accuracy by combining high-level features with contextual

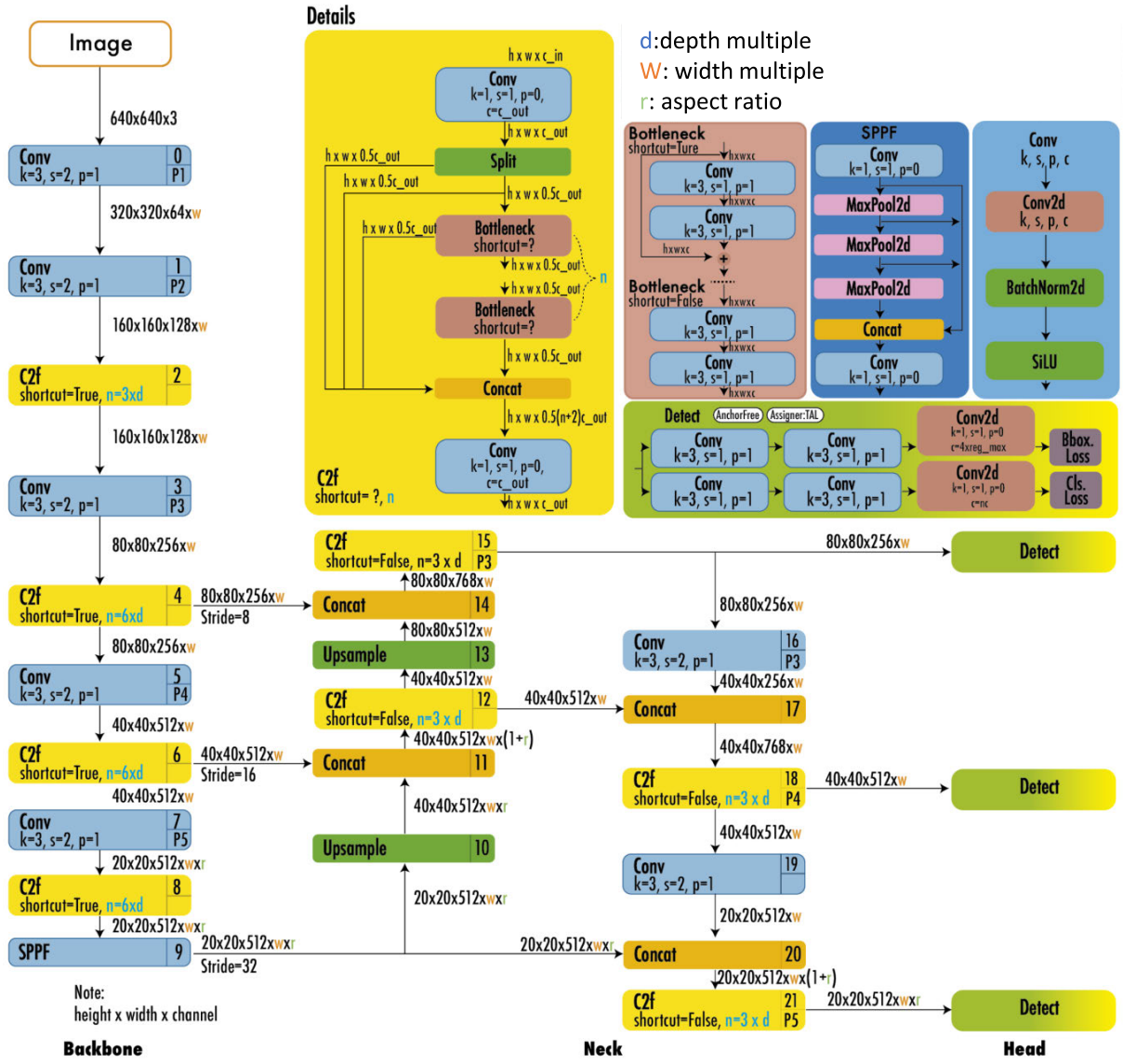

FIGURE 16. The architecture of YOLOv8.

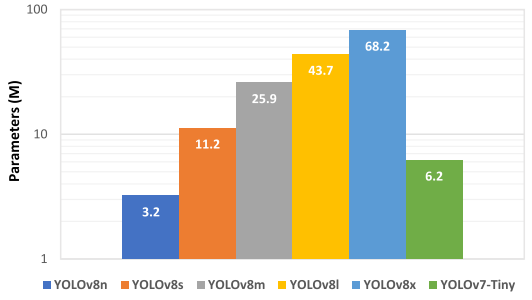

FIGURE 17. Parameters of target models.

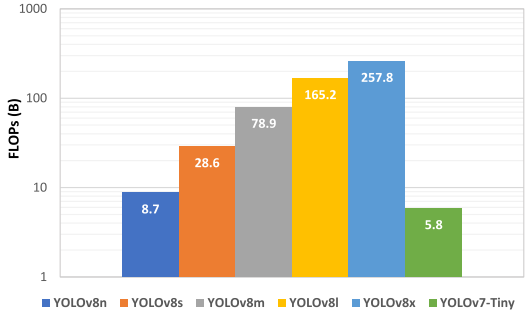

FIGURE 18. FLOPs of target models.

information. This fusion of features enhances the model's ability to recognize and locate objects within images [70]. Also, the use of C2f module leads to enhanced accuracy, especially for detecting small objects.

b: ANCHOR-FREE MODEL WITH DECOUPLED HEAD

One of the hallmark features of YOLOv8 is its transition to an anchor-free model. Where anchor boxes are predefined bounding boxes of various sizes and aspect ratios. These

anchor boxes serve as reference points that the model uses to predict object locations and sizes. In contrast to previous iterations that relied on anchor boxes for object localization, YOLOv8 directly predicts object centers. Furthermore, it adopts a decoupled head architecture, which means that the model independently handles objectness, classification, and regression tasks. This decoupling allows each branch to specialize in its respective task, thereby refining the model's overall accuracy.

c: ACTIVATION FUNCTIONS

In the output layer of YOLOv8, distinct activation functions are employed for different purposes. The sigmoid function is used for the objectness score, which represents the probability that a bounding box contains an object. This score assists in determining whether an object is present within the box. On the other hand, the softmax function is applied to compute class probabilities. These probabilities indicate the likelihood of an object belonging to each possible class within the dataset [70].

d: LOSS FUNCTIONS

To train and optimize the YOLOv8 model effectively, it employs specialized loss functions. For bounding box loss, it utilizes the Complete IoU (CIoU) and Distribute Focal Loss (DFL) loss functions. These loss functions are tailored to improve object detection performance, particularly when dealing with smaller objects. For classification loss, binary cross-entropy (BCE) is employed to differentiate between various object classes. These loss functions contribute significantly to the model's ability to accurately detect and classify objects in images.

2) YOLOv7-TINY MODEL

Furthermore, for a comprehensive comparative analysis of YOLOv8 models against their predecessors, YOLOv7-Tiny is trained targeting the created dataset. YOLOv7-Tiny is the most compact model within the YOLOv7 series. YOLOv7, renowned for its exceptional efficiency as an anchor-based object detection algorithm, excels in delivering rapid detection without compromising on accuracy. YOLOv7-Tiny model is structured around three integral components: the backbone, the neck, and the head.

