Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Review article

Object tracking and detection techniques under GANN threats: A systemic review

Saeed Matar Al Jaberi [a,*], Asma Patel [a], Ahmed N. AL-Masri [b]

[a] School of Digital Technologies and Arts, Staffordshire University, UK
[b] Ministry of Energy and Infrastructure, Dubai, United Arab Emirates

ARTICLE INFO

Article history:
Received 15 June 2022
Received in revised form 22 December 2022
Accepted 12 March 2023
Available online 23 March 2023

Keywords:
Object detection
Tracking techniques
GANN threats in object detection
Adversarial attack and defence

ABSTRACT

Current developments in object tracking and detection techniques have directed remarkable improvements in distinguishing attacks and adversaries. Nevertheless, adversarial attacks, intrusions, and manipulation of images/ videos threaten video surveillance systems and other object-tracking applications. Generative adversarial neural networks (GANNs) are widely used image processing and object detection techniques because of their flexibility in processing large datasets in real-time. GANN training ensures a tamper-proof system, but the plausibility of attacks persists. Therefore, reviewing object tracking and detection techniques under GANN threats is necessary to reveal the challenges and benefits of efficient defence methods against these attacks. This paper aims to systematically review object tracking and detection techniques under threats to GANN-based applications. The selected studies were based on different factors, such as the year of publication, the method implemented in the article, the reliability of the chosen algorithms, and dataset size. Each study is summarised by assigning it to one of the two predefined tasks: applying a GANN or using traditional machine learning (ML) techniques. First, the paper discusses traditional applied techniques in this field. Second, it addresses the challenges and benefits of object detection and tracking. Finally, different existing GANN architectures are covered to justify the need for tamper-proof object tracking systems that can process efficiently in a real-time environment.

Contents

[*] Corresponding author.
E-mail addresses: a030340i@student.staffs.ac.uk (S.M. Al Jaberi), asma.patel@staffs.ac.uk (A. Patel), ahmedalmasri@ieee.org (A.N. AL-Masri).

https://doi.org/10.1016/j.asoc.2023.110224

1. Introduction

Object tracking and detection techniques remained a dynamic area of research for a long time, but it is exceptionally developed in recent years. The increased research on object tracking and detection is driven by its diverse applications, such as video surveillance, human-machine interactions, traffic surveillance, and malicious object or human behaviour detection. Malicious cyber, criminal, and adversarial attacks on artificial intelligence (AI) based applications are increasing daily, threatening the security and safety of nations worldwide [1]. For instance, cities such as Abu Dhabi and Dubai in the UAE faced around 86 cyber-attacks at the beginning of 2018, wherein the famous cab service Careem was also among the victims [2], these attacks increased by 23% in the same year. The consequences of these attacks are severe on AI-based surveillance systems because more than 14 million customers' data leaked and were exposed on the Internet. These malicious and adversarial attacks not only interrupt systems, business operations and services to citizens but also impose danger to the economy and security on the national level.

Current advances demand tamper-proof object tracking, and detection techniques have become common adversarial attacks, intrusions, and hacks. Furthermore, real-time object tracking has become necessary to rapidly process images when detecting malicious objects during surveillance [3]. The security prerequisites are increasingly demanding nowadays, and supervision relying on human actors is insufficient. Thus, efficient object tracking and detection techniques for security systems rely on independence or automation of security measures, such as using ML algorithms that enable feature learning and image/object generation but are also critical to ensure the security and robustness of the deployed algorithms. To guide the research in this field, several challenges of video surveillance systems [4,5], including occlusion, viewpoint variations, and the problem of illumination are discussed in the latter part of this paper. The primary emphasis of this study is to examine the vulnerabilities of object tracking and detection techniques towards generative adversarial neural network (GANN) threats. It then discusses various types of GANN threats and examines distinct types of adversarial attacks based on the threat model, such as black-box, white-box, grey box attacks, and other adversarial attacks. Finally, this paper provides a systematic review of object tracking and detection techniques under GANN threats, including implications and concluding remarks for future work.

1.1. Purpose and contributions

The purpose of this study is two fold: to provide the researchers with an overview of the traditional machine learning algorithms applied in object tracking and detection; and to emphasise the need to explore different types of GANNs to add the feature of demand for these proposed tamper-proof technologies. Below are the major contributions of this study:

(i) A systematic literature review of different proposed object detection and tracking algorithms
(ii) A summary of 48 studies grouped into GANN and non-GANN related tasks, mainly adversarial attacks and defences.
(iii) A discussion on the challenges, gaps, and future research directions

1.2. Research methodology

For the systematic literature review, this study searched for the techniques through well-known journals and conferences related to the field for over eight years (2014–2021). The keyword queries 'object detection', 'object tracking', 'GANN', 'threats' and 'adversarial attacks' have been used in the google scholar academic research database for search filtration. The search were limited to the first 98 pages, or 980 out of the 188 K total results, because the database does not seem to operate after that amount of data. We filtered the suitable papers by reading the abstracts and excluding those unrelated to the scope of this research. Furthermore, through backward snowballing related works, we included a few papers that were not retrieved by the database. The final number of studies presented in this paper is 48.

1.3. Scope

The paper is organised as mentioned: the starting section contains the study's introduction and an overview of object tracking and detection. Then, the study provides background about object detection techniques and object tracking methods that produce state-of-the-art results. Delivering an evaluation and a systematic review of various types of GANN threats and adversarial attacks against object detection and tracking performance is included in this study. Challenges and recommendations concerning object detection and tracking techniques under GANN attacks are also discussed. In the end, the conclusions and implications for future research are discussed.

2. Overview of object tracking and detection

Object tracking and detection techniques are extensively used in various applications, such as video surveillance, traffic monitoring, security cameras, and vehicle recognition system. For object detection and tracking, a video consists of diverse information, such as the detected object's shape, size, colour, texture, and, sometimes, the object's motion also aids in its detection and tracking. Therefore, many authors have proposed incorporating features' statistical analysis and motion information of the object. Hence, nowadays, surveillance systems employ high-resolution cameras and sensors in different applications, such as security surveillance and vehicle detection systems.

2.1. Techniques and process for object detection and tracking

Typical object tracking and detection include three stages: detection, classification, and tracking. The first stage is object detection to confirm and locate the presence of the items in an image/video. Afterwards, the detected object, such as birds, vehicles, human beings, and other objects, is classified. Object tracking, on the contrary, refers to the detection of an object in the occurrence of occlusions, spatial items, and other changes. The object's colour, shape, texture, and location are essential during object tracking [6]. The straightforward object tracking process includes four steps [7]: video sequence, object detection, object recognition, and object tracking. Then, the image sequences in a video are analysed to locate and detect objects, followed by the aforementioned stages: detection, classification, and tracking processes.

A video is a sequence of images, known as frames, that may comprise still and moving objects. Koraqi and Idrizi [8] mentioned that object detection requires prerequisites in a video system, such as a basic data model, hypothesis, detect, and hypothesis verifier. In addition, the region of interest (ROI) is expressed in many object detection processes [9]. After defining the ROI, an ID is given to the target object or continuous tracking and counting. Then, another entry or object in the video sequence is given a new ID. The tracking of the object with IDs is stopped once it exits from the video sequence.

2.2. Challenges in object detection, classification, and tracking

The most common challenges of object detection and tracking include the following: variations in viewpoints, angles, and dimensions of the object; occlusion, such as objects appearing similar to humans; and illumination in the image sequence.

2.2.1. Viewpoint variations

The objects are identified from various angles, and directions/poses correctly. In the past, several methods have been used to address the problem of viewpoint variations. For instance, handcrafted features are employed using a discriminative distance to separate faces (i.e. the distance would be smaller for the same people's faces). However, handcrafted features are considered to provide less efficient results than deep learning techniques [10]. Noord and Postma [11] developed a CNN to address the problem of viewpoint variation through the deep learning of the image characteristic. Similarly, Keceli [12] used a pre-trained CNN model to map different views of 3D images and translate them into 2D features that facilitate distinguishing face images.

2.2.2. Occlusion

Occlusion also poses a significant challenge during object detection and tracking in videos and images. The occlusion problem affects the accuracy of object detection and tracking because objects hidden by other objects in an image produce further complexities. Previously, a sparse representation-based classification (SRC) was considered a robust technique to deal with the issue of occlusion in which an identity matrix is obtained and used to reference the developed occlusion database with presumed features [13]. However, Ou et al. [14] argued against the complexity and computational challenges of the SRC model and proposed a structured SRC (SSRC). The SSRC model uses data instead of an identity matrix that reduces the overall computational complexity. Computational complexity is an important consideration when designing and developing machine learning systems. Cao et al. [15] developed a framework to address the problem of complexity but failed as it reduced the accuracy rate (0.6) for object detection in the presence of occlusion.

3. Literature review

Countries, such as UAE and others in Gulf Cooperation Council (GCC), are continuously looking for novel technologies and tools to combat cyber and malicious attacks [16]. The old surveillance technologies used for security video surveillance are identified to be ineffective and inefficient [17]. The new surveillance applications use object detection, and tracking techniques based on deep learning and are vulnerable to threats and adversaries. Therefore, the UAE and many other countries are searching for an optimum solution that can be helpful in the long term.

Fig. 1 provides a synopsis of the previous surveillance system. The old systems used scattered monitoring systems based on scattered data and federalised control units with decentralised expenses. Fig. 2 shows the benefits of using novel surveillance technologies based on AI and machine learning. The AI-based system employs unified data where the centralised control unit is used for optimisation and smart surveillance. These novel surveillance systems are efficient, smart, cost-effective solutions to object detection and tracking issues. Recent studies on new surveillance technologies have widely used ML approaches for object tracking and detection [18,19]. The new surveillance technology relies upon AI-based/ ML-based technologies that are accurate, robust, and efficient. However, these technologies have highlighted the need to prepare against threats and adversarial attacks that can impact the network's performance significantly. Performance and accuracy improve when a network identifies and distinguishes the attacks from real input images.

