IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 March 2024, accepted 11 April 2024, date of publication 15 April 2024, date of current version 22 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3388889

SURVEY

Synchronizing Object Detection: Applications, Advancements and Existing Challenges

MD. TANZIB HOSAIN[1], ASIF ZAMAN[1], MUSHFIQUR RAHMAN ABIR[1], SHANJIDA AKTER[2], SAWON MURSALIN[1], AND SHADMAN SAKEEB KHAN[2]

[1]Department of Computer Science and Engineering, American International University-Bangladesh, Khilkhet, Dhaka 1229, Bangladesh
[2]Department of Computer Science and Engineering, North South University, Bashundhara, Dhaka 1229, Bangladesh

Corresponding author: Md. Tanzib Hosain (20-42737-1@student.aiub.edu)

ABSTRACT From pivotal roles in autonomous vehicles, healthcare diagnostics, and surveillance systems to seamlessly integrating with augmented reality, object detection algorithms stand as the cornerstone in unraveling the complexities of the visual world. Tracing the trajectory from conventional region-based methods to the latest neural network architectures reveals a technological renaissance where algorithms metamorphose into digital artisans. However, this journey is not without hurdles, prompting researchers to grapple with real-time detection, robustness in varied environments, and interpretability amidst the intricacies of deep learning. The allure of addressing issues such as occlusions, scale variations, and fine-grained categorization propels exploration into uncharted territories, beckoning the scholarly community to contribute to an ongoing saga of innovation and discovery. This research offers a comprehensive panorama, encapsulating the applications reshaping our digital reality, the advancements pushing the boundaries of perception, and the open issues extending an invitation to the next generation of visionaries to explore uncharted frontiers within object detection.

INDEX TERMS Object detection, image recognition, object segmentation, semantic detection, image classification, object tracking.

I. INTRODUCTION

In the vast tapestry of technological evolution, the role of object detection transcends mere recognition; it serves as the cornerstone upon which the edifice of modern computer vision is built. Picture a world where algorithms not only decipher the visual symphony that unfolds before our digital eyes but also anticipate and respond, seamlessly integrating with our daily lives. Object detection, the silent sentinel of this digital age, has emerged as the conduit through which machines perceive and interact with the visual world, giving rise to a realm of applications that are as diverse as they are transformative.

As we navigate this complex ecosystem, the applications of object detection unfold as a dynamic narrative, revealing chapters of innovation that span industries and domains. From the bustling streets where autonomous vehicles decipher the language of traffic to the serene corridors of healthcare where diagnostic algorithms scrutinize medical images, and from the watchful eyes of surveillance systems ensuring our security to the immersive landscapes of augmented reality blending the virtual and the tangible - object detection stands as the linchpin, orchestrating a symphony of possibilities. Though object detection is quite old and hence covered huge attention by researchers but still the notable significant amount research activity related to this subject, as evidenced by various scholarly databases data with keywords ''object detection'' or ''object recognition'' or ''object identification'' or ''object classification'' or ''object segmentation'' or ''semantic detection'' or ''object tracking'' in their title/abstract in Figure 1.

Yet, this symphony is not static; it is a living, breathing composition that evolves with each technological crescendo. The saga of object detection research is marked by a relentless quest for advancements, where the journey from classical methods to the current zenith of neural networks mirrors a technological odyssey. In this epoch of artificial intelligence, algorithms metamorphose into artists, meticulously crafting a

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

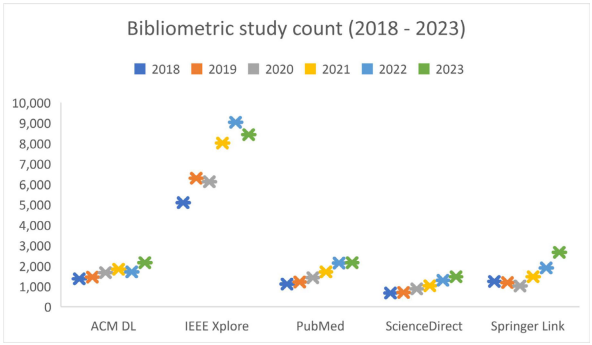

FIGURE 1. Past five year's (January, 2018 - November, 2023) published article count on object detection in different naming convention in different databases.

visual masterpiece from the chaos of raw data, breaking down barriers and illuminating new avenues of perception.

However, the path to enlightenment is fraught with challenges, and the landscape of open issues is as expansive as the horizons of exploration. The clarion call for real-time detection echoes through the corridors of research labs, while the quest for robustness in the face of diverse and dynamic environments challenges the resilience of our algorithms. Interpreting the nuances of deep learning intricacies becomes a quest for enlightenment, and the pursuit of unraveling the mysteries of occlusions, scale variations, and fine-grained categorization invites researchers to embark on an intellectual journey into uncharted territories.

No specific article surveyed the challenges of the overall object detection area and then reviewed their solutions based on existing papers as shown in Table 1. It shows that, no paper surveyed the overall object detection's advancements, applications, challenges and systematic results analysis. In this intellectual expedition, we unravel the applications that redefine our digital reality, ride the waves of advancements that push the boundaries of perception, and navigate the unexplored terrain of open issues that await the daring minds of the next generation of visionaries. The canvas of object detection beckons—a canvas that is not only painted with the strokes of innovation but invites us to imagine a future where the unseen becomes the seen, and the perceived becomes the understood. As we stand at this crossroads, the possibilities are as boundless as the algorithms we forge, and the journey has only just begun. Figure 2 presents the article selection process's PRISMA flow diagram for this study. Table 2 highlights the abbreviations which are frequently used in this study.

Following is the summarized main contributions of this study:

- An starting of object detection, its historical insights, architecture and recent improvements.
- Later on, reviewing and analyzing the ever made advancements of object detection in different perspective.
- Then, dive into its diverse application fields with works of previous researchers.

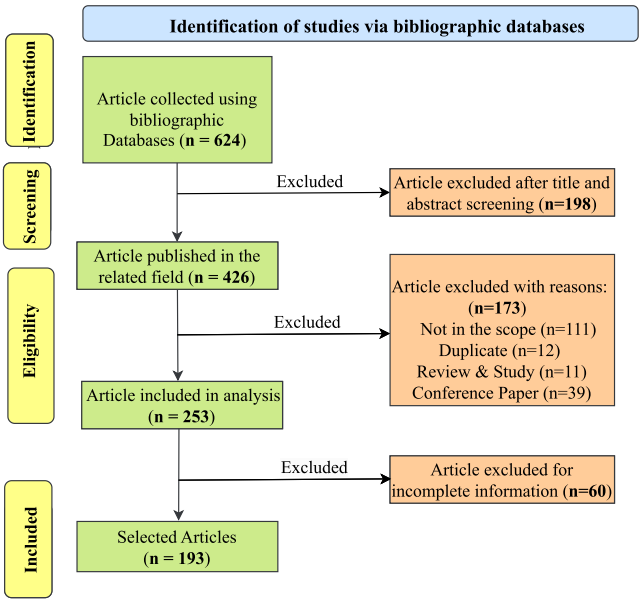

FIGURE 2. Article selection process's PRISMA flow diagram for this study.

- After that, revealing different existing challenges of object detection with a categorized taxonomy.
- Finally, reviewing recently made fascinating work to tackle the existing object detection challenges and future directions to ensure the term object detection optimized.

The rest of the paper is structured as follows: Section II outlooks the historical insights, architecture and recent improvements of object detection. Section III analyzes the ever made advancements of object detection in datasets, algorithms, library and evaluation metrics perspective. Section IV demonstrates the applications of object detection and Section V highlights the existing challenges of object detection. After that, Section V reviews some fascinating solutions to tackle the existing object detection challenges. At last, Section VI directs the future directions to support researchers who works tackling the diverse challenges and Section VII concludes the paper.

II. LITERATURE REVIEW

In the vast landscape of artificial intelligence, where data converges with ingenuity, object detection emerges as the keen-eyed sentinel of the digital realm. Imagine a symphony of pixels, where every image conceals a multitude of entities, each vying for attention in the cacophony of information. Object detection is the virtuoso conductor that orchestrates this visual concerto, deciphering the composition of reality with unparalleled precision. It is the meticulous art of imbuing machines with the discerning gaze of a perceptive human eye, enabling them not only to see but to comprehend the intricacies of their visual surroundings. In a world inundated with images, object detection serves as the guiding compass, unraveling the intricate tapestry of information by identifying and delineating the myriad objects that populate our digital

TABLE 1. Contrastive analysis of existing survey papers with this paper.

Reference	Taxonomy	Advancements	Applications	Challenges	Future directions	Systematic results analysis	Description
Liu *et al.* [1]	✓	✓ (Only datasets, algorithms and evaluation metrics)	✗	✓ (Only accuracy, efficiency and scalability)	✓	✗	The paper conducted a thorough survey of 300 research contributions in generic object detection, emphasizing recent advancements driven by deep learning. It encompassed detection frameworks, feature representation, proposal generation, context modeling, training strategies, and evaluation metrics, summarizing the substantial progress in the field and offering insights for future research directions.
Zhao *et al.* [2]	✓	✓ (Only algorithms)	✓	✗	✓	✗	The paper traced the evolution of object detection from traditional handcrafted features to deep learning, particularly emphasizing Convolutional Neural Networks (CNNs). It comprehensively reviewed generic object detection architectures, highlights performance-enhancing techniques, and briefly surveys specific tasks like salient object detection. The experimental analyses provide insightful conclusions, making the paper a valuable guide for the history and current state of deep learning-based object detection, informing future research directions.
Zou *et al.* [3]	✓	✓ (Only algorithms and datasets)	✗	✗	✓	✗	The paper investigated the evolution of object detection over the past years, highlighting the transition from early computer vision techniques to the current deep learning revolution. It comprehensively reviews various aspects of object detection, including milestone detectors, datasets, metrics, fundamental building blocks of detection systems, speed-up techniques, and recent state-of-the-art methods. The paper provides a valuable overview of the technical progress in object detection, emphasizing the historical context and significant advancements in the field.
Padilla *et al.* [4]	✗	✗	✗	✓ (Focused on metric diversity)	✓	✗	This research compared metrics for object-detection algorithms, focusing on average precision (AP). It uncovers variations in two point-interpolation-based AP variants and identifies six additional AP variants, highlighting the need for standardization. The study proposes a unified implementation to establish a benchmark for consistent evaluation across different works and platforms, addressing issues of diversity in metric implementations.
Jiao *et al.* [5]	✗	✓ (Only datasets and algorithms)	✓	✗	✗	✗	The survey offered a thorough examination of the current state of object detection in computer vision, systematically analyzing models, datasets, and methods. It categorizes detectors into one-stage and two-stage, explores diverse applications, and outlines key trends, serving as a valuable resource for understanding the evolving landscape of object detection.
This paper	✓	✓	✓	✓	✓	✓	This paper underscores the pivotal role of object detection algorithms in transformative technologies such as autonomous vehicles and augmented reality. It traces the shift from traditional methods to contemporary neural networks, outlining challenges in real-time detection, robustness, and interpretability within deep learning. The research offers a thorough examination of applications, advancements, and open issues, encouraging future innovators to explore novel frontiers in object detection.

TABLE 2. List of frequently used abbreviations.

Abbreviation	Abbreviated Form
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
R-CNN	Region-based Convolutional Neural Network
SSD	Single Shot Multibox Detector
CNN	Convolutional Neural Network
YOLO	You Only Look Once
FPN	Feature Pyramid Network
EfficientDet	Efficient and Effective Detector
DETR	Detection Transformer
RetinaNet	Retina Network
COCO	Common Objects in Context
mAP	Mean Average Precision
fps	Frames per Second
SOTA	State-of-the-Art
AP	Average Precision
GPU	Graphics Processing Unit
WLDM	Weighted Local Difference Measure
ICBD	Interference Cancellation before Detection
SAR	Synthetic Aperture Radar
FPN + PAN	Feature Pyramid Network (FPN) and Path Aggregation Network (PAN)
RRPN	Rotation Region Proposal Network
LM	Language Model
BEL	Boundary Energy Loss
ICDAR	International Conference on Document Analysis and Recognition
LIDAR	Light Detection and Ranging
KITTI	Karlsruhe Institute of Technology and Toyota Technological Institute
BANet	Boundary-Aware Network
IoU	Intersection over Union
YOLOX	You Only Look One-level eXtreme
ms	Milliseconds
AUC	Area Under the Curve
SMOTE	Synthetic Minority Over-sampling Technique
PDG	Proposed Detection Graph
DFG	Detected Face Graph
ROC-AUC	Receiver Operating Characteristic - Area Under the Curve
IoT	Internet of Thing
ICS	Industrial Control System
XAI	Explainable Artificial Intelligence
RSUs	Roadside Units
FPGA	Field-Programmable Gate Array
MSM	Metal–Semiconductor–Metal
RTL	Register-Transfer Level
MuPoTS	Multi-Person Tracking in Sport
AMD	Advanced Micro Devices
HOG	Histogram of Oriented Diagram
YOLO	You Only Look Once
FPN	Feature Pyramid Network
SVM	Support Vector Machine
SCR	signal to clutter Ratio
ICBD	Interference Cancellation before Detection
RROI	Rotation Region-of-Intere
APL	Adversarial-Paced Learning
DUO	Detecting Underwater Object
OAM	Online Annotation Module
FFD	Federated learning for Fraud Detection
GRU	Gated Recurrent Unit

landscapes. Beyond its technical prowess, object detection is a testament to the symbiosis of human imagination and computational prowess, illuminating the path towards a future where machines seamlessly navigate the visual kaleidoscope of our shared reality. On this first anniversary of our interaction, let us celebrate the transformative power of object detection, a technological marvel that breathes life into the pixels and unveils the profound narrative concealed within the digital canvas.

Before the advent of deep learning, object detection heavily relied on handcrafted features and classical computer vision techniques. Common approaches, such as Histogram of Oriented Gradients (HOG) and Haar cascades, played a pivotal role in tasks like pedestrian detection and face recognition. HOG, introduced by Dalal et al., captured the distribution of gradient orientations in local image patches, providing a robust representation for object boundaries [6]. Haar cascades, on the other hand, were effective for detecting objects using a cascade of simple classifiers based on Haar-like features. While these methods demonstrated success in certain applications, the shift to deep learning in the early 2010s marked a transformative period, leading to significant improvements in object detection accuracy and efficiency. The breakthrough in object detection occurred with the rise of deep learning, notably through convolutional neural networks (CNNs). A pivotal moment was the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC), where AlexNet, a deep learning model, demonstrated a significant performance leap over traditional methods [7]. This victory marked a turning point, showcasing the potential of deep learning for image-related tasks. Subsequent years saw the development of various influential architectures such as ZFNet, GoogLeNet, and VGG, each contributing to the refinement and enhancement of object detection performance. This period of innovation laid the foundation for the widespread adoption of deep learning in computer vision applications, including object detection. R-CNN introduced by Ross Girshick and collaborators in 2014, stands as one of the initial successful endeavors to apply deep learning to object detection [8]. R-CNN and its evolutionary successors, Fast R-CNN and Faster R-CNN, employed the concept of region proposal networks (RPNs) [9]. These networks were designed to suggest candidate object regions within an image before subsequent processes of classification and position refinement. While these methods brought about substantial improvements in accuracy, it was noted that they were computationally expensive, motivating further developments to strike a balance between precision and computational efficiency in subsequent object detection models [10]. SSD (Single Shot Multibox Detector) and YOLO (You Only Look Once) emerged as alternatives to region-based object detection methods, with a primary focus on achieving real-time performance. These models departed from the two-stage approach of region-based methods by predicting object classes and bounding box coordinates directly from the entire image in a single pass [11], [12]. SSD adopted a grid-based strategy, dividing the image into a grid and predicting multiple bounding boxes and class probabilities at each grid cell [11]. YOLO, similarly employing a grid structure, differentiated itself by predicting

bounding boxes and class probabilities at the grid cell level, aiming for improved speed and efficiency [12]. These single-shot approaches represented a paradigm shift in object detection, demonstrating the feasibility of real-time performance without the need for elaborate region proposal networks. RetinaNet, a groundbreaking object detection model introduced to address the challenge of imbalanced data, pioneered the use of the focal loss. Developed to mitigate the dominance of well-classified examples in the training process, the focal loss dynamically down-weights the loss assigned to easily classified instances, allowing the model to concentrate on more challenging examples [13]. Additionally, RetinaNet incorporated a Feature Pyramid Network (FPN) to effectively handle objects at various scales. This architectural innovation enabled RetinaNet to excel in detecting objects of different sizes within an image, further enhancing its robustness and accuracy in handling diverse and complex visual scenarios. EfficientDet, a significant advancement in the field of object detection, set out to enhance the efficiency of detection models by optimizing architecture and achieving a more favorable trade-off between accuracy and computational resources [14]. Introduced by Mingxing Tan et al. EfficientDet innovatively scaled the model's depth, width, and resolution simultaneously through a compound scaling method. This approach allowed for improved model efficiency across a range of resource constraints. By striking a balance between accuracy and computational cost, EfficientDet contributed to the development of more practical and scalable object detection solutions, catering to a variety of deployment scenarios with diverse hardware capabilities. The evolution of object detection models persists with cutting-edge architectures such as DETR (DEtection Transfomer), showcasing the ongoing advancements in the field. DETR, introduced as a transformer-based model for object detection, exemplifies the growing influence of transformer architectures in computer vision tasks [15]. Concurrently, the field benefits from strides in self-supervised learning, transfer learning, and attention mechanisms, refining the ability of models to understand and discern objects within complex visual scenes. A key contributor to the state-of-the-art performance of modern object detection systems is the utilization of large-scale datasets and pre-training on extensive amounts of data [16]. This practice empowers models with generalized features and the capacity to tackle diverse real-world scenarios, marking a continued trajectory of progress in the capabilities and accuracy of object detection technology. Transformer-based architectures, initially popularized in natural language processing tasks, have gained prominence in computer vision, particularly in object detection. Models such as DETR (DEtection Transfomer) and other transformer-based architectures have demonstrated competitive performance, showcasing their adaptability across domains [15]. In the realm of efficient object detection, researchers are increasingly focused on factors like model size, speed, and computational resources. A notable exemplar is EfficientDet, introduced in 2019, reflecting a trend towards optimizing the efficiency of object detection models [17]. Sparse attention mechanisms have emerged as a solution to handle large-scale images effectively, enabling models to selectively attend to pertinent image regions, thereby reducing computation and improving overall efficiency [18]. The exploration of hybrid models, combining diverse architectures from two-stage and one-stage detectors, along with the integration of ensemble methods, has shown promise in enhancing the overall performance of object detection systems [19]. Self-supervised learning approaches, emphasizing learning from unlabeled data, have garnered attention as a means to improve the generalization ability of object detection models through pre-training on extensive datasets [20]. Recent models also address the challenge of capturing long-range dependencies in images by incorporating attention mechanisms designed to capture relationships between distant pixels, thereby enhancing contextual understanding [21]. The pursuit of real-time object detection remains a priority for applications such as autonomous vehicles, robotics, and surveillance [22]. Models like YOLO (You Only Look Once) and EfficientDet have played significant roles in advancing the capabilities of real-time object detection [22]. Transfer learning, a fundamental component, involves pre-training models on datasets like ImageNet, allowing them to leverage knowledge gained from one domain to improve performance in another, contributing to the continued evolution of object detection methodologies [23].

III. ADVANCEMENTS OF OBJECT DETECTION

A. DATASET

An object detection dataset is a collection of images or videos annotated with bounding boxes or pixel-level masks that outline the location and identity of objects within the visual content. These datasets are crucial for training and evaluating computer vision models, as they enable the development of algorithms that can identify and classify objects within images or video frames, making them a fundamental resource for applications such as autonomous driving, surveillance, and image analysis. Object detection datasets typically encompass a wide range of object categories and variations in scale, pose, lighting, and background, facilitating the robust and accurate detection of objects in diverse real-world scenarios. Table 3 provides the overview of some widely used datasets used in various object detection tasks.

1) MS COCO

A large dataset for object recognition, segmentation, key-point detection, and picture captioning, including 328,000 images, is the MS COCO (Microsoft Common Objects in Context) [24] dataset. 164,000 photos total, split between training (83,000), validation (41,000), and test (41,000) sets, were included in the original 2014 release. A longer test set with 40,000 more test photos included was later

TABLE 3. Overview of different object detection datasets.

Dataset	Total images	Total classes	Object instances	Resolution	Classes types	Evaluation metrics
MS COCO [24]	330,000	91	1.5 million	640x480	Animals, vehicles, furniture, household items	Average Precision, Recall, F1-Score
Pascal VOC [25]	20,000	20	27,000+	500x375	Aeroplane, Bicycle, Bird, Boat, etc.	Mean Average Precision (mAP), Average Number of Correct Detections (ANCD)
ILSVRC [26]	14,197,122	1000	>14 million	256x256	Common objects, abstract concepts	Top-1 accuracy, Top-5 accuracy
KITTI [27]	7,481 (train) and 7,518 (test)	8	80,256	1248x384	Car, Van, Truck, Pedestrian, etc.	Average Precision (AP), Precision/Recall curve, Detection rate
CIFAR100 [28]	60,000	100	-	32x32	Fine-grained (100) and Coarse-grained (20)	Top-1 and Top-5 accuracy

made available in 2015. It included all of the earlier test photographs. 2017 saw a change in the training/validation split to 118,000/5,000 pictures in response to community feedback, while keeping the same image and annotation data. A collection of 123,000 photos without annotation was also included in the 2017 edition. A variety of tasks are covered by annotations: identification of objects with bounding boxes and per-instance segmentation masks for 80 object categories; captioning of images; detection of keypoints for over 200,000 images and 250,000 person instances; segmentation of stuff images with 91 stuff categories; panoptic segmentation with 80 thing categories and a subset of 91 stuff categories; and dense pose annotations for over 39,000 images and 56,000 person instances, restricted to training and validation data, offering extensive body part mapping to a 3D model for each labeled individual.

2) PASCAL VOC

Specifically, the PASCAL Visual Object Classes (VOC) [25], [29] The 2010 dataset includes 20 different object types, such as automobiles, bikes, buses, aircraft, boats, and more, in addition to objects like vehicles, household goods, and animals. Pixel-by-pixel segmentation, bounding box, and class annotations are added to every picture in this collection. As a common benchmark for assessing object detection, semantic segmentation, and classification techniques, the PASCAL VOC dataset has garnered a lot of attention throughout time. The 1,464 photos in the training subset, the 1,449 images in the validation subset, and the separate private testing set for evaluation are the three subsets from which the dataset is divided.

3) ILSVRC

14,197,122 photos that have been tagged using the WordNet hierarchy make up the ImageNet dataset. This dataset has been the basis for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [26], an established benchmark for object identification and picture classification since 2010. Annotated training photos are included in the publically available dataset, however an unannotated collection of test images is offered separately. The ILSVRC annotations may be divided into two groups: annotations at the picture level that indicate whether object classes are present in the image or not, and annotations at the object level that provide precise class labels and tight bounding boxes for specific object instances in the image. It should be noted that only thumbnails and image URLs are made available because the ImageNet project does not own copyright to the photographs. The dataset includes 1.2 million photos linked to SIFT features from 1,000 synsets and a broad range of 21,841 non-empty WordNet synsets, totaling 14,197,122 images. Of these, 1,034,908 images include bounding box annotations.