FIGURE 19 illustrates the architecture of YOLOv7-Tiny model [71]. YOLOv7-Tiny's backbone predominantly consists of convolutional layers, and it integrates several pivotal modules:'

a: EXTENDED-ELAN (E-ELAN) MODULE

An extension of the original Efficient Long-Range Aggregation Network (ELAN) module, which is a network module employed in deep learning models to improve the efficiency of training deep neural networks. It does so by facilitating the flow of gradient information across network layers, promoting faster convergence and improved training in tasks such as object detection. However, the E-ELAN module introduces modifications to the calculation block while preserving the core structure of ELAN. These adjustments enhance the network's learning capabilities by expanding, shuffling, and merging cardinality.

b: MAX POOLING CONV (MPCONV) MODULE

This module employs parameters with varying precisions for convolutional operations, effectively balancing computational complexity against accuracy.

c: SPPCSPC MODULE

Comprising two crucial sub-modules, the SPPCSPC module augments the expressive potential of convolutional neural networks. The Spatial Pyramid Pooling (SPP) module facilitates multi-scale object detection and classification by segmenting the input feature map into sub-regions and pooling them to obtain fixed-size feature vectors. The Cross-Stage Partial Network (CSP) module divides the network into two segments, one dedicated to feature extraction and the other to feature processing and fusion. This partition optimizes the network by reducing parameters and computational complexity while simultaneously enhancing expressive power and computational efficiency.

The neck module serves as the epicenter for feature fusion, harmoniously combining feature maps from various levels to generate multi-scale feature maps. This fusion significantly amplifies the model's ability to accurately detect objects of diverse sizes. The head network operates on the multi-scale feature maps generated by the neck network, steering the object detection process. It employs anchor boxes to predict object location, size, and class within the input image. Subsequently, post-processing steps, such as Non-Maximum Suppression (NMS), refine the predicted object boxes by eliminating redundancies, ultimately enhancing the model's precision. The specifications of YOLOv7-Tiny in terms of network parameters and required FLOPs are presented in FIGURE 17 and FIGURE 18.

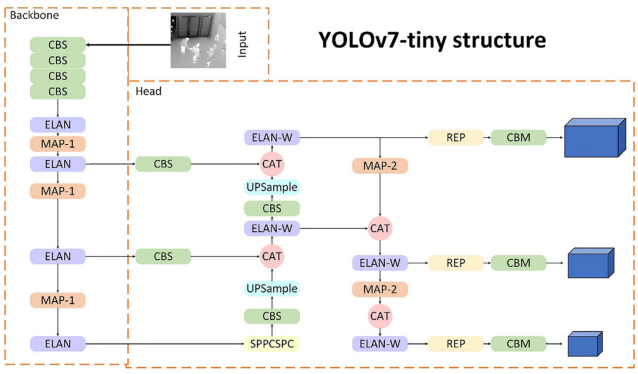

FIGURE 19. The architecture of YOLOv7-Tiny model.

C. TRAINING AND VALIDATION

The collected dataset is split into two distinct subsets: a training dataset, consisting of 15,418 images, accompanied

by 81,154 annotations, accounting for 90% of the initial dataset, and a validation dataset, which comprised 1,715 images containing 9,715 annotations, representing 10% of the original dataset. Table 3 presents the specifications of training and validation datasets. To ensure a fair and accurate comparison of model performance, training sessions have been conducted using this divided dataset for the five variations of the YOLOv8 model, along with the YOLOv7-Tiny model. These Python-based models have been trained in *PyTorch* with identical hyperparameters, enabling precise performance assessments. The training process spanned 300 epochs, utilizing a batch size of 32 and images down-scaled to dimensions of 640 × 640 pixels. The learning rate, a pivotal hyperparameter governing the training process and model convergence, has been set at the commonly recommended initial value of 0.01. The learning rate is adjusted automatically over training time where the final training rate reaches 0.0001. The first three epochs are utilized for learning rate warmup to increase the learning rate from a low value to the initial chosen value. In order to avoid overfitting, L2 regularization is applied. The training process has been conducted on GeForce RTX 3080 graphical processing unit from NVIDIA with 8704 CUDA cores and 10 GB memory. Several Data augmentation (DA) methods are applied during the training process. The hue is adjusted to attain generalization across different lighting and environmental conditions. Translation is applied in order to learn the model to recognize and detect partially visible objects. Scaling of images is also applied in order to simulate objects at different distances from the camera. Also, images are left and right flipped randomly in order to increase the diversity of the dataset along with long training process. Mosaic is applied by combining four images into one novel image before passing it to the model input. FIGURE 20 presents sample DA methods applied on images during the training process.

TABLE 3. Specifications of training and validation datasets.

Dataset	Training	Validation
Number of images	15,418	1,715
Number of annotations	81,154	9,715

To comprehensively evaluate the performance of the detection models, the well known object detection metrics are employed. Precision P signifies the proportion of correctly identified individuals among all proposed objects classified as people. Meanwhile, recall R gauges the percentage of individuals who were correctly detected among the total number of labeled individuals in the validation dataset.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

where TP, FP and FN stand for True Positive, False Positive and False Negative respectively.

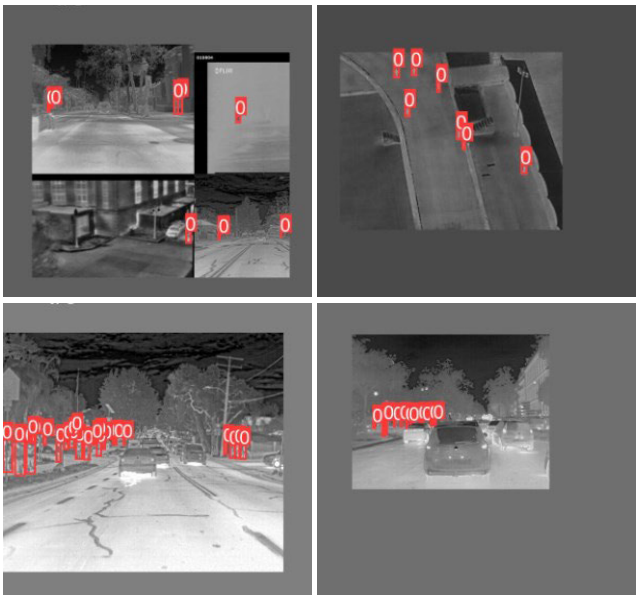

FIGURE 20. Sample DA methods applied on images during the training process.

The F1-score metric evaluates model accuracy by harmonizing P and R as expressed in the following equation:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{3}$$

Additionally, the average precision (AP) representing the area under the precision-recall curve for assessing the model's performance across trade-offs is integrated. $AP@50$ specifically evaluates precision and recall at an IoU threshold of 50%, measuring overlap between predicted and actual bounding boxes. The $AP@50 - 95$ range extends the assessment across IoU thresholds from 50% to 95%, enhancing comprehension of the model's robustness across various detection scenarios. The IoU is a fundamental metric used in object detection tasks to evaluate the accuracy of bounding box predictions. It measures the degree of overlap or intersection between the predicted bounding box and the ground truth bounding box for an object within an image. IoU is calculated as the ratio of the area of overlap between these two bounding boxes to the area of their union.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{4}$$

IoU values typically range from 0 to 1, where $IoU = 0$ indicates no overlap between the predicted and ground truth bounding boxes, meaning they are completely disjoint and $IoU = 1$ signifies a perfect match, where the predicted bounding box precisely matches the ground truth bounding box. In object detection tasks, a threshold IoU value is defined to determine whether a prediction is considered a true positive or a false positive. Predictions with IoU values above this threshold are considered correct detections, while those with IoU values below the threshold are treated as errors or