Tracking–learning–detection (TLD) is a well-recognised technique, proposed by Kalal et al. [20], which integrated three separate tasks, namely tracking, learning, and detecting to track objects from frame to frame in a video feed. In TLD, continuous tracking allows the retrieval of information and features to contribute to localisation. In a TLD model, the error related to the detector is computed and updated in the learning step, followed by the detecting step, which consistently reduces the total error. However, similar to the issue in struck technique, Valestin et al. [21] argued that TLD was also less satisfying because of its low accuracy. Thus, deep learning-based neural networks have become prominent in providing promising object detection and tracking results.

Sun et al. [22] developed an object tracking model using a combination of CNN and extreme machine learning. Large datasets are used to train CNN-based models with a sufficient computational requirement. This hybrid model is considered robust for tracking and detecting objects even in the presence of occlusions, viewpoint variations, and illumination. This model is running in a real-time environment, demonstrating an improved tracking speed. Denis et al. [23] presented another object-tracking technique that required high-resolution video processing to detect small-size targeted objects. This technique reduced computational complexity by incorporating the block-wise processing of images. Dong et al. [24] addressed the object detection of moving objects using the deep learning technique in video sequences. Using perspective transformation, the background motion is computed from a moving camera. The patches around the moving objects are classified using deep learning by subtracting the sparse points from the image or the background. The temporal

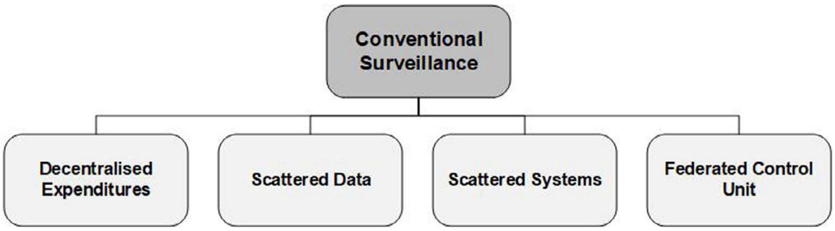

Fig. 1. Old Surveillance technologies.

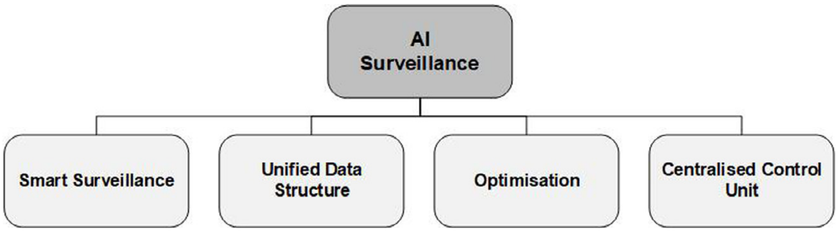

Fig. 2. AI Surveillance technologies.

Table 1
Experimentation results of 45 video sequences for accuracy of object detection of HD images.

	Motion difference only	Appearance computation using deep learning
Recall	0.766	0.798
F-Score	0.684	0.806
Precision	0.630	0.819

analysis is performed using Kalman filter to increase the consistency of moving objects. Table 1 shows the accuracy of object detection of HD images from 45 video sequences used during the experimentation by Dong et al. The videos are processed with 30 frames per second with one minute of length and recorded by an HD resolution camera, such as 1920 × 1080. With appearance information, actual video sequences showed improved results.

Gokhan and Süsstrunk [25] introduced another technique that could detect fast-moving objects with enhanced accuracy by reducing colours and developing two components to estimate object saliency accurately. The first component is related to the colours and space centre variation using effective filtering. The second component computes the contrast of the entire image. Next, a saliency map is calculated using 30 fps of HD videos using these two components. Another work by Kim et al. [26] proposed a novel framework that combines deep NN and background subtraction to address fast-moving object detection. At first, background subtraction is performed in all image frames to detect the objects. Then, the CNN classification practice recognises and categorises image frames into different groupings. This technique further decreases the computational complexity compared to other object tracking approaches [26]. According to Martins et al. [27] many studies have used different object detection techniques that address the problem of detecting intrusions and adversaries. On the contrary, Kanimozhi and Jacob [28] argued that a wide range of object detection and tracking models were developed and discussed in the academic space, yet, their practicality concerning real-life situations has not been established. Arguably, defence mechanisms and enhanced security against GANN threats and adversarial attacks have gained little attention, whereas most studies discuss the resistance of these models towards attacks.

3.1. You-only-look-once

YOLO is a one-step approach to spotting and categorising objects. Usually, the bounding box method is used to evaluate the input image [29]. Many different models use the YOLO method for object tracking and detection [30]. Redmon et al. [31] noted that YOLO was a straightforward model; compared to CNN, its application in a real-time environment is promising. Furthermore, YOLO is trained on complete images and shows a simplified representation of the target objects suitable for fast object detection. However, the lack of availability of large datasets and complete images is the biggest limitation in object detection when using YOLO. Hence, a developer of YOLO models must be an expert and professional to work with manual labels during training and handle the above limitations.

The YOLOv3 [32] detected multiple small objects efficiently and accurately through the deep-sort approach. A study [33] argued against the deep-sort approach concerning its efficiency in a real-time environment and supported an alternative approach called kernerlised correlation filter (KCF). Moreover, Yadav and Payandeh [34] also supported the use of KCF as it reduces overall computational complexity during object tracking and detection.

3.2. AdaBoost and Haar cascade function

Phuc et al. [35] performed AdaBoost training and classification using a Haar cascade classifier for object recognition and detection. The cascade function trains the algorithm on negative and positive sample images. The four steps involved in the proposed algorithm are the selection of Haar features, creation of images, training using AdaBoost, and classification using the Cascade function. Ulfa and Widyantoro [36] used Haar cascade classification to detect vehicles, whereas, Cuimei et al. [37] detected objects of three different classes. Cruz et al. [38] demonstrated the results of Haar cascade-based object tracking and detection compared with the HOG method and LBP technique of object detection. Haar cascade method showed improved results and performance compared with the other mentioned techniques.

3.3. Convolutional neural network family

CNN is amongst the deep learning neural networks that are flexible in training large datasets. CNN is praised for its improved results in object detection and tracking in still images. Zhu

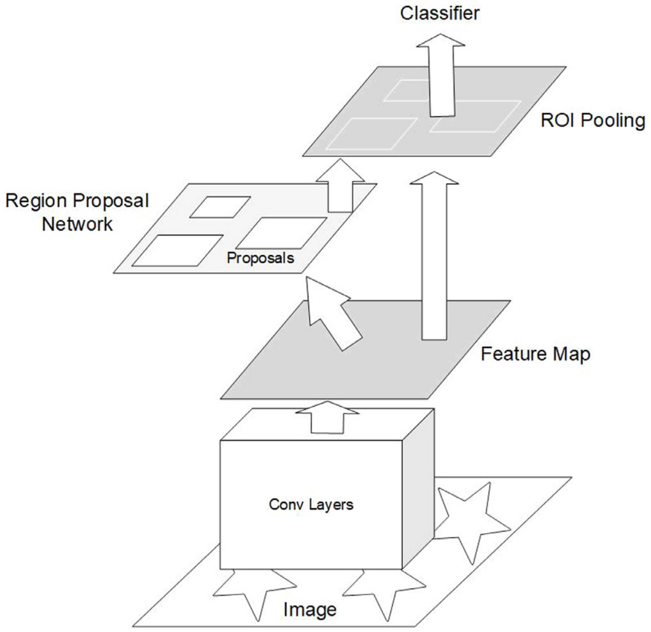

Fig. 3. Faster R-CNN Architecture.

et al. [39] used CNN to achieve the object detection of moving objects in a real-time environment. Deep CNNs were used to extract information from moving regions in the video frame, and the model was trained using video clips. The objects were accurately detected in moving frames using a framework-based fine-grained and coarse-grained approach. The major limitation of this approach is that it relies on high-resolution videos, which might not be possible with all surveillance cameras used for security. V̀ah́akainu and Lehto [40] stated that lighting effects, camera type, lens, and other factors could affect the quality of video clips, thereby affecting the accuracy of object detection.

Girshick et al. [41] established a region-based CNN (R-CNN), which trained CNNs on proposal regions from the input images in terms of categories or background using an end-to-end classifier. Hosang et al. [42] noted that R-CNN accuracy of subject detection highly depends on the results of the regional proposal module. On the contrary, Ren et al. [43] argued against this lack of accuracy and efficiency in R-CNN and thus proposed a Faster R-CNN. It used two units: first, a fully convolutional network (FCN) to determine proposal regions, and the second unit, which is a finder in the Faster R-CNN. The system works as a unit to detect objects, as presented in Fig. 3. The region proposal network (RPN) serves as the consideration to guide the Fast R-CNN module regarding the object to locate.

Kuan et al. [44] stated that Faster R-CNN was suitable for detecting objects from a large area, such as for surveillance on parking lots. Alom et al. [45] praised CNNs for their flexibility and adaptability to various applications, where the authors used CNN to address the extraction problem and detected co-saliency in the images. CNN models are also known to solve the problems of viewpoint variations.