4) KITTI

One of the most well-known datasets in mobile robotics and autonomous driving is the KITTI [27] dataset, which includes long recordings of traffic scenes taken with several types of sensors, such as RGB and grayscale stereo cameras as well as a 3D laser scanner. Notably, KITTI has been enhanced by hand annotations from many research groups, while lacking intrinsic semantic segmentation ground truth. For example, in the road recognition challenge, Álvarez et al. presented ground truth data for 323 photos and classified items into three classes: road, vertical components, and sky. Similar to this, Zhang et al. painstakingly annotated 252 acquisitions for testing and training, distinguishing between 10 item categories: buildings, sky, roads, greenery, sidewalks, automobiles, people, bicycles, signs/poles, and fences. In addition, Ros et al. provided annotations for 170 training and 46 testing photos from the visual odometry challenge. These annotations classified items into 11 different categories, which included skies, buildings, trees, cars, signs, roads, pedestrians, fences, poles, sidewalks, buildings, and bicycles.

5) CIFAR100

The CIFAR-100 [28] dataset is a subset of the Tiny Pictures dataset that contains 60,000 color, 32×32 pixel images. There are one hundred classes available at the Canadian Institute for Advanced Research, or CIFAR. Twenty superclasses are created from these 100 classes, and each class is given 600 photos. Every image has two tags: the specific class

(represented by the label) and the superclass to which the image belongs. For every class, there are 500 training photos and 100 test images in the dataset. What is this image of, and how likely is it that the class name will be a reasonable response to the question? For an image, a specific category has been selected. Line drawings were rejected as the primary emphasis, and a certain degree of photorealism was also necessary. In every image, there should be one distinct, easily observable example of the object the class is discussing; it doesn't matter if it is partially hidden or seen from an odd angle; the labeler should be able to identify the object even in these situations.

B. ALGORITHM

Mainly, there are three types of object detection architecture; traditional machine learning, new deep learning one-stage and two-stage architecture as presented in Figure 3. Traditional object detection architectures, often associated with handcrafted features, typically involve multi-step processes, which visualized in Figure 4. In this approach, an image is initially processed to extract features using methods like Histogram of Oriented Gradients (HOG) or Haar-like features. Subsequently, these features are fed into a classifier, such as a Support Vector Machine (SVM), to distinguish between object and non-object regions. Finally, post-processing steps, like non-maximum suppression, are applied to refine and consolidate the detected bounding boxes. On the other hand, one-stage object detection architectures drawn in Figure 5, exemplified by YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector) etc., streamline the process by simultaneously predicting object classes and bounding box coordinates across the entire image in a single forward pass. These models employ dense sampling and anchor boxes to handle object size and aspect ratio variations. In contrast, two-stage architectures, like Faster R-CNN (Region-based Convolutional Neural Network), employ a region proposal network (RPN) in the first stage to suggest potential object regions, followed by a second stage that refines these proposals and classifies objects as shown in Figure 6. The use of region proposals allows for improved localization accuracy, especially for small objects, and facilitates the integration of deep learning techniques for end-to-end training. Table 4 provides the insights of well-known algorithms of object detection.

1) TRADITIONAL MACHINE LEARNING ARCHITECTURE

- **VJ:** Without requiring any limitations, such as skin color segmentation, P. Viola and M. Jones developed the first real-time human face detector in 2001 [32], [37]. Powered by a 700MHz Pentium III CPU, the detector operated at rates tens to hundreds of times faster than existing technologies, while maintaining equal detection accuracy. In order to find windows containing human faces at all imaginable sizes and locations inside a picture, the VJ detector uses a straightforward sliding window approach. This procedure may appear simple, but the calculation required was greater than what the computers of the day could handle. The VJ detector was able to drastically improve its target identification speed by employing three key techniques: detection cascades, feature selection, and integral images.
- **HOG:** In 2005, N. Dalal and B. Triggs created a feature descriptor known as Histogram of Oriented Gradients (HOG) [33]. HOG is considered a major breakthrough in shape contexts [38] and scale-invariant feature transform [39], [40] of its era. The HOG descriptor is calculated on a dense grid of uniformly spaced cells using overlapping local contrast normalization to balance feature invariance with nonlinearity. HOG's creation was primarily motivated by the need to recognize pedestrians, even though it can detect a wide variety of objects. In order to recognize objects of different sizes, the HOG detector rescales the input picture many times while keeping the detection window size constant. For a considerable amount of time, the HOG detector functioned as an essential component of many object detectors [34], [41] [42] and different computer vision applications.
- **DPM:** DPM was the gold standard for traditional object detection methods after winning the VOC-07, -08, and -09 detection contests. DPM was used for the enhancement of the HOG detector; P. Felzenszwalb [34] originally proposed this technique in 2008. Divide and conquer detection theory views training as just learning how to correctly dissect an object, and inference as an ensemble of detections on different object components. This clarifies the current situation. For example, one has to be able to recognize an automobile's window, wheels, and body in order to distinguish it from another. The star model, as it is sometimes called, was given in this portion of the study by P. Felzenszwalb and colleagues [34]. Then, to improve the star model even more, Girshick included mixing models to account for objects in the actual world with larger variations. In [41], [43], [44], and [45], despite significant advancements in detection accuracy in many contemporary object detecting systems, the deep insights offered by DPM nonetheless have a long-lasting influence. Some examples of these tactics are context priming, mixture models, bounding box regression, hard negative mining, and others. 2010 saw the PASCAL VOC Lifetime Achievement Award given to Girshick et al.

2) MODERN DEEP LEARNING ARCHITECTURE: ONE-STAGE MODEL

- **YOLO:** YOLO was first presented in 2015 by R. Joseph et al. According to [12], it was the first one-stage detector in the deep learning period. With a VOC07 mAP=63.4% for its enhanced version and a VOC07 mAP=52.7% for its fast version, YOLO

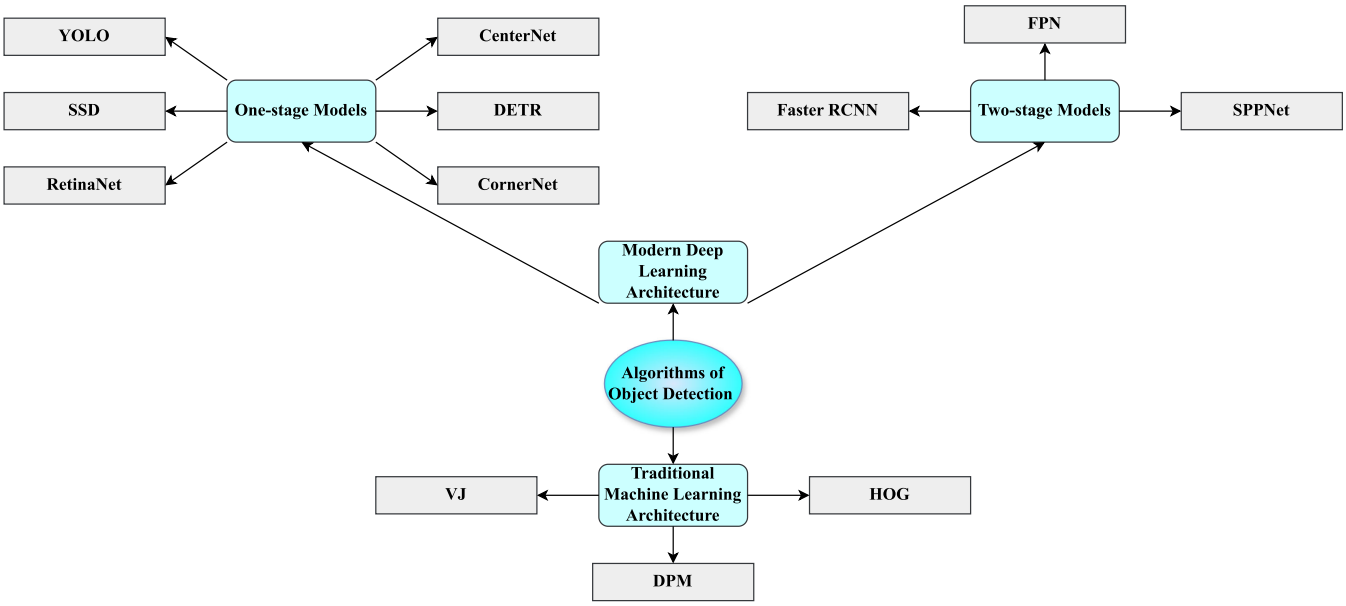

FIGURE 3. Algorithms of object detection.

TABLE 4. Comparison of different object detection algorithms.

Algorithm	Features	Performance metrics	Advantages	Disadvantages
YOLO [12]	1-stage detector, simultaneous prediction	VOC07 mAP=63.4% (enhanced), 52.7% (fast)	Fast, real-time detection.	Lower localization accuracy, especially for small objects.
SSD [11]	Single Shot Multibox Detector, multi-reference	COCO mAP@.5=46.5%, 59 fps (fast)	High detection speed, multi-reference improves accuracy.	May still have challenges with small object detection.
RetinaNet [13]	Focal loss, one-stage detector with improved accuracy	COCO mAP@.5=59.1%	Improved accuracy with focused loss.	One-stage detectors generally have lower accuracy compared to two-stage detectors.
CornerNet [30]	Key points decouple and re-groups corners	COCO mAP@.5=57.8%	Address category imbalance and convergence time issues.	Performance might be sensitive to keypoint detection accuracy.
CenterNet [31]	End-to-end networks treat objects as a single point	COCO mAP@.5=61.1%	Simple and elegant end-to-end detection.	Basic detection method may lack complexity for certain scenarios.
DETR [15]	Detection as set prediction, transformers	COCO mAP@.5=71.9% (Deformable DETR)	Transformer-based, no need for anchor points.	Long convergence time, performance challenges on small objects.
VJ [32]	Real-time face detection, sliding window, cascades	-	Real-time detection.	May struggle with variations in lighting conditions and object orientations.
HOG [33]	Histogram of Oriented Gradients, dense grid	-	Effective feature descriptor.	Computationally expensive, may not handle scale variations well.
DPM [34]	Divide and conquer, ensemble, star model	VOC-07, -08, -09 winner	Ensemble approach, context priming, bounding box regression.	Complexity and computational cost.
SPPNet [35]	Spatial Pyramid Pooling Networks, fixed-length	VOC07 mAP=59.2%	Fixed-length representation, better performance than R-CNN.	Multi-stage training, refinement limited to fully connected layers.
Faster RCNN [10]	Region Proposal Network, end-to-end framework	COCO AP@.5=42.7%, VOC07 mAP=73.2%	End-to-end training, improved accuracy.	Computation redundancy at the proposal stage.
FPN [36]	Feature Pyramid Network, top-down architecture	COCO mAP@.5=59.1%	Utilizes feature maps at different levels, improved object identification.	-

is incredibly fast. It works at 45 frames per second for the enhanced version. By using a single neural network to analyze the entire image, YOLO operates on a totally different paradigm than two-stage detectors. With simultaneous prediction of bounding boxes and probability for each zone, this network divides the picture into regions. Although YOLO can identify items more quicker than two-stage detectors, it still has poorer localization accuracy, especially when it comes to tiny objects. Additional thought has been given to this problem by later versions of YOLO [46], [47], [48] and the proposed SSD of today [11]. The YOLOv7 [49] team has been proposed as a follow-up to the work of the YOLOv4 team. It achieves higher speeds and higher accuracy (varying from 5 FPS to 160 FPS) than most existing object detectors by introducing optimized

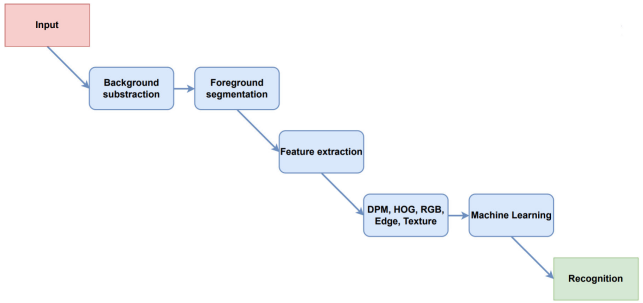

FIGURE 4. Traditional model of object detection.

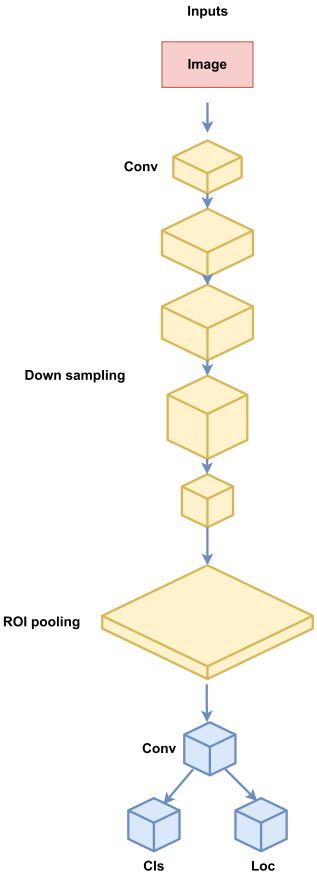

FIGURE 5. One-stage model of object detection.

structures such as dynamic label assignment and model structure reparameterization.

- **SSD:** SSD was introduced by Liu et al. [11]. The main benefit of SSD is the addition of multi-reference and multiresolution detection techniques (to be covered in Section II-C1). These approaches significantly improve the detection accuracy of a one-stage detector, especially for some tiny objects. Accuracy and detection speed are two areas where SSD shines (COCO mAP@.5=46.5%; fast version runs at 59 frames per second). SSD can recognize objects at different sizes across several network levels, while the older detectors could only detect items on their topmost layers. This is the main difference between SSD and the earlier detectors.
- **RetinaNet:** Despite its incredible speed and ease of use, one-stage detectors have never been as accurate as two-stage detectors. After looking into the causes, Lin et al. [13] suggested RetinaNet. The main finding they made was that there is a noticeable discrepancy between the background and foreground classes while dense detectors are being trained. In order to do this, RetinaNet presents a unique loss function known as focused loss, which alters the traditional cross-entropy loss to encourage the detector to focus more during training on difficult, misidentified examples. One-stage detectors may detect at a very high rate (COCO mAP@.5=59.1%) while keeping accuracy levels comparable to two-stage detectors by applying focused loss.
- **CornerNet:** In previous methods, the primary method of supplying references for classification and regression was through anchor boxes. It is usual for an object's quantity, position, scale, ratio, etc. to vary. They need to keep installing a lot of reference boxes so that ground facts better match in order to get higher performance. Still, there would be more category imbalance, a long convergence time, and a lot of hand-designed hyper-parameters in the network. To address these problems, Law et al. [30] reject the previous paradigm of detection and consider the work as a prediction problem involving key points, i.e., the corners of a box. Once the key points are gathered, it will use the extra embedding information to decouple and re-group the corner points in order to construct the bounding boxes. CornerNet outperforms the majority of one-stage detectors at that moment (COCO mAP@.5=57.8%).
- **CenterNet:** CenterNet [31] was presented by X. Zhou and colleagues in 2019. This completely end-to-end detection network eliminates costly post-processes such as group-based keypoint assignment and NMS (found in CornerNet [30], ExtremeNet [50], etc.) while yet adhering to the same keypoint-based detection paradigm. CenterNet treats an object as a single point (the object's center) and regresses all of its attributes (size, orientation, position, pose, and so on) based on the reference center point. The model is simple and elegant, capable of encapsulating several tasks including optical flow learning, depth estimation, 3-D object identification, and human location estimation into a single framework. It is possible for CenterNet to get similar detection results (COCO mAP@.5=61.1%) even with a basic detection method.
- **DETR:** In recent years, deep learning has been greatly influenced by transformers, particularly in the field of computer vision. Transformers avoid the traditional convolution operator in favor of attention-alone calculation, which allows them to overcome CNN limitations and reach a global-scale receptive field. Carion et al. suggested DETR [15] in 2020. They addressed object

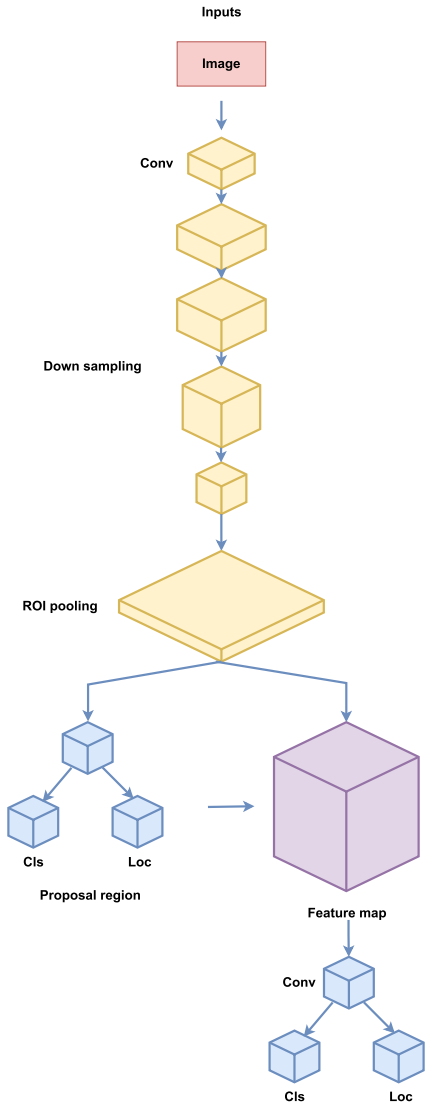

FIGURE 6. Two-stage model of object detection.

detection as a set prediction issue and demonstrated an end-to-end detection network with transformers. Up to now, object detection has entered a new phase where objects may be recognized without the need for boxes or anchor points. Deformable DETR was subsequently proposed by Zhu et al. [51] as a solution to the DETR's lengthy convergence time and poor performance on tiny object detection. It achieves state-of-the-art performance (COCO mAP@.5=71.9%) on the MSCOCO dataset.

3) MODERN DEEP LEARNING ARCHITECTURE: TWO-STAGE MODEL

- **SPPNet:** He et al. [35] introduced Spatial Pyramid Pooling Networks (SPPNet) in 2014. A fixed-size input was required by earlier CNN models; for instance, AlexNet needed an image with dimensions of 224 by 224 [52]. The main component of SPPNet is the Spatial Pyramid Pooling (SPP) layer, which enables a CNN to generate a fixed-length representation regardless of Whenever SPPNet is used for object recognition, fixed-length representations of any region may be created to train the detectors without having to compute the convolutional features again. This enables a single computation of the feature maps from the entire picture. More than 20 times (VOC07 mAP=59.2%) better performance is achieved by SPPNet than R-CNN without sacrificing detection accuracy. Despite the huge boost in detection speed, SPPNet still has certain limitations. Firstly, training is still multi-stage; secondly, SPPNet only refines its entirely connected layers, ignoring all other layers. Later that year, with the introduction of Fast RCNN [9], these problems were fixed.
- **Faster RCNN:** In 2015, Ren et al. introduced the Faster CNN detector, shortly after the Fast RCNN [10], [53]. This faster CNN (COCO AP@.5=42.7%, VOC07 mAP=73.2%) is the first near-realtime deep learning detector, achieving 17 frames per second with ZF-Net 48. The main feature of Faster-RCNN is the introduction of the Region Proposal Network (RPN), which essentially enables cost-free region suggestions. With the exception of proposal detection, feature extraction, bounding box regression, and TC, the majority of discrete object detection system components have been gradually incorporated into a single, end-to-end earning framework, beginning with R-CNN and concluding with Faster RCNN. Computation redundancy remains at the following detection stage even after Faster RCNN has over the speed constraint of Fast RCNN. Light head RCNN [54] and RFCN [55] are two further improvements that have been proposed since then.
- **FPN:** Lin et al. [36] introduced FPN in 2017. Before the introduction of FPN, most deep learning-based detectors only employed feature maps at the top layer for detection. The deeper layers of a CNN do not help with object localization, even though they do contain features that are helpful for classifying data. To accomplish so, FPN builds a top-down architecture for high-level semantics creation at all sizes, complete with lateral links. Given that the forward propagation of a CNN naturally generates a feature pyramid, the FPN shows notable gains in object identification over a wide variety of sizes. By integrating FPN in a basic Faster R-CNN system, it achieves state-of-the-art single model identification performance on the COCO dataset without further bells and whistles (COCO mAP@.5=59.1%).

C. LIBRARY

An object detection library is a software framework or toolkit designed to facilitate the automated identification and localization of objects within digital images or video streams. It typically includes pre-trained models, computer vision algorithms, and tools for training custom models, allowing

developers to build applications that can detect and classify objects in real-world scenes. These libraries are crucial for a wide range of applications, from autonomous vehicles and surveillance systems to augmented reality and image analysis, enabling the extraction of valuable information from visual data by accurately recognizing and delineating objects of interest. Table 5 provides a brief comparison of these libraries based on some criteria.

1) IMAGEAI

The ImageAI library is a comprehensive toolkit designed to empower developers with a wide range of computer vision algorithms and deep learning techniques for various tasks in object detection and image processing. Its core mission is to streamline the development of object detection projects by simplifying the coding process to just a few lines. ImageAI offers extensive support for operations such as image recognition, image object detection, video object detection, video detection analysis, custom image recognition training and inference, and custom object detection training and inference. With its image recognition capabilities, it can identify up to 1000 distinct objects within an image, while for image and video object detection, it can efficiently spot 80 of the most commonly encountered objects in everyday scenarios. Furthermore, the library enables the training of custom object recognition and detection models using own datasets, allowing for the inclusion of a broader array of objects through the utilization of new images and datasets.

2) GLUONCV

GluonCV stands out as a leading library framework for deep learning in computer vision, offering a powerful arsenal of state-of-the-art algorithms to expedite results in the field. With an extensive range of tasks supported, including image classification, object detection in images, videos, and real-time scenarios, semantic and instance segmentation, pose estimation, and action recognition, GluonCV proves itself as a versatile tool. This framework accommodates both MXNet and PyTorch, bolstered by a wealth of tutorials and additional resources to facilitate exploration of various concepts. It boasts a rich repository of pre-trained models, allowing users to craft tailored machine learning models for specific tasks with ease.

3) YOLOV3_TENSORFLOW

YOLOv3 represents a significant advancement in the YOLO series, boasting improved performance in both speed and accuracy over its predecessors. What sets it apart is its ability to effectively detect smaller objects with precision. However, it faces a tradeoff between speed and accuracy when compared to other prominent algorithms. YOLOv3_TensorFlow, an early implementation of the YOLO architecture for object detection, is known for its swift GPU computations, efficient results, streamlined data pipelines, weight conversions, faster training times, and a host of other benefits.