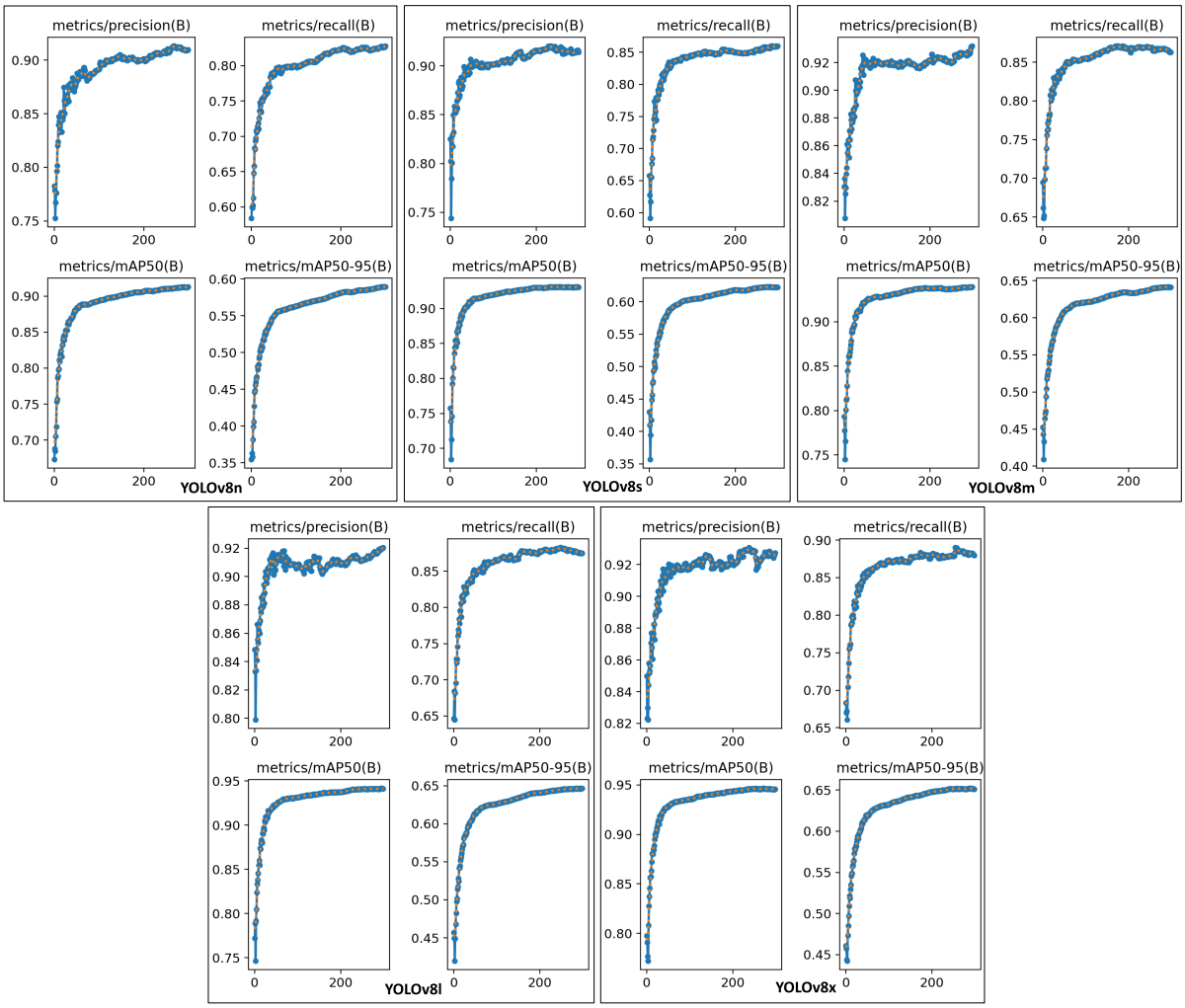

FIGURE 21. Training and validation performances of trained YOLOv8 models.

false positives. The choice of *IoU* threshold can influence the precision and recall of the object detection model and is an important parameter in model evaluation and training.

IV. RESULTS AND DISCUSSION

A. DETECTION PERFORMANCE

While conducting training, the models are also validated using the validation dataset. FIGURE 21 presents the training and validation performances of all trained YOLOv8 models. The figure shows that all models exhibit an initial rapid, exponential rise in its metrics, followed by a stabilization phase in the graph. Remarkably, all models consistently maintain precision and *AP*@50 scores exceeding 90%, highlighting their high detection performance. FIGURE 22 presents a sample of detections on the images in the validation dataset. It can be observed that the humans are accurately detected in the thermal images. In the figure, the '0' appearing on each human represents the class index, which is zero in this case, as only one class is being targeted. FIGURE 23 presents the precision-recall curves of YOLOv8 trained models, which demonstrate a tradeoff between precision and recall for different thresholds. The high areas under all curves represent the obtained high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate.

Table 4 presents the obtained performance metrics of the five trained models of YOLOv8 using our created dataset. Comparing the models, all consistently exhibit strong accuracy in human detection in diverse scenarios, with precision values ranging from 0.910 to 0.932, minimizing false positives. Their achieved recall values (0.828 to 0.879) ensure robustness in capturing actual positives. Notably, they achieve high *AP*) scores across varying *IoU* thresholds (0.589 to 0.651 for strict *IoU*, and 0.913 to 0.946 for lenient *IoU*), showcasing adaptability. Further comparison shows that YOLOv8x and YOLOv8m stand out with particularly the highest precision values (92.7% and 93.2% respectively), suggesting a superior ability to accurately discern humans from other objects in the scene. Additionally, as the size of the model decreases, the *AP* value decreases. This trend starts with the highest *AP* of YOLOv8x at 95% and the lowest *AP* of 91% with YOLOv8n.

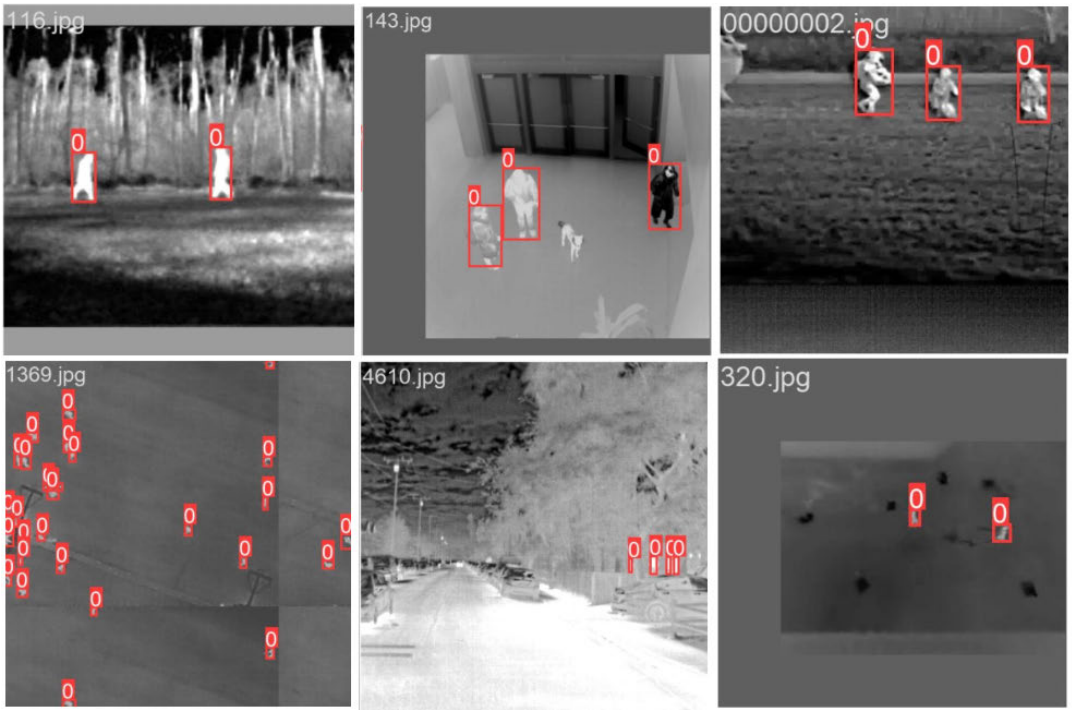

FIGURE 22. Sample detection results on the images of the validation dataset.