3.4. Generative adversarial networks

A typical GANN architecture uses two different convolutional networks. The first network is known as a generator because it generates adversaries to deceive the system, whereas the second network, known as the discriminator, continuously operates to identify these adversaries and separate them from real images [46]. GANNs are more suitable for the real-time environment than CNN because the latter is good at detecting small objects, whereas the former is more scalable. GANNs are also used to generate high-resolution images as the generator of the network gradually learns to improve the pixel values and produce a high-quality image. GANNs are a suitable choice in the presence of obstructions and resolution issues. A wide range of studies has praised GANNs for their flexibility, training for a tamper-proof system, adaptability to work with other tools/applications [47] and ability to train large datasets of high-resolution images [48, 49]. Du et al. [50] stated that GANNs improved the performance and accuracy as compared with previous methods, such as CNN family, YOLO, and other object tracking and detection methods. Liu et al. [51] highlighted that GANNs were beneficial because multiple generative adversarial networks (GANs) can be used for multiple object detection and tracking to increase efficiency and accuracy. Liu et al. [52] developed a model based on multiple GANs by training them on images to generate adversaries for the tamper-proof object detection systems. Similar to CNN, various variations of GANNs, such as DCGAN, cGAN, Cross-GAN, and IDSGAN, have been developed in the past [53]. Lin et al. [54] proposed a GAN framework called IDSGAN to address the false-positive problem produced by adversarial attacks and obtained effective results. Zhang et al. [55] proposed a GANN called multi-task GAN or MTGAN, which detected small objects effectively. The proposed network can be used with other detectors because the network generates and distinguishes images simultaneously like a conventional GANN. More significantly, the loss is computed throughout, and backpropagating technique is used to guide the network to produce high-resolution images for enhanced localisation and classification. In MTGAN, the generator can capture in-depth details from the input images and produces a high-quality image that enhances the detection accuracy. Peng et al. [56] noted that GANNs were used for translating images, but the details are not preserved with quality. Concerning the use of the facial expression, Aggarwal et al. [57] stated that GANNs had limited applicability and are best suited for detecting and tracking objects. In this regard, Wang et al. [58] proposed the evaluation metrics for GANNs and concluded that they are complex, require a larger dataset for training, and cannot generate a quality image as perceived by humans. With these limitations, GANNs are still widely used in computer vision, machine learning and object detection fields for various sensitive purposes. Lee et al. [59] mentioned that GANNs can engage in a powerful learning and do not require supervision. GANNs generate high-resolution images and videos as adversaries that train the system to be tamper-proof against attacks and threats. Although GANNs do not require supervision, Pier Davide et al. [60]demonstrated that supervised learning in GANN could aid in discriminating different objects in space. Another benefit of GANNs for object detection and tracking is that they can enhance the overall resolution of images that help in making the small objects clearer, thereby deceiving the discriminator to consider it a real-looking object. Kalirajan and Sudha [61] stated that further research regarding the use of GANNs was needed as it could contribute to enhanced, cost-effective and efficient object detection and tracking along with improved security under GANN threats. Several authors discuss the limitations and drawbacks of GANNs. For example, Simao et al. [62] evaluated the GANN framework and argued that the model was beneficial in producing high-quality images. However, the model was less effective in detecting gestures that were not in the dataset. GANNs are poor in generalising and have less ability to determine the distribution across the dataset. The application of GANN for untrained gestures is limited. Donahue et al. [63] criticised GANN for its complexity, thereby making the training harder compared with other deep neural networks (DNNs). Due to its complexity, training on a large dataset may increase the computational complexity, which can affect the overall performance of the system. Once GANNs are successfully trained on a larger dataset, it generates an ample number of adversaries with low computation. Based on this, it is argued that object detection and tracking under the GANN threat is beneficial because it may increase the complexity but can produce anomalies and prepare the system against them. The current research on GANNs for object detection and tracking systems reveals that most studies, such as Zhang et al. [64], used images for training rather than video clips. Considering that training under GANN threats is a novel area of research, tamper-proof object detection and tracking real-time systems have not been extensively investigated before [65]. A cost-effective solution that can detect small objects in video feed accurately and efficiently is required for a real-time video surveillance system. Furthermore, with high accuracy, the benchmark of computational complexity must be achieved because a tamper-proof system should be robust and efficient. Table 2 is included below to provide a systematic review and comparison of the above-discussed techniques. The table includes techniques and algorithms used by different researchers for object detection and tracking in images and videos. The accuracy of each proposed model concerning object detection is shown along with real-time object tracking performance on videos. The speed of the algorithm is also analysed by examining tracking speed using the frames per second (fps) evaluation parameter. In addition to that, the performance of each model is evaluated based on the presence of different challenges like occlusion, illumination, scale variance, multi-objects, anomalies, etc.

4. Threats to GANN in object detection

Identifying and detecting dangerously problematic objects and threats in object recognition plays an important role in ensuring and guaranteeing the security and safety of systems. Due to the complex nature of the task, the human expert detection performance is only about 80%–90% accurate. Deep convolutional neural networks have already shown good results, but not at the security level and adversarial attacks. Our goal in this section is to conduct a literature review covering different GANN types that might add value to this known issue and open opportunities for promising future research work. Object tracking and detection systems are often deceived by adversarial attacks that influence their detection accuracy and produce complications for applications that rely on accurate object detection. An adversary can be anything from a manipulated image, a similar-looking image, noise in the image, or intrusions. Almost all neural networks are susceptible to attacks and threats that can deceive the system into generating false-positive results and affecting its performance. The most common example of an adversarial attack is strategically established noise to add to the input image and deceive the neural network. The classification algorithm has clear decision-making boundaries in a neural network system, and an adversary corrupts the process. Fig. 4 shows the real input data points, the classified data points, and the adversaries. The green points represent the accurately classified data points using the parameters and features used to design the decision boundary based on which decision is taken. The orange data points represent the adversaries wrongly classified as green dots.

The decision is manipulated because of attacks and perturbations in the system. Today, various applications use neural networks, and machine learning approaches widely, and adversarial attacks are the most significant challenge. In systems, where human recognition, crime prevention and security surveillance are involved, adversarial attacks must be resolved. For instance, a DNN in an autonomous car and self-driving applications may predict the occurrence of an adversarial attack falsely, thereby raising several concerns regarding safety and trust in these systems [79]. Pan et al. [80] noted that the problem of false-positive in GANNs and other machine learning approaches was essential to address; otherwise, the consequences could be severe. In the applications of surveillance video, the first attack of the adversary was executed by manipulating the pixel values of the input image, leading to its misclassification [1]. Another way to create an adversary is by applying 'patches' to the objects in the image to deceive the network towards misclassification. Therefore, security surveillance systems are susceptible to adversarial attacks, which can affect the overall accuracy and performance of the system. Machine learning attacked by adversaries may not identify anomalies accurately and may require human intelligence and intervention. The surveillance video system can also face malicious attacks that change or hide the content to deceive the network and the human actors that monitor the security video feed, such as security personnel. Ullah et al. [81] argued that human actors monitoring the video feed of security surveillance were prone to several issues, such as missing the changed content and may not be able to detect threats in the presence of multiple video feeds and cameras. Henceforth, using technology is an operative and competent method to mitigate the challenge of adversarial attacks and false positives in machine learning systems. The attacks are classified into subsequent categories based on the information accessible to the attacker or adversaries, as explained in the next sub-sections.

4.1. White box attack

In a white box attack, the attacker has information regarding the architecture, modelling of the system, details about the training set, weights, or the samples on which the system was trained. The function used for the classification is prone to adversarial attacks in a white-box setting because the attackers have adequate knowledge to damage the system. Concerning neural networks, a backpropagating technique, such as DeepFool and the fast gradient sign method (FGSM), is used to conduct an attack because the gradients are known to the attacker. FGSM aims to ensure that the system misclassifies the data points and introduces intrusions to influence the system's accuracy and performance. The FGSM algorithm can be employed to conduct a white-box adversarial attack. An example input image is shown in Fig. 5. An adversarial image is generated by using the formula shown in Eq. (1)

$$Advx = x + \epsilon . \sin(\nabla x J(\Theta, x, y)) \quad (1)$$

The input is represented by x; the label is denoted by y; ϵ is the multiplier, which will keep the noise or perturbation unnoticeable, small, and effective. J is used to calculate the total loss, and Θ is for the model's limitations. Using the FGSM method, the gradients are selected to ensure that loss is as minimum as possible based on the pixels manipulated from the real input image. The chain rule is used to monitor and determine the loss value, whilst perturbation is introduced to each pixel, maximising loss. The white-box attacks can deceive an already trained system, leading to misclassifying the object in an image. An example is given in Fig. 5, where the original image x of a panda is manipulated using small perturbations that deceive the network and impact decision-making. The input image is misclassified as a gibbon and is not accurately classified as a panda, hence impacting the confidence level.

Basic iterative refers to the extended process of FGSM, and a series of adversaries are produced in every iteration. The white-box attack is more serious when extensive iterations are used, and the model does not depend on system estimation [83]. Jacobian Saliency Map Approach (JSMA) is another variation of FGSM, in which the input image is manipulated to achieve the goal

Table 2
Review of different object detection and tracking techniques in challenging scenes.

	Technique	Cited	Obj. Detect/ Track accuracy	Total frames	Tracking speed	Performance during challenging scenes
TLD	Improved TLD algorithm	[66]	80%	500	26.32 fps	Performed well in the presence of a moving camera, motion blur, similar objects, scale change, illumination changes, and occlusions
YOLO	with SORT algo.	[67]	85.1%	930	–	–
	YOLO-ACN	[68]	55.8% (Average Precision, AP)	4491	16 fps	Performed well to detect occluded and small size objects in real-time videos
	with hard-example mining	[69]	90.49%	–	–	Tested on images only. Detected objects accurately in the presence of occlusion
Adaboost	with Haar training	[70]	85.9%	–	20 fps	Performed well in the presence of occlusions
CNN	Faster R-CNN	[71]	75.2%	–	25 fps	Performed efficiently in a real-time environment and detected multi-objects in a single frame
	with KCF	[72]	86.6% (Precision)	576	–	Detected human objects in the presence of occlusion and scale variance (SV)
	with Correlation Filters	[73]	49% (59.7% SV, 59.2% occluded)	–	8 fps	The model was slower but effectively detected objects in the presence of occlusions and SV
GANN	NM-GAN	[74]	90.7%	–	0.031 fps	Detected objects in the presence of anomalies and detected anomalies as well
	OPGAN	[75]	84.2%	–	–	Tested on images only, efficient in detecting small objects
	Deep learning with GAN	[76]	91.9% (precision)	–	14.8 fps	Performed successfully in 11 challenging scenes: occlusion, deformation, scale variance, etc. The algorithm is slow, with poor tracking performance in the presence of long-term occlusion and similar objects.
	GAN with Faster-RCNN	[77]	89.65%	–	–	Detected small objects in images High complexity time consuming and high computation
	GAN-Do	[78]	67.47%	–	–	Addressed the problem of object detection in reduced quality images. Performed well to detect images in a camera-shake blur setting, Gaussian Blur, Defocus Blur, and additive white Gaussian noise