4) DETECTRON2

Detectron2, an advanced framework created by Facebook's AI research team (FAIR), stands as a cutting-edge library supporting a wide array of state-of-the-art techniques for object detection and segmentation, all grounded in PyTorch. This versatile and extensible library offers users access to top-notch implementation algorithms and methods, making it a go-to choice for numerous applications and production projects at Facebook. Detectron2's ability to be trained on single or multiple GPUs delivers rapid and highly effective results, empowering users to employ a variety of high-quality object detection algorithms, including innovations like DensePose, panoptic feature pyramid networks, and various iterations of the Mask R-CNN model family.

5) DARKFLOW

Darkflow is a Python-based adaptation of the Darknet framework, originally written in C and CUDA, designed to make object detection more accessible to a broader audience using TensorFlow. To utilize Darkflow effectively, one will need prerequisites like Python 3, TensorFlow, Numpy, and Opencv. With these essential dependencies, Darkflow empowers users to perform various object detection tasks. This framework grants access to YOLO models and facilitates the downloading of custom weights for diverse models. Its capabilities encompass parsing annotations, configuring networks, visualizing flow graphs, training new models, custom dataset training, real-time or video analysis, and leveraging Darkflow for similar applications. Furthermore, Darkflow allows users to save these models in the protobuf (.pb) format for future use.

D. EVALUATION METRICS

Table 6 provide a comparison of different metrics to evaluation object detection.

1) INTERSECTION OVER UNION (IOU)

IoU is one of the most fundamental metrics used in object detection. It measures the overlap between the predicted bounding box and the ground truth bounding box. The IoU is calculated as the ratio of the area of intersection between the two bounding boxes to the area of their union:

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \tag{1}$$

2) PRECISION AND RECALL

Precision and recall are used to assess the accuracy and completeness of object detection. Precision measures the proportion of true positive detections among all positive predictions, while recall measures the proportion of true positive detections among all actual positive instances. The equations for precision and recall are as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{3}$$

TABLE 5. Comparison of different object detection libraries.

Criteria	ImageAI [56]	GluonCV [57]	YOLOv3_TensorFlow [58]	Detectron2 [59]	DarkFlow [60]
Ease of use	Very easy, few lines of code	User-friendly interfaces and tutorials	Suitable for advanced users familiar with TensorFlow and YOLO	User-friendly interfaces and tutorials	Suitable for advanced users familiar with YOLO
Speed	Slower, relies on pre-trained models, no GPU acceleration	Efficient GPU computations, real-time performance	Fast GPU computations, real-time performance	Efficient GPU computations, real-time performance	Fast GPU computations, real-time performance
Accuracy	Detects common objects with reasonable accuracy	High accuracy on challenging datasets and tasks	Accurate, especially for small objects	High accuracy on challenging datasets and tasks	Accurate, especially for small objects
Customization	Limited options for model architecture and hyperparameters	Highly extensible and modular, supports customization	Supports training custom models using custom datasets and weights	Highly extensible and modular, supports customization	Supports training custom models using custom datasets and weights

TABLE 6. Comparison of different detection evaluation metrics.

Metric	Advantages	Disadvantages
IoU	Simple and intuitive. Provides a clear measure of overlap.	Sensitive to small variations. May not capture all aspects of detection quality. Binary nature (threshold-based).
P & R	Balances trade-off between relevance and completeness. Suitable for imbalanced datasets.	Measures the quality and quantity of predictions. May not be suitable for tasks where FPs or FNs are crucial independently. Can mislead based on the number of predictions.
AP	Provides a comprehensive evaluation at various confidence levels. Sensitive to the choice of confidence thresholds.	May not work for tasks with strict precision or recall requirements.
mAP	A comprehensive metric. Aggregates performance across multiple classes.	Can mask poor performance in specific classes. Sensitive to class imbalance. Involves complex calculations and is computationally expensive.
F1 Score	Balances precision and recall. Good choice for imbalanced datasets.	Ignores true negatives, so may not work for tasks where they are crucial. Sensitive to the selected threshold.

3) AVERAGE PRECISION (AP)

AP is a common metric for summarizing the precision-recall trade-off across different IoU thresholds. It involves calculating the precision-recall curve and computing the area under this curve. A high AP indicates a better object detection model.

4) MEAN AVERAGE PRECISION (MAP)

mAP is a more comprehensive metric that calculates the Average Precision (AP) for each class and then averages them. It is often used to evaluate the overall performance of an object detection model across multiple object categories. The equations for mAP is as follows:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{Average Precision}_i \tag{4}$$

Here N and AP_i are total classes number and i^{th} class of the Average Precision.

5) F1 SCORE

F1 score is the harmonic mean of precision and recall. It provides a single value that balances the trade-off between precision and recall. The formula for F1 score is:

$$\text{F1 Score} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \tag{5}$$

IV. APPLICATIONS OF OBJECT DETECTION

Object detection plays a crucial role in various applications, including autonomous driving, surveillance, and medical imaging etc. It involves identifying and locating specific objects within images or video frames. Object detection algorithms use deep learning models, such as Convolutional Neural Networks (CNNs), to detect and draw bounding boxes around objects of interest, making it a powerful tool for tasks like pedestrian detection in self-driving cars, identifying anomalies in security footage, and locating tumors in medical scans. This technology has the potential to enhance efficiency, safety, and accuracy in a wide range of fields by enabling automated and real-time object recognition and tracking. In Table 7, the distributions of the selected articles on various application domains of object detection are indexed. Whereas in Table 8 the data of various scholarly articles about the applications of object detection is shown.

A. FACE DETECTION

Face detection is a compelling application of object detection, where the goal is to identify and locate human faces within images or video frames. This technology leverages convolutional neural networks and deep learning to recognize facial features and determine their positions accurately. It has widespread applications in various fields, including security for surveillance systems, photography for autofocus and facial recognition, as well as in the development

TABLE 7. Papers count of different application domains in different scholarly databases.

Keyword	Paper count
Face Detection	10
Arial Target Detection	19
Text Detection	10
Pedestrian and Traffic Detection	11
Human Violence and Sport's Foul Detection	11
Plant and Human disease detection	11
Astronomical Object Detection	9

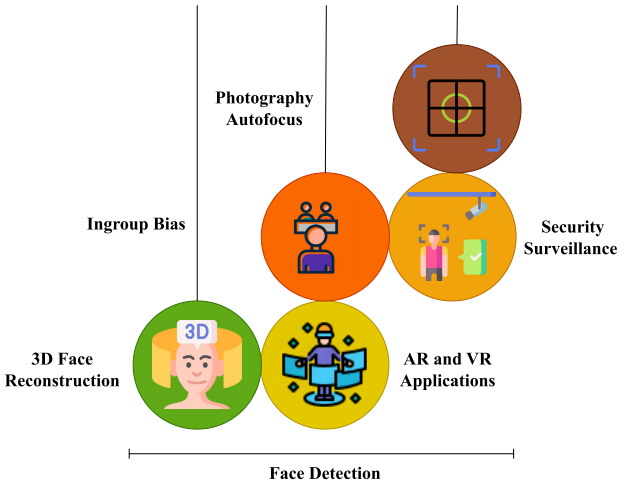

FIGURE 7. Face detection fields.

of augmented reality and virtual reality applications. The ability to swiftly and accurately detect faces in diverse contexts contributes to improved human-computer interaction and enhances the efficiency and security of many technological systems. Figure 7 shows various applicable fields of face detection (For figures, icons were taken from https://www.flaticon.com/).

Yang et al. introduced the WIDER FACE [61] dataset illustrated it as an effective training source in the field of face detection. Through their work, the authors proposed a multi-scale two-stage cascade framework which uses divide and conquer technique using WIDER FACE dealing large scale variation. Qi et al. implemented YOLO5Face [62] face detector on the basis of YOLOv5 object detector and WiderFace dataset. The authors also implemented a backbone for mobile devices based on ShuffleNetV2, which also provided the SOTA performance and fast execution speed. Jiang, Huaizu, and Erik Learned-Miller experimented based on the dataset WIDER FACE and two benchmarks, FDDB and IJB-A to use Faster R-CNN [63]. Zhu et al. invented TinaFace [64] as a baseline method dealing face detection. As the backbone with ResNet-50 and dataset WIDER FACE, TinaFace attained 92.4% AP which outperformed the state-of-the-art method that time. Mamieva et al. build a single-stage face detector, RetinaNet [65] baseline. Their work on WIDER FACE and FDDB datasets showed that the method achieves high Average Precision (AP) scores, with an accuracy of 95.6% for successfully detected faces. Boyd et al. introduced a training strategy, CYBORG [66], which uses human-annotated saliency maps to guide deep learning models in focusing on image regions that humans find salient for a given task. The authors found that CYBORG significantly improves generalization and accuracy on unseen samples in synthetic face detection compared to traditional training methods. Hangaragi et al. introduced a face detection and recognition model using Face mesh [67]. The model is trained on Labeled Wild Face (LWF) dataset images and real-time captured images. While testing, the model compares face landmarks of the test image with those of the training images and achieves an accuracy of 94.23% for face recognition. Prunty et al. showed that humans exhibit an ingroup bias at the earliest stage of face processing, where they detect ingroup faces (Black and White) more quickly and accurately than outgroup faces (Asian, Black, and White) in everyday scenes [68]. According to their findings, this bias in face detection is independent of the color of faces and can be attributed to both visual and social factors. Sandhya et al. proposed a smart criminal detection and identification system [69]. They combined a Single Shot Multibox Detector for face detection and an auto-encoder model for matching captured facial images with criminals in a database. The authors found that the system achieves a confidence rate of 0.75 and above, making it effective in identifying individuals with a history of felonies based on facial images. Al-Neama et al. build a GPU-based system for real-time face recognition [70]. According to the findings, the system significantly outperforms traditional CPU-based methods, with a 19.72x improvement in the detection phase and a remarkable 1573x improvement in the recognition phase when implemented on an NVidia GTX 570 graphics card.

B. ARIAL TARGET DETECTION

Arial target detection, as an application of object detection, involves the identification and localization of specific objects or subjects within an image or video stream. This technology is widely employed in various domains, such as surveillance, autonomous vehicles, and military applications, to recognize and pinpoint predefined targets of interest, which could be vehicles, pedestrians, wildlife, or even specific objects like weapons. Object detection algorithms use deep learning and computer vision techniques to draw bounding boxes around

TABLE 8. Scholarly articles about the applications of object detection.

Application	Year	Title work and reference	Datasets	Methods	Metrics and Results
Face detection	2016	Efficient Face Detection [61]	WIDER FACE	Multi-scale, two-stage cascade framework	Notable performance improvement on large-scale variation.
	2022	Mobile-Friendly Face Detection Implementation. [62]	WIDER FACE	YOLO5Face with ShuffleNetV2 backbone	SOTA performance and fast execution on mobile devices.
	2017	Advancing Face Detection Benchmarks. [63]	WIDER FACE, FDDB, IJB-A	Faster R-CNN	Notable results on WIDER FACE, FDDB, and IJB-A benchmarks.
	2020	Innovating Baseline for Face Detection. [64]	WIDER FACE	TinaFace	92.4% Average Precision on WIDER FACE.
	2023	Single-Stage Precision in Face Detection. [65]	WIDER FACE, FDDB	RetinaNet	High Average Precision scores on WIDER FACE and FDDB.
	2023	Enhanced Synthetic Face Detection. [66]	-	CYBORG	Improved generalization and accuracy using saliency maps.
	2023	High Accuracy Face Recognition. [67]	Labeled Wild Face (LWF)	Face mesh	94.23% accuracy for face recognition on LWF dataset.
	2023	Human Perception: Ingroup Face Bias. [68]	-	-	Ingroup bias in early face processing.
	2023	Smart Criminal Detection with SSD and Autoencoder. [69]	-	SSD for face detection, Autoencoder for matching	Smart criminal detection and identification system.
	2023	GPU-based Real-time Face Recognition Outperforms CPU. [70]	-	GPU-based system	Real-time face recognition, significant performance improvement over CPU-based methods.
Arial target detection	2018	Improved Point Cloud Filtering: Reduction in Errors. [71]	-	Progressive TIN Densification (PTD)	Reduction in type I errors by 7.53% and total errors by 4.09%.
	2022	Underwater Target Detection: mAP with Modified Swin Transformer. [72]	-	Modified Swin Transformer	87.2% mean average precision (mAP) for underwater target detection.
	2022	Real-time Target Detection: Lightweight CNN. [73]	-	Lightweight CNN with depthwise separable modules	Real-time target detection with improved speed.
	2016	Small Target Detection in Infrared: WLDM Scheme. [74]	Infrared images	WLDM-based scheme	Improved accuracy and robustness in small target detection.
	2023	Underwater Target Detection: mAP Improvement. [75]	-	YOLOv5s-CA	2.4% increase in mAP for underwater target detection.
	2023	ISignal Detection with ICBD Method. [76]	-	ICBD method	Improved signal detection in the presence of noise and interference.
	2023	Multiscale SAR Ship Detection: Upgraded YOLOv5s. [77]	-	Upgraded YOLOv5s with C3, FPN + PAN, and attention	Multiscale SAR ship detection with enhanced structures and attention.

TABLE 8. *(Continued.)* **Scholarly articles about the applications of object detection.**

	2023	Improved Small Target Detection. [78]	-	KPE-YOLOv5	5.3% improvement in detection mAP, 7% increase in precision for small target detection.
	2023	Underwater Object Detection. [79]	-	Lightweight YOLOv5	1.1% increase in mAP for underwater object detection.
	2023	Damaged Building Detection. [80]	Post-disaster UAV images	DB-YOLOv5	High accuracy and efficiency for real-time detection.
	2020	Human Detection in Disasters. [81]	-	Image processing for human detection	Improved disaster management through video analysis.
	2018	Humanitarian Relief Coordination. [82]	-	Machine learning	Coordination of humanitarian relief efforts in disasters.
	2021	Building Extraction for Covid-19. [83]	Pléiades satellite imagery	Mask R-CNN	High recall, precision, and F1 score in building extraction.
	2019	Automated Bridge Detection. [84]	UAV-generated multispectral images	Multi-stage approach	Faster processing and high accuracy in bridge detection.
	2020	Local-Level Poverty Prediction. [85]	High-resolution satellite imagery	Object detectors	High Pearson's r^2 in predicting village-level poverty.
	2023	Flood Detection Fusion. [86]	Satellite remote sensing data, social media data	Ensemble classification technique	Enhanced flood detection and assessment through data fusion.
	2023	Power Outage Detection. [87]	Space-borne remote sensing imagery	PODM	Successful detection of power outages in meteorological disasters.
	2023	Algae Bloom and Fish Extinction. [88]	Satellite imagery, soundings	-	Linked algae bloom to fish extinction using satellite imagery.
	2023	Progressive Image Classification. [89]	Satellite and aerial imagery	PICA	Improved speed and accuracy in flood disaster detection.
Text detection	2019	Complex Text Shapes: Scene Text Detection. [90]	-	Scene text detection method	High flexibility for complex text shapes.
	2018	Arbitrary-Oriented Text Detectio. [91]	-	RRPN for arbitrary-oriented text detection	Generation of inclined text proposals with orientation angle information.
	2017	Multi-Oriented Text Detection: FCN Approach. [92]	-	Fully convolutional network for multi-oriented text detection	Pixel-wise classification and direct regression for text boundary coordinates.
	2017	Arbitrary-Oriented Text Identification. [93]	-	R2CNN for text identification	Application of R2CNN on Faster R-CNN for arbitrary-oriented text detection.
	2023	DetectGPT Framework for Text Detection. [94]	-	DetectGPT framework	Text detection without separate classifiers, datasets, or watermarking.

TABLE 8. *(Continued.)* **Scholarly articles about the applications of object detection.**

	2023	Scene Text Detection: DPText-DETR Network. [95]	-	DPText-DETR network	Improved training efficiency and detection performance using explicit point coordinates.
	2023	Benchmarking MGTs: MGTBench Framework. [96]	-	MGTBench benchmark framework	Benchmarking for detecting Machine-Generated Texts.
	2023	Unified Text Detection Framework: Coarse-to-Fine. [97]	-	Unified coarse-to-fine framework	Accurate and efficient text boundary localization for arbitrary shapes.
	2023	AI-Generated Tweets Detection. [98]	Twitter dataset	Stylometric signals algorithm	Enhanced detection of AI-generated tweets with improved accuracy.
	2023	Combined text spotting framework. [99]	-	DeepSolo framework	End-to-end text spotting with combined detection and recognition using explicit point queries.
Pedestrian and traffic detection	2018	Pedestrian detection insight. [100]	Diverse datasets	R-CNN with AlexNet and transfer learning	Insights into R-CNN performance across different datasets.
	1999	Adaptive Object Detectio. [101]	-	Flexible object detection system	Adaptability to different scenarios without manual design.
	2007	Real-time Obstacle Detection up to 50 meters range. [102]	-	Real-time obstacle and pedestrian detection	Reliable detection up to 50 meters at 64 frames per second.
	2016	Improved Pedestrian Detection. [103]	KITTI dataset	CNN with LIDAR and color imagery	Improved detection using LIDAR and color imagery on the KITTI dataset.
	2014	Faster Object Detection with improved speed, minimal accuracy loss. [104]	-	Faster object detection method	Significant speed improvements with minimal loss in accuracy.
	2010	Multiresolution Object Recognition: Deformable part-based modeling. [105]	Caltech Pedestrian benchmark	Multiresolution model with deformable part-based modeling	Significant improvement in detection rates on Caltech Pedestrian benchmark.
	2020	Traffic Signal Control with ML-based adaptive switching [106]	-	Machine learning-based traffic signal control system	Real-time, adaptive signal switching, reducing waiting times.
	2019	On-road Vehicle Detection. [107]	On-road vehicle datasets	YOLOv3 algorithm with convolution layers	Vehicle detection and tracking in on-road datasets.
	2010	Traffic Object Detection System. [108]	Traffic video scenes	Object detection and segmentation system	Around 90% accuracy on four traffic video scenes.
	2022	Tiny Object Detection. [109]	Surveillance videos	YOLOv2 with DenseNet-201	97.51% average precision in vehicle detection and recognition.
	2023	Small and Multi-Object Detection [110]	Several datasets	BANet (Bidirectional Attention Network)	Outperformed YOLOX in mean average precision on several datasets.

TABLE 8. *(Continued.)* **Scholarly articles about the applications of object detection.**

Human violence and sports foul detection	2020	Real-time Character Tracking. [111]	Super Smash Brothers Melee	Real-time character tracking	Developed real-time detection model for character tracking.
	2022	Activity Recognition. [112]	-	Activity recognition	Refined features for activity and sub-object classification.
	2019	Football Foul Feature Extraction. [113]	Football competitions	Deep learning-based foul feature extraction	Enhanced accuracy in foul identification.
	2023	Deep Learning in Football Video Analysi. [114]	Football video analysis	Various deep learning techniques	Contribution to sports video analysis.
	2023	Automated Ball Possession Extraction. [115]	Sports analytics	Temporal Convolutional Networks (TCNs)	Improved possession estimation and classification accuracy.
	2023	Fall Detection with LIDAR Robot. [116]	-	2D LIDAR-equipped cleaning robot	High accuracy in fall detection and detection of prone positions.
	2019	Violence Detection in Movies. [117]	-	Shot segmentation, saliency-based frame selection	Classification of violence and non-violence shots.
	2020	Violence Detection Dataset Creation. [118]	Violence detection dataset	-	New dataset for testing violence detection techniques.
	2022	Video Surveillance Violence Detection. [119]	Real-world security camera footage	U-Net-like network with MobileNet V2	Good results on violence detection in real-world footage.
	2019	Efficient Violence Detection. [120]	Surveillance videos	Triple-staged deep learning framework	Efficient violence detection in surveillance videos.
	2020	Multimodal Violence Detection. [121]	XD-Violence dataset	Comprehensive violence detection approach	Positive impact of multimodal input and relationship modeling.
Plant and human disease detection	2023	Crop Disease Diagnosis. [122]	-	Image Captioning and Object Detection	High BLEU score for Image Captioning. Object Detection improvement needed.
	2023	Plant Disease Retrieval. [123]	-	Object Detection and Deep Metric Learning	Plant disease identification across various scenarios.
	2022	Rice Plant Disease Detection. [124]	-	Image Processing and Deep Learning	Early detection and prevention of rice plant diseases.
	2022	Medical Image Domain Adaptation. [125]	-	CLU-CNNs	Domain adaptation for medical image data.
	2019	Skin Disease Diagnosis. [126]	Skin-10, Skin-100	CNNs and Ensemble Approach	Improved accuracy for certain skin disease classes.
	2019	Generative Model: Object Detection in Limited Annotated Data. [127]	-	Generative model for object detection	Outperformed existing methods in limited annotated data scenarios.
	2023	Object Detection in Brain MRI. [128]	Brain MRI	Ensemble strategies for object detection	Potential mAP increase and improved AP for anatomical parts.

TABLE 8. *(Continued.)* **Scholarly articles about the applications of object detection.**

	2023	Medical Data Enhancement. [129]	-	Logistic regression and YOLOv4	Significant performance improvements for medical data classification and image detection.
	2023	Industrial Object Detection. [130]	-	ResNet18-based image segmentation	Precise recognition in industrial object detection.
	2023	Transformer-based Leukocyte Detection. [131]	-	Transformer-based object detection network	Superior mean average precision for leukocyte detection.
	2021	Concealed Object Detection. [132]	COD10K	SINet	Surpassed twelve contemporary baselines in concealed object detection.
Astronomical object and air index detection	2017	Urban Planning with LSTM. [133]	-	LSTM-based deep learning model	Promising results for air quality prediction in smart cities.
	2017	PM2.5 concentration estimation using deep CNNs [134]	Beijing dataset	Deep CNNs	Effective PM2.5 concentration estimation for air quality analysis.
	2023	Occupant-Centric Control: Enhancing Thermal Comfort and Energy Savings. [135]	-	OCC strategies with real-time occupancy detection	Enhanced thermal comfort and energy savings in office environments.
	2023	Leftover Item Detection. [136]	-	Computer vision prediction model	89% accuracy for detecting leftover items in shared vehicles.
	2023	Low-Cost Air Quality Monitoring. [137]	-	LPAQD	Real-time detection of particles with low-cost monitoring device.
	2023	IoT Air Quality Monitoring: Real-time, Cost-effective Data for Urban Management. [138]	-	IoT-based air quality monitoring system	Real-time, cost-effective, and precise data for urban air quality management.
	2023	Combatting Pollution with IoT: Real-time Monitoring and Alerts. [139]	-	IoT-based air quality monitoring system	Real-time monitoring and alerts for rising pollution levels.
	2023	Deep Learning in Radio Astronomy. [140]	-	Deep learning for radio astronomy	Insights into automatic object detection in radio astronomy.
	2023	Time-series Analysis in Astronomy. [141]	-	BLS periodogram analysis, neural network classifier	Detection and characterization of stars, exoplanets, and galaxies using time-series data analysis.

the recognized targets, enabling precise tracking, classification, and subsequent decision-making processes, thereby enhancing safety and situational awareness in complex real-world scenarios. Figure 8 shows various applicable fields of arial target detection.

