IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 July 2023, accepted 19 July 2023, date of publication 28 July 2023, date of current version 8 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3299878

SURVEY

From Pixel to Peril: Investigating Adversarial Attacks on Aerial Imagery Through Comprehensive Review and Prospective Trajectories

SYED M. KAZAM ABBAS KAZMI, NAYYER AAFAQ, MANSOOR AHMED KHAN, MOHSIN KHALIL, AND AMMAR SALEEM

College of Aeronautical Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan

Corresponding author: Nayyer Aafaq (naafaq@cae.nust.edu.pk)

This work was supported by the National University of Science and Technology (NUST), Islamabad, Pakistan.

ABSTRACT Deep models' feature learning capabilities have gained traction in recent years, driving significant progress in various Artificial Intelligence (AI) domains. The use of Deep Neural Networks (DNNs) has expanded the scope of Computer Vision (CV) and revealed their vulnerability to deliberate adversarial attacks. These attacks involve the careful introduction of perturbations crafted through complex optimization problems. Exploiting vulnerabilities in advanced deep neural network algorithms present security concerns, particularly in practical applications with high stakes like unmanned aerial vehicles (UAVs) and satellite imagery in computer vision. Adversarial attacks, both in digital and physical dimensions, pose a serious threat in the field. This research provides a comprehensive examination of state-of-the-art adversarial attacks specific to aerial imagery using autonomous platforms such as UAVs and satellites. This review covers fundamental concepts, techniques, and the latest advancements, identifying research gaps and suggesting future directions. It aims to deepen researchers' understanding of the challenges and threats related to adversarial attacks on aerial imagery, serving as a valuable resource to guide future research and enhance the security of computer vision systems in aerial environments.

INDEX TERMS Aerial imagery, adversarial attacks, adversarial perturbations, autonomous systems, remote sensing, AI-applications.

I. INTRODUCTION

Owing to the ongoing technological advancements in aerial imagery, its viability has experienced a notable upsurge in obtaining a substantial quantity of exceptional and high-resolution aerial images. The utilization of technology has presented significant prospects in the domain of imagery [1], [2], [3], [4] and remote-sensing [5], [6], [7]. These advancements have contributed to the creation of several crucial applications including critical infrastructure resilience [8], defense systems [9] and public safety [10]. Deep learning models have significantly contributed to most state-of-the-art methodologies where they demonstrated noteworthy success across various domains [11], [12], [13], [14], [15], [16], [17]. Typically, these models obtain a hierarchical representation of attributes and features. Moreover, empirical studies have indicated that the effectiveness of DNNs exhibits a positive correlation in terms of their architecture. Computer Vision (CV) has become increasingly important in various applications. Image segmentation [18], [19] and object detection [20], [21], [22] are among the most commonly used CV techniques that play a crucial role in these applications.

In 2014, Szegedy et al. [23] made a novel discovery, which they referred to as adversarial samples. This phenomenon has the capacity to fool deep neural networks (DNN),

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

leading to erroneous predictions and significant deterioration in the performance of state-of-the-art deep learning techniques. Adversarial samples manifest as imperceptible perturbations or noise within images, which have the potential to introduce bias in Convolutional Neural Networks (CNNs). This pioneering research has illustrated the vulnerability of Deep Neural Networks. Subsequent to this work, various approaches have been investigated by researchers to produce adversarial examples, as documented in the literature [24], [25], [26], [27], [28]. In response to the emergence of security breaches, scholars have endeavored to safeguard deep neural networks against such vulnerabilities by devising defensive techniques and formulating countermeasures, which are commonly referred to as adversarial defenses [29], [30], [31], [32]. Whereby, defense mechanisms have been proposed at the same time work on adversarial training of DNN also caught attention. These two lines of work surged and went side by side providing a defense mechanism for each attack, researchers have enhanced Deep Neural Networks (DNNs) by incorporating resistive phenomena during the training phase through adversarial training techniques [33], [34], [35]. Therefore, deep neural networks (DNNs) have undergone an evolutionary process as depicted in Figure 1, involving adversarial attacks, countermeasures, and ultimately adversarial training, in order to enhance their resilience against such attacks. Adversarial attacks are utilized during the deployment stage of the deep neural network (DNN) model life cycle. The occurrence of such attacks has instigated the creation of adversarial defense mechanisms, whereby model's training phase is embedded with adversarial training.