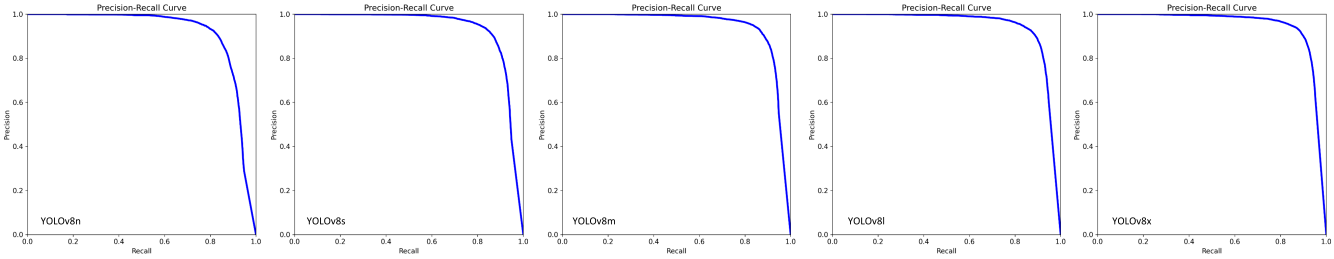

FIGURE 23. Precision-Recall curves of trained YOLOv8 models.

TABLE 4. Performance results of trained YOLOv8 and YOLOv7-Tiny models.

Model	Parameters (M)	FLOPs (B)	Precision	Recall	AP@50	AP@50-95	F1-score
YOLOv8x	68.2	257.8	0.927	0.879	0.946	0.651	0.902
YOLOv8l	43.7	165.2	0.921	0.875	0.941	0.646	0.897
YOLOv8m	25.9	78.9	0.932	0.863	0.939	0.641	0.896
YOLOv8s	11.2	28.6	0.914	0.860	0.931	0.622	0.886
YOLOv8n	3.2	8.7	0.910	0.828	0.913	0.589	0.867
YOLOv7-Tiny	6.2	5.8	0.893	0.848	0.917	0.555	0.869

The $F1$-score curve of three YOLO models, namely v8x, v8m, and v8n are displaced in FIGURE 24 reflecting that all have remarkable balanced performance. But YOLOv8n exhibits a rapid decrease in $F1$-score, which may lead to decreased detection accuracy in scenarios characterized by unpredictable lighting, weather, or heavy object occlusions. These challenges could result in missed detections or false positives especially in crowded or cluttered environments.

However, it's worth noting that YOLOv8n's trade-off for its smaller model size and faster processing speed might make these compromises justifiable in situations where computational efficiency or real-time processing is paramount. To mitigate its limitations, options include fine-tuning, ensemble learning, dynamic model selection, hybrid models, data augmentation, and advanced post-processing techniques, enabling users to balance speed and accuracy based on specific applications. YOLOv8m follows with

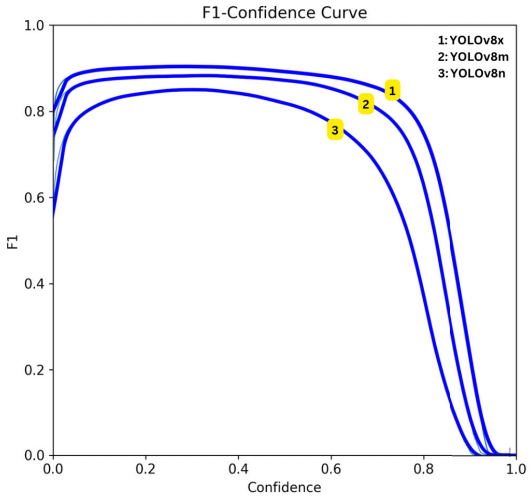

FIGURE 24. Comparative analysis of F1-Scores for 3 models of YOLOv8.

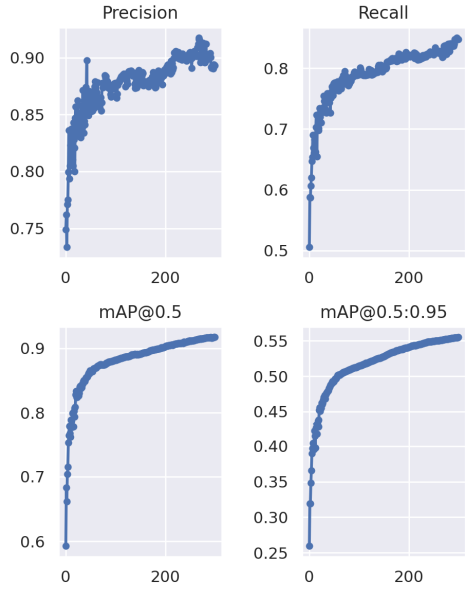

FIGURE 25. Training and validation performance of YOLOv7-Tiny model.

a more gradual decline, showcasing better stability than YOLOv8n. YOLOv8x, with the slowest rate of decrease, maintains a consistently high $F1$-score over varying conditions, suggesting its robustness and adaptability. The extent of the gaps between the curves underscores the disparities in performance. Considering the context, YOLOv8n's compromise in $F1$-score might be justifiable owing to its smaller size and faster processing. Still, it is important to note that even though YOLOv8n is approximately 95.31% smaller than YOLOv8x in terms of the number of parameters, as shown in FIGURE 17, YOLOv8n is only 3.49% less precise than YOLOv8x in terms of $AP@50$.

B. COMPARISON BETWEEN YOLOv8 AND YOLOv7

FIGURE 25 portrays the training and validation performances of the YOLO7-Tiny model as it is applied to human detection using our meticulously curated dataset. The graph captures a captivating phenomenon characterized by a rapid and exponential rise in a range of evaluation metrics. These metrics ultimately achieve commendable values, indicating the model's strong performance. However, a notable feature is the irregularity observed in the precision metric. While it displays an overall upward trajectory, it is marred by frequent fluctuations, raising questions about the model's stability in producing precise results.

Table 4 also provides a comprehensive overview of YOLO7-Tiny's performance metrics upon completing the training process. The outcomes highlight the model's robust performance, exemplified by an impressive $AP@50$ score of 91.7%. Nonetheless, upon closer scrutiny and a comparison with the metrics of YOLOv8, it becomes evident that YOLO7-Tiny lags in terms of precision, registering a score of 89.3%. However, it is worth noting that other metrics exhibit a closer alignment with YOLOv8n, while achieving an $F1$-score being nearly identical at 87%. Additionally, the $AP@50$ values for YOLOv8n and YOLO7-Tiny, at 91.3% and 91.7% respectively, are also quite close. This congruence is visually emphasized in FIGURE 26, which displays the $F1$-score curves for both YOLOv8 and YOLO7-tiny illustrating closely overlapping trends.

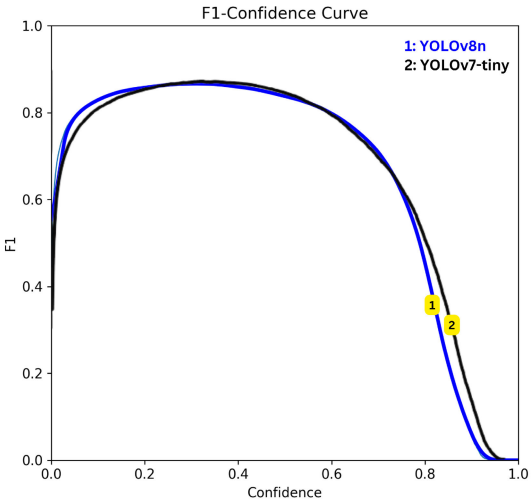

FIGURE 26. $F1$-Score for YOLOv8n vs. YOLOv7-Tiny.