Dong et al. introduced an improved Progressive TIN Densification (PTD) filtering algorithm for point clouds with high density and standard variance [71]. The results demonstrated that the improved PTD algorithm significantly reduces type I errors and total errors in the point clouds compared to the original PTD method by 7.53% and 4.09%, respectively. Lei et al. experimented that by using the Swin Transformer as the backbone network, enhancing multi-scale feature fusion, and improving the confidence loss function, the modified model achieved a mean average precision (mAP) of 87.2%, making it highly effective for

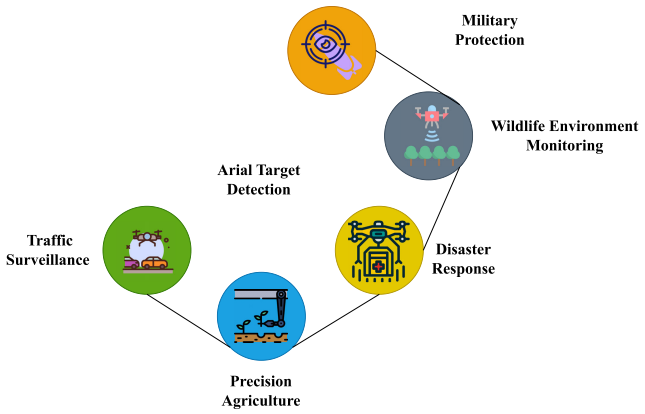

FIGURE 8. Arial target detection fields.

detecting underwater targets [72]. Yun et al. developed a real-time target detection method based on a lightweight convolutional neural network [73]. The method utilized depthwise separable residual modules and depthwise separable convolutions to reduce the number of model parameters and improve detection speed. According to their findings, it introduced 1×3 and 3×1 convolution kernels to enhance feature extraction. Deng et al. proposed a ''weighted local difference measure (WLDM)-based scheme'' for detecting small targets in infrared images against complex cloudy-sky backgrounds [74]. The studies found that this method enhances target visibility while suppressing background clutter and noise, leading to improved small target detection accuracy, robustness across various backgrounds and target movements, and significantly improved signal-to-clutter ratio (SCR) values. Wen et al. discovered a modified YOLOv5s network, YOLOv5s-CA [75], for underwater target detection. This modified network incorporated a Coordinate Attention (CA) module and a Squeeze-and-Excitation (SE) module to enhance target detection accuracy. For increasing the number of bottlenecks in the initial C3 module and embedding the CA module and SE layer, the model's ability to focus on targets and extract shallow features is improved. As the results of their studies on data from the 2019 China Underwater Robot Competition show a 2.4% increase in mean Average Precision (mAP) compared to the baseline YOLOv5s network. Liu et al. introduced the Interference Cancellation before Detection (ICBD) [76] method for signal detection in the presence of unknown Gaussian noise and subspace interference. The study found that ICBD effectively addresses the detection issue by projecting data into an interference-orthogonal subspace, enabling efficient detection with minimal training data. The study also found that, ICBD offered lower computational burden and can work even when interference is present in the training data. Yasir et al. developed an upgraded YOLOv5s technique for multiscale SAR ship detection [77]. This model incorporated C3 and FPN + PAN structures, as well as an attention mechanism, to improve ship detection accuracy in SAR imaging. Through the results using SAR ship detection datasets and satellite images, the authors showed that the proposed model is highly applicable for maritime surveillance. Yanget al. invented the KPE-YOLOv5 [78] algorithm, which enhances small target detection by addressing the limitations of existing methods, including low accuracy, high false detection rates, and missed detections. The algorithm achieved more accurate anchor box sizes for small targets through K-means++ clustering, integrates the scSE attention module to prioritize small target feature information in the backbone network, and improves small target feature extraction by adding a dedicated detection layer. As their evaluation on the VisDrone-2020 dataset which demonstrated that KPE-YOLOv5 outperforms YOLOv5, achieving a 5.3% improvement in detection mAP and a 7% increase in precision for small target detection. Liang, Heng, and Tingqiang Song proposed a lightweight underwater object detection algorithm based on YOLOv5 [79]. In underwater object detection, the algorithm addressed, including the reduction of model parameters and computational complexity through the use of depth-wise separable convolution and Ghost convolution. The incorporation of RepVgg and Rep-ECA modules enhances feature extraction and channel attention for small objects in blurred images. Results on the URPC underwater object detection dataset showed a 39% reduction in model parameter count, a 42% decrease in computational complexity, a 24% improvement in frame rate, and a 1.1% increase in mAP, improving detection precision while maintaining a lightweight model suitable for deployment in underwater equipment.

C. TEXT DETECTION

Text detection as an application of object detection involves the identification and localization of textual elements within images or scenes. It plays a crucial role in various domains, such as document analysis, image captioning, and autonomous navigation. Object detection models are employed to detect and outline regions containing text, enabling subsequent text recognition or analysis. Accurately identifying text objects within diverse visual data, this application facilitates tasks like automated transcription, signage interpretation, and the extraction of valuable information from images, contributing to improved accessibility, data retrieval, and overall comprehension in the digital age. Figure 9 some text detection applicable fields. Also,

Baek et al. developed a scene text detection method that effectively detects text areas by considering individual characters and their affinities, especially for complex text shapes like arbitrarily-oriented, curved, or deformed texts [90]. Achieving this by using character-level annotations for synthetic images and estimating character-level ground-truths for real images, their method offering high flexibility in detecting complex scene text images. Ma et al. introduced the ''Rotation Region Proposal Networks (RRPN)'' [91] framework for arbitrary-oriented text detection. They showed that the framework generates inclined text proposals with orientation angle information and adapts this angle

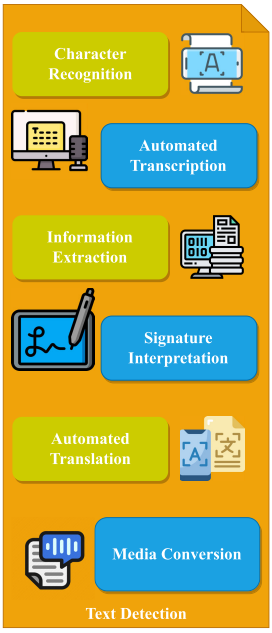

FIGURE 9. Text detection fields.

information for accurate bounding box regression. They also introduced the ''Rotation Region-of-Interest (RRoI) pooling layer'' [91] to project arbitrary-oriented proposals onto a feature map for text region classification. According to their analysis, this region-proposal-based architecture enhances the computational efficiency and effectiveness of arbitrary-oriented text detection compared to previous systems. He et al. stated a new approach to object detection, specifically for multi-oriented scene text involving a fully convolutional network with pixel-wise classification and direct regression for quadrilateral text boundary coordinates [92]. The method achieved F-measure of 81% on the ICDAR2015 Incidental Scene Text benchmark. Jiang et al. introduced a novel text detection method called Rotational Region CNN (R2CNN) [93] designed for identifying text in natural scene images with arbitrary orientations. The approach was built upon the Faster R-CNN architecture and involving inclined non-maximum suppression to produce the final detection results. Through their work, the authors demonstrated competitive performance on text detection benchmarks, including ICDAR 2015 and ICDAR 2013. Mitchell et al. developed a framework, DetectGPT [94] for detecting text generated by large language models (LLMs). They found that text generated by LLMs tends to occupy regions of negative curvature in the model's log probability function. DetectGPT leverages this observation to judge if a passage is generated by a specific LLM, without the need for training a separate classifier, collecting datasets, or watermarking generated text. Ye et al. introduced the DPText-DETR [95] network for scene text detection. DPText-DETR improves training efficiency and detection performance by using explicit point coordinates to generate and dynamically update position queries. As their results, DPText-DETR also enhanced the spatial inductive bias of non-local self-attention with an Enhanced Factorized Self-Attention module. He et al. build MGTBench [96], a benchmark framework for detecting Machine-Generated Texts (MGTs) generated by powerful Language Model (LLMs). The authors found that most existing detection methods are ineffective against MGTs, except for ChatGPT Detector and LM Detector. Model-based detection methods show promise with fewer training samples and transferability. The authors also explored text attribution, finding that the LM Detector is the best at identifying the originating model of a given text. Zhang et al. invented a unified coarse-to-fine framework for arbitrary shape text detection that accurately and efficiently locates text boundaries without the need for complex post-processing [97]. Results stated that, guided by a boundary proposal module, the module refined coarse boundary proposals using an encoder-decoder structure and introduces a novel boundary energy loss (BEL) to optimize and stabilize the learning of boundary refinement. Kumarage et al. developed a novel algorithm that uses stylometric signals [98] to enhance the detection of AI-generated tweets on Twitter. This work stated that these stylometric features effectively improve the accuracy of AI-generated text detectors while discriminating between human and AI-generated tweets, and detecting when AI begins generating tweets in a user's Twitter timeline. Ye et al. introduced DeepSolo [99], a DETR-like framework for end-to-end text spotting. DeepSolo combines text detection and recognition in a single decoder using explicit point queries for character sequences. The authors showed that DeepSolo also supports line annotations, reducing annotation costs.

D. PEDESTRIAN AND TRAFFIC DETECTION

Object detection plays a pivotal role in plant and animal disease detection by enabling the automated identification of infected or unhealthy specimens in agriculture and wildlife conservation. Utilizing computer vision and machine learning algorithms, this technology can identify specific symptoms or anomalies in plants or animals, such as discolored leaves, lesions, or abnormal behavior, allowing for early detection and intervention. Accurately pinpointing the areas or individuals affected, object detection contributes to more precise monitoring, timely responses, and ultimately, the preservation of crop yields and the well-being of wildlife populations. Figure 10 shows various applicable fields of pedestrian and traffic detection.

Masita et al. assessed the performance of R-CNN for pedestrian detection using two diverse dataset [100]. They employed AlexNet as a feature extraction model and fine-tuned it through transfer learning for dataset-specific classification. The study revealed insights into the R-CNN detector's performance across different datasets. Papageorgiou et al. created a flexible object detection system for automotive vision, with a focus on pedestrian detection [101]. They emphasized learning from examples, making it adaptable to different scenarios without requiring manual design. Their

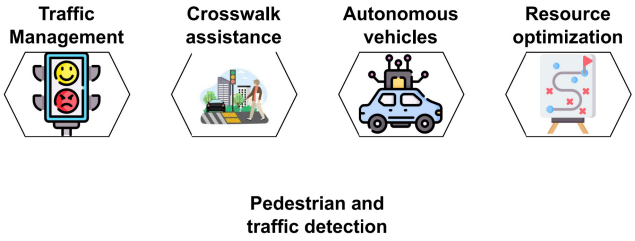

FIGURE 10. Pedestrian and traffic detection fields.

approach was not reliant on motion data or scene assumptions. They also discussed video processing improvements and integration into a DaimlerChrysler test vehicle. Ma et al. created a real-time obstacle and pedestrian detection system for vehicles using a single monochrome camera [102]. They identified obstacles above the ground plane with a ''virtual stereo system'' via inverse perspective mapping and used digital image stabilization for accuracy. They introduced a pedestrian segmentation method for bounding box extraction and a ''pedestrian detection strip'' to improve calculation speed. The system reliably detected obstacles and pedestrians up to 50 meters away at 64 frames per second on a standard PC. Schlosser et al. improved pedestrian detection using CNNs by incorporating LIDAR and color imagery [103]. They converted LIDAR data to depth maps, including three 3D scene features as extra image channels. The KITTI dataset was used for validation. Dollár et al. developed a faster object detection method by approximating multi-resolution image features through scale extrapolation, resulting in significant speed improvements with minimal loss in detection accuracy [104]. This approach is broadly applicable to various vision algorithms requiring fine-grained multi-scale analysis, particularly for images with broad spectra. Park et al. explored the challenges of recognition at different object scales. They argued against the idea of strict scale-invariance and proposed a multiresolution model that adapts to the size of detection windows, using deformable part-based modeling for large objects and rigid templates for small ones [105]. Their research, demonstrated on the Caltech Pedestrian benchmark, significantly improved detection rates, reducing missed detection from 86%-37% to 29% compared to recent state-of-the-art methods. Ng et al. addressed the challenge of traffic congestion in Hong Kong, where high traffic flow is a common issue. They developed a new traffic signal control system that employs machine learning with object detection and an evolutionary algorithm [106]. This system allows real-time, adaptive signal switching at intersections, reducing waiting times for pedestrians and vehicles and enhancing the overall travel experience. Jain et al. explored single object detection using convolution layers in a neural network, focusing on on-road vehicle datasets with varying illuminations [107]. They also performed multiple object detection using the YOLOv3 algorithm with the KITTI dataset, specifically for classes like car, bus, truck, motorcycle, and train. Additionally, they implemented vehicle tracking by using centroid positions in subsequent video frames, utilizing OpenCV and Python in conjunction with the YOLOv3 algorithm. Low et al. created a system to detect and segment objects in video frames for traffic surveillance [108]. They used MATLAB to apply a shadow removal method to improve detection accuracy. By distinguishing background from foreground, the system segmented foreground regions as objects, enabling potential use in object recognition and classification. Testing on four traffic video scenes achieved around 90% accuracy. Akhtar et al. improved YOLOv2 for precise tiny object detection in surveillance videos [109]. They used DenseNet-201 for compact feature extraction, achieving an average precision of 97.51% in vehicle detection and recognition, outperforming other methods. Wang et al. addressed the challenge of small and multi-object detection in traffic environments by introducing BANet, a bidirectional attention network featuring multichannel attention blocks, alpha-effective IoU loss, and multiple attention fusion [110]. BANet outperformed YOLOX, achieving improved mean average precision on several datasets and demonstrating better speed, reducing forward time by 0.97 ms.

E. HUMAN VIOLENCE AND SPORT'S FOUL DETECTION

Object detection plays a pivotal role in various applications, including human violence and sports foul detection. In the realm of security and public safety, object detection can be employed to identify acts of violence or aggression, such as physical altercations or suspicious behavior, through real-time analysis of video footage in public spaces. Similarly, in the context of sports, object detection can be utilized to identify fouls and rule violations in games, ensuring fair play and enhancing the accuracy of referee decisions. Analyzing video feeds and recognizing specific actions and objects, object detection contributes to the safety and integrity of both public spaces and sports events. Figure 11 shows various applicable fields of human violence and sport's foul detection.

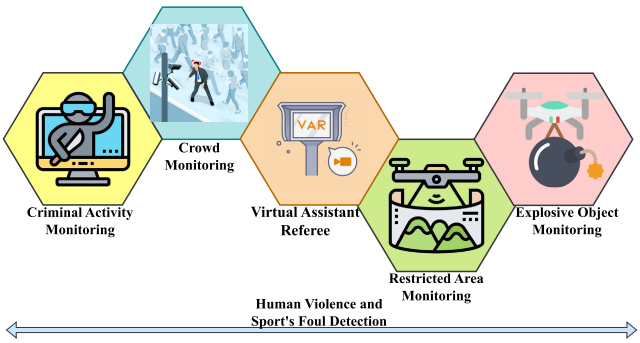

FIGURE 11. Human violence and sport's foul detection fields.

Thamaraimanalan et al. explored the application of computer vision and object detection in gaming and sports, with a particular focus on character tracking in the game Super Smash Brothers Melee [111]. They developed a real-time

detection model for this purpose, leading to the creation of a basic bot capable of taking actions based on character locations. Ryu el al. introduced an activity recognition model that blends spatial and temporal analysis of video activities [112]. They employed a novel technique called sequential object feature accumulation, refining deep neural network features for activity and sub-object classification. Ma et al. introduced a football foul feature extraction method based on deep learning algorithms. The approach aims to enhance the accuracy of foul identification during normal football competitions by eliminating background elements from input images, employing human motion tracking, and utilizing star skeleton features for foul action extraction [113]. The experimental results indicated that the method's target detection accuracy requires further improvement. Liu et al. examined deep learning techniques for football video analysis, covering player and ball detection, tracking, event detection, and game analysis [114]. Their study contributes to the growing demand for video analysis in sports. The study referenced Borghesi et al. and focused on utilizing Temporal Convolutional Networks (TCNs) for automating ball possession data extraction from tracking data in sports analytics [115]. The paper explored various classification approaches using TCNs to categorize game states. Performance evaluation on professional soccer tracking data demonstrated significant improvements over state-of-the-art methods, achieving 86.2% accuracy in possession estimation and 89.2% accuracy in dead-alive classification. The study also conducted ablation studies to understand the contributions of input data to the final prediction. A study by Bouazizi et al. explored using a cleaning robot equipped with a 2D LIDAR for fall detection and monitoring in environments with furniture obstructions [116]. By continuously collecting and processing LIDAR data, the robot can identify falls and individuals on the ground, achieving an 81.2% accuracy in fall detection and a 99% accuracy in detecting individuals in prone positions through simulations. This approach outperforms static LIDAR methods with lower accuracies. Khan et al. proposed a violence detection scheme for movies that involves three key steps: shot segmentation, frame selection based on saliency, and the classification of violence and non-violence shots using a fine-tuned deep learning model to create violence-free versions of movies, suitable for children and individuals sensitive to violence [117]. Bianculli et al. created a new dataset containing 350 high-resolution video clips (1920×1080 pixels, 30 fps) for the purpose of testing violence detection techniques [118]. The dataset consisted of 230 clips depicting violent behaviors and 120 clips representing non-violent behaviors, including actions that may lead to false positives in violence detection due to their resemblance to violent actions. Vijeikis et al. introduced a novel architecture for violence detection in video surveillance cameras [119]. The proposed model utilized a U-Net-like network with MobileNet V2 as an encoder for spatial feature extraction, followed by LSTM for temporal feature extraction and classification. This architecture is computationally lightweight yet achieves good results, with an average accuracy of 0.82 ± 2% and an average precision of 0.81 ± 3% when tested on a complex real-world security camera footage dataset (RWF-2000). As the authors stated, the effectiveness of this model in efficiently and accurately detecting violent events in surveillance footage. Ullah et al. developmed a triple-staged deep learning framework for violence detection in surveillance videos [120]. The model was optimized using an Intel toolkit for efficient execution. Wu et al. proposed a comprehensive approach to violence detection in videos [121]. The authors introduced a substantial and diverse dataset (XD-Violence) containing long untrimmed videos with audio signals and weak labels. The authors also emphasized the positive impact of multimodal (audio-visual) input and modeling relationships in violence detection.

F. PLANT AND HUMAN DISEASE DETECTION

Plant and human disease detection leverage object detection technology to identify and diagnose diseases in respective domains. In the context of plant disease detection, object detection algorithms are employed to locate and classify symptoms or anomalies on leaves, fruits, or other plant parts, aiding farmers in early disease identification and management. Similarly, in human disease detection, object detection can be used to identify abnormal regions in medical images, such as X-rays or MRIs, enabling timely diagnosis and treatment. Automating the identification process, object detection enhances the accuracy and efficiency of disease detection in both plant and human contexts, ultimately contributing to improved agricultural yields and better healthcare outcomes. Figure 12 shows various applicable fields of plant and human disease detection.

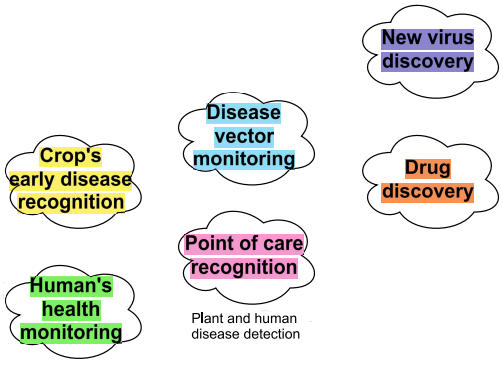

FIGURE 12. Plant and human disease detection fields.

Lee et al. previously developed a crop disease diagnosis solution that combines Image Captioning and Object Detection using deep learning methods [122]. The Image Captioning model achieved a high performance with a 64.96% average BLEU score, while the Object Detection model, with an mAP50 of 0.382, needs further improvement [122]. This solution enhances assistance to novice farmers by offering detailed descriptions of disease symptoms based

on severity, improving diagnosis reliability. Peng et al. developed an image retrieval system for plant disease identification. This system combines object detection and deep metric learning, offering flexibility in recognition categories and data requirements [123]. It effectively addresses plant disease detection across various scenarios. Poornappriya et al. previously applied Image Processing techniques and Deep Learning models to detect rice plant diseases, addressing the critical need for early disease detection and prevention in agriculture to mitigate yield loss and resource wastage [124]. Li et al. introduced CLU-CNNs, a domain adaptation framework tailored for medical images [125]. This framework addressed the data distribution disparity between source and target domains, enhancing domain adaptation without the need for domain-specific training. Leveraging the work by He et al., the study explored skin disease diagnosis using CNNs [126]. Two datasets were created: ''Skin-10'' and ''Skin-100,'' containing 10 and 100 skin disease classes, respectively. The research revealed lower accuracy for ''Skin-100'' compared to ''Skin-10.'' An ensemble approach was implemented, achieving 79.01% accuracy for ''Skin-10'' and 53.54% for ''Skin-100.'' The introduction of object detection to the ''Skin-10'' dataset improved accuracy for certain disease classes [126]. Liu et al. focused on object detection in scenarios with limited annotated bounding box data, a common issue in domains like medical imaging. They introduced a generative model that optimizes both image generation and object detection simultaneously. This approach outperformed existing methods, notably achieving a 20% relative improvement in average precision and a 50% relative increase in localization accuracy on challenging datasets such as disease detection and small-data pedestrian detection [127]. Terzi et al. proposed ensemble strategies to enhance deep learning object detection models for anatomical and pathological detection in brain MRI [128]. It assessed nine models, identified anatomical and pathological regions, and employed ensembles, achieving a potential 10% mAP increase and improved AP for anatomical parts. The ensemble approach outperformed individual models, especially for detecting small anatomical objects. Awad et al. addressed the challenges of big-medical-data classification and image detection by enhancing logistic regression and YOLOv4 algorithms. Their approach incorporated advanced parallel k-means pre-processing to identify data patterns and structures and utilized a neural engine processor for improved speed and efficiency [129]. Evaluation of large medical datasets confirmed the method's accuracy and reliability, demonstrating significant performance improvements for logistic regression and YOLOv4, offering a more robust solution for medical data classification and image detection in healthcare applications. Chotikunnan et al. enhanced industrial object detection via ResNet18-based image segmentation, favoring dual image processing for precise recognition [130]. They highlighted strengths and limitations in automation, with room for future improvements. Leng et al. served as a reference for the development of a transformer-based object detection network for leukocyte detection. The model, based on the DETR architecture, integrates pyramid vision transformers and deformable attention modules, achieving superior mean average precision compared to convolutional neural networks [131]. This research offers a valuable approach to leukocyte detection. Fan et al. conducted the first systematic study on concealed object detection (COD), introducing the COD10K dataset with rich annotations [132]. Their research unveiled SINet, a powerful baseline for COD, surpassing twelve contemporary baselines. The study offered significant insights and potential applications for this challenging field.