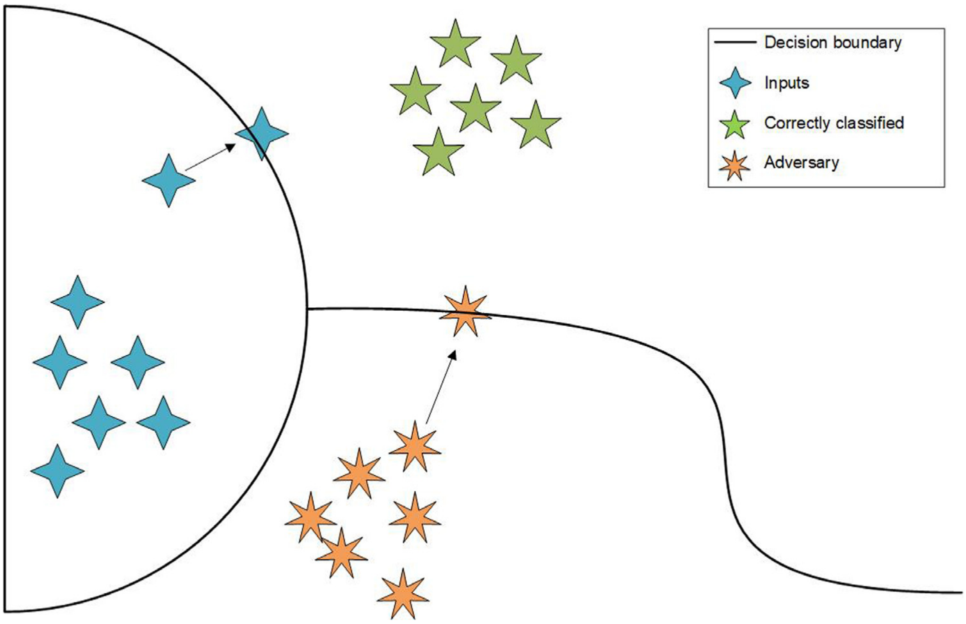

Fig. 4. Adversarial attack example.

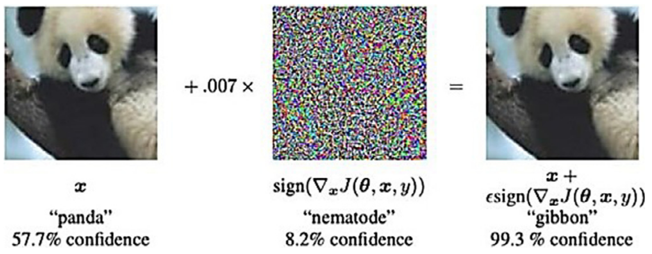

Fig. 5. Adversarial attack example [82].

of misclassification. A sample image is taken as input x with a target denoted by t; a Jacobian matrix is implemented where derivatives are forwarded gradually in each step. The pixel values are incremented with each passing iteration, and gradually the target class denoted by $t*$ also increases. Thus, the value of t is changed to $t*$, becoming the cause of misclassification. This type of attack is limited to the training models that use supervised training and targets. Therefore, this method of adversarial attack generation is not suitable for all approaches and has limitations when it comes to white-box attacks [84]. The benefit of creating an adversarial attack in a white-box environment is that the adversary is comparatively easier and simpler to create. In addition, the white box setting allows the attacker to gain more access and knowledge about the system, thus facilitating the generation of adversaries. The time consumed in the overall process is also less than other adversary generation approaches. The white-box setting generates adversarial attacks and analyses the system under GANN threats to ensure that the system is tamper-proof. In less time, the system is prepared against the threats and GANN attacks using white-box settings, such as by determining all the system's vulnerabilities. A white-box attack has limitations such that the environment is not practical to be applied to real-life machine learning-based application systems. Furthermore, the computational cost gradually increases when the white box set is used to conduct an attack or test the system's vulnerability because small perturbations are constructed to manipulate the input samples.

4.2. Black box attack

Compared with white-box attacks, black-box attacks have information about the outputs of the model instead of the key internal/functional information (i.e., the architectural model, weights, and other training details). The black-box attacks are conducted in several ways based on the knowledge of outputs and other sub-areas, as mentioned below.

4.2.1. Score-based attack

In this type of attack, the output layer is accessible and known to the attacker. Using the information and knowledge regarding the queries related to the output, the attacker determines the classification approach used for the system. In a black-box environment, the adversarial images are generated by manipulating loss, confidence level and pixel values. The feedback is obtained by interpreting the confidence level and score, and the pixel values in increments gradually change the value of the loss. A change in the pixel value contributes to a change in the confidence level, which then manipulates the system to misclassify the input images. Using the score-based attack, the adversarial samples can be generated without gradients. For instance, Brendel et al. [85] developed a genetic-based algorithm that generated adversaries without manipulating the gradients.

4.2.2. Transfer-based attack

In this attack, a model is developed by using the available knowledge and information regarding the original system. Then, an auxiliary or imitated model is constructed to mimic the operations and outcomes generated by the original system. This model is now used for transfer-based attacks using a white-box setting. In other words, the attacker knows the auxiliary model's training details, architecture, and other parameters and uses this retrieved information to attack the imitated model. If the attack on the secondary model is successful, then the same attack episode is conducted on the original model. Compared with a score-based attack, the transfer-based attack is complex and time-consuming because the attacker must design and imitate the original system. In addition, the computational cost is also high for this type of attack, given that the auxiliary model mimics complex operations.

4.2.3. Decision-based attack

The outcome and output values are accessible in this attack, and the attacker commences a decision-based attack using this knowledge. The decision-based attack is more relatable and relevant than a score-based attack in real-world scenarios as logits are rarely known or accessible to the attackers. Furthermore, compared with other attacks, decision-based black-box attacks are resistant against common defences, such as gradient masking and robust adversarial training. Some benefits of conducting a black-box attack to prepare and train the object tracking and detection system under GANN threats are noted. The first benefit is that the black-box environment is near to real-life situations; thus, it is applicable for various applications, such as self-driving autonomous cars. Another benefit of conducting a black-box attack is that the attacker works with limited available information whilst utilising all the available resources. Comparatively, decision-based and score-based attacks are easier because transfer-based attack requires the imitation of the original model, which increases the time and complexity of the process. The systems trained and tested under black-box testing are more secured and robust than those trained only for gradient-based and score-based attacks. The limitations of black-box attack testing and training include hard labels that complicate the process. The adversaries are less likely to be generated in an environment where the hard label is used. In such a situation, the attacker attempts to use a series of queries, and convergence is uncertain, further complicating the process. This attempt to use several queries is referred to as the random walk approach. If the random walk approach does not produce the desired results, then the attacker moves to the optimisation-based technique. However, the success of the optimisation technique relies heavily on the dimensionality of the input and training dataset.

4.3. Grey box attack

In a grey box attack, also known as a semi-white box attack, the attacker uses a generative model for training to generate adversaries in a white box setting. Once the new model based on generative training is developed successfully, the attacker does not require the original model of the system and uses the generative model to create adversaries whilst keeping the black-box attack.

5. Training under GANN threats

The real-life applications of using object detection and tracking techniques must be trained against adversarial attacks and GANN threats. Appiah et al. [86] noted that security is required in highly crowded areas and cities to prevent crime and terrorist attacks, including cyber threats; thus, the security officials are interested in employing AI in human behaviour and threat detection. In 2017, the malicious attack on UAE's Telecommunication Regulatory Authority (TRA) software and website interrupted operations. According to Chandra et al. [87], the Dubai police reported that almost every one of five residents in the UAE faced a cyber-attack in 2015. Threats like these, intrusions and malicious attacks can interrupt business operations and impact sustainability. Therefore, training under GANN threats must be conducted for object detection and tracking systems to ensure tamper-proofing and enhance the security of these systems. The object detection and tracking technique used in video surveillance and security systems mainly deal with crowded spaces and low-resolution images. Various reasons behind developing object detection and tracking systems, such as crowd estimation, face recognition, human identification and behaviour detection, can be noted in [88]. GANN-based object detection and tracking ensure that the system is prepared against possible attacks and threats. The GANN model supports the training but generates different adversaries and trains the network to distinguish the original inputs from adversaries.

5.1. *Generating adversarial samples*

Several algorithms can be used to generate adversaries, which are discussed below.

5.2. *Fast Gradient Sign Method (FGSM)*

This algorithm utilises a single-step technique in which perturbation and adversarial noise are computed by using the gradient method discussed in Huang et al. [89]. The gradient value is given by $\nabla x\Theta$ because this objective function is used to train the network. x represents the original input sample image, whereas ϵ is the noise or perturbation that is added to the input (x). The adversarial image noted by $xadv$ is calculated by adding the noise to the input image as shown in Eq. (2)

$$xadv = x + \epsilon.\sin(\nabla x \int \Theta(x, y)) \tag{2}$$

Every input image is given a label noted by y in Eq. (2) for its correct classification. The noise in the image aims to attack the system to misclassify the image. The noise is kept as small as possible, which is performed by using Eq. (3)

$$\|x - x_{adv}\| \infty < \epsilon \tag{3}$$

5.2.1. *Broyden–Flecther–Goldfarb–Shanno (l-BFGS)*

This algorithm views the problem of generating an adversary similar to an optimisation issue and uses Eq. (4) to address the problem. $minus\delta c.\,\|\delta\|\,2 + \Theta(x + \delta, t)$

Subject to

$$Lm \leq x + \delta \leq Um \tag{4}$$

where θ is the objective function used by the algorithm. In this case, the adversarial noise is denoted by δ, and the label for each input image x is given by t. The range of the pixel values is between L to U for the maximum loss computation denoted by m. Furthermore, the evaluation is conducted with the help of the line-search method and recorded in c, as shown in the above equation. The evaluation value should be greater than 0.