A crucial aspect to emphasize is the considerable discrepancy in model parameter sizes. YOLO7-Tiny boasts a substantial parameter count of 6.2 millions, which is twice the number of parameters in YOLOv8n, standing at 3.2 millions as shown in FIGURE 17 and FIGURE 18. This substantial difference suggests that YOLOv8n is likely to operate at a significantly faster pace than YOLO7-Tiny. Consequently, when two models deliver comparable accuracy in object detection but one demonstrates superior efficiency in terms of processing speed,

FIGURE 27. Sample prediction taken from an aerial video captured using a IR thermal camera for humans moving in a forest: (a) snapshot from the video (b) resulting prediction using the trained YOLOv8 model.

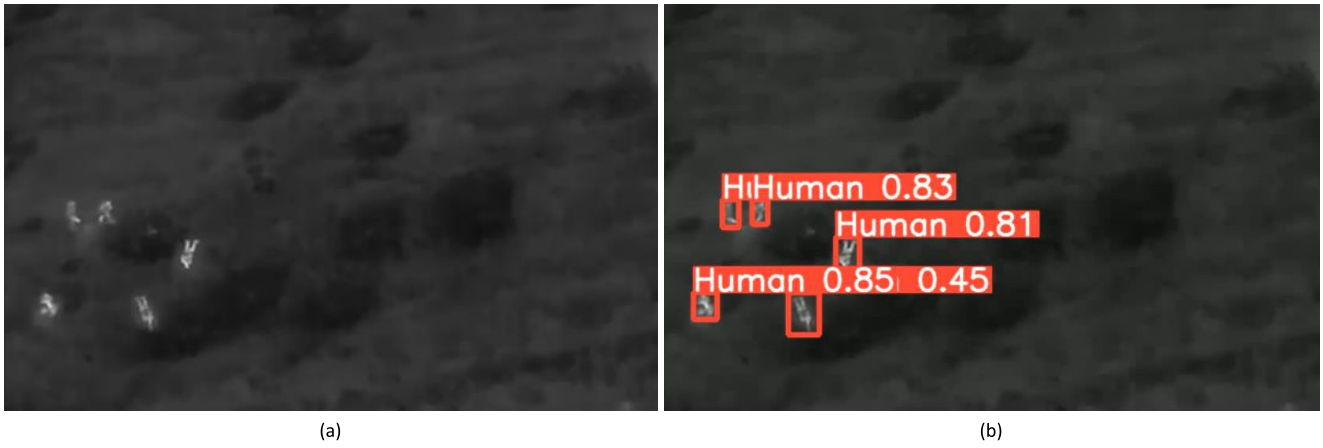

FIGURE 28. Sample predictions taken from an aerial video captured using a IR thermal camera for humans laying in a forest: (a) snapshot from the video (b) resulting prediction using the trained YOLOv8 model.

as exemplified by YOLOv8n, it positions the latter as a more appealing choice over YOLO7-Tiny for real-time applications and resource-constrained environments.

In summary, the conducted evaluation of YOLOv7-Tiny has also yielded commendable results achieving an *AP*@50 score of 91.7%. However, it is crucial to acknowledge that compared to YOLOv8 models, YOLOv7-Tiny exhibits slightly lower metric values with the exception of YOLOv8n. In the case of YOLOv8n, it is found that it offers a comparable performance while boasting a smaller model size, rendering it a more practical choice for real-world deployment. This decision is underpinned by the delicate balance that strikes between performance and efficiency.

The achieved results verify that YOLOv8 model architecture has undergone enhancements to improve its object detection capabilities. The anchor-free design of YOLOv8 architecture eliminates the need for the pre-defined anchor boxes, which simplifies the training process and enhances the model's ability to detect objects of different sizes and aspect ratios effectively. Moreover, the adoption of multi-scale prediction approach enables YOLOv8 models to perform predictions at multiple resolutions, which improves the capability to recognize objects at various scales, leading to improved object detection accuracy. As a summary, the refined architecture of YOLOv8 its performance improvements make it a superior choice over YOLOv7 for real-time object detection applications.

V. EVALUATION

The detection performances of the trained models are evaluated using a set of test video sequences captured by TIR cameras in diverse environments and showing human in different positions and scales. These video sequences are not provided to the models during training. FIGURE 27 and FIGURE 28 illustrate snapshots of predictions using the

FIGURE 29. Sample predictions using the trained YOLOv8 model for an aerial video capturing humans in urban environment.

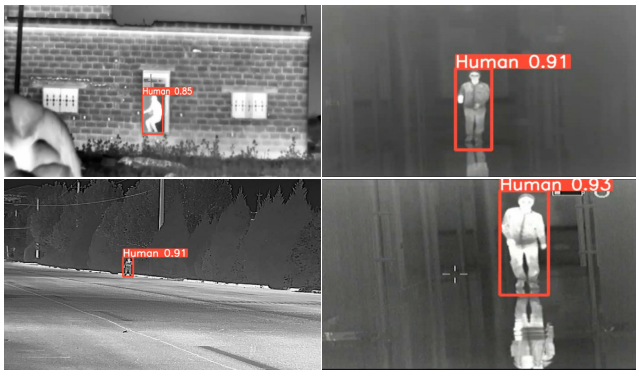

FIGURE 30. Sample predictions using the trained YOLOv8 model from several videos captured using IR thermal cameras showing humans in urban environment in different positions and scales.

FIGURE 31. Sample predictions of humans showing in several video sequences captured by thermal cameras from FLIR.

trained YOLOv8 model from an aerial video showing humans moving and laying in a forest respectively. Both figures demonstrate the capability of the trained models to detect all humans while achieving high confidence ratios. Also, no false negatives are recorded. The resultant predictions emphasize the ability to detect concurrently more than one human in the same scene despite their diverse size, position, rotation, and distinctness.

FIGURE 29 shows a set of resulting predictions in aerial videos showing humans in urban environment. It can be noticed that humans are detected and distinguished from other objects in the surrounding such as cars, trees, fence, lamppost, etc. In addition, FIGURE 30 presents several snapshots of predictions taken from several videos captured using IR thermal cameras showing humans in different urban environments (outdoor on the street, indoor inside a room and in a hall and standing exactly at the door). The humans are perfectly detected although they present in different scales and positions. Also, it can be noticed that the thermal reflections of humans are not detected falsely as humans. For further evaluation, FIGURE 31 shows the sample predictions using the trained YOLOv8 model of humans showing in several video sequences captured by thermal cameras from FLIR.

The models are also tested for video sequences showing the movement of humans between buildings as illustrated in FIGURE 32. The figure demonstrates that the models can predict the presence of the human with high confidence ratio

FIGURE 32. Sample predictions using the trained YOLOv8 model for an aerial video capturing human moving between buildings in an urban environment: (a) snapshot from the video (b) resulting prediction using the trained YOLOv8 model.

FIGURE 33. Several snapshots from the TIR video showing the human predicted while walking in the city using YOLOv8 trained model.

FIGURE 34. Sample predictions using the trained YOLOv8 model for an aerial video capturing human moving between cars: (a) snapshot from the video (b) resulting prediction using the trained YOLOv8 model.

even if the video scenes are captured from high elevation and the size of the human is small compared to the objects in the surrounding. FIGURE 33 provides additional snapshots from the TIR video showing the human predicted while walking in the city. FIGURE 34 provides an example of detecting human in an aerial TIR video capturing a human running beside cars towards the forest.