G. ASTRONOMICAL OBJECT DETECTION

Object detection, a powerful computer vision technique, finds applications in diverse fields, including the identification of astronomical objects detection. In the realm of astronomy, object detection assists in the automated recognition and tracking of celestial bodies, such as stars, planets, and asteroids, enabling astronomers to analyze their movements and characteristics more efficiently. Also, in the context of environmental monitoring, object detection can be employed to identify and classify airborne particulate matter, gases, and pollution sources, contributing to real-time air quality assessment. Harnessing object detection algorithms, these applications offer critical insights for both astronomical research and environmental protection. Figure 13 shows various applicable fields of astronomical object detection.

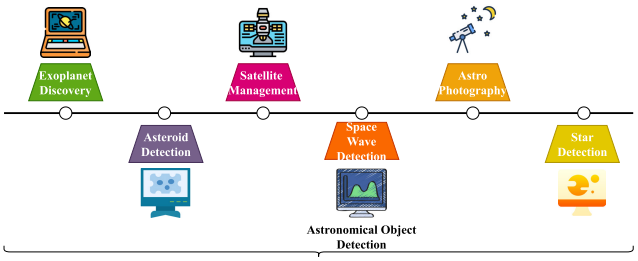

FIGURE 13. Astronomical object detection fields.

Kök et al. (2021), introduced a novel LSTM-based deep learning model for predicting air quality in IoT-based smart cities [133]. The model offers promising results in air quality prediction and has the potential to address various predictive challenges in smart cities, showcasing the applicability of deep learning in urban planning and management. Chakma et al. developed an efficient image-based method for PM2.5 concentration estimation using deep CNNs [134]. Their approach, based on a dataset of 591 images and corresponding PM2.5 concentrations from Beijing, proves effective for air quality analysis and pollution control. The study by Yang et al. explored the implementation of occupant-centric control (OCC) strategies in office environments, combining real-time indoor occupancy detection with OCC to enhance thermal comfort, maintain air quality, and reduce energy consumption (2.3%-8.1% savings) [135].

The research revealed that lower occupancy levels result in more significant improvements in comfort and energy efficiency, indicating the potential for more occupant-friendly and energy-efficient building management systems. The study by Jayawickrama et al. introduced a novel computer vision prediction model for detecting leftover items in shared vehicles, addressing cleanliness concerns. The model, utilizing convolutional neural networks (CNNs), achieved an accuracy of 89% for distinct classes of items and 91% for general classes of trash and valuables [136]. Additionally, an indoor air quality (IAQ) unit was implemented for monitoring specific air pollutants within the vehicle. Future research will focus on integrating these systems to assess cleanliness levels and expand the dataset with data from real shared vehicles in operation. Baker et al., introduced a portable, low-cost air-quality monitoring device (LPAQD) that detects particles from micron-sized down to 100 nm-sized particles in real-time. It accurately measures particulate matter densities as low as 3 μg m-3. The LPAQD uses a vapor-condensed film to enhance particle detection, making it capable of tracking even sub-pixel and sub-diffraction-limit-sized particles [137]. Its high dynamic range and affordability empower individuals to monitor air quality for better health decisions. Paithankar et al. introduced an innovative air quality monitoring system using IoT technology to provide real-time, cost-effective, and precise data, addressing concerns from industrial and transportation activities [138]. The system combines portable sensors, a low-power wide area network, and IoT for data analysis, demonstrating accuracy in air quality monitoring and pattern recognition, and contributing to urban air quality management. Dhanush et al. introduced an IoT-based air quality monitoring system to combat rising pollution levels driven by factors like population growth and industrialization [139]. The system displays air quality in parts per million (PPM) on LCD and web platforms for easy monitoring. It triggers alarms when harmful gases like CO2, smoke, alcohol, aromatic hydrocarbon, and NH3 reach critical levels. Moreover, it can activate devices or send alerts when pollution exceeds set thresholds. Sortino et al. analyzed deep learning approaches for automatic object detection and instance segmentation in the domain of radio astronomy, with a focus on the emerging need for such techniques in the era of Big Data and the Square Kilometre Array (SKA) telescope [140]. Prasad et al. A utilizaed time-series data analysis, primarily based on light-curves, for the detection and characterization of stars, exoplanets, and galaxies, with a focus on data accessible from the NASA Exoplanet Archive [141]. The authors highlighted the use of Box Least Squares (BLS) periodogram analysis to identify potential transiting signals from exoplanets and the application of a neural network classifier to determine the likelihood of an object being an exoplanet. This research leveraged NASA's wealth of data from missions like Kepler and TESS and provides an interactive means of exploring and analyzing space data, making it a valuable tool for astronomers and researchers in the field.

V. MULTI-NODAL CHALLENGES OF OBJECT DETECTION

Object detection architectures undergo extensive training on vast datasets to facilitate effortless identification of objects within input images. However, the efficacy of detection algorithms is influenced by diverse sets of parameters during the detection process. Figure 14 illustrates a taxonomy of some object detection challenges discussed following which interfare its efficiency to detect objects.

A. ANNOTATING TRAINING DATA

The process of annotating training data for object detection presents a myriad of challenges. Achieving accuracy and consistency in bounding box annotations is a primary concern, as variations can impact model performance. Subjectivity in determining object boundaries, coupled with the labor-intensive nature of manual annotation, results in time and cost implications. Handling complex object shapes, dealing with ambiguous object categorization, addressing data imbalance, and selecting appropriate annotation tools are critical considerations. Ensuring data privacy, maintaining annotation quality across annotators, and facilitating data augmentation while preserving accurate annotations are additional complexities. Moreover, tasks like 3D object annotation and video annotation pose specialized challenges, and the need for clear documentation, transparency, and the ability to adapt to evolving object categories further complicate the process. Overcoming these challenges necessitates well-defined guidelines, training, quality control measures, and suitable tools to create high-quality datasets essential for the development of accurate object detection models.

Zhang et al. introduced an adversarial-paced learning (APL) framework to reduce manual annotations. APL, inspired by self-paced learning, employed data-driven adversarial learning to establish a robust learning pace. Experiments showcased competitive results, achieved using only 1,000 human-annotated training images across four datasets [142]. Shao et al. introduced Objects365, a large object detection dataset with 365 categories and over 600K training images. It serves as a valuable resource for feature learning in object detection and semantic segmentation tasks, outperforming ImageNet models on the COCO benchmark [143]. Liu et al. tackled underwater object detection for robot picking, an increasingly important field. They identified challenges, such as the lack of test set annotations in existing datasets, leading to self-divided test sets and hindering benchmark comparisons [144]. To address this, they introduced the Detecting Underwater Objects (DUO) dataset and benchmark. DUO offers diverse underwater images with improved annotations, serving as an efficient benchmark for academic and industrial applications, including robot-embedded environments. Ponce et al. addressed the importance of datasets in object recognition research. They highlighted current dataset limitations and shared insights from existing efforts, along with innovative methods for acquiring large and diverse annotated datasets [145]. The

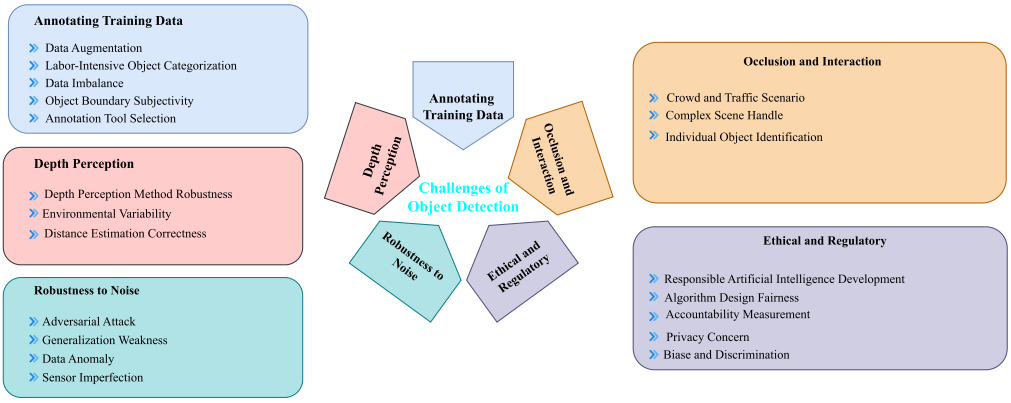

FIGURE 14. Taxonomy of object detection challenges.

paper also proposed criteria for gathering future datasets. Biffi et al. discussed the challenges of object detection relying on time-consuming manual annotations [146] They introduced an online annotation module (OAM) that generates reliable annotations from weakly labeled images. This OAM can enhance the performance of Fast(er) R-CNN, improving mAP by 17% and AP50 by 9% on PASCAL VOC 2007 and MS-COCO benchmarks, outperforming other methods using mixed supervision.

B. DEPTH PERCEPTION

The challenge of depth perception in object detection involves accurately estimating the distance or depth of objects from the sensor or camera, which is crucial for understanding the three-dimensional environment and ensuring the safety and reliability of autonomous systems. Accurate depth perception is essential for tasks like obstacle avoidance, scene understanding, and object recognition in autonomous vehicles, robotics, and various computer vision applications. However, challenges in depth estimation, such as handling occlusions, low-textured surfaces, and adverse weather conditions, can significantly affect the performance of object detection systems. Developing robust and accurate depth perception methods is vital to address these challenges and advance the capabilities of object detection systems in real-world scenarios.

Chen et al. addressed the challenge of depth perception affected by optical illusions in single-lens imaging. They presented a compact and intelligent depth-sensing meta-device, using a 3600-achromatic meta-lens array, measuring 1.2×1.2 mm2, capable of depth measurement and structured light projection [147]. The depth information is processed with deep learning and has implications for applications like autonomous driving, machine vision, augmented reality, and biometric identification. Chen et al. addressed challenges in RGB-D salient object detection. They introduced a unified feature fusion module, enhancing both semantic and spatial information, and a multi-scale contextual perception module [148]. Their method, outperforming 14 state-of-the-art approaches on eight datasets, effectively combines RGB and depth features. Wu et al. addressed the challenge of monocular 3D object detection in autonomous driving and mobile robotics. They introduced Depth Dynamic Center Difference Convolution (DDCDC) to improve depth estimation using surrounding pixel cues with distinct convolution weights for each pixel [149]. Their end-to-end detection network with DDCDC modules achieved significant improvements in evaluation results on KITTI and nuScenes datasets. Hou et al. aimed to boost RGB-D object detection by effectively coordinating depth with RGB data [150]. They introduced a two-stage learning framework, involving property derivation and fusion, to comprehensively describe objects by deriving visual properties from color/depth and their pairs. Their experimental results on NYUD2 and SUN RGB-D datasets outperformed baselines. Aguilar et al. created a real-time system for object detection and depth estimation using a micro-UAV's onboard camera and convolutional neural networks [151]. They emphasized that their approach avoids the need for complex SLAM visual systems, making it resource-efficient and faster. Training the neural networks with stereo images enables accurate real-time obstacle detection. Gupta et al., conducted a survey on deep learning in autonomous vehicles, covering theory, implementations, and evaluations [152]. They aim to bridge the gap between deep learning and autonomous driving, discussing self-driving car fundamentals, deep learning, and computer vision. The survey explores techniques for image perception and evaluates recent implementations, concluding with research recommendations.

C. ROBUSTNESS TO NOISE

Robustness to noise is a critical challenge in object detection, primarily due to environmental variability, sensor imperfections, adversarial attacks, object occlusion, data anomalies, class imbalance, and generalization issues. Real-world conditions introduce noise through changing lighting, sensor imperfections, and adversarial attacks can manipulate input data. Additionally, object occlusion, data anomalies,

class imbalances, and difficulties in generalization further compound this challenge. Addressing robustness to noise in object detection requires strategies like data augmentation, adversarial training, and the development of robust model architectures. Furthermore, the collection of diverse datasets that incorporate noisy and real-world examples is crucial for training object detection models that perform reliably in practical scenarios.

Liu et al., addressed noisy annotations in Domain Adaptive Object Detection (DAOD) with their Noise Latent Transferability Exploration (NLTE) framework, improving mAP on benchmark DAOD datasets when 60% of annotations were corrupted [153]. Volk et al. addressed the need for robust convolution Neural Networks (CNNs) in automotive object detection. They automated data augmentation to enhance robustness against natural distortions caused by different weather conditions, such as rain. Their approach outperformed existing techniques like Gaussian noise or Salt-and-Pepper noise when validated against real rain datasets [154]. Adhikari et al., explored label noise's effect on object detection loss functions, essential for system robustness [155]. They focused on missing labels, simulated during training, and compared cross-entropy loss and focal loss. They discovered that adjusting focal loss hyperparameters can enhance its robustness, allowing for up to 50% missing labels. Sabater et al. focused on improving object recognition in videos, vital for applications like autonomous driving and surveillance [156]. They introduced an innovative post-processing method using learning-based similarity evaluation, enhancing state-of-the-art video detectors, especially for fast-moving objects, while maintaining low computational requirements. When applied to efficient still image detectors like YOLO, it achieved comparable results to computationally intensive alternatives.

D. OCCLUSION AND INTERACTION

Occlusion and interaction pose significant challenges in object detection. Occlusion occurs when objects are partially or completely blocked by other objects or obstacles in the scene, making it difficult for detectors to accurately identify and locate them. Interactions involve objects that are in close proximity or contact with each other, such as in a crowd or traffic scenario. Detecting and distinguishing individual objects within these complex scenes is a demanding task for object detection algorithms. Accurate handling of occlusion and interaction is crucial for applications like autonomous driving, surveillance, and robotics, as it directly impacts the reliability and safety of these systems. Researchers are continuously working to improve object detection models' ability to address these challenges and enhance their performance in real-world, cluttered environments.

Song et al. addressed the challenge of recognizing partially occluded faces using a mask learning strategy [157]. By establishing a mask dictionary through innovative methods, they effectively identified and discarded corrupted feature elements during recognition. This approach significantly outperformed existing systems in experiments with occluded face datasets. The VOT2020 challenge, organized by Kristan et al. assessed 58 trackers across five sub-challenges, spanning various tracking domains [158]. These sub-challenges covered short-term tracking in RGB, real-time short-term tracking in RGB, long-term tracking with target disappearance and reappearance, short-term tracking in RGB and thermal imagery, and long-term tracking in RGB and depth imagery. Notably, VOT-ST2020 introduced a new evaluation methodology and replaced bounding boxes with segmentation ground truth. Wang et al. introduced a novel approach for human-object interaction (HOI) detection that focuses on interactions between human-object pairs [159]. This fully convolutional method predicts interaction points, which directly localize and classify the interactions. It is the first approach to framing HOI detection as a key point detection and grouping problem. The method was tested on the V-COCO and HICO-DET benchmarks. Wu et al. addressed video object detection (VID) challenges related to appearance degradation in fast-motion frames [160]. They introduced the Sequence Level Semantics Aggregation (SELSA) module, which aggregates features at the full-sequence level, improving performance on ImageNet VID and EPIC KITCHENS datasets while simplifying the pipeline. Liu et al. introduced Position Embedding Transformation (PETR) for multi-view 3D object detection, achieving state-of-the-art results with 50.4% NDS and 44.1% mAP on the nuScenes dataset [161].

E. ETHICAL AND REGULATORY

Ethical challenges in object detection encompass privacy concerns stemming from pervasive surveillance and data collection, potential biases and discrimination in algorithmic decision-making, and the misuse of surveillance technologies. They raise questions about the infringement of civil liberties, transparency, and accountability in data collection practices. Object detection also presents ethical dilemmas related to autonomous weapons, job displacement, and concerns about unauthorized surveillance or harassment. Addressing these challenges requires clear policies, accountability measures, fairness in algorithm design, and the importance of informed consent. Responsible AI development and regulatory safeguards are essential to ensure the ethical use of object detection technology while respecting individual rights and societal values.

Chattopadhyay et al. addressed security concerns related to adversarial attacks on models and data [162]. The study explored dimensionality and spatial patterns to improve adversarial robustness. It also investigated protecting model ownership using watermarking, extending the concept to natural language processing. Additionally, the research focused on data privacy in decentralized setups, utilizing federated learning and differential privacy to prevent information leakage and malicious attacks. The findings have advanced

the reliability and privacy of machine learning systems. Kagan et al. analyzed privacy risks in video conferencing. They examined 15,700 collage images and 142,000 face images from public meetings, demonstrating the ease of extracting personal information, including age, gender, usernames, and full names [163]. The study emphasized potential risks when facial images are linked with social network data and the need to address privacy concerns in virtual meetings, affecting various user groups. Lin et al. introduced HS-YOLO, a method to enhance small object detection in power safety monitoring [164]. It is based on HRNet and uses parallel branches to process feature maps of various scales and maintain microfeature information. HS-YOLO achieved an mAP of 87.2%, surpassing YOLOv5 by 3.5%, and notably improved the detection of small objects in power operation scenarios. Sirisha et al., delved into object detection in computer vision, comparing two-stage and one-stage detectors. While two-stage detectors excel in detection accuracy, one-stage detectors like YOLO have made significant strides in this regard. The study explored performance metrics, regression formulations, and different YOLO variations, shedding light on their design, performance, and applications [165]. BinDarwish, Abdulaziz et al. addressed the rising issue of ATM-related crimes by proposing a system for bank ATMs [166]. This system detects dangerous objects like weapons and employs facial recognition to identify potential repeat offenders. Using object, face, and action recognition algorithms, their approach effectively detects threats.

VI. REMARKABLE INVENTIONS AND FUTURE RESEARCH DIRECTIONS

A. REMARKABLE TECHNIQUES TO TACKLE CHALLENGES

1) FEW SHOT LEARNING BASED DETECTION

Few-shot learning in object detection involves training a model to recognize objects with limited annotated examples. Figure 15 illustrates the methodology, which leverages a pre-trained model on a large dataset to capture generic features and knowledge about objects. The few-shot learning process is then introduced by fine-tuning the model on a small dataset containing only a limited number of examples for each object class. Transfer learning techniques, such as meta-learning or episodic training, are commonly employed to enhance the model's ability to generalize from the small dataset to unseen objects. During training, the model learns to adapt quickly to new classes and instances with minimal labeled data. Techniques like episodic memory, where the model is trained on episodes comprising a small support set and a query set, enable the model to generalize effectively to novel classes. Few-shot object detection methodologies often aim to strike a balance between utilizing the knowledge from a large dataset and adapting to specific, limited examples, enabling the model to perform well on new, unseen object classes with only a handful of annotated samples. The rapid progress in deep learning offers effective solutions for remote sensing image interpretation, but limited labeled samples

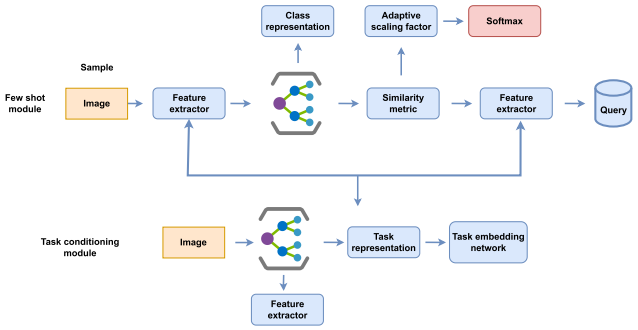

FIGURE 15. Few shot object detection model.

in the field underscore the need for few-shot learning. Figure 15 illustrates the generic architecture of few shot object detection.

Sun et al. presented a bibliometric analysis, introduced two few-shot learning methods, and outlined typical remote sensing applications, serving as a valuable reference for scholars in this domain [167]. Quellec et al. introduced a few-shot learning framework, merging convolutional neural networks (CNNs) with an unsupervised probabilistic model. Outperforming other frameworks on a dataset of 164,660 screening exams, it detected 37 out of 41 conditions with an average AUC of 0.938 [168]. This advancement has the potential to automate eye pathology screening, revolutionizing clinical practice in ophthalmology. Wang et al. introduced FSL-SCNN, a few-shot learning model with a Siamese CNN, addressing limited labeled data challenges for intelligent anomaly detection in industrial cyber-physical systems. The model aims to enhance accuracy and mitigate overfitting issues by measuring distances between input samples based on optimized feature representations. A robust cost function with three specific losses is proposed, demonstrating significant improvements in false alarm rates (FAR) and F1 scores for detecting intrusion signals in industrial cyber-physical security [169]. Zhang et al. introduced G-FSDet, a Generalized Few-Shot Detector for remote sensing images (RSIs), addressing the limitations of current few-shot object detection methods. G-FSDet prevents forgetting previous knowledge and achieves competitive novel class performance with minimal base class degradation, demonstrating state-of-the-art overall performance on DIOR and NWPU VHR-10.v2 datasets. Source code is available [170]. Yang et al. introduced HESFOL, a spike-based framework for few-shot learning in neural networks. Using entropy theory, HESFOL establishes a gradient-based few-shot learning scheme in a recurrent SNN architecture. Evaluation on few-shot tasks and motor control shows improved accuracy and robustness, emphasizing the application of entropy-based methods in spike-driven learning [171]. This offers new perspectives for enhancing SNN learning and applied developments in neuromorphic systems. Hu et al., addressed few-shot learning (FSL) in computer vision, focusing on a practical and effective pipeline. They explored neural

architecture, employed a three-stage approach involving pre-training on external data, meta-training on labeled few-shot tasks, and task-specific fine-tuning. The study investigated the benefits of pre-training on external data, the utilization of state-of-the-art transformer architectures, and optimal fine-tuning strategies [172]. Demonstrating strong performance on benchmarks like Mini-ImageNet, CIFAR-FS, CDFSL, and Meta-Dataset, the simple transformer-based pipeline proved effective. Kang et al. introduced a few-shot object detector to address limited bounding box annotations for rare categories. The model quickly adapted to novel classes using fully labeled base classes, a meta feature learner, and a reweighting module within a one-stage detection architecture [173]. Trained end-to-end with an episodic few-shot learning scheme, the model significantly outperformed baselines in few-shot object detection across multiple datasets, inspiring future research in this area. Koizumi et al. addressed the challenge of overlooking anomalies in anomaly detection systems by proposing a method for training a cascaded specific anomaly detector using few-shot samples (1 to 3). This approach aims to reduce false negatives by decreasing the false-positive rate while maintaining a true-positive rate of 1. Experimental results demonstrated the superiority of the proposed method over conventional cross-entropy-based few-shot learning methods, providing an effective solution for updating systems to avoid overlooking observed anomalies [174]. Gamal et al. proposed CNN-IDS, a Few-Shot Deep Learning-based Intrusion Detection System for IoT networks [175]. The system automatically identifies zero-day attacks from the network edge using a two-stage approach: 1) a filtered Information Gain method for feature selection, and 2) a one-dimensional Convolutional Neural Network (CNN) algorithm for recognizing new attack types. Trained on UNSW-NB15 and Bot-IoT datasets, the model exhibited improved detection rates and reduced false-positive rates, enhancing IoT system security.