Machine learning models can be subjected to adversarial attacks in two distinct ways, namely digital [36] and physical attacks. Evasion attacks [37] and poisoning attacks [38] are frequently observed in the context of digital attacks. Evasion attacks pertain to the act of manipulating input data in order to fool the model's predictions, whereas poisoning attacks involve the introduction of malevolent data during the training process to undermine the model's integrity. The aforementioned attacks have the potential to result in mis-classifications and erroneous patterns acquired by the model. Conversely, physical attacks [27] are centered on the manipulation of tangible entities in the physical realm. Adversarial images are generated through the introduction of slight modifications or patterns to physical objects, with the aim of misleading computer vision systems. Adversarial inputs pertaining to the manipulation of sensor readings or the introduction of noise to impede the decision-making process of the model. Both forms of attacks pose a challenge to the durability and dependability of machine learning models, necessitating continued research into defense mechanisms to bolster their resilience against these adversarial threats.

Distinguishing between digital and physical adversarial attacks in computer vision can be achieved based on the timing of their execution. Digital attacks are defined as attacks that are carried out solely within the digital realm, occurring only after the camera has captured an image. Conversely, physical attacks involve the manipulation or alteration of tangible objects prior to their capture by the camera. Although DNN models have been successfully attacked digitally by various methods [35], [39], [40], [41]; implementing these attacks in the physical world presents a significantly greater challenge. Although digital perturbations are generally both universal and imperceptible, accurately recording them with sensors can be challenging. The presence of these drawbacks serves as a driving force for researchers to investigate novel, practical methods of carrying out attacks in actual settings. Notwithstanding, physical attacks are associated with significant difficulties [42].

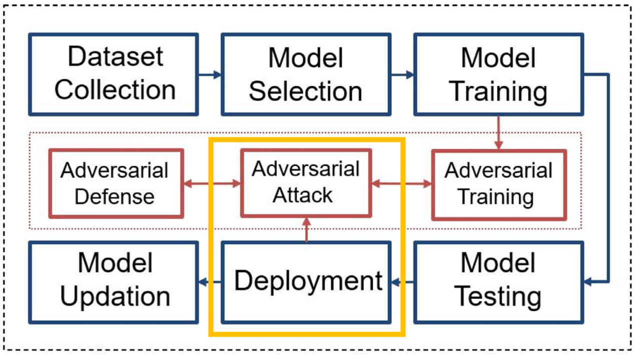

FIGURE 1. The evolutionary progression of adversarial and counter techniques during DNN's life-cycle. Illustration explains the instance of attacks and countermeasures; where attacks are carried out during the model deployment stage and adversarial training is incorporated during the training phase of DNN model. The region surrounded by the yellow-box emphasizes the review's central theme.

The adversary instance in practical scenarios must possess the capacity to endure the impacts of imaging devices, which are predominantly determined by the sensors and processing devices. Furthermore, the physical adversary must exhibit resilience towards multiple modifications, encompassing fluctuations in shot distance, perspective, and portrayal. The adversary's physical manifestation should possess an unobtrusive quality. The manipulation of digital images occurs at the level of individual pixels, rendering them challenging to detect. Nevertheless, the concealment of physical attacks poses a challenge. Wei et al. [42] have demonstrated the utilization of physical and digital attacks at different stages of a conventional CV pipeline. This is exemplified through the use of a traffic-sign example, as depicted in Figure 2.

Prior research has predominantly concentrated on terrestrial-based scenarios and applications, such as person detection [43], [44], [45], [46], [47], facial recognition [48], [49], [50], [51], security surveillance [52], [53], [54] and autonomous vehicles [55], [56], [57], [58], [59]. The literature preceding this study has provided broader insights into the potential of adversarial attacks, as demonstrated by the works of Wei et al. [42], Wang et al. [60], Akhtar and Mian [61], and Akhtar et al. [62]. The focus of this study is solely on adversarial attacks that are directed towards aerial-based imagery captured from larger distances, specifically those obtained through the utilization of drones and satellites, as opposed to scenarios and use cases that take place on the ground.