FIGURE 35 presents the prediction of a firefighter in an aerial video captured using TIR camera of a firefighting mission on a road passing in a forest. FIGURE 36 provides more snapshots showing the predictions of human from several angles of view. It can be noticed that the trained model can identify and localize efficiently the firefighter. FIGURE 37 illustrates the capability of the trained YOLOv8 model to distinguish between human body and other creatures with TIR immersions such as dogs. These figures demonstrate the capability of the trained model to detect and localize humans in different positions, various scales and diverse environments. Also, these predictions prove the ability to detect effectively more than one human in the same scene with high confidence ratios. Such evaluation cases, prove the importance of the proposed method to use deep learning with thermal images for SAR missions.

(a) (b)

FIGURE 35. Sample predictions using the trained YOLOv8 model for an aerial video capturing human during a firefighting mission: (a) snapshot from the video (b) resulting prediction using the trained YOLOv8.

FIGURE 36. Several snapshots showing the human predicted using YOLOv8 trained model in the TIR aerial video capturing human during a firefighting mission.

(a) (b)

FIGURE 37. Capability of the trained YOLOv8 model to distinguish a human from a dog.

(a) (b)

FIGURE 38. Miss-classification case.

VI. MISS-DETECTIONS AND LIMITATIONS

In order to evaluate the miss-detections, a thorough investigation has been conducted to the validation results. The number of false positives, the number of objects that are predicted as humans falsely is counted across the validation dataset. The investigation shows that a total count of 649 objects are predicted as humans from all 1709 images including 9715 humans. This illustrate the high achieved precession. Note that the prediction confidence ration is set to the default value of 0.25. Increasing its value will lead to the decrease of miss-detections. FIGURE 38 demonstrates a sample of false positive predictions. FIGURE 38(a) shows the real labels of humans in the image. FIGURE 38(b) shows the resultant predictions. It can be noticed that the model predicts falsely the presence human in the left middle of the image. however,

(a) (b)

FIGURE 39. Miss-classification case.

TABLE 5. Specifications of the used test video sequences.

Video Sequence	Aspect Ratio	Frame Rate	Resolution width × hight
Video 1	16:9	30	1280 × 720
Video 2	16:9	30	640 × 360
Video 3	16:9	30	480 × 270

TABLE 6. Specifications of targeted GPUs.

GPU Platform	CUDA Cores	Boost Clock	Memory Size	Memory Type
Nvidia Geforce RTX 3080	8704	1.71 GHz	10 GB	GDDR6X
Nvidia Geforce RTX 3060	3584	1.78 GHz	12 GB	GDDR6
Nvidia Geforce GTX 1080	2560	1.73 GHz	8 GB	GDDR5X

the confidence ratio is 0.4. Thus, this false prediction can be skipped when the confidence ration is increased.

In order to illustrate the potential failure cases, a large set of frames showing animals capturing using thermal cameras are passed through the trained network. FIGURE 39 shows a that the trained model predicts falsely a wolf as a human. However, the trained model does not fail to predict other wolves in the image (one laying down wolf and one standing wolf) and the cows in the upper part of the image. In fact, the miss-classified wolf has a similar shape of a human as the image shows its front facet. The wolf is posed a standing human. The only difference is its ears. This illustrates a challenging aspect to achieve optimal detection performance. A proposed solution is to fine tune the training of the models by adding a set of thermal images showing several species of animals in different positions.

VII. DEPLOYMENT ON GPU PLATFORMS

There video sequences with different resolutions are used to evaluate the inference speed, in terms of FPS, of the trained YOLOv8 models while deployed on set of high-end GPU platforms with various computational resources. The specifications of the videos are presented in Table 5. Table 6 summarizes the specifications of the targeted GPUs. The evaluation in conducted for different input image sizes noting that initial training of all models considers an image size of 640 × 640. During inference, the video frames are preprocessed to meet the selected dimension of input image size. Table 7, Table 8 and Table 9 presents respectively the achieved inference speeds in terms of FPS of the models when deployed on GeForce RTX 3080 GPU, GeForce RTX 3060 GPU and GeForce GTX 1080 GPU from NVIDIA. The tables show that YOLOv8 models can achieve high inference speeds that meet with the requirement of fast detection of humans in SAR missions.

TABLE 7. Average inference speed of trained YOLOv8 models when deployed on NVIDIA GeForce RTX 3080 GPU.

Image Size	YOLO Model	Average inference speed (FPS) Video1	Video2	Video3
320	YOLOv8n	119.78	141.84	141.06
	YOLOv8s	126.01	139.97	138.99
	YOLOv8m	109.2	119.72	117.43
	YOLOv8l	92.38	100.54	99.71
	YOLOv8x	92.89	100.29	99.48
480	YOLOv8n	119.0	131.34	137.42
	YOLOv8s	113.96	129.93	134.9
	YOLOv8m	107.01	111.14	114.69
	YOLOv8l	93.45	95.74	99.0
	YOLOv8x	89.75	92.07	93.96
640	YOLOv8n	119.67	133.59	123.91
	YOLOv8s	118.77	129.31	121.34
	YOLOv8m	103.08	111.34	105.13
	YOLOv8l	91.01	97.45	92.31
	YOLOv8x	53.1	54.44	52.88
800	YOLOv8n	109.22	117.6	119.02
	YOLOv8s	107.82	113.88	114.4
	YOLOv8m	85.42	89.49	91.0
	YOLOv8l	61.1	62.04	62.67
	YOLOv8x	46.71	47.11	47.29

TABLE 8. Average inference speed of trained YOLOv8 models when deployed on NVIDIA GeForce RTX 3060 GPUs.

Image Size	YOLO Model	Average inference speed (FPS)		
		Video1	Video2	Video3
320	YOLOv8n	243.49	273.6	260.25
	YOLOv8s	211.87	226.55	223.01
	YOLOv8m	167.13	163.87	158.09
	YOLOv8l	111.53	104.27	103.24
	YOLOv8x	83.56	81.21	79.7
480	YOLOv8n	229.0	235.7	244.05
	YOLOv8s	185.24	177.71	182.34
	YOLOv8m	103.25	102.26	103.26
	YOLOv8l	64.35	68.07	68.12
	YOLOv8x	49.2	48.51	48.84
640	YOLOv8n	213.3	203.77	186.7
	YOLOv8s	134.39	143.23	129.54
	YOLOv8m	78.09	76.36	74.08
	YOLOv8l	48.91	48.34	47.33
	YOLOv8x	32.76	32.29	31.77
800	YOLOv8n	156.64	151.04	144.43
	YOLOv8s	102.2	100.39	97.67
	YOLOv8m	52.7	52.09	51.82
	YOLOv8l	36.68	35.86	35.76
	YOLOv8x	21.82	21.77	21.67

TABLE 9. Average inference speed of trained YOLOv8 models when deployed on NVIDIA GeForce GTX 1080 GPUs.