Heidari et al. introduced a few-shot error detection framework with effective data augmentation to minimize human involvement. The approach used an expressive model and data augmentation on a small set of clean records, achieving 94% average precision and 93% average recall across diverse datasets [176]. Outperformed traditional methods, showing a 20-point average F1 improvement with 3x fewer labeled examples. In addressing the scarcity of abnormal samples in network intrusion detection, a Few-Shot Learning (FSL) method was introduced. With less than % of the NSL-KDD KDDTrain+ dataset utilized for training, high accuracy was achieved—92.34% for KDD-Test+ and 85.75% for KDD-Test-21. In contrast, lower accuracy was observed with traditional methods (J48, Naive Bayes, Random Forest, Support Vector Machine, recurrent neural network, and deep convolutional neural network), despite using 20% of the KDDTrain+ dataset [177]. Detection rates for Dos, U2R, R2L were improved, notably increasing U2R and R2L detection rates from 13% to 81.50% and 44.41% to 75.93%, respectively, on the UNSW-NB15 dataset.

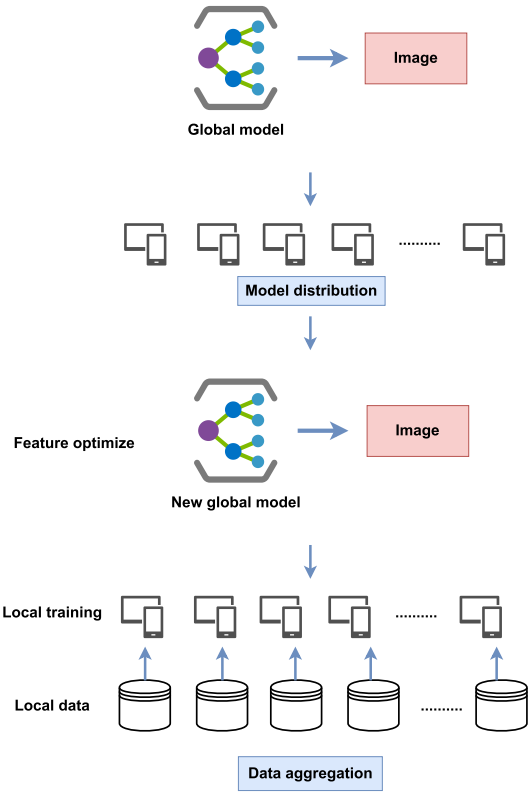

FIGURE 16. Federated object detection model.

2) FEDERATED LEARNING BASED DETECTION

Federated learning-based object detection is a collaborative machine learning approach that aims to train robust and accurate object detection models across decentralized and privacy-sensitive devices. In this methodology shown in Figure 16, multiple edge devices, such as smartphones or Internet of Things (IoT) devices, collaboratively participate in the training process without sharing their raw data centrally. The process begins with the distribution of a pre-trained model to these devices. Each device then refines the model locally using its own data, extracting relevant features and updating the model parameters based on its observations. Subsequently, these locally updated models are aggregated or federated in a secure and privacy-preserving manner. The federated averaging or similar aggregation techniques ensure that the collective knowledge from all devices contributes to the global model without exposing individual data instances. This iterative process of local training and global aggregation continues until the object detection model achieves satisfactory performance. Federated learning, by preserving data privacy at the edge and promoting collaborative model improvement, addresses concerns related to centralized data storage and enhances the scalability and efficiency of object detection systems across diverse and distributed environments. Figure 16 demonstrates the generic model of FL based object detection.

Mothukuri et al. introduced a Federated Learning (FL)-based anomaly detection approach for IoT security,

prioritizing user data privacy. The method uses federated training rounds on gated recurrent units (GRUs) models, sharing only learned weights with the central server. The ensembler aggregates updates from multiple sources to optimize the global ML model's accuracy [178]. Experimental results show superior performance over classic/centralized machine learning (non-FL) versions, ensuring user data privacy and achieving optimal accuracy in attack detection within IoT networks. Yang et al. introduced FFD (Federated learning for Fraud Detection) to enhance credit card fraud detection, using federated learning for privacy-preserving model training on locally distributed data. The shared Fraud Detection System (FDS) aggregates locally computed updates, and an oversampling approach balances the skewed dataset. In real-world credit card transactions, FFD achieved an average test AUC of 95.5%, surpassing traditional FDS by approximately 10% [179]. Tian et al. introduced DC-Adam, an asynchronous federated learning-based detection approach for resource-limited IoT devices. Addressing the gradient delay problem in non-IID data patterns, DC-Adam utilizes a Taylor Expansion-based scheme for compensation and a pre-shared data training strategy [180]. The method demonstrates stable convergence and outperforms benchmarks, showing a significant improvement in accuracy, precision, recall, and F1 score compared to barrier-free asynchronous federated learning. Preuveneers et al. introduced a blockchain-based federated learning method to enhance accountability in cybersecurity. Addressing the challenge of potential poisoning in federated learning setups, the solution integrates federated learning with a permissioned blockchain, allowing incremental updates on the distributed ledger. Experiments in intrusion detection show a limited performance impact (5-15%) while providing transparency over the distributed training process [181]. The blockchain-based federated learning solution is applicable to various neural network architectures and use cases. Liu et al. introduced FedVision, a platform for developing federated learning-powered computer vision applications, addressing privacy concerns and data transmission costs. Deployed in smart city applications, FedVision achieved efficiency improvement and cost reduction, eliminating the need to transmit sensitive data for three major customers [182]. This marks the first real application of federated learning in computer vision-based tasks. Huong et al. introduced FedeX, an architecture for distributed anomaly detection in IoT-based Industrial Control Systems (ICSs) for Smart Manufacturing. FedeX achieves high detection performance, fast learning of new data patterns, and lightweight deployment on resource-constrained edge devices. In experiments, it outperformed 14 existing solutions on all metrics with liquid storage and SWAT datasets, demonstrating fast training (7.5 minutes) and lightweight hardware requirements (14% memory consumption) [183]. FedeX incorporates Explainable AI (XAI) for interpreting predicted anomalies, enhancing decision-making trust in real-time on edge computing infrastructure. Huong et al. proposed a Federated Learning-based anomaly detection for Industrial IoT in Smart Manufacturing, outperforming existing solutions. The architecture is efficient on edge computing hardware, saving 35% bandwidth, with realistic resource consumption (max CPU 85%, avg. memory 37%) [184]. Liu et al. introduced a cooperative intrusion detection mechanism for vehicular networks, leveraging distributed edge devices like connected vehicles and RSUs. The federated-based approach offloads model training to enhance efficiency, utilizing blockchain for secure storage and sharing of models. The scheme ensures cooperative privacy-preservation for vehicles, reducing communication overhead and computation costs while maintaining security [185]. Jahromi et al. introduced a scalable deep federated learning-based method for Industrial Control System (ICS) security, addressing IT-ICS network differences and data privacy issues. In this method, clients train local unsupervised deep neural network models, share parameters with a server, which aggregates them to create a generalized public model [186]. Evaluation on a real-world ICS dataset in a water treatment system demonstrates superior performance compared to non-federated learning-based methods, with similar computational complexity to existing deep neural network-based approaches in the literature.

3) EXPLAINABLE ARTIFICIAL INTELLIGENCE

Explainable Artificial Intelligence (XAI) in the context of object detection involves employing transparent and interpretable models, as well as developing post-hoc explanations to enhance the comprehensibility of the detection process. Initially, interpretable models, such as decision trees or rule-based systems, are favored over complex black-box models like deep neural networks. These models facilitate understanding by providing explicit rules governing object detection. Additionally, post-hoc explanation methods, such as saliency maps or attention mechanisms, are applied to illuminate the key features influencing model predictions. These visualizations help users discern why a specific object was detected or provide insights into potential biases. By combining both interpretable models and post-hoc explanations, the methodology aims to make object detection systems more transparent, trustworthy, and accessible for users, fostering confidence in the decisions made by AI algorithms. Figure 17 shows an architecture of XAI enabled object detection.

4) RECONFIGURABLE COMPUTING BASED DETECTION

Reconfigurable computing-based object detection leverages the flexibility and adaptability of reconfigurable hardware, typically using Field-Programmable Gate Arrays (FPGAs) or similar technologies. The process begins with the selection of an appropriate object detection algorithm, often a convolutional neural network (CNN), which is then tailored and optimized for deployment on reconfigurable hardware. The FPGA's reconfigurability allows for dynamic adjustments to

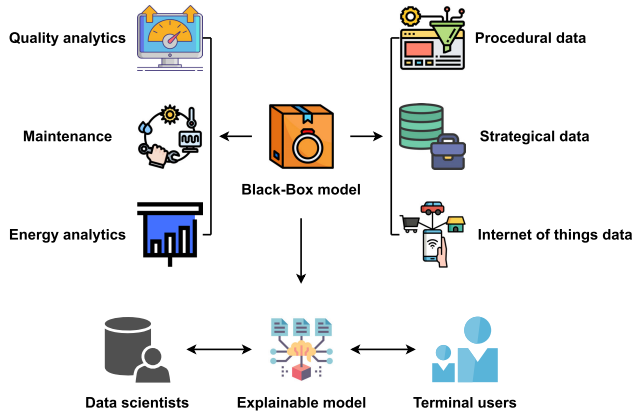

FIGURE 17. Explainable Artificial Intelligence model.

the hardware architecture, enabling efficient parallelization and acceleration of the algorithm. This adaptability is particularly advantageous in scenarios where real-time processing or low-latency requirements are critical. The methodology involves the design and mapping of the algorithm onto the FPGA, exploiting parallelism and optimizing resource utilization. Additionally, reconfigurable computing facilitates iterative refinement and optimization of the object detection model based on specific application requirements. This methodology offers a balance between the performance benefits of hardware acceleration and the flexibility to adapt to evolving detection tasks, making it suitable for applications such as video surveillance, autonomous vehicles, and other real-time image processing tasks.

Li et al. [187] demonstrated a reconfigurable and non-volatile neuromorphic device based on two-dimensional semiconducting metal sulfides, functioning as a photovoltaic detector. The device, utilizing a metal-semiconductor-metal (MSM) two-terminal structure with pulse-tunable sulfur vacancies, achieved highly tunable responsivity and concurrent non-volatile storage of image data. Additionally, a convolutional neuromorphic network was designed for image processing and object detection using the same device, showcasing its potential as a key component in visual perception hardware. Zhao et al. [188] introduced a novel method for optimizing CNN-based object detection algorithms on embedded FPGA platforms, addressing limited computation resources and power constraints. The key contributions include parameterized CNN hardware modules, an optimization flow that considers network architectures and resource constraints, and achieved results demonstrating over 85% accuracy and a 49.6 times speed-up compared to software implementation with optimized configurations. Chan et al. [189] explored the use of modern FPGAs for accelerating intelligent vision-guided crop detection in agricultural field robots, adapting the YOLOv3 object detection neural network for broccoli and cauliflower detection. The FPGA implementation achieved a 92% mAP with efficient quantization, demonstrating superior power efficiency and throughput compared to an embedded GPU. Specifically, the FPGA solution is 4.12 times more power-efficient and offers 6.85 times higher throughput, leading to faster and longer operation of battery-powered field robots. Kim et al. [190] introduced an optimized convolutional neural network (CNN) accelerator design on a mobile FPGA, specifically focusing on limited-resource edge computing environments. The reconfigurable accelerator design was implemented at the register-transfer level (RTL), employing low-power techniques to enhance programming speed. The optimization techniques included clock gating to eliminate residual signals and deactivate unnecessary blocks. The proposed design, tested with Resnet-20 on the CIFAR-10 dataset, demonstrated a significant improvement in power efficiency consumption (16%), hardware utilization (up to 58%), and throughput (15%) based on experimental results. Na et al. [191] proposed a novel active learning algorithm. The algorithm considers both classification and localization informativeness of unannotated video frames, leveraging temporal information to measure localization informativeness. The evaluation on the MuPoTS and FootballPD datasets demonstrates the effectiveness of the proposed algorithm in selecting informative frames for annotation in object detection training. Baczmanski et al. [192] implemented a perception system for autonomous vehicles by leveraging the MultiTaskV3 detection-segmentation network on the AMD Xilinx Kria KV260 Vision AI embedded platform. The system demonstrated high efficiency in obstacle recognition, real-time performance, and energy efficiency, achieving over 97% mean average precision for object detection and above 90% mean intersection over union for image segmentation. Additionally, the implementation on the FPGA platform resulted in significant power savings (5 watts on average) and a compact form factor (119mm x 140mm x 36mm), making it suitable for space-constrained applications.

5) QUANTUM COMPUTING BASED DETECTION

Quantum computing-based object detection employs the principles of quantum mechanics to enhance traditional object detection algorithms. Unlike classical computers that process information using bits, which can exist in a state of 0 or 1, quantum computers use qubits, which can exist in multiple states simultaneously due to superposition. This inherent parallelism allows quantum algorithms to explore multiple possibilities simultaneously, potentially speeding up computations for object detection tasks. Quantum entanglement, another quantum phenomenon, enables qubits to be correlated in a way that the state of one qubit is directly related to the state of another, facilitating complex computations. Quantum parallelism and entanglement can be leveraged to optimize the processing of large datasets commonly encountered in object detection, leading to more efficient and faster identification of objects within images or videos. However, the field is still in its infancy, and practical implementations of quantum computing for object detection face numerous

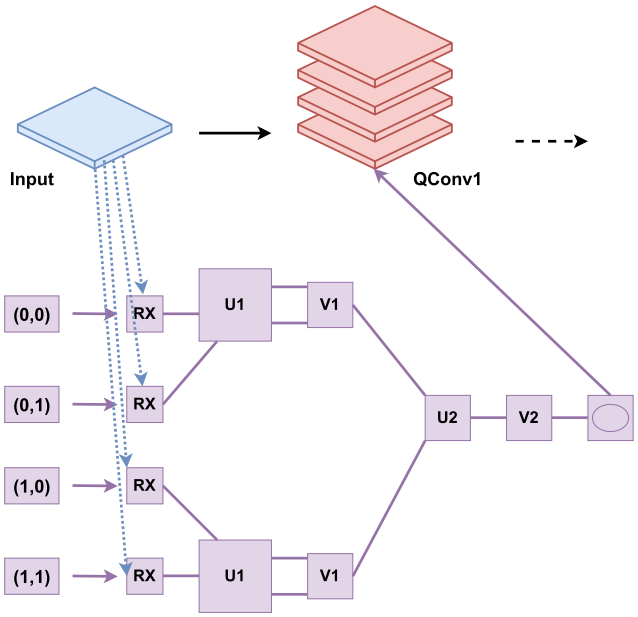

FIGURE 18. Quantum architecture for object detection.

challenges, including error correction, decoherence, and the need for scalable quantum hardware. Ongoing research aims to address these issues and unlock the full potential of quantum computing in revolutionizing object detection methodologies. Figure 18 shows the quantum model object detection.

Zaech et al. [193] presented a novel formulation of Multi-Object Tracking (MOT) tailored for Adiabatic Quantum Computing (AQC) by utilizing an Ising model. The proposed approach demonstrated competitiveness with state-of-the-art optimization methods, even when compared to traditional integer programming solvers. Additionally, the research showcased the solvability of the MOT problem on current quantum computers for small instances, highlighting its potential for addressing NP-hard optimization challenges in the field. Rajesh et al. [194] explored the application of Quantum Convolutional Neural Networks (QCNN) in computer vision, specifically focusing on image recognition and object detection. The research demonstrated that QCNN can enhance computational speeds and outperform classical methods, with potential applications in computer vision, signal processing, pharmaceuticals, cryptography, and other fields, highlighting the significance of ongoing developments in quantum computing algorithms and hardware support. Furthermore, Meedinti et al. [195] also explored the efficacy of Quantum Convolutional Neural Networks (QCNNs) compared to classical Convolutional Neural Networks (CNNs) and Artificial/Classical Neural Network (ANN) models for object detection and classification. Through comprehensive evaluations, it was found that QCNNs, leveraging qubits in a quantum environment, exhibited superior accuracy and efficiency, particularly in handling large datasets and real-time processing, demonstrating their potential as a promising advancement in machine learning. Hu et al. [196] presented a novel quantum automated object detection algorithm designed for urban surveillance systems. The algorithm demonstrated high accuracy in detecting objects in images and exhibited the capability to handle measurement errors arising from quantum measurements during the image retrieval process. The research contributed to the exploration of quantum computing applications in the field of computer vision, showcasing its potential for enhancing object detection in urban surveillance scenarios.

6) MIXED ARCHITECTURAL DETECTION: BLOCKCHAIN INTEGRATION

The blockchain based object detection model which integrates federated learning, blockchain, and few-shot learning methodologies as illustrated in 19 can be a powerful detection techniques. Federated learning allows the model to train across decentralized devices, preserving data privacy by keeping data localized. Blockchain technology ensures a transparent and secure record of model updates and transactions, enhancing trust in the training process. Few-shot learning empowers the model to generalize from a limited number of examples, improving adaptability to novel objects. The combination of these approaches results in a robust and privacy-preserving object detection model with a transparent and secure training process, capable of learning from minimal examples for enhanced versatility. Figure 19 demonstrates the generic model of blockchain integrated object detection.

She et al. introduced a blockchain trust model (BTM) for detecting malicious nodes in wireless sensor networks (WSNs). The model ensured fairness and traceability in the detection process by presenting a framework, constructing a blockchain data structure, and implementing detection in 3D space through blockchain smart contracts and the WSNs' quadrilateral measurement localization method [197]. Simulation results demonstrated effective malicious node detection in WSNs with traceability assurance. Guha Roy et al. proposed a decentralized security mechanism for IoT in mobile edge and fog computing. The system integrates SDN and blockchain for continuous monitoring and analysis of system traffic, providing an attack identification model. Blockchain addresses failure issues, delivering a decentralized attack identification scheme that detects and reduces attacks in the fog and edge nodes [198]. BAD (Blockchain Anomaly Detection) by Signorini et al. is a pioneering solution tailored for anomaly detection in blockchain-based systems [199]. This distributed, tamper-proof, trusted, and private framework leverages blockchain meta-data, providing effective protection against attacks specific to blockchain systems, as validated through experimental results and analysis. Mirsky et al. introduced a novel anomaly detection approach and a lightweight blockchain-based framework to secure IoT devices against vulnerabilities and adversarial attacks. The framework, tested on a distributed IoT simulation platform, utilizes blockchain for incremental

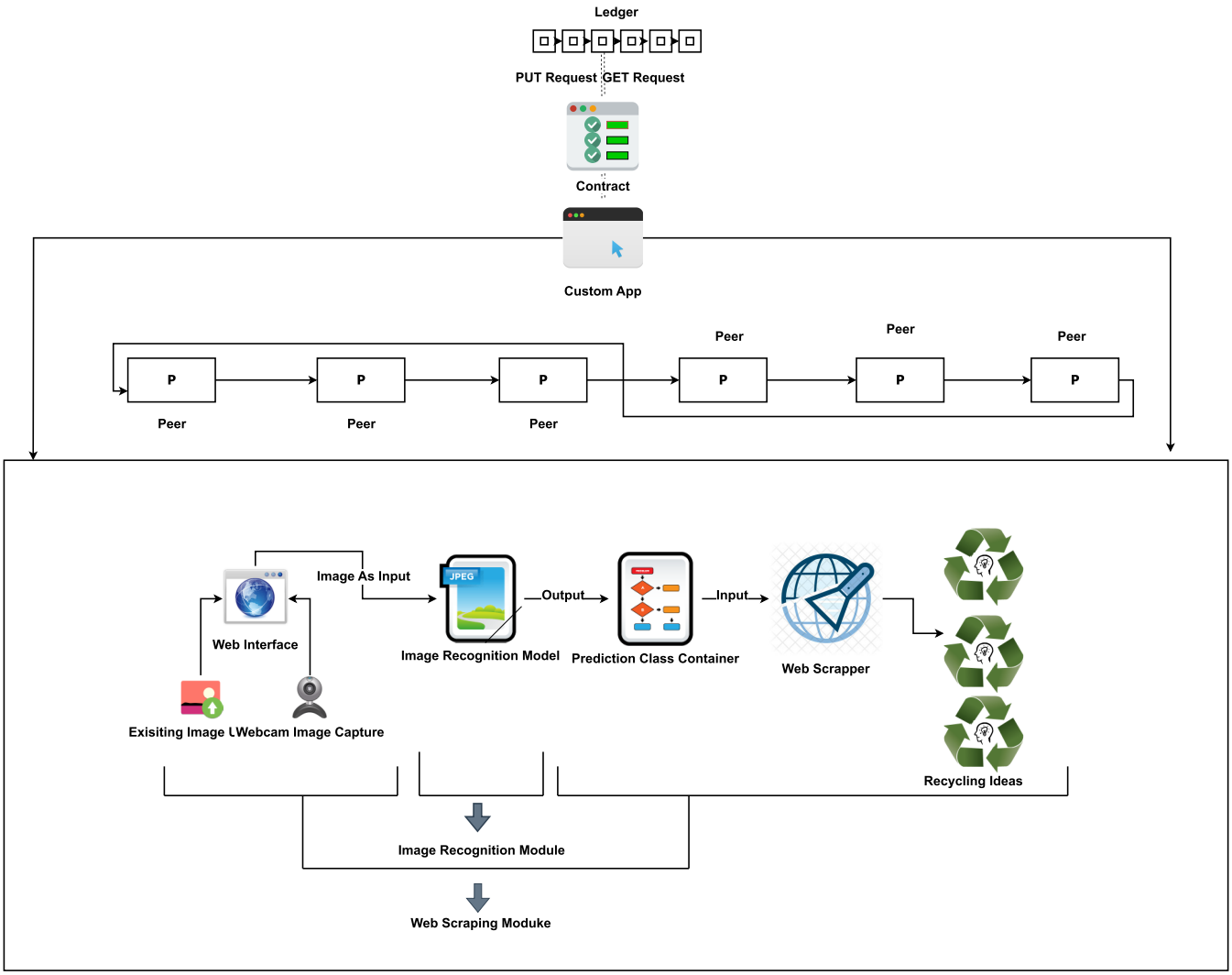

FIGURE 19. Blockchain integrated object detection model.

updates to a trusted anomaly detection model, enhancing the security of individual devices and the entire network [200]. Ashfaq et al. proposed a secure fraud detection model for the Bitcoin network, addressing evolving fraud methods in e-banking and online transactions. The model integrates machine learning algorithms (XGboost and random forest) for transaction classification, training on fraudulent and integrated transaction patterns [201]. Blockchain technology is employed to enhance security. The proposed smart contract undergoes a security analysis, and an attacker model is introduced for system protection, demonstrating robustness against attacks and vulnerabilities. Yu et al. introduced a DNS Cache Resources Trusted Sharing Model using consortium blockchain to enhance DNS resolution credibility [202]. It employs a trust-based incentive mechanism, addressing free-riding, and maintains efficiency with a decentralized storage mechanism. The model proves advantageous in ensuring credibility and efficiency in domain name resolution. Han et al. proposed a blockchain-based medical information-sharing model to address security threats in medical data distribution [203]. Through actual implementation, the model demonstrates reliability, traceability, and a data recovery function to prevent forgery and alteration of medical information. Wang et al. proposed a blockchain-based risk management system for network public opinion (NPO) to enhance risk prediction accuracy and credibility detection. The study utilizes smart contracts and risk association tree technology within the blockchain environment, allowing for traceability of public opinion through a smart ledger [204]. The experimental results demonstrate the effectiveness of the proposed model in optimizing the network environment and enhancing control measures. Yazdinejad et al. introduced a fuzzy blockchain framework for secure IoT environments. The approach integrates fuzzy logic, ANFIS-based attack detection, and fuzzy matching to enhance threat detection and fraud prevention [205]. Results confirm

improved security metrics in both blockchain and IoT networks.