5.2.2. *Carlini–Wagner Method (CW)*

In this algorithm, the distance metrics are utilised to calculate the differences given by $L0$, $L2$, and $L\infty$. CW proposed this algorithm to generate an adversarial attack denoted by L2. Eq. (5) is important in the process, as shown below: $min \,||\, \frac{1}{2}(\tanh(w)) + 1 - x \left\| \binom{2}{2} \right\| + c.f\left(\frac{1}{2}(\tanh(w) + 1)\right)$ with

$$\left(x^{adv}\right) = \max\left(\max\left\{Z\left(x_i^{adv}\right) : i \neq t\right\} - Z\left(x^{adv}\right)_t, K\right) \tag{5}$$

As per the above equation, the main function of the algorithm is given by f, whereas $Z(x)_t$ represents the logits. The optimisation variable is represented by the letter w in the first part of the equation. k is the confidence parameter. As opposed to the above equation, the following equation is concerning the perturbations, the distance metric ∞ is thus calculated.

$$mincf\,(x + \delta) + \sum_i \left[(\delta_i - \tau)^+\right] \tag{6}$$

Eq. (6) shows that the threshold level τ is used to monitor the adversarial perturbation or noise introduced to the input image to train the network on adversaries.

The algorithms mentioned above are widely used to generate adversaries to train the DNNs and GANN on the adversaries for creating a tamper-proof object detection system. The machine learning-based deep neural networks attempt to use several traditional techniques to ensure that the model is robust against adversarial attacks. The traditional methods include addressing the problem of overfitting by using the weight decay techniques, such as keeping the weights as small as possible. Another traditional technique is a dropout, in which the selected random neurons are ignored during learning and training. However, the traditional technique does not provide an efficient and practical security mechanism against adversarial attacks. Currently, only two methods are known to provide an effective defence against adversarial attacks and samples. Defensive distillation is a strategy, in which the machine learning model is trained to produce output based on the probability of different classes instead of making a difficult decision. An earlier model that provides the probabilities of the outcomes belonging to the respective class is developed. Another method that is widely used and found significant is adversarial training, which prevents the addition of any weight updates of neurons that were removed on a backward pass.

5.3. *Generating adversarial samples*

Adversarial training is a brute force solution in which numerous adversarial sample images are generated, and the model is trained on these generated images so that these samples do not deceive the network. Previous work in this area includes GAN, which was examined and used extensively for different applications. GANNs were initially introduced by Goodfellow et al. [90] to generate examples by learning the probability distribution of the training set effectively. The ability of GANN to discriminate between the real image and adversary was well-received, and its application for object detection and tracking was identified by many researchers [91,92]. However, the adversarial attacks have severely affected the object detection and tracking models that use GANNs for video surveillance and other security applications. According to Kalbo et al. [93], GANN threats could influence the accuracy of object detection and tracking, hence contributing to vulnerabilities in the video surveillance security system. As a result, more object detection and tracking models trained under GANN threats will be needed. Massoli et al. [94] developed a model that could detect adversaries effectively and trained the model on adversaries to increase the robustness of the model (see Fig. 6). A detection approach was employed by the classifier

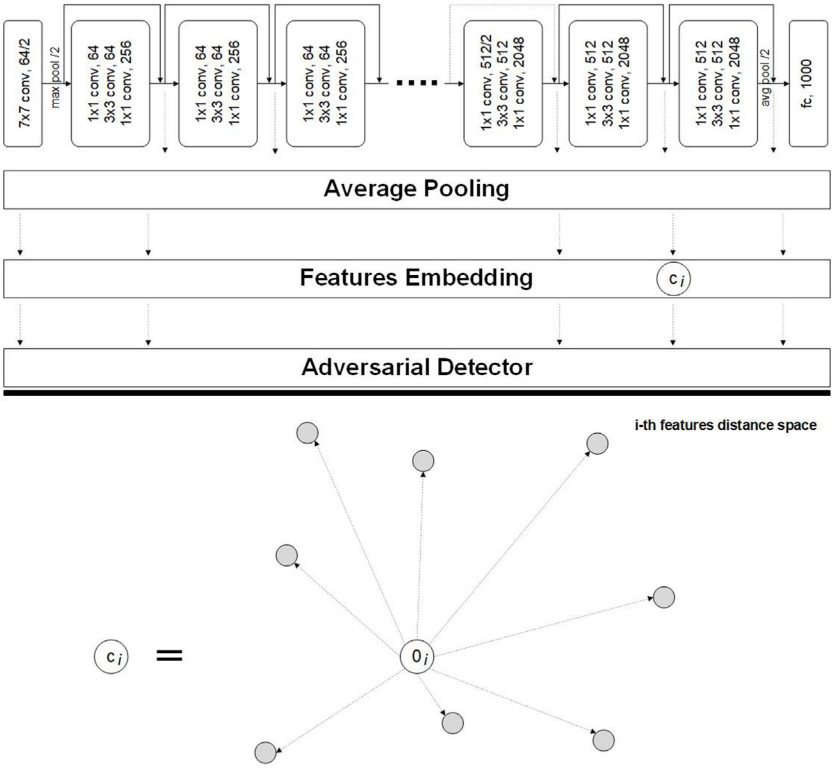

Fig. 6. Detection model (above), embedding process (below, c_j).

$f\theta(x) : X \to C$, C is the label given to input images in this case, where x is the input image. About the above classifier, x belongs to the different dimensions of the input image, such that, given by $X \subseteq Rd$

Eq. (7) is important for the model because the classifier receives the input image x and performs operations on it.

$$f\Theta\ (x) = f^n\ (0_n - 1;\ \Theta_n)^o \cdots {}^o f^o\ (x;\ \Theta 0)\ , \tag{7}$$

The values of centroids and medoids are calculated by using the following Eqs. (8) and (9), respectively:

$$p_i^j = c_i^j = \frac{1}{|B_c|}.\sum_{n}^{|B_c|} o_{i.n}^j \tag{8}$$

$$p_i^j = m_i^j = argmin \sum_{n}^{|B_c|} \left\| o - o_{i,n}^j \right\| \binom{2}{2} \tag{9}$$

In the above equations, the class j is denoted by $o(i, n)j$ in the ith layer to represent the output value of this class j in the ith layer. The cardinality of the class j is shown by $|Bc|$, which refers to the number of elements or properties related to that class. This proposed method used the CW method, Basic iterative method, and MI-FGSM technique to produce adversaries and train the system to learn to distinguish between adversaries and real inputs. As mentioned before, it is suggested that noise or perturbation should be small so that its loss is minimum. In this case, the noise is denoted by ϵ, the authors kept the values of perturbation, $\epsilon = (0.03, 0.07, 0.1, 0.3)$. The value of noise or perturbation is as low as 0.07 to as high as 0.3. It is kept at less than 1. The adversarial images are generated using the mentioned algorithms, and the process is described using Fig. 7.

The authors used a kNN classifier to view the problem of generating an adversary as an optimisation issue and to guide the source, as shown in Fig. 7. The above part of the figure shows the source and guide before the attacks, whereas the below part shows the results after the adversarial noise has been added. SotA, one of the state-of-the-art feature extraction models, is used to extract the features from the input image. The threshold values have been applied during the adversary generation. The adversarial noise limits or threshold values were set between 5.0 and 10.0 for this case, such that $\delta\epsilon$ 5.0, 7.0 and 10.0, 5.0.

The perturbation value is kept at 10.0 for the upper row of images in Fig. 8. This example shows that with an increased perturbation value of 10.0, the input image still looks similar to the output image. The accuracy to detect perturbations in such images was found to be 96.3% for supervised learning and 96.8% for the unsupervised learning environment. Table 3 shows a comparison of various adversarial attacks along with the kind of adversarial attacks or other threats to fool the network. The training or the core method used to generate the threat is discussed, followed by the nature of the attack, such as targeted, untargeted, or both. The information required to be accessed by the generated adversarial attacks is also highlighted, such as either model parameters information or logits, which implies inputs to the softmax layer of the model. The distance metric used for each adversarial generation method is also identified, followed by the vulnerability of the models, which are found to be easily fooled by the respective adversarial attacks or threats.

6. Discussion

This section presents discussions regarding the outlook of the current research direction based on the literature systematically reviewed in the earlier sections. In all, this review has examined and demonstrated several different facts and useful insights related to the use of GANNs in tamper-proof object detection and tracking. The previous sections provided a brighter and broader prospect for the readers, including technical details.

Table 3
Comparison of different adversarial attacks and vulnerable models.