Image Size	YOLO Model	Average inference speed (FPS)		
		Video1	Video2	Video3
320	YOLOv8n	100.95	119.59	117.81
	YOLOv8s	110.57	119.53	116.74
	YOLOv8m	75.27	76.76	75.92
	YOLOv8l	54.14	55.15	54.91
	YOLOv8x	41.95	42.45	42.25
480	YOLOv8n	108.42	113.19	114.8
	YOLOv8s	93.06	97.56	98.54
	YOLOv8m	58.27	59.63	60.43
	YOLOv8l	38.89	39.57	39.77
	YOLOv8x	28.19	28.54	28.67
640	YOLOv8n	103.53	112.21	103.22
	YOLOv8s	80.23	84.56	81.21
	YOLOv8m	45.37	47.0	45.96
	YOLOv8l	33.42	33.98	33.46
	YOLOv8x	23.8	24.18	23.82
800	YOLOv8n	98.37	105.61	103.18
	YOLOv8s	65.66	68.13	67.73
	YOLOv8m	38.76	39.55	39.3
	YOLOv8l	23.76	23.92	23.94
	YOLOv8x	16.74	16.84	16.81

VIII. CONCLUSION

This paper tackles the use of deep learning models in the realm of human detection with thermal images, paving the way for applications in critical scenarios like search and rescue missions. A pivotal component of this work centers on the creation of an innovative and meticulously annotated dataset, encompassing a diverse array of thermal images capturing humans in various scenarios and environmental conditions. All available YOLOv8 models are trained and validated using the collected dataset. The obtained results underscore the versatility of YOLOv8 models, showcasing their consistent ability to accurately identify humans across a broad spectrum of scenarios. Particularly noteworthy is the precision achieved by YOLOv8x of 93% and an $AP@50$ value of 95%. Remarkably, this high level of performance remains uncompromised even as the model size is reduced, with precision and $AP@50$ values consistently exceeding 91%. The detection performance of the models are evaluated using several TIR real-life video sequences capturing humans in diverse scenarios and angles of view. The trained YOLOv8 models are deployed on GPUs to examine the inference rates. The obtained inference rates along with the detection performance affirm the reliability of the proposed approach in enhancing the efficiency of search and rescue missions.

Future work will focus on deploying the trained models on embedded edge devices with reduced computational resources and power budget in order to enable on-site and field real-time detection during search and rescue missions.

ACKNOWLEDGMENT

The authors would like to thank Zahraa Hadwan for proofreading and editing this article. An earlier version of this paper was presented in part at the 2023 IEEE International Conference on Advances in Biomedical Engineering (ICABME) [DOI: 10.1109/ICABME59496.2023.10293139].

REFERENCES

[1] M. Rizk and I. Bayad, "Human detection in thermal images using YOLOv8 for search and rescue missions," in *Proc. 7th Int. Conf. Adv. Biomed. Eng. (ICABME)*, Oct. 2023, pp. 210–215.

[2] M. Rizk, F. Slim, and J. Charara, "Toward AI-assisted UAV for human detection in search and rescue missions," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Dec. 2021, pp. 781–786.

[3] M. Ivašić-Kos, M. Krišto, and M. Pobar, "Human detection in thermal imaging using YOLO," in *Proc. 5th Int. Conf. Comput. Technol. Appl.*, Apr. 2019, pp. 20–24.

[4] R. Ippalapally, S. H. Mudumba, M. Adkay, and H. R. N. Vardhan, "Object detection using thermal imaging," in *Proc. IEEE 17th India Council Int. Conf. (INDICON)*, Dec. 2020, pp. 1–6.

[5] M. Krišto, M. Ivasic-Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using YOLO," *IEEE Access*, vol. 8, pp. 125459–125476, 2020.

[6] A. N. Wilson, K. A. Gupta, B. H. Koduru, A. Kumar, A. Jha, and L. R. Cenkeramaddi, "Recent advances in thermal imaging and its applications using machine learning: A review," *IEEE Sensors J.*, vol. 23, no. 4, pp. 3395–3407, Feb. 2023.

[7] G. Batchuluun, J. K. Kang, D. T. Nguyen, T. D. Pham, M. Arsalan, and K. R. Park, "Deep learning-based thermal image reconstruction and object detection," *IEEE Access*, vol. 9, pp. 5951–5971, 2021.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Computer Vision—ECCV*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham, Switzerland: Springer, 2014, pp. 740–755.

[9] K. J. Havens and E. J. Sharp, "Chapter 8—Imager selection," in *Thermal Imaging Techniques to Survey and Monitor Animals in the Wild*, K. J. Havens and E. J. Sharp, Eds. London, U.K.: Academic Press, 2016, pp. 121–141.

[10] M. Kristo and M. Ivasic-Kos, "Thermal imaging dataset for person detection," in *Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectronics (MIPRO)*, May 2019, pp. 1126–1131.

[11] X. Wang, "Human detection in a sequence of thermal images using deep learning," M.S. thesis, Fac. Geo-Inf. Sci. Earth Observ., Univ. Twente, 2019.

[12] X. Wang and S. Hosseinyalamdary, "Human detection based on a sequence of thermal images using deep learning," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 117–122, Jan. 2019.

[13] B. B. Lahiri, S. Bagavathiappan, T. Jayakumar, and J. Philip, "Medical applications of infrared thermography: A review," *Infr. Phys. Technol.*, vol. 55, no. 4, pp. 221–235, Jul. 2012.