B. FUTURE DIRECTIONS

The future landscape of object detection research is poised for dynamic evolution as scholars and practitioners strive to enhance the precision, efficiency, and adaptability of detection models. One pivotal avenue of exploration is in the realm of few-shot learning, an area dedicated to empowering object detectors to glean meaningful insights from a limited number of examples per class. This pursuit gains particular significance in scenarios where labeled data is scarce, prompting researchers to investigate novel methodologies and frameworks for robust learning with minimally labeled instances.

Attention mechanisms have become a cornerstone of research aimed at enhancing object detectors' ability to focus on salient features within an image. Recent works demonstrate the potential of attention mechanisms in refining spatial relationships, thus contributing to heightened accuracy and contextual awareness in object detection systems.

Another salient trajectory of future research lies in the domain of cross-domain object detection, focusing on the critical task of adapting models to operate seamlessly in diverse environments beyond their initial training data. Techniques devised for domain adaptation have shown promise in augmenting the generalization capabilities of object detectors, thereby enhancing their applicability across a spectrum of real-world scenarios characterized by varying domains.

The quest for real-time efficiency remains an ongoing and significant research focus. Researchers are continuously exploring methodologies to strike an optimal balance between the speed and accuracy of object detection algorithms, especially in applications where real-time responsiveness is imperative. This pursuit involves refining existing algorithms and exploring innovative architectures that can meet the demands of dynamic, time-sensitive environments. The integration of semantic segmentation and object detection is emerging as a compelling avenue, promising a more nuanced and holistic understanding of visual scenes. Approaches such as panoptic segmentation exemplify the synthesis of these two tasks, offering the potential for object detectors to leverage richer contextual information, thereby elevating their overall performance.

Another future direction of object detection is poised to intertwine with the principles of green computing, driven by a growing awareness of environmental sustainability. As the demand for more sophisticated object detection models continues to rise, researchers are increasingly focusing on developing energy-efficient algorithms and hardware architectures. Green computing in object detection involves optimizing model architectures to minimize computational complexity, exploring lightweight neural network designs, and leveraging quantization techniques to reduce computational and energy requirements. Additionally, researchers

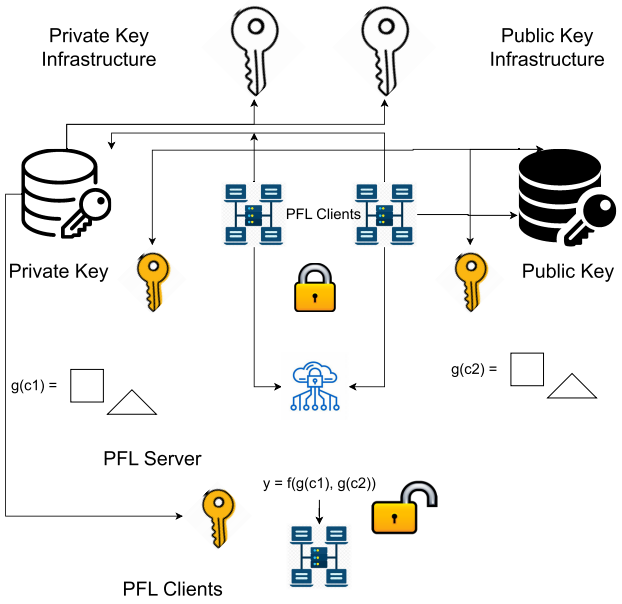

FIGURE 20. HE integrated with FL in object image classification task.

are investigating the integration of renewable energy sources and energy-aware scheduling strategies to power data centers that handle object detection tasks. The pursuit of eco-friendly computing solutions aims to mitigate the environmental impact of large-scale AI deployments while ensuring that future object detection systems remain efficient and sustainable. As the intersection of object detection and green computing evolves, the field is expected to witness innovations that not only enhance detection accuracy but also contribute to a more energy-conscious and environmentally responsible approach to computational tasks.

The foray into 3D object detection is gaining prominence, propelled by the growing relevance of applications in autonomous driving, robotics, and augmented reality. Researchers are actively engaged in advancing techniques that extend traditional object detection paradigms to seamlessly operate in the three-dimensional space, harnessing point cloud data for a more comprehensive and accurate perception of the environment. The imperative to fortify object detectors against adversarial attacks is shaping a critical research direction. Ensuring the robustness of models in the face of deliberate attempts to deceive or compromise their functionality is paramount. This involves the exploration of novel architectures, training methodologies, and evaluation frameworks to withstand adversarial challenges and instill confidence in the reliability of object detection systems across diverse, and at times, adversarial environments. Homomorphic Encryption integrated with Federated Learning (FL) here can plays an active role due to its complex encryption architecture to overcome these challenges. Figure 20 shows the workflow of HE combined with FL to classify objects.

As AI systems are increasingly employed for tasks such as image and video analysis, it becomes imperative to provide

users with insights into the decisions made by these models. The implementation of XAI techniques in object detection not only fosters user trust but also helps ensure the responsible and ethical use of AI in various applications. However, adapting these techniques to object detection scenarios comes with its own set of challenges. The dynamic nature of visual data, the need to balance interpretability with performance, and the requirement for user-friendly explanations all pose significant hurdles. It is essential to conduct research focused on developing XAI methods tailored specifically to the intricacies of object detection data and applications. Real-time and interactive explanations that seamlessly integrate with visual experiences need to be explored. Additionally, efforts should be directed toward creating decentralized and privacy-preserving XAI solutions, upholding user trust and comprehension in the ever-evolving landscape of AI-driven object detection. In the generation module, captions are produced from input images through the utilization of an encoder-decoder architecture. Simultaneously, the explanation module generates a weight matrix that corresponds to specific regions in the input image and words in the generated caption. The model is designed to produce two distinct loss values, denoted as $Loss_g$ and $Loss_e$. The interplay between the generation and explanation modules, governed by these loss values, facilitates the model's ability to effectively incorporate region information in the generation of captions.

Mainly, The future trajectory of object detection research unfolds across a spectrum of interconnected domains, encompassing advancements in few-shot learning, cross-domain adaptation, real-time efficiency, 3D detection, semantic integration, attention mechanisms, and robustness against adversarial challenges. As researchers continue to delve into these multifaceted dimensions, the collective goal is to push the boundaries of object detection capabilities, rendering them more versatile, accurate, and resilient across an ever-expanding array of applications and environmental conditions. The iterative process of innovation, guided by these diverse research directions, promises to propel object detection into new frontiers of capability and applicability.

VII. CONCLUSION

In conclusion, the odyssey through the landscape of object detection unveils a tapestry woven with technological marvels, where algorithms emerge as the architects of visual understanding. The applications explored, ranging from autonomous vehicles to augmented reality, underscore the transformative impact of object detection on diverse domains. The narrative of advancements, from traditional methods to cutting-edge neural network architectures, mirrors a continual striving for precision and efficiency. As we peer into the future, the persistent open issues serve as compass points guiding researchers toward uncharted territories, beckoning them to unravel the remaining intricacies of real-time detection, robustness in dynamic environments, and the interpretability of increasingly complex models. This exploration is a testament to the dynamic synergy between innovation and challenges, urging the research community to persist in their pursuit of refining object detection algorithms for a future where the unseen is laid bare and the visual world becomes an open book for intelligent systems.

REFERENCES

[1] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Feb. 2020.
[2] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.
[3] 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.
[4] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jul. 2020, pp. 237–242.
[5] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, and R. Qu, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019.
[6] J. M. Rodríguez Fernández, "Computer vision for pedestrian detection using histograms of oriented gradients," Polytech. Univ. Catalonia, Barcelona, Spain, Tech. Rep., 2014.
[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–12.
[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
[9] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–19.
[11] W. Liu, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf.*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Oct. 2016, pp. 21–37.
[12] 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.
[13] 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. 2999–3007.
[14] 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. 10778–10787.
[15] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 213–229.
[16] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 12116–12128.
[17] L. Aziz, M. S. B. H. Salam, U. U. Sheikh, and S. Ayub, "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review," *IEEE Access*, vol. 8, pp. 170461–170495, 2020.
[18] W. Li, Y. Gao, A. Li, X. Zhang, J. Gu, and J. Liu, "Sparse subgraph prediction based on adaptive attention," *Appl. Sci.*, vol. 13, no. 14, p. 8166, Jul. 2023.
[19] Á. Casado-García and J. Heras, "Ensemble methods for object detection," in *Proc. ECAI*, 2020, pp. 2688–2695.
[20] I. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6706–6716.
[21] J. Yang, "Focal attention for long-range interactions in vision transformers," in *Proc. NeurIPS*, 2021, pp. 30008–30022.
[22] Y. Chen, X. Yuan, R. Wu, J. Wang, Q. Hou, and M.-M. Cheng, "YOLO-MS: Rethinking multi-scale representation learning for real-time object detection," 2023, *arXiv:2308.05480*.
[23] M. Iman, H. R. Arabnia, and K. Rasheed, "A review of deep transfer learning and recent advancements," *Technologies*, vol. 11, no. 2, p. 40, Mar. 2023.

[24] T. Lin, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.
[25] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
[26] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[27] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.
[28] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. Toronto,Toronto, ON, Canada, Tech. Rep., 2009.
[29] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, Jan. 2015.
[30] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 734–750.
[31] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.
[32] 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. CVPR*, vol. 1, Dec. 2001, p. 1.
[33] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
[34] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[36] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[37] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
[38] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.
[39] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.
[40] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[41] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2241–2248.
[42] T. Malisiewicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-SVMs for object detection and beyond," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 89–96.
[43] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Mar. 2009.
[44] R. Girshick, P. Felzenszwalb, and D. McAllester, "Object detection with grammar models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 24, 2011, pp. 1–6.
[45] R. B. Girshick, *From Rigid Templates to Grammars: Object Detection With Structured Models*. Chicago, IL, USA: Univ. Chicago, 2012.
[46] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[47] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[48] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
[49] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.
[50] X. Zhou, J. Zhuo, and P. Krähenbühl, "Bottom-up object detection by grouping extreme and center points," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 850–859.
[51] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," 2020, *arXiv:2010.04159*.
[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[53] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland. Cham, Switzerland: Springer, 2014, pp. 818–833.
[54] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "Light-head R-CNN: In defense of two-stage object detector," 2017, *arXiv:1711.07264*.
[55] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–7.
[56] O. Moses. *Imageai Documentation*. Accessed: Mar. 12, 2024. [Online]. Available: https://imageai.readthedocs.io/en/latest/
[57] *Gluoncv Documentation*. Accessed: Mar. 12, 2024. [Online]. Available: https://cv.gluon.ai/
[58] WizYoung. *YOLOv3 TensorFlow Implementation*. Accessed: Mar. 12, 2024. [Online]. Available: https://github.com/wizyoung/YOLOv3_TensorFlow
[59] *Detectron2: A PyTorch-based Modular Object Detection Library—Ai.meta.com*. Accessed: Mar. 12, 2024. [Online]. Available: https://ai.meta.com/blog/-detectron2-a-pytorch-based-modular-object-detection-library-/
[60] *GitHub—Thtrieu/Darkflow: Translate Darknet to Tensorflow. Load Trained Weights, Retrain/Fine-Tune Using Tensorflow, Export Constant Graph Def to Mobile Devices— Github.com*. Accessed: Mar. 12, 2024. [Online]. Available: https://github.com/thtrieu/darkflow
[61] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5525–5533.
[62] D. Qi, W. Tan, Q. Yao, and J. Liu, "YOLO5Face: Why reinventing a face detector," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland. Springer, 2022, pp. 228–244.
[63] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 650–657.
[64] Y. Zhu, H. Cai, S. Zhang, C. Wang, and Y. Xiong, "TinaFace: Strong but simple baseline for face detection," 2020, *arXiv:2011.13183*.
[65] D. Mamieva, A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo, "Improved face detection method via learning small faces on hard images based on a deep learning approach," *Sensors*, vol. 23, no. 1, p. 502, Jan. 2023.
[66] A. Boyd, P. Tinsley, K. Bowyer, and A. Czajka, "CYBORG: Blending human saliency into the loss improves deep learning-based synthetic face detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 6097–6106.
[67] S. Hangaragi, T. Singh, and N. N, "Face detection and recognition using face mesh and deep neural network," *Proc. Comput. Sci.*, vol. 218, pp. 741–749, Jan. 2023.
[68] J. E. Prunty, R. Jenkins, R. Qarooni, and M. Bindemann, "Ingroup and outgroup differences in face detection," *Brit. J. Psychol.*, vol. 114, no. S1, pp. 94–111, May 2023.
[69] S. Sandhya, A. Balasundaram, and A. Shaik, "Deep learning based face detection and identification of criminal suspects," *Comput., Mater. Continua*, vol. 74, no. 2, pp. 2331–2343, 2023.
[70] M. W. Al-Neama, A. A. M. Alshiha, and M. G. Saeed, "A parallel algorithm of multiple face detection on multi-core system," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 29, no. 2, p. 1166, Feb. 2023.
[71] Y. Dong, X. Cui, L. Zhang, and H. Ai, "An improved progressive TIN densification filtering method considering the density and standard variance of point clouds," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 10, p. 409, Oct. 2018.
[72] F. Lei, F. Tang, and S. Li, "Underwater target detection algorithm based on improved YOLOv5," *J. Mar. Sci. Eng.*, vol. 10, no. 3, p. 310, Feb. 2022.
[73] J. Yun, D. Jiang, Y. Liu, Y. Sun, B. Tao, J. Kong, J. Tian, X. Tong, M. Xu, and Z. Fang, "Real-time target detection method based on lightweight convolutional neural network," *Frontiers Bioeng. Biotechnol.*, vol. 10, Aug. 2022, Art. no. 861286.
[74] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4204–4214, Jul. 2016.

[75] G. Wen, S. Li, F. Liu, X. Luo, M.-J. Er, M. Mahmud, and T. Wu, "YOLOv5s-CA: A modified YOLOv5s network with coordinate attention for underwater target detection," *Sensors*, vol. 23, no. 7, p. 3367, Mar. 2023.

[76] W. Liu, J. Liu, T. Liu, H. Chen, and Y.-L. Wang, "Detector design and performance analysis for target detection in subspace interference," *IEEE Signal Process. Lett.*, vol. 30, pp. 618–622, 2023.

[77] M. Yasir, L. Shanwei, X. Mingming, S. Hui, M. S. Hossain, A. T. I. Colak, D. Wang, W. Jianhua, and K. B. Dang, "Multi-scale ship target detection using SAR images based on improved Yolov5," *Frontiers Mar. Sci.*, vol. 9, Jan. 2023, Art. no. 1086140.

[78] R. Yang, W. Li, X. Shang, D. Zhu, and X. Man, "KPE-YOLOv5: An improved small target detection algorithm based on YOLOv5," *Electronics*, vol. 12, no. 4, p. 817, Feb. 2023.

[79] H. Liang and T. Song, "Lightweight marine biological target detection algorithm based on YOLOv5," *Frontiers Mar. Sci.*, vol. 10, Jul. 2023, Art. no. 1219155.

[80] Y. Wang, W. Feng, K. Jiang, Q. Li, R. Lv, and J. Tu, "Real-time damaged building region detection based on improved YOLOv5s and embedded system from UAV images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4205–4217, 2023.

[81] V. Ellore, L. N. Anand, and S. Tripathi, "Image deblurring and localization for disaster response and humanitarian assistance," *Proc. Comput. Sci.*, vol. 171, pp. 1614–1623, Jan. 2020.

[82] J. A. Quinn, M. M. Nyhan, C. Navarro, D. Coluccia, L. Bromley, and M. Luengo-Oroz, "Humanitarian applications of machine learning with remote-sensing data: Review and case study in refugee settlement mapping," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 376, no. 2128, Sep. 2018, Art. no. 20170363.

[83] D. Tiede, G. Schwendemann, A. Alobaidi, L. Wendt, and S. Lang, "Mask R-CNN-based building extraction from VHR satellite data in operational humanitarian action: An example related to COVID-19 response in Khartoum, Sudan," *Trans. GIS*, vol. 25, no. 3, pp. 1213–1227, Jun. 2021.

[84] H. S. Munawar, J. Zhang, H. Li, D. Mo, and L. Chang, "Mining multi-spectral aerial images for automatic detection of strategic bridge locations for disaster relief missions," in *Trends and Applications in Knowledge Discovery and Data Mining*, Macau, China. Cham, Switzerland: Springer, Apr. 2019, pp. 189–200.

[85] K. Ayush, B. Uzkent, M. Burke, D. Lobell, and S. Ermon, "Generating interpretable poverty maps using object detection in satellite images," 2020, *arXiv:2002.01612*.

[86] R. I. Jony, "Multimodal data fusion of remote sensing and social media using machine learning for natural disaster detection and assessment," Ph.D. thesis, School Comput. Sci., Queensland Univ. Technol., Brisbane City QLD, Australia, 2023.

[87] H. Cui, S. Qiu, Y. Wang, Y. Zhang, Z. Liu, K. Karila, J. Jia, and Y. Chen, "Disaster-caused power outage detection at night using VIIRS DNB images," *Remote Sens.*, vol. 15, no. 3, p. 640, Jan. 2023.

[88] D. Absalon, M. Matysik, A. Woźnica, and N. Janczewska, "Detection of changes in the hydrobiological parameters of the oder river during the ecological disaster in July 2022 based on multi-parameter probe tests and remote sensing methods," *Ecolog. Indicators*, vol. 148, Apr. 2023, Art. no. 110103.

[89] R. F. Mansour and E. Alabdulkreem, "Disaster monitoring of satellite image processing using progressive image classification," *Comput. Syst. Sci. Eng.*, vol. 44, no. 2, pp. 1161–1169, 2023.

[90] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9357–9366.

[91] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3111–3122, Nov. 2018.

[92] W. He, X.-Y. Zhang, F. Yin, and C.-L. Liu, "Deep direct regression for multi-oriented scene text detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 745–753.

[93] Y. Jiang, X. Zhu, X. Wang, S. Yang, W. Li, H. Wang, P. Fu, and Z. Luo, "R2CNN: Rotational region CNN for orientation robust scene text detection," 2017, *arXiv:1706.09579*.

[94] E. Mitchell, Y. Lee, A. Khazatsky, C. D. Manning, and C. Finn, "DetectGPT: Zero-shot machine-generated text detection using probability curvature," 2023, *arXiv:2301.11305*.

[95] M. Ye, J. Zhang, S. Zhao, J. Liu, B. Du, and D. Tao, "DPText-DETR: Towards better scene text detection with dynamic points in transformer," in *Proc. Conf. Artif. Intell. (AAAI)*, 2023, pp. 3241–3249.

[96] X. He, X. Shen, Z. Chen, M. Backes, and Y. Zhang, "MGTBench: Benchmarking machine-generated text detection," 2023, *arXiv:2303.14822*.

[97] S.-X. Zhang, C. Yang, X. Zhu, and X.-C. Yin, "Arbitrary shape text detection via boundary transformer," *IEEE Trans. Multimedia*, vol. 26, pp. 1747–1760, 2023.

[98] T. Kumarage, J. Garland, A. Bhattacharjee, K. Trapeznikov, S. Ruston, and H. Liu, "Stylometric detection of AI-generated text in Twitter timelines," 2023, *arXiv:2303.03697*.

[99] M. Ye, J. Zhang, S. Zhao, J. Liu, T. Liu, B. Du, and D. Tao, "DeepSolo: Let transformer decoder with explicit points solo for text spotting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 19348–19357.

[100] K. L. Masita, A. N. Hasan, and S. Paul, "Pedestrian detection using R-CNN object detector," in *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, Nov. 2018, pp. 1–6.

[101] C. Papageorgiou and T. Poggio, "Trainable pedestrian detection," in *Proc. Int. Conf. Image Process.*, Oct. 1999, pp. 35–39.

[102] G. Ma, S.-B. Park, A. Ioffe, S. Müller-Schneiders, and A. Kummert, "A real time object detection approach applied to reliable pedestrian detection," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2007, pp. 755–760.

[103] J. Schlosser, C. K. Chow, and Z. Kira, "Fusing LiDAR and images for pedestrian detection using convolutional neural networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 2198–2205.

[104] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1532–1545, Aug. 2014.

[105] D. Park, D. Ramanan, and C. Fowlkes, "Multiresolution models for object detection," in *Proc. Eur. Conf. Comput. Vis.*, Heraklion, Greece. Cham, Switzerland: Springer, 2010, pp. 241–254.

[106] S.-C. Ng and C.-P. Kwok, "An intelligent traffic light system using object detection and evolutionary algorithm for alleviating traffic congestion in Hong Kong," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, p. 802, 2020.

[107] N. Jain, S. Yerragolla, T. Guha, and Mohana, "Performance analysis of object detection and tracking algorithms for traffic surveillance applications using neural networks," in *Proc. 3rd Int. Conf. I-SMAC (IoT Social, Mobile, Anal. Cloud) (I-SMAC)*, Dec. 2019, pp. 690–696.

[108] C. H. Low, M. K. Lee, and S. W. Khor, "Frame based object detection–An application for traffic monitoring," in *Proc. 2nd Int. Conf. Comput. Res. Develop.*, May 2010, pp. 322–325.

[109] M. J. Akhtar, R. Mahum, F. S. Butt, R. Amin, A. M. El-Sherbeeny, S. M. Lee, and S. Shaikh, "A robust framework for object detection in a traffic surveillance system," *Electronics*, vol. 11, no. 21, p. 3425, Oct. 2022.

[110] S.-Y. Wang, Z. Qu, C.-J. Li, and L.-Y. Gao, "BANet: Small and multi-object detection with a bidirectional attention network for traffic scenes," *Eng. Appl. Artif. Intell.*, vol. 117, Jan. 2023, Art. no. 105504.

[111] T. Thamaraimanalan, D. Naveena, M. Ramya, and M. Madhubala, "Prediction and classification of fouls in soccer game using deep learning," *Ir. Interdiscip. J. Sci. Res*, vol. 4, pp. 66–78, Jan. 2020.

[112] K. Ryu, H. Kim, and S. Lee, "A deep learning model based on sequential object feature accumulation for sport activity recognition," *Multimedia Tools Appl.*, vol. 82, no. 24, pp. 37387–37406, Oct. 2023.

[113] W. Ma and Y. Lv, "Feature extraction method of football fouls based on deep learning algorithm," *Int. J. Inf. Commun. Technol.*, vol. 1, no. 1, pp. 404–421, 2021.

[114] C. Liu and L. Cao, "Automatic detection of sports injuries based on multimedia intelligent 3D images," *Adv. Multimedia*, vol. 2023, pp. 1–9, Mar. 2023.

[115] M. Borghesi, L. D. Costa, L. Morra, and F. Lamberti, "Using temporal convolutional networks to estimate ball possession in soccer games," *Expert Syst. Appl.*, vol. 223, Aug. 2023, Art. no. 119780.