A. MOTIVATION FOR REVIEW PUBLICATION

The field of computer vision has witnessed significant maturation in this direction in recent years. Since the inception of adversarial concept by [23], numerous surveys and review papers on adversarial attacks have been published in both digital and physical disciplines of adversarial attacks as well as defenses. The majority of the reviews exhibit a wider scope of applicability [42], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], some are tailored to specific computer vision tasks [72], [73], [74], [75], while others are geared towards particular problems [76], [77], [78], or type of attacks [79]. Based on already published reviews and cited articles, it is found that this study exhibits several distinctions from the previously conducted reviews. This review article possesses a distinctive quality as it serves as a follow-up to our previous work [80], which happens to be the first-ever and only peer-reviewed literature survey conducted on this specific topic. Drawing upon the seminal work of [80] and subsequent research pursuits, we have provided more refined conceptualizations of the specialized vocabulary pertaining to this rapidly evolving field. As a consequence, a more cohesive and systematic review framework has been established, wherein we present succinct discussions grounded on the contemporary comprehension of terminologies within the research community. Additionally, our attention is directed towards scholarly articles that have undergone peer-review and have been published in esteemed research outlets. By concentrating on the primary contributions, we have been able to offer a more lucid perspective of this area for academicians and researchers. Moreover, the necessity of conducting research in this domain is underscored by several significant factors and driven by various motivations:

- The utilization of aerial imagery is prevalent in crucial domains such as military defense, monitoring, emergency management, and infrastructure strategizing. Comprehending and mitigating the susceptibilities and hazards linked with adversarial attacks is imperative in upholding the dependability, confidentiality, and efficacy of said applications. Moreover, the potential repercussions of adversarial attacks on aerial imagery can be substantial in practical applications.
- The manipulation or deception of machine learning models that analyze aerial imagery has the potential to cause disruption or compromise of critical operations, thereby posing a threat to human lives and property. Through the implementation of research endeavors in this domain, proficient professionals can formulate sturdy defensive strategies to safeguard against such malevolent attacks, thereby augmenting the tenacity of aerial imagery systems.
- In general, investigation and dissemination of knowledge regarding vulnerabilities of aerial imaging are imperative for guaranteeing the dependability and safety of its implementations in critical and diverse sectors.

B. RESEARCH CONTRIBUTIONS

This study presents a noteworthy research contribution through a comprehensive examination and analysis of the landscape of adversarial attacks that are directed toward aerial imagery. The manuscript encompasses multiple pivotal aspects, furnishing a thorough comprehension of the topic. To the best of our knowledge, it is the first-ever effort till date consolidating adversarial attacks targeting aerial imagery. This paper consolidates the various components of research conducted by researchers, thereby establishing a cohesive framework for researchers interested in pursuing this domain of study. Our endeavor involves addressing the deficiencies and presenting a strategic plan that would be advantageous to scholars aspiring to undertake research on adversarial attacks focusing on imagery through satellites, drones, and UAVs. This review has made specific contributions, which are outlined as follows:

- **Provision of Terminologies, Key-Concepts and Definitions:** To insure cohesion in literature, it is considered important to furnish precise depictions of the recurring and particular terminologies that are present in this paper. The present study also provides an exposition of the frequently used terminologies in the pertinent literature, as construed by the research community. The section II provides an overview of key definitions, terminologies, and concepts.
- **Emphasis on Vulnerabilities of Deep Neural Networks (DNNs):** Deep Neural Networks (DNNs) have become ubiquitous in a variety of industries, including healthcare, finance, and transportation. These models have been developed to learn intricate patterns and make precise predictions on new data, which has resulted in their pervasive acceptance. DNNs are vulnerable to an array of adversarial attacks and understanding these vulnerabilities is essential for constructing improved models and securing DNN-based systems. In this context, the section III will address a few of the most prevalent vulnerabilities of DNNs to adversarial attacks and emphasize certain approaches utilized to take advantage of these vulnerabilities.
- **Perspective-Based Distinction of Adversarial Attack Methodologies:** This study provides a succinct overview of the two adversarial attack domains, digital and physical, that are utilized to impede the functionality of machine learning models. The emergence of innovative technologies and methodologies in machine learning models has resulted in concurrent progress in attack techniques, which present multifaceted risks. The section IV presents a succinct and consolidated analysis of both attack domains. This section will also discuss