[14] R. Ishimwe, K. Abutaleb, and F. Ahmed, "Applications of thermal imaging in agriculture—A review," *Adv. Remote Sens.*, vol. 3, no. 3, pp. 128–140, Jan. 2014.
[15] A. S. Bhadoriya, V. Vegamoor, and S. Rathinam, "Vehicle detection and tracking using thermal cameras in adverse visibility conditions," *Sensors*, vol. 22, no. 12, p. 4567, Jun. 2022.
[16] C. Tian, Z. Zhou, Y. Huang, G. Li, and Z. He, "Cross-modality proposal-guided feature mining for unregistered RGB-thermal pedestrian detection," *IEEE Trans. Multimedia*, vol. 26, pp. 6449–6461, 2024.
[17] Z. Zhou, S. Wu, G. Zhu, H. Wang, and Z. He, "Channel and spatial relation-propagation network for RGB-thermal semantic segmentation," 2023, *arXiv:2308.12534*.
[18] Q. Liu, X. Li, Z. He, C. Li, J. Li, Z. Zhou, D. Yuan, J. Li, K. Yang, N. Fan, and F. Zheng, "LSOTB-TIR: A large-scale high-diversity thermal infrared object tracking benchmark," in *Proc. 28th ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, Oct. 2020, pp. 3847–3856.
[19] A. Gomaa, M. M. Abdelwahab, and M. Abo-Zahhad, "Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis," *Multimedia Tools Appl.*, vol. 79, nos. 35–36, pp. 26023–26043, Sep. 2020.
[20] H. Zhang, E. Fromont, S. Lefevre, and B. Avignon, "Guided attentive feature fusion for multispectral pedestrian detection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 72–80.
[21] M. Salem, A. Gomaa, and N. Tsurusaki, "Detection of earthquake-induced building damages using remote sensing data and deep learning: A case study of Mashiki Town, Japan," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2023, pp. 2350–2353.
[22] A. Gomaa and A. Abdalrazik, "Novel deep learning domain adaptation approach for object detection using semi-self building dataset and modified YOLOv4," *World Electric Vehicle J.*, vol. 15, no. 6, p. 255, Jun. 2024.
[23] M. Rizk, D. Heller, R. Douguet, A. Baghdadi, and J.-P. Diguet, "Optimization of deep-learning detection of humans in marine environment on edge devices," in *Proc. 29th IEEE Int. Conf. Electron., Circuits Syst. (ICECS)*, Oct. 2022, pp. 1–4.
[24] E. Valldor, "Person detection in thermal images using deep learning," Dept. Inf. Technol., Institutionen för Informationsteknologi, Uppsala , Sweden, Tech. Rep., 2018. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1275338/FULLTEXT01.pdf
[25] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.
[26] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2001, pp. 1–11.
[27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, San Diego, CA, USA, Jun. 2005, pp. 886–893.
[28] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[29] D. Heller, M. Rizk, R. Douguet, A. Baghdadi, and J-P. Diguet, "Marine objects detection using deep learning on embedded edge devices," in *Proc. IEEE Int. Workshop Rapid Syst. Prototyping (RSP)*, Oct. 2022, pp. 1–7.
[30] D. Kalita, "Basics of CNN in deep learning," Tech. Rep., 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/
[31] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
[32] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[33] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[35] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
[36] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[37] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[38] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.
[39] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
[40] G. Jocher and A. Chaurasia, "Ultralytics YOLOv8 docs," Tech. Rep., Dec. 2023. [Online]. Available: https://docs.ultralytics.com/
[41] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Computer Vision—ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham, Switzerland: Springer, 2016, pp. 21–37.
[42] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[43] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10781–10790.
[44] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6568–6577.
[45] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 642–656, Mar. 2020.
[46] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Comput. Vission—ECCV*, 2020, pp. 213–229.
[47] M. A. Ansari and D. K. Singh, "Human detection techniques for real time surveillance: A comprehensive survey," *Multimedia Tools Appl.*, vol. 80, no. 6, pp. 8759–8808, 2021.
[48] D. T. Nguyen, W. Li, and P. O. Ogunbona, "Human detection from images and videos: A survey," *Pattern Recognit.*, vol. 51, pp. 148–175, Mar. 2016.
[49] J. W. Davis, V. Sharma, A. Tyagi, and M. Keck, *Human Detection and Tracking*. Boston, MA, USA: Springer, 2009, pp. 708–712.
[50] M. Rizk, F. Slim, A. Baghdadi, and J.-P. Diguet, "Towards real-time human detection in maritime environment using embedded deep learning," in *Advances in System-Integrated Intelligence*, M. Valle, D. Lehmhus, C. Gianoglio, E. Ragusa, L. Seminara, S. Bosse, A. Ibrahim, and K.-D. Thoben, Eds., Cham, Switzerland: Springer, 2023, pp. 583–593.
[51] W. Wang, J. Zhang, and C. Shen, "Improved human detection and classification in thermal images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2313–2316.
[52] W. Li, D. Zheng, T. Zhao, and M. Yang, "An effective approach to pedestrian detection in thermal imagery," in *Proc. 8th Int. Conf. Natural Comput.*, May 2012, pp. 325–329.
[53] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION)*, Jan. 2005, pp. 364–369.
[54] I. Riaz, J. Piao, and H. Shin, "Human detection by using CENTRIST features for thermal images," in *Proc. Int. Conf. Comput. Graph., Visualizat., Comput. Vis. Image Process.*, 2013, pp. 1–11.
[55] J. Wu and J. M. Rehg, "CENTRIST: A visual descriptor for scene categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1489–1501, Aug. 2011.
[56] A. Lakshmi, A. G. J. Faheema, and D. Deodhare, "Pedestrian detection in thermal images: An automated scale based region extraction with curvelet space validation," *Infr. Phys. Technol.*, vol. 76, pp. 421–438, May 2016.
[57] R. Kalita, A. K. Talukdar, and K. K. Sarma, "Real-time human detection with thermal camera feed using YOLOv3," in *Proc. IEEE 17th India Council Int. Conf. (INDICON)*, Dec. 2020, pp. 1–5.
[58] P.-F. Tsai, C.-H. Liao, and S.-M. Yuan, "Using deep learning with thermal imaging for human detection in heavy smoke scenarios," *Sensors*, vol. 22, no. 14, p. 5351, Jul. 2022.
[59] K. R. Akshatha, A. K. Karunakar, S. B. Shenoy, A. K. Pai, N. H. Nagaraj, and S. S. Rohatgi, "Human detection in aerial thermal images using faster R-CNN and SSD algorithms," *Electronics*, vol. 11, no. 7, p. 1151, Apr. 2022.
[60] S. Yeom, "Thermal image tracking for search and rescue missions with a drone," *Drones*, vol. 8, no. 2, p. 53, Feb. 2024.
[61] M. Li, Z. Zhang, L. Lei, X. Wang, and X. Guo, "Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD," *Sensors*, vol. 20, no. 17, p. 4938, Aug. 2020.
[62] J.-A. Kim, J.-Y. Sung, and S.-H. Park, "Comparison of faster-RCNN, YOLO, and SSD for real-time vehicle type recognition," in *Proc. IEEE Int. Conf. Consum. Electron.*, Nov. 2020, pp. 1–4.

[63] A. C. Rios, D. H. dos Reis, R. M. da Silva, M. A. S. L. Cuadros, and D. F. T. Gamarra, "Comparison of the YOLOv3 and SSD MobileNet v2 algorithms for identifying objects in images from an indoor robotics dataset," in *Proc. 14th IEEE Int. Conf. Ind. Appl. (INDUSCON)*, Aug. 2021, pp. 96–101.
[64] M. Kristo, M. Ivasic-Kos, and M. Pobar, "Thermal image dataset for person detection—UNIRI-TID," *IEEE Dataport*, 2020, doi: 10.21227/yec9-yy29.
[65] (Jan. 2022). *Teledyne FLIR*. Accessed: Mar. 1, 2024. [Online]. Available: https://adas-dataset-v2.flirconservator.com/
[66] J. Suo, T. Wang, X. Zhang, H. Chen, W. Zhou, and W. Shi, "HIT-UAV: A high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection," *Sci. Data*, vol. 10, no. 1, p. 227, Apr. 2023.
[67] Q. Liu, Z. He, X. Li, and Y. Zheng, "PTB-TIR: A thermal infrared pedestrian tracking benchmark," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 666–675, Mar. 2020.
[68] *The Person and Dog Thermal Images Dataset*, Mar. 2021. Accessed: Mar. 1, 2024. [Online]. Available: https://public.roboflow.com/object-detection/thermal-dogs-and-people
[69] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics," Tech. Rep., 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
[70] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," 2023, *arXiv:2304.00501*.
[71] F. Wang, J. Jiang, Y. Chen, Z. Sun, Y. Tang, Q. Lai, and H. Zhu, "Rapid detection of Yunnan Xiaomila based on lightweight YOLOv7 algorithm," *Frontiers Plant Sci.*, vol. 14, Jun. 2023, Art. no. 1200144.

MOSTAFA RIZK received the Maîtrise-ès degree in electronics, the M.Sc. degree in biomedical physics, and the M.Sc. degree in signal, telecom, image, and speech from Lebanese University, in 2007, 2008, and 2010, respectively, the Ph.D. degree in sciences and technologies of information and communication (STIC) from IMT-Atlantique (former Telecom Bretagne), in 2014, the Ph.D. degree in electronics and telecommunication from Lebanese University, and the Habilitation to Direct Research/Habilitation à diriger des recherches (HDR) degree in STIC from the University of South Brittany (UBS), France, in 2022. He was a Research and Development Engineer and later a Postdoctoral Fellow at UBS University, France. He was an Associate Researcher at IMT-Atlantique, France, from 2017 to 2020. In 2021 and 2022, he joined the Lab-STICC, French National Center of Scientific Research (CNRS), as a Researcher, which is a research unit historically recognized in France in the field of ICT. He has been an Associate Professor at Lebanese International University and Lebanese University, since 2016. His general research interests include both algorithm development and corresponding hardware/software implementations and digital circuit design; NoC design; new MPSoC architectures based on emerging non-volatile memory technologies; embedded machine learning; embedded intelligence; and embedded computer vision.

ISRAA BAYAD received the B.S. degree in general physics and the M.Sc. degree in medical physics and life imaging from Lebanese University, in 2020 and 2023, respectively. Her research interest includes the application of artificial intelligence, particularly deep learning techniques, in biomedical physics for the development of detection tools and data processing methods.