Cited	Attacks/Threats	Training/ Core method	Target/ untargeted	Accessed Info.	Distance metric	Vulnerable models
[95]	DeepFool/ perturbations to fool the network	Iterative linearisation	Both	Model parameters	Lp, $p\epsilon[1,\infty]$	Deep Neural Networks
[96]	Universal adversarial perturbations	Generalising DeepFool to create universal adversarial attacks	Both	Logits	L2 (Universal perturbation)	Deep neural network classifier
[97]	High confidence adversarial examples, targeted misclassification	Adam optimiser	Both	Logits	$L0$, $L2$, $L\infty$	Distilled/ undistilled Neural Networks
[98]	JSMA, adversarial perturbations	Jacobian Saliency	Both	Model parameters	$L0$	Deep Neural Networks
[99]	Misclassification by rotation and /or translation	Natural transformations	Both	Logits	n/a	CNN
[100]	Image, and universal perturbations	GAN based adversarial generation	Both	Logits	Lp (Universal perturbation)	FCN
[101]	Spatially transformed adversarial examples, high-quality sophisticated adversarial attacks	Minimising adversarial and Lflow loss	Both	Logits	$L_f low$ (Measuring geometric distortion)	Deep Neural Networks
[102]	Few-Feature-Attack-GAN, black-box attack	Mask mechanism, GAN training	Targeted	Logits	$L0$, $L1$	Machine learning models

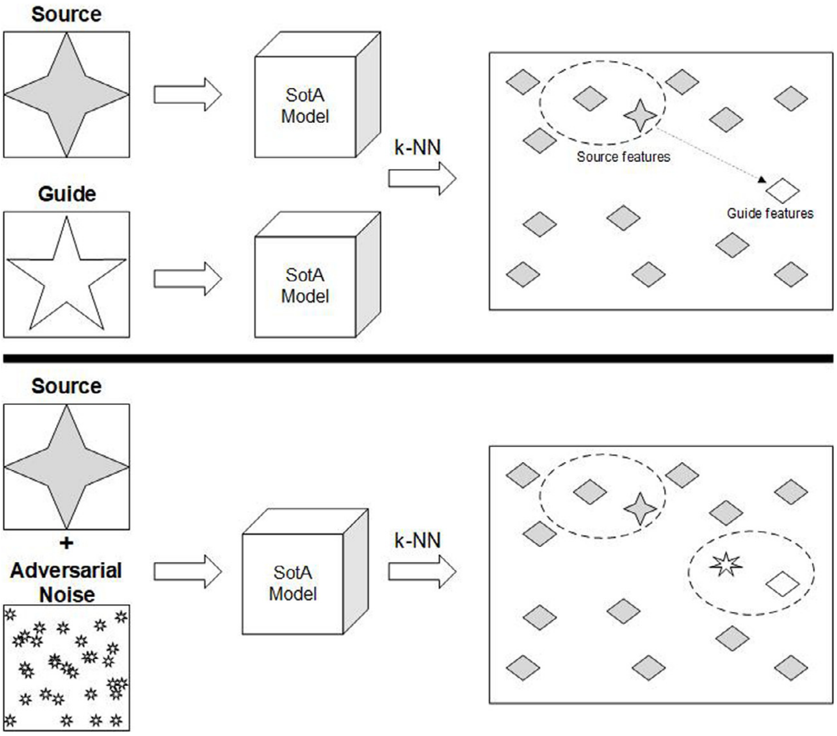

Fig. 7. Generating adversarial samples.

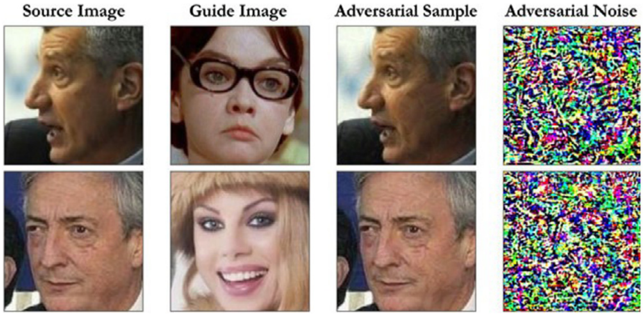

Fig. 8. Adversarial Samples and threshold [96].

Adaptability. Adaptability is significant for object detection and tracking method using machine learning. Some models, such as YOLO, defend against one type of attack and leave room for another kind of open attack for the attacker who knows the defence mechanism. It is analysed that machine learning-based algorithms and models widely used for object detection, such as R-CNN, YOLO, Faster RCNN, and CNNs, are broken and vulnerable to adversarial attacks in surprising ways. The failure to defend against the adversarial examples revealed that even the most straightforward algorithm behaves differently in an attack scenario than what the algorithm intended to do. It shows that adversaries can affect a model's learning and training outcomes in the presence of an attack. Thus, machine learning models for object detection and tracking should be trained under GANN threats to reduce the gap between what developers intended and how the algorithm performs.

This study has reviewed different techniques and found GANNs to be useful for object detection and tracking in viewpoint variations and occlusions. GANNs are complex and require more computational capacity, whereas adversarial training produces robust outcomes and provides a better mechanism against adversaries. Furthermore, once GANN has been trained on a large dataset, it can generate desirable results with low computation and more adversarial samples. Further research is required to review the existing object detection and tracking technique under GANN threats. It is suggested that GAN-based detectors should be developed for robust object detection and tracking. Using GANs for data augmentation and generating more adversarial samples would be beneficial, whereas the efficiency of GANNs to work with large datasets can be exploited to obtain a robust and efficient detector. Thus, GANN based detector is likely to produce robust results with strong generalisation with the help of using a combination of real-world images and simulated images generated by GAN.

Adversarial attacks. Adversarial attacks, such as white-box, black-box and grey-box, were reviewed. It is assessed that white-box attacks have less plausibility of occurring in real-life scenarios. On the contrary, black-box attacks are relatable and relevant to real-world scenarios because less information is accessible for the attacker in this type of attack. Therefore, the object detection and tracking technique under GANN threats must be trained on all adversarial attacks, emphasising black-box attacks. The models trained under black-box attacks are robust and provide a strong defence against adversaries. Adversarial training can be performed using the object detection and tracking technique to generate and train the system on adversarial examples. Several algorithms, such as FGSM, L-BFGS and CW, are utilised to generate adversarial instances for the training of the model. A dataset is a significant element of adversarial training; currently, COCO, OpenImages and WIDER face are the benchmark datasets for evaluating different object detection models [103–105]. The problem of the large and useful datasets can also be solved by using GANs to generate more adversarial samples for adversarial training.

Threats to deep learning model. Adversarial attacks are serious and are a major concern for machine learning-based object detection and tracking. Although few studies claim that adversaries are not a serious issue, a wide range of studies suggest the opposite. The literature discussed in Section 3 suggests severe and grave consequences of adversarial attacks on deep learning. The review shows that deep neural networks, such as CNNs and RNNs, can be deceived in various ways, such as detection and recognition. In general, deep learning models are susceptible to adversarial attacks and threats.

Transfer property of adversaries. The adversarial examples transfer well amongst different neural networks. This observation is valid, particularly for the models or networks with similar construction designs and architecture. Concerning this, black-box attacks are often determined to exploit adversaries' generalisation.

The notion of linearity. The design of neural networks forces the model to behave linearly, which makes the model susceptible to threats and adversaries. Although this notion is argued against and criticised, the literature review suggests that linearity is one of the weaknesses of neural networks in adversarial attacks.

Further investigation. This study examined improper training, computational capacity, architecture and design of the model, and weak defence mechanism amongst various viewpoints. However, the viewpoints lack alignment in a single direction. Thus, further investigation should be conducted in this direction to reveal the causes behind the weaknesses of neural networks towards adversarial threats.

7. Conclusion

In this paper, the existing object detection and tracking techniques were discussed and reviewed. In addition, image classification, object localisation, and detection techniques were reviewed, such as SVM, Adaboost, HOG, Haar cascade, CNN family, YOLO and GANNs. GANNs and CNNs produce better results than other object detection and tracking models in real-time. However, under adversarial training, GANNs are more suitable for object detection/tracking because they can work with large datasets and generate more data from the samples. Object detection and tracking have several different applications in real-life, such as face detection and recognition, medical imaging, traffic monitoring, weapon detection, vehicle recognition and video surveillance security systems. The real-life applications of object detection and tracking come with certain challenges. The real-world images vary in terms of light, angle, variations, and occlusions, such as objects that look like humans, etc. These challenges are considered significant in machine learning-based object detection and tracking. CNNs are deemed effective in addressing the problem of viewpoints variations where the handcrafted features technique failed to provide desired results. Similarly, CNNs have also been praised for addressing the problem of illumination and occlusions in addition to viewpoint variations. Concerning the challenges faced by machine learning and deep neural network (DNN) for object detection and tracking, the adversarial attack problem is among the most significant.

Adversarial samples are not easy to defend because the machine learning models must generate good output value for possible input images. Most machine learning models, such as YOLO, AdaBoost and Haar cascade-based object detection, are perceived to work well. Still, they tend to work on a small amount of dataset or possible inputs the model may encounter. Various object detection and tracking techniques have been reviewed in this paper, and the practicality of these models in real-world applications is questioned. Furthermore, the consideration of adversarial attacks and defence against adversarial examples is lacking in most state-of-the-art object detection and tracking techniques. Additionally, after assessing the literature on white-box, black-box and grey-box adversarial attacks, it was inferred that white-box attacks have less plausibility of occurring in real-life scenarios. Finally, the adaptability of an ML model is significant for object detection and tracking method in the presence of an attack, and it should be trained against adversarial threats to reduce the gap between what developers intended and how the algorithm performs.

8. Future developments

The future work related to object detection and tracking under GANN threats is dynamic and wide-ranging. The implications of the susceptibility of deep learning-based neural networks object towards adversaries in object detection, and tracking tasks are gaining recent attention. Future developments and applications of object detection and tracking under GANN threats are abundant. Researchers in this field may focus on complex data sources and dynamic targets concerning objects for the future. For example, the current study has focused on the case of tamper-proof object detection and tracking model for video processing in real-time under GANN threats. However, video object detection in real-life scenarios is challenging because remote sensing issues, blurring, occlusion, intense movement of target objects, and motion ambiguity lead to complications.

Real-time detection using remote sensing is useful for agricultural settings and crowded spaces, such as security surveillance in public spaces. The automatic object detection and tracking software with integrated hardware has opened a window of opportunity for the UAE in this area. Another promising future development application and improvement is multi-domain object detection. Previous literature has established that detection performance for the domain-related model is high for a specific domain/dataset. Thus, universal detection should be focused on the future, such as a multi-domain detection that can work on several datasets/domains without having preceding information about new domains. Consequently, domain transfer is a challenging yet beneficial area for future development.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Its a review paper with scholar references.

Acknowledgments

The author would like to thank Dubai Police for funding this project under fund contract 41524.