[116] M. Bouazizi, A. Lorite Mora, and T. Ohtsuki, "A 2D-LiDAR-equipped unmanned robot-based approach for indoor human activity detection," *Sensors*, vol. 23, no. 5, p. 2534, Feb. 2023.

[117] S. U. Khan, I. U. Haq, S. Rho, S. W. Baik, and M. Y. Lee, "Cover the violence: A novel deep-learning-based approach towards violence-detection in movies," *Appl. Sci.*, vol. 9, no. 22, p. 4963, Nov. 2019.

[118] M. Bianculli, N. Falcionelli, P. Sernani, S. Tomassini, P. Contardo, M. Lombardi, and A. F. Dragoni, "A dataset for automatic violence detection in videos," *Data Brief*, vol. 33, Dec. 2020, Art. no. 106587.

[119] R. Vijeikis, V. Raudonis, and G. Dervinis, "Efficient violence detection in surveillance," *Sensors*, vol. 22, no. 6, p. 2216, Mar. 2022.
[120] F. U. M. Ullah, A. Ullah, K. Muhammad, I. U. Haq, and S. W. Baik, "Violence detection using spatiotemporal features with 3D convolutional neural network," *Sensors*, vol. 19, no. 11, p. 2472, May 2019.
[121] P. Wu, J. Liu, Y. Shi, Y. Sun, F. Shao, Z. Wu, and Z. Yang, "Not only look, but also listen: Learning multimodal violence detection under weak supervision," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2020, pp. 322–339.
[122] D. I. Lee, J. H. Lee, S. H. Jang, S. J. Oh, and I. C. Doo, "Crop disease diagnosis with deep learning-based image captioning and object detection," *Appl. Sci.*, vol. 13, no. 5, p. 3148, Feb. 2023.
[123] Y. Peng and Y. Wang, "Leaf disease image retrieval with object detection and deep metric learning," *Frontiers Plant Sci.*, vol. 13, Sep. 2022, Art. no. 963302.
[124] T. Poornappriya and R. Gopinath, "Rice plant disease identification using artificial intelligence approaches," *Int. J. Electr. Eng. Technol.*, vol. 11, no. 10, pp. 392–402, 2022.
[125] Z. Li, M. Dong, S. Wen, X. Hu, P. Zhou, and Z. Zeng, "CLU-CNNs: Object detection for medical images," *Neurocomputing*, vol. 350, pp. 53–59, Jul. 2019.
[126] X. He, S. Wang, S. Shi, Z. Tang, Y. Wang, Z. Zhao, J. Dai, R. Ni, X. Zhang, X. Liu, Z. Wu, W. Yu, and X. Chu, "Computer-aided clinical skin disease diagnosis using CNN and object detection models," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 4839–4844.
[127] L. Liu, M. Muelly, J. Deng, T. Pfister, and L.-J. Li, "Generative modeling for small-data object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6072–6080.
[128] R. Terzi, "An ensemble of deep learning object detection models for anatomical and pathological regions in brain MRI," *Diagnostics*, vol. 13, no. 8, p. 1494, Apr. 2023.
[129] F. H. Awad, M. M. Hamad, and L. Alzubaidi, "Robust classification and detection of big medical data using advanced parallel K-means clustering, YOLOv4, and logistic regression," *Life*, vol. 13, no. 3, p. 691, Mar. 2023.
[130] P. Chotikunnan, T. Puttasakul, R. Chotikunnan, B. Panomruttanarug, M. Sangworasil, and A. Srisiriwat, "Evaluation of single and dual image object detection through image segmentation using ResNet18 in robotic vision applications," *J. Robot. Control (JRC)*, vol. 4, no. 3, pp. 263–277, Apr. 2023.
[131] B. Leng, C. Wang, M. Leng, M. Ge, and W. Dong, "Deep learning detection network for peripheral blood leukocytes based on improved detection transformer," *Biomed. Signal Process. Control*, vol. 82, Apr. 2023, Art. no. 104518.
[132] D.-P. Fan, G.-P. Ji, M.-M. Cheng, and L. Shao, "Concealed object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6024–6042, Oct. 2022.
[133] I. Kök, M. U. Simsek, and S. Özdemir, "A deep learning model for air quality prediction in smart cities," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 1983–1990.
[134] A. Chakma, B. Vizena, T. Cao, J. Lin, and J. Zhang, "Image-based air quality analysis using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3949–3952.
[135] B. Yang, Y. Liu, P. Liu, F. Wang, X. Cheng, and Z. Lv, "A novel occupant-centric stratum ventilation system using computer vision: Occupant detection, thermal comfort, air quality, and energy savings," *Building Environ.*, vol. 237, Jun. 2023, Art. no. 110332.
[136] N. Jayawickrama, E. P. Ollé, J. Pirhonen, R. Ojala, K. Kivekäs, J. Vepsäläinen, and K. Tammi, "Architecture for determining the cleanliness in shared vehicles using an integrated machine vision and indoor air quality-monitoring system," *J. Big Data*, vol. 10, no. 1, pp. 1–31, Feb. 2023.
[137] M. Baker, F. Gollier, J. E. Melzer, and E. McLeod, "Lensfree air-quality monitoring of fine and ultrafine particulate matter using vapor-condensed nanolenses," *ACS Appl. Nano Mater.*, vol. 6, no. 13, pp. 11166–11174, Jul. 2023.
[138] D. N. Paithankar, A. R. Pabale, R. V. Kolhe, P. William, and P. M. Yawalkar, "Framework for implementing air quality monitoring system using LPWA-based IoT technique," *Meas., Sensors*, vol. 26, Apr. 2023, Art. no. 100709.
[139] A. Dhanush, S. Panimalar, K. L. Chowdary, and S. Supreeth, "IoT-based air quality monitoring system," Tech. Rep., 2023, pp. 175–184.
[140] R. Sortino, D. Magro, G. Fiameni, E. Sciacca, S. Riggi, A. DeMarco, C. Spampinato, A. M. Hopkins, F. Bufano, C. Bordiu, and C. Pino, "Radio astronomical images object detection and segmentation: A benchmark on deep learning methods," *Experim. Astron.*, vol. 56, no. 1, pp. 293–331, Aug. 2023.
[141] M. S. Prasad, S. Verma, and Y. A. Shichkina, "Astronomical image processing: Exoplanet detection," in *Proc. 26th Int. Conf. Soft Comput. Meas. (SCM)*, May 2023, pp. 336–340.
[142] D. Zhang, H. Tian, and J. Han, "Few-cost salient object detection with adversarial-paced learning," in *Proc. NIPS*, vol. 33, 2020, pp. 12236–12247.
[143] S. Shao, Z. Li, T. Zhang, C. Peng, G. Yu, X. Zhang, J. Li, and J. Sun, "Objects365: A large-scale, high-quality dataset for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8429–8438.
[144] C. Liu, H. Li, S. Wang, M. Zhu, D. Wang, X. Fan, and Z. Wang, "A dataset and benchmark of underwater object detection for robot picking," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2021, pp. 1–6.
[145] J. Ponce, T. L. Berg, M. Everingham, D. A. Forsyth, M. Hebert, S. Lazebnik, M. Marszalek, C. Schmid, B. C. Russell, and A. Torralba, "Dataset issues in object recognition," in *Toward Category-Level Object Recognition*. Berlin, Germany: Springer, 2006, pp. 29–48.
[146] C. Biffi, S. McDonagh, P. Torr, A. Leonardis, and S. Parisot, "Many-shot from low-shot: Learning to annotate using mixed supervision for object detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Nov. 2020, pp. 35–50.
[147] M. K. Chen, X. Liu, Y. Wu, J. Zhang, J. Yuan, Z. Zhang, and D. P. Tsai, "A meta-device for intelligent depth perception," *Adv. Mater.*, vol. 35, no. 34, Aug. 2023, Art. no. 2107465.
[148] Z. Chen, M. Zhu, S. Chen, L. Lu, H. Tang, X. Hu, and C. Ji, "Discriminative feature fusion for RGB-D salient object detection," *Comput. Electr. Eng.*, vol. 106, Mar. 2023, Art. no. 108579.
[149] X. Wu, D. Ma, X. Qu, X. Jiang, and D. Zeng, "Depth dynamic center difference convolutions for monocular 3D object detection," *Neurocomputing*, vol. 520, pp. 73–81, Feb. 2023.
[150] S. Hou, Z. Wang, and F. Wu, "Object detection via deeply exploiting depth information," *Neurocomputing*, vol. 286, pp. 58–66, Apr. 2018.
[151] W. G. Aguilar, F. J. Quisaguano, G. A. Rodríguez, L. G. Alvarez, A. Limaico, and D. S. Sandoval, "Convolutional neuronal networks based monocular object detection and depth perception for micro UAVs," in *Proc. Int. Conf. Intell. Sci. Big Data Eng.*, Lanzhou, China. Cham, Switzerland: Springer, Aug. 2018, pp. 401–410.
[152] A. Gupta, A. Anpalagan, L. Guan, and A. S. Khwaja, "Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues," *Array*, vol. 10, Jul. 2021, Art. no. 100057.
[153] X. Liu, W. Li, Q. Yang, B. Li, and Y. Yuan, "Towards robust adaptive object detection under noisy annotations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14187–14196.
[154] G. Volk, S. Müller, A. v. Bernuth, D. Hospach, and O. Bringmann, "Towards robust CNN-based object detection through augmentation with synthetic rain variations," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 285–292.
[155] B. Adhikari, J. Peltomäki, S. B. Germi, E. Rahtu, and H. Huttunen, "Effect of label noise on robustness of deep neural network object detectors," in *Proc. Int. Conf. Comput. Saf., Rel., Secur.*, Cham, Switzerland. Springer, 2021, pp. 239–250.
[156] A. Sabater, L. Montesano, and A. C. Murillo, "Robust and efficient post-processing for video object detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 10536–10542.
[157] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential Siamese network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 773–782.
[158] M. Kristan, "The eighth visual object tracking VOT2020 challenge results," in *Proc. Comput. Vis. ECCV Workshops*. Glasgow, U.K. Cham, Switzerland: Springer, Aug. 2020, pp. 547–601.
[159] T. Wang, T. Yang, M. Danelljan, F. S. Khan, X. Zhang, and J. Sun, "Learning human-object interaction detection using interaction points," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4115–4124.
[160] H. Wu, Y. Chen, N. Wang, and Z.-X. Zhang, "Sequence level semantics aggregation for video object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9216–9224.

[161] Y. Liu, T. Wang, X. Zhang, and J. Sun, "PETR: Position embedding transformation for multi-view 3D object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 531–548.
[162] N. Chattopadhyay, "Robust AI: Security and privacy issues in machine learning," Nanyang Technol. Univ., Singapore, Tech. Rep., 2023.
[163] D. Kagan, G. F. Alpert, and M. Fire, "Zooming into video conferencing privacy," *IEEE Trans. Computat. Social Syst.*, vol. 11, no. 1, pp. 933–944, Feb. 2024.
[164] Z. Lin, W. Chen, L. Su, Y. Chen, and T. Li, "HS-YOLO: Small object detection for power operation scenarios," *Appl. Sci.*, vol. 13, no. 19, p. 11114, Oct. 2023.
[165] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi, and A. K. Bhoi, "Statistical analysis of design aspects of various YOLO-based deep learning models for object detection," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, p. 126, Aug. 2023.
[166] A. BinDarwish, S. Alhammadi, and A. SALEHI, "Crime detection and suspect identification system," Tech. Rep., 2023.
[167] X. Sun, B. Wang, Z. Wang, H. Li, H. Li, and K. Fu, "Research progress on few-shot learning for remote sensing image interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2387–2402, 2021.
[168] G. Quellec, M. Lamard, P.-H. Conze, P. Massin, and B. Cochener, "Automatic detection of rare pathologies in fundus photographs using few-shot learning," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101660.
[169] Z.-M. Wang, J.-Y. Tian, J. Qin, H. Fang, and L.-M. Chen, "A few-shot learning-based Siamese capsule network for intrusion detection with imbalanced training data," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–17, Sep. 2021.
[170] T. Zhang, X. Zhang, P. Zhu, X. Jia, X. Tang, and L. Jiao, "Generalized few-shot object detection in remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 195, pp. 353–364, Jan. 2023.
[171] S. Yang, B. Linares-Barranco, and B. Chen, "Heterogeneous ensemble-based spike-driven few-shot online learning," *Frontiers Neurosci.*, vol. 16, May 2022, Art. no. 850932.
[172] S. X. Hu, D. Li, J. Stühmer, M. Kim, and T. M. Hospedales, "Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9058–9067.
[173] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell, "Few-shot object detection via feature reweighting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8419–8428.
[174] Y. Koizumi, S. Murata, N. Harada, S. Saito, and H. Uematsu, "SNIPER: Few-shot learning for anomaly detection to minimize false-negative rate with ensured true-positive rate," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 915–919.
[175] M. Gamal, H. M. Abbas, N. Moustafa, E. Sitnikova, and R. A. Sadek, "Few-shot learning for discovering anomalous behaviors in edge networks," *Comput., Mater. Continua*, vol. 69, no. 2, pp. 1823–1837, 2021.
[176] A. Heidari, J. McGrath, I. F. Ilyas, and T. Rekatsinas, "HoloDetect: Few-shot learning for error detection," in *Proc. Int. Conf. Manage. Data*, 2019, pp. 829–846.
[177] Y. Yu and N. Bian, "An intrusion detection method using few-shot learning," *IEEE Access*, vol. 8, pp. 49730–49740, 2020.
[178] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, and G. Srivastava, "Federated-learning-based anomaly detection for IoT security attacks," *IEEE Internet Things J.*, vol. 9, no. 4, pp. 2545–2554, Feb. 2022.
[179] W. Yang, Y. Zhang, K. Ye, L. Li, and C.-Z. Xu, "FFD: A federated learning based method for credit card fraud detection," in *Proc. Int. Conf. Big Data*. Cham, Switzerland: Springer, 2019, pp. 18–32.
[180] P. Tian, Z. Chen, W. Yu, and W. Liao, "Towards asynchronous federated learning based threat detection: A DC-adam approach," *Comput. Secur.*, vol. 108, Sep. 2021, Art. no. 102344.
[181] D. Preuveneers, V. Rimmer, I. Tsingenopoulos, J. Spooren, W. Joosen, and E. Ilie-Zudor, "Chained anomaly detection models for federated learning: An intrusion detection case study," *Appl. Sci.*, vol. 8, no. 12, p. 2663, Dec. 2018.
[182] Y. Liu, "FedVision: An online visual object detection platform powered by federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 13172–13179.
[183] T. T. Huong, T. P. Bac, K. N. Ha, N. V. Hoang, N. X. Hoang, N. T. Hung, and K. P. Tran, "Federated learning-based explainable anomaly detection for industrial control systems," *IEEE Access*, vol. 10, pp. 53854–53872, 2022.
[184] T. T. Huong, T. P. Bac, D. M. Long, T. D. Luong, N. M. Dan, L. A. Quang, L. T. Cong, B. D. Thang, and K. P. Tran, "Detecting cyberattacks using anomaly detection in industrial control systems: A federated learning approach," *Comput. Ind.*, vol. 132, Nov. 2021, Art. no. 103509.
[185] H. Liu, S. Zhang, P. Zhang, X. Zhou, X. Shao, G. Pu, and Y. Zhang, "Blockchain and federated learning for collaborative intrusion detection in vehicular edge computing," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6073–6084, Jun. 2021.
[186] A. N. Jahromi, H. Karimipour, and A. Dehghantanha, "Deep federated learning-based cyber-attack detection in industrial control systems," in *Proc. 18th Int. Conf. Privacy, Secur. Trust (PST)*, Dec. 2021, pp. 1–6.
[187] T. Li, J. Miao, X. Fu, B. Song, B. Cai, X. Ge, X. Zhou, P. Zhou, X. Wang, D. Jariwala, and W. Hu, "Reconfigurable, non-volatile neuromorphic photovoltaics," *Nature Nanotechnol.*, vol. 18, no. 11, pp. 1303–1310, Nov. 2023.
[188] R. Zhao, X. Niu, Y. Wu, W. Luk, and Q. Liu, "Optimizing CNN-based object detection algorithms on embedded FPGA platforms," in *Proc. Int. Symp. Appl. Reconfigurable Comput.*, Delft, The Netherlands. Cham, Switzerland: Springer, Apr. 2017, pp. 255–267.
[189] C. Wing-Hei Chan, P. H. W. Leong, and H. Kwok-Hay So, "Vision guided crop detection in field robots using FPGA-based reconfigurable computers," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Oct. 2020, pp. 1–5.
[190] V. H. Kim and K. K. Choi, "A reconfigurable CNN-based accelerator design for fast and energy-efficient object detection system on mobile FPGA," *IEEE Access*, vol. 11, pp. 59438–59445, 2023.
[191] J. Na and V. De-Silva, "Uncertainty aware active learning for reconfiguration of pre-trained deep object-detection networks for new target domains," in *Proc. IEEE IAS Global Conf. Emerg. Technol. (GlobConET)*, May 2023, pp. 1–7.
[192] M. Baczmanski, M. Wasala, and T. Kryjak, "Implementation of a perception system for autonomous vehicles using a detection-segmentation network in SoC FPGA," in *Proc. Int. Symp. Appl. Reconfigurable Comput.* Cham, Switzerland: Springer, 2023, pp. 200–211.
[193] J.-N. Zaech, A. Liniger, M. Danelljan, D. Dai, and L. Van Gool, "Adiabatic quantum computing for multi object tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8801–8812.
[194] V. Rajesh, U. P. Naik, and Mohana, "Quantum convolutional neural networks (QCNN) using deep learning for computer vision applications," in *Proc. Int. Conf. Recent Trends Electron., Inf., Commun. Technol. (RTEICT)*, Aug. 2021, pp. 728–734.
[195] G. N. Meedinti, K. Sai Srirekha, and R. Delhibabu, "A quantum convolutional neural network approach for object detection and classification," 2023, *arXiv:2307.08204*.
[196] L. Hu and Q. Ni, "Quantum automated object detection algorithm," in *Proc. 25th Int. Conf. Autom. Comput. (ICAC)*, Sep. 2019, pp. 1–4.
[197] W. She, Q. Liu, Z. Tian, J.-S. Chen, B. Wang, and W. Liu, "Blockchain trust model for malicious node detection in wireless sensor networks," *IEEE Access*, vol. 7, pp. 38947–38956, 2019.
[198] D. G. Roy and S. N. Srirama, "A blockchain-based cyber attack detection scheme for decentralized Internet of Things using software-defined network," *Softw., Pract. Exper.*, vol. 51, no. 7, pp. 1540–1556, Jul. 2021.
[199] M. Signorini, M. Pontecorvi, W. Kanoun, and R. Di Pietro, "BAD: A blockchain anomaly detection solution," *IEEE Access*, vol. 8, pp. 173481–173490, 2020.
[200] Y. Mirsky, T. Golomb, and Y. Elovici, "Lightweight collaborative anomaly detection for the IoT using blockchain," *J. Parallel Distrib. Comput.*, vol. 145, pp. 75–97, Nov. 2020.
[201] T. Ashfaq, R. Khalid, A. S. Yahaya, S. Aslam, A. T. Azar, S. Alsafari, and I. A. Hameed, "A machine learning and blockchain based efficient fraud detection mechanism," *Sensors*, vol. 22, no. 19, p. 7162, Sep. 2022.
[202] Z. Yu, D. Xue, J. Fan, and C. Guo, "DNSTSM: DNS cache resources trusted sharing model based on consortium blockchain," *IEEE Access*, vol. 8, pp. 13640–13650, 2020.
[203] S.-H. Han, J.-H. Kim, W.-S. Song, and G.-Y. Gim, "An empirical analysis on medical information sharing model based on blockchain," *Int. J. Adv. Comput. Res.*, vol. 9, no. 40, pp. 20–27, Jan. 2019.
[204] Z. Wang, S. Zhang, Y. Zhao, C. Chen, and X. Dong, "Risk prediction and credibility detection of network public opinion using blockchain technology," *Technol. Forecasting Social Change*, vol. 187, Feb. 2023, Art. no. 122177.

[205] A. Yazdinejad, A. Dehghantanha, R. M. Parizi, G. Srivastava, and H. Karimipour, "Secure intelligent fuzzy blockchain framework: Effective threat detection in IoT networks," *Comput. Ind.*, vol. 144, Jan. 2023, Art. no. 103801.

MD. TANZIB HOSAIN is currently pursuing the Bachelor of Science (B.Sc.) degree in computer science and engineering (CSE) with American International University-Bangladesh. He is also a Competitive Programmer and participated many programming competitions specially the ACM International Collegiate Programming Contest (ACM-ICPC). He is familiar with various programming languages, including, C/C++, Python, TensorFlow, PyTorch, MATLAB, and Java. His research interests include cognitive science, artificial intelligence, data science, natural language processing, and graph theory.

ASIF ZAMAN is currently pursuing the Bachelor of Science degree in computer science and engineering (CSE) with American International University-Bangladesh (AIUB). He is also a Certified Node JS Developer and a Project Coordinator with EBS Solution. He is familiar with various programming languages, including C/C++, Java, JavaScript, Python, and Node JS. His research interests include artificial intelligence (AI), deep learning (DL), and data science (DS).

MUSHFIQUR RAHMAN ABIR is currently pursuing the degree in computer science with American International University-Bangladesh (AIUB). He passionate about learning new technologies and coding, he is proficient in C++, Python, and TypeScript. He enjoys creating software and websites. He is an efficient Linux User, contributing to the open source community and building tools for linux like ffmpeg-coder, and AppNotEx. He has contributed to projects like StudyFolderOrganizer, ALib, and AIUB-Discobot. He is also a member and a photographer with AIUB Photography Club. His diverse programming skills include C/C++, .NET Framework, JavaScript, and Python. His research interests include artificial intelligence, natural language processing, and data science.

SHANJIDA AKTER is currently pursuing the Bachelor of Science (B.Sc.) degree in computer science and engineering (CSE) with North South University (NSU). She is also a Mathematics Enthusiast and a Teaching Assistant (TA) with the Department of Mathematics and Physics, NSU. She has developed her skills by learning different programming languages as C/C++, Java, JavaScript, and Python. Her research interests include data science, deep learning, natural language processing, and artificial intelligence.

SAWON MURSALIN is currently pursuing the Bachelor of Science degree in computer science and engineering (CSE) with American International University-Bangladesh (AIUB). He is also a Designer and Fontaine Developer. So, he is also familiar with various programming languages and designing platforms, including C/C++, Java, Python, Abobe Illustrator, and Adobe XD. His research interests include artificial general intelligence (AGI), software engineering, networking, and data science (DS).

SHADMAN SAKEEB KHAN is currently pursuing the Bachelor of Science (B.Sc.) degree in computer science and engineering (CSE) with North South University (NSU). He is also a passionate and skilled individual with expertise in machine learning, deep learning, natural language processing, and coding languages, such as C/C++, Java, and Python. He is also a Research Assistant (RA) with NSU. He excels in bridging Front-end development with machine learning through React JS. His research interests include artificial intelligence and robotics.