References

[1] K. Ren, T. Zheng, Z. Qin, X. Liu, Adversarial attacks and defenses in deep learning, Engineering 6 (3) (2020) 346–360.
[2] M. Alrawi, Ride sharing platform careem says hit by cyber attack with data of up to 14 million users stolen, 2018, https://www.thenationalnews.com/uae/ride-sharing-platform-careem-says-hit-by-cyber-attack-with-data-of-up-to-14-million-users-stolen-1.723927. (Accessed 18 November 2022).
[3] S. Abdurrahman, Smart video-based surveillance: Opportunities and challenges from image processing perspectives, in: 2016 3rd Int. Conf. on Information Technology, Computer, and Electrical Engineering, ICITACEE, IEEE, 2016, p. 10.
[4] B. Ma, L. Huang, J. Shen, L. Shao, M.-H. Yang, F. Porikli, Visual tracking under motion blur, IEEE Trans. Image Process. 25 (12) (2016) 5867–5876.
[5] W. Kim, C. Jung, Illumination-invariant background subtraction: Comparative review, models, and prospects, IEEE Access 5 (2017) 8369–8384.
[6] J. Han, D. Zhang, G. Cheng, N. Liu, D. Xu, Advanced deep-learning techniques for salient and category-specific object detection: a survey, IEEE Signal Process. Mag. 35 (1) (2018) 84–100.
[7] S.S.A. Rajjak, A. Kureshi, Recent advances in object detection and tracking for high resolution video: Overview and state-of-the-art, in: 5th Int. Conf. on Computing, Communication, Control and Automation, ICCUBEA, IEEE, 2019, pp. 1–9.
[8] L. Koraqi, F. Idrizi, Detection, identification and tracking of objects during the motion, in: 3rd Int. Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT, IEEE, 2019, pp. 1–6.
[9] M.M. Jan, N. Zainal, S. Jamaludin, Region of interest-based image retrieval techniques: a review, IAES Int. J. Artif. Intell. 9 (3) (2020) 520.
[10] C.-H. Chen, R. Chellappa, Face recognition using an outdoor camera network, in: Human Recognition in Unconstrained Environments, Elsevier, 2017, pp. 31–54.
[11] N. Van Noord, E. Postma, Learning scale-variant and scale-invariant features for deep image classification, Pattern Recognit. 61 (2017) 583–592.
[12] A.S. Keçeli, Viewpoint projection based deep feature learning for single and dyadic action recognition, Expert Syst. Appl. 104 (2018) 235–243.
[13] D. Zeng, R. Veldhuis, L. Spreeuwers, A survey of face recognition techniques under occlusion, IET Biom. 10 (6) (2021) 581–606.
[14] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, Z. Zhu, Robust face recognition via occlusion dictionary learning, Pattern Recognit. 47 (4) (2014) 1559–1572.
[15] W. Cao, J. Yuan, Z. He, Z. Zhang, Z. He, Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection, IEEE Access 6 (2018) 8990–8999.
[16] zawya.com, UAE ranks 8th globally and 1st regionally in UN's 2016 e-Smart services index, 2016, https://www.zawya.com/en/press-release/uae-ranks-8th-globally-and-1st-regionally-in-uns-2016-e-smart-services-index-bmty66du. (Accessed 18 November 2022).
[17] M. Cayford, W. Pieters, The effectiveness of surveillance technology: What intelligence officials are saying, Inf. Soc. 34 (2) (2018) 88–103.
[18] I. Sharma, A More Responsible Digital Surveillance Future, Federation of American Scientists, 2021.
[19] S. Feldstein, The Global Expansion of AI Surveillance, Vol. 17, Carnegie Endowment for International Peace Washington, DC, 2019.
[20] E.M. Abdali, A.W. Hanniche, M. Pelcat, J.-P. Diguet, F. Berry, Hardware acceleration of the tracking learning detection (TLD) algorithm on FPGA, in: Procs. of the 11th Int. Conf. on Distributed Smart Cameras, 2017, pp. 180–185.
[21] S.A. Velastin, R. Fernández, J.E. Espinosa, A. Bay, Detecting, tracking and counting people getting on/off a metropolitan train using a standard video camera, Sensors 20 (21) (2020) 6251.
[22] R. Sun, X. Wang, X. Yan, Robust visual tracking based on convolutional neural network with extreme learning machine, Multimedia Tools Appl. 78 (6) (2019) 7543–7562.
[23] D. Vorobjov, I. Zakharava, R. Bohush, S. Ablameyko, An effective object detection algorithm for high resolution video by using convolutional neural network, in: Int. Symposium on Neural Networks, Springer, 2018, pp. 503–510.
[24] D.H. Ye, J. Li, Q. Chen, J. Wachs, C. Bouman, Deep learning for moving object detection and tracking from a single camera in unmanned aerial vehicles (UAVs), Electron. Imaging 2018 (10) (2018) 466–1.
[25] G. Yildirim, S. Süsstrunk, FASA: fast, accurate, and size-aware salient object detection, in: Asian Conf. on Computer Vision, Springer, 2014, pp. 514–528.
[26] C. Kim, J. Lee, T. Han, Y.-M. Kim, A hybrid framework combining background subtraction and deep neural networks for rapid person detection, J. Big Data 5 (1) (2018) 1–24.
[27] N. Martins, J.M. Cruz, T. Cruz, P.H. Abreu, Adversarial machine learning applied to intrusion and malware scenarios: a systematic review, IEEE Access 8 (2020) 35403–35419.
[28] V. Kanimozhi, T.P. Jacob, Artificial intelligence based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing, in: Int Conf. on Communication and Signal Processing, ICCSP, IEEE, 2019, pp. 0033–0036.
[29] J. Yu, H. Choi, YOLO MDE: Object detection with monocular depth estimation, Electronics 11 (1) (2021) 76.
[30] W. Zhihuan, C. Xiangning, G. Yongming, L. Yuntao, Rapid target detection in high resolution remote sensing images using yolo model, Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42 (3) (2018).
[31] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Procs. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[32] S. Song, Y. Li, Q. Huang, G. Li, A new real-time detection and tracking method in videos for small target traffic signs, Appl. Sci. 11 (7) (2021) 3061.
[33] J. Shin, H. Kim, D. Kim, J. Paik, Fast and robust object tracking using tracking failure detection in kernelized correlation filter, Appl. Sci. 10 (2) (2020) 713.
[34] S. Yadav, S. Payandeh, Understanding tracking methodology of kernelized correlation filter, in: IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conf., IEMCON, IEEE, 2018, pp. 1330–1336.
[35] L.T.H. Phuc, H. Jeon, N.T.N. Truong, J.J. Hak, Applying the haar-cascade algorithm for detecting safety equipment in safety management systems for multiple working environments, Electronics 8 (10) (2019) 1079.
[36] D.K. Ulfa, D.H. Widyantoro, Implementation of haar cascade classifier for motorcycle detection, in: IEEE Int. Conf. on Cybernetics and Computational Intelligence, CyberneticsCom, IEEE, 2017, pp. 39–44.
[37] L. Cuimei, Q. Zhiliang, J. Nan, W. Jianhua, Human face detection algorithm via haar cascade classifier combined with three additional classifiers, in: 13th IEEE Int. Conf. on Electronic Measurement & Instruments, ICEMI, IEEE, 2017, pp. 483–487.
[38] J. Cruz, E. Shiguemori, L. Guimaraes, A comparison of haar-like, LBP and HOG approaches to concrete and asphalt runway detection in high resolution imagery, Int. Sci. J. Comp. Int. Sci 6 (61) (2015) 121–1363.
[39] H. Zhu, X. Yan, H. Tang, Y. Chang, B. Li, X. Yuan, Moving object detection with deep CNNs, IEEE Access 8 (2020) 29729–29741.

[40] J. Jain, Artificial intelligence in the cyber security environment, in: Artificial Intelligence and Data Mining Approaches in Security Frameworks, Wiley Online Library, 2021, pp. 101–117.
[41] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Procs of the IEEE Conf. on Computer Vision and Pattern Recognition, 2014, pp. 580–587.
[42] J. Hosang, R. Benenson, P. Dollár, B. Schiele, What makes for effective detection proposals? IEEE Trans. Pattern Anal. Mach. Intell. 38 (4) (2015) 814–830.
[43] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Adv. Neural Inf. Process. Syst. 28 (2015).
[44] K. Kuan, G. Manek, J. Lin, Y. Fang, V. Chandrasekhar, Region average pooling for context-aware object detection, in: 2017 IEEE Int. Conf. on Image Processing, ICIP, IEEE, 2017, pp. 1347–1351.
[45] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C. Van Essen, A.A. Awwal, V.K. Asari, A state-of-the-art survey on deep learning theory and architectures, Electronics 8 (3) (2019) 292.
[46] C.B. Murthy, M.F. Hashmi, N.D. Bokde, Z.W. Geem, Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review, Appl. Sci. 10 (9) (2020) 3280.
[47] S. Roheda, B.S. Riggan, H. Krim, L. Dai, Cross-modality distillation: A case for conditional generative adversarial networks, in: 2018 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2018, pp. 2926–2930.
[48] J. Peng, H. Wang, F. Xu, X. Fu, Cross domain knowledge learning with dual-branch adversarial network for vehicle re-identification, Neurocomputing 401 (2020) 133–144.
[49] T. Yu, L. Wang, H. Gu, S. Xiang, C. Pan, Deep generative video prediction, Pattern Recognit. Lett. 110 (2018) 58–65.
[50] Y. Du, Y. Yan, S. Chen, Y. Hua, Object-adaptive LSTM network for real-time visual tracking with adversarial data augmentation, Neurocomputing 384 (2020) 67–83.
[51] W. Liu, R. Yao, G. Qiu, A physics based generative adversarial network for single image defogging, Image Vis. Comput. 92 (2019) 103815.
[52] W. Liu, Z. Luo, S. Li, Improving deep ensemble vehicle classification by using selected adversarial samples, Knowl.-Based Syst. 160 (2018) 167–175.
[53] P.S. Bhat, A. Dharani, Methodologies in face recognition for surveillance, in: 3rd Int. Conf. on Computational Systems and Information Technology for Sustainable Solutions, CSITSS, IEEE, 2018, pp. 105–113.
[54] Z. Lin, Y. Shi, Z. Xue, Idsgan: Generative adversarial networks for attack generation against intrusion detection, in: Pacific-Asia Conf. on Knowledge Discovery and Data Mining, Springer, 2022, pp. 79–91.
[55] Y. Zhang, Y. Bai, M. Ding, B. Ghanem, Multi-task generative adversarial network for detecting small objects in the wild, Int. J. Comput. Vision 128 (6) (2020) 1810–1828.
[56] C. Peng, N. Wang, J. Li, X. Gao, Soft semantic representation for cross-domain face recognition, IEEE Trans. Inf. Forensics Secur. 16 (2020) 346–360.
[57] A. Aggarwal, R. Rathore, P. Chattopadhyay, L. Wang, EPD-net: A GAN-based architecture for face de-identification from images, in: IEEE Int. IOT, Electronics and Mechatronics Conf., IEMTRONICS, IEEE, 2020, pp. 1–7.
[58] B. Wang, F. Zou, X. Liu, New algorithm to generate the adversarial example of image, Optik 207 (2020) 164477.
[59] Y. Lee, J. Yun, Y. Hong, J. Lee, M. Jeon, Accurate license plate recognition and super-resolution using a generative adversarial networks on traffic surveillance video, in: 2018 IEEE Int. Conf. on Consumer Electronics-Asia, ICCE-Asia, IEEE, 2018, pp. 1–4.
[60] P.D. Ciampa, B. Nagel, AGILE paradigm: the next generation collaborative MDO for the development of aeronautical systems, Prog. Aerosp. Sci. 119 (2020) 100643.
[61] K. Kalirajan, M. Sudha, Moving object detection for video surveillance, Sci. World J. 2015 (2015).
[62] M. Simao, P. Neto, O. Gibaru, Improving novelty detection with generative adversarial networks on hand gesture data, Neurocomputing 358 (2019) 437–445.
[63] C. Donahue, J. McAuley, M. Puckette, Adversarial audio synthesis, 2018, arXiv preprint arXiv:1802.04208.
[64] G. Zhang, Y. Pan, L. Zhang, R.L.K. Tiong, Cross-scale generative adversarial network for crowd density estimation from images, Eng. Appl. Artif. Intell. 94 (2020) 103777.
[65] X. Chen, H. Xie, D. Zou, G.-J. Hwang, Application and theory gaps during the rise of artificial intelligence in education, Comput. Educ. : Artif. Intell. 1 (2020) 100002.
[66] X. Zhen, S. Fei, Y. Wang, W. Du, A visual object tracking algorithm based on improved TLD, Algorithms 13 (1) (2020) 15.
[67] A. Bathija, G. Sharma, Visual object detection and tracking using yolo and sort, Int. J. Eng. Res. Technol. 8 (11) (2019).
[68] Y. Li, S. Li, H. Du, L. Chen, D. Zhang, Y. Li, YOLO-ACN: Focusing on small target and occluded object detection, IEEE Access 8 (2020) 227288–227303.
[69] S.-E. Ryu, K.-Y. Chung, Detection model of occluded object based on YOLO using hard-example mining and augmentation policy optimization, Appl. Sci. 11 (15) (2021) 7093.
[70] Y. Park, L.M. Dang, S. Lee, D. Han, H. Moon, Multiple object tracking in deep learning approaches: A survey, Electronics 10 (19) (2021) 2406.
[71] G. Khan, Z. Tariq, M.U.G. Khan, P. Mazzeo, S. Ramakrishnan, P. Spagnolo, Multi-person tracking based on faster R-CNN and deep appearance features, in: Visual Object Tracking with Deep Neural Networks, IntechOpen London, UK, 2019, pp. 1–23.
[72] F. Feng, B. Shen, H. Liu, Visual object tracking: In the simultaneous presence of scale variation and occlusion, Syst. Sci. Control Eng. 6 (1) (2018) 456–466.
[73] Y. Yuan, J. Chu, L. Leng, J. Miao, B.-G. Kim, A scale-adaptive object-tracking algorithm with occlusion detection, EURASIP J. Image Video Process. 2020 (1) (2020) 1–15.
[74] D. Chen, L. Yue, X. Chang, M. Xu, T. Jia, NM-GAN: Noise-modulated generative adversarial network for video anomaly detection, Pattern Recognit. 116 (2021) 107969.
[75] Z. Ruiqiang, Z. Yu, J. Xin, Optimization of small object detection based on generative adversarial networks, in: E3S Web of Conferences, 245, EDP Sciences, 2021, p. 03062.
[76] X. Cheng, C. Song, Y. Gu, B. Chen, Learning attention for object tracking with adversarial learning network, EURASIP J. Image Video Process. 2020 (1) (2020) 1–21.
[77] W. Huang, M. Huang, Y. Zhang, Detection of traffic signs based on combination of GAN and faster-RCNN, in: J. Phys.: Conf. Ser., 1069, (1) IOP Publishing, 2018, 012159.
[78] C.D. Prakash, L.J. Karam, It GAN DO better: GAN-based detection of objects on images with varying quality, IEEE Trans. Image Process. 30 (2021) 9220–9230.
[79] X. Huang, D. Kroening, W. Ruan, J. Sharp, Y. Sun, E. Thamo, M. Wu, X. Yi, A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability, Comp. Sci. Rev. 37 (2020) 100270.
[80] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, Y. Zheng, Recent progress on generative adversarial networks (GANs): A survey, IEEE Access 7 (2019) 36322–36333.
[81] Z. Ullah, F. Al-Turjman, L. Mostarda, R. Gagliardi, Applications of artificial intelligence and machine learning in smart cities, Comput. Commun. 154 (2020) 313–323.
[82] TensorFlow, Adversarial example, 2022, https://www.tensorflow.org/tutorials/generative/adversarial-fgsm. (Accessed 1 December 2022).
[83] Q. Wang, L. Zhang, Y. Li, K. Kpalma, Overview of deep-learning based methods for salient object detection in videos, Pattern Recognit. 104 (2020) 107340.
[84] F. Marra, D. Gragnaniello, L. Verdoliva, On the vulnerability of deep learning to adversarial attacks for camera model identification, Signal Process., Image Commun. 65 (2018) 240–248.
[85] W. Brendel, J. Rauber, M. Bethge, Decision-based adversarial attacks: Reliable attacks against black-box machine learning models, 2017, arXiv preprint arXiv:1712.04248.
[86] G. Appiah, J. Amankwah-Amoah, Y.-L. Liu, Organizational architecture, resilience, and cyberattacks, IEEE Trans. Eng. Manage. (2020).
[87] G.R. Chandra, B.K. Sharma, I.A. Liaqat, UAE's strategy towards most cyber resilient nation, Int. J. Innov. Technol. Explor. Eng. (IJITEE) 8 (12) (2019) 2803–2809.
[88] J. Li, H. Yang, L. Chen, J. Li, C. Zhi, An end-to-end generative adversarial network for crowd counting under complicated scenes, in: 2017 IEEE Int. Symposium on Broadband Multimedia Systems and Broadcasting, IEEE, 2017, pp. 1–4.
[89] T. Huang, V. Menkovski, Y. Pei, M. Pechenizkiy, Bridging the performance gap between fgsm and pgd adversarial training, 2020, arXiv preprint arXiv:2011.05157.
[90] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.
[91] A. Senthil Murugan, K. Suganya Devi, A. Sivaranjani, P. Srinivasan, A study on various methods used for video summarization and moving object detection for video surveillance applications, Multimedia Tools Appl. 77 (18) (2018) 23273–23290.
[92] S. Qiu, Q. Liu, S. Zhou, C. Wu, Review of artificial intelligence adversarial attack and defense technologies, Appl. Sci. 9 (5) (2019) 909.
[93] N. Kalbo, Y. Mirsky, A. Shabtai, Y. Elovici, The security of IP-based video surveillance systems, Sensors 20 (17) (2020) 4806.
[94] F.V. Massoli, F. Carrara, G. Amato, F. Falchi, Detection of face recognition adversarial attacks, Comput. Vis. Image Underst. 202 (2021) 103103.

[95] S.-M. Moosavi-Dezfooli, A. Fawzi, P. Frossard, Deepfool: a simple and accurate method to fool deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2574–2582.

[96] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, P. Frossard, Universal adversarial perturbations, in: Procs. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2017, pp. 1765–1773.

[97] N. Carlini, D. Wagner, Towards evaluating the robustness of neural networks, in: IEEE Symposium on Security and Privacy, Ieee, 2017, pp. 39–57.

[98] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z.B. Celik, A. Swami, The limitations of deep learning in adversarial settings, in: IEEE European Symposium on Security and Privacy, EuroS&P, IEEE, 2016 pp. 372–387.

[99] L. Engstrom, B. Tran, D. Tsipras, L. Schmidt, A. Madry, A rotation and a translation suffice: Fooling cnns with simple transformations, 2018.

[100] O. Poursaeed, I. Katsman, B. Gao, S. Belongie, Generative adversarial perturbations, in: Procs of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018, pp. 4422–4431.

[101] C. Xiao, J.-Y. Zhu, B. Li, W. He, M. Liu, D. Song, Spatially transformed adversarial examples, 2018, arXiv preprint arXiv:1801.02612.

[102] C. Feng, Y. Shang, H. Jincheng, X. Bo, Few features attack to fool machine learning models through mask-based GAN, in: 2020 Inter. Joint Conf. on Neural Networks, IJCNN, IEEE, 2020, pp. 1–7.

[103] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: European Conf. on Computer Vision, Springer, 2014, pp. 740–755.

[104] S. Yang, P. Luo, C.-C. Loy, X. Tang, Wider face: A face detection benchmark, in: Procs. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 5525–5533.

[105] Google APIs, Open images dataset V7 and extensions, 2022, https://storage.googleapis.com/openimages/web/index.html. (Accessed 5 December 2022).