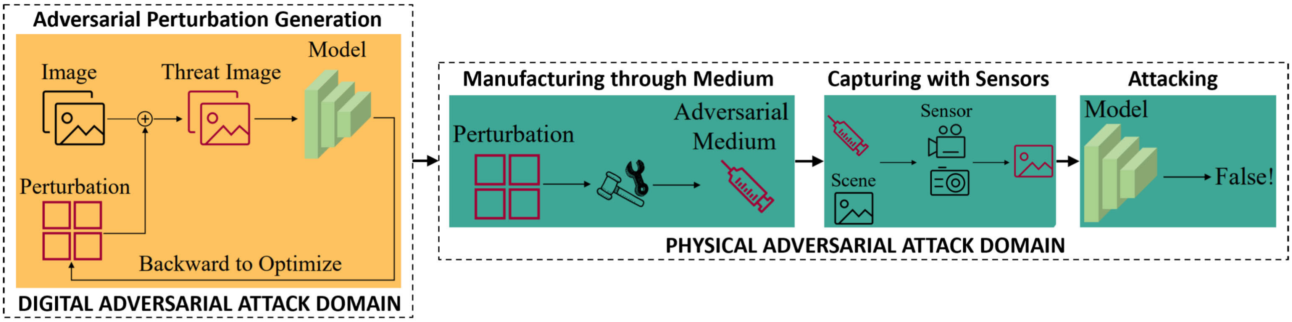

FIGURE 2. An illustration of traditional framework of digital adversarial attacks leading to physical manufacturing and incorporating perturbations in physical space [63].

fundamentals which are the most impactful contributions serving as a source of influence for a multitude of the latest techniques.

- **Insights of Adversarial Attacks on Computer Vision Tasks pertaining to Aerial Imagery:** The authors have comprehensively incorporated a variety of viewpoints related to the field of adversarial attacks, with a specific focus on their implications for computer vision applications including object detection, image segmentation, and image classification; that are linked to aerial imagery. This paper showcases the vulnerabilities and challenges posed by adversarial attacks in aerial image analysis systems by integrating existing literature. Section V delves into the various techniques and methodologies proposed for generating adversarial examples in aerial imagery, providing a thorough evaluation of their effectiveness and limitations.
- **Constraints, Challenges and Future Directions:** The present analysis examines the constraints of prior studies and suggests potential directions for future inquiry in this domain. The authors have duly acknowledged the limitations and challenges associated with this research domain. The constraints are related to the insufficiency of annotated adversarial datasets that are specifically designed for aerial imagery, the complexity of developing effective defense mechanisms, and the need for standardized evaluation metrics to assess the impact of attacks. Furthermore, the authors highlight the importance of collaborative endeavors and the adoption of shared standards among researchers to facilitate the progress and evaluation of protective measures. Section VI highlights the limitations and way forward is discussed in Section VII.

The remaining article is organized in the following manner. Section II of this paper presents definitions and concepts pertaining to recurring terminologies of this field. In section III, we will examine some of the primary susceptibilities of deep neural networks (DNNs) to adversarial attacks and highlight specific methods employed to exploit these vulnerabilities. Section IV provides a concise and integrated examination of the two attack domains. In Section V, an in-depth analysis is presented on the different techniques and methodologies proposed so far for producing adversarial examples in aerial imagery. The effectiveness and limitations of these approaches are thoroughly evaluated. Section VI and VII presents a thorough examination of constraints, recommendations, and prospects for future research. In conclusion, this review is summarized in Section VIII.

II. TERMINOLOGIES, KEY-CONCEPTS, AND DEFINITIONS

In order to insure cohesion in literature, it is considered important to furnish precise depictions of the recurring and particular terminologies that are present in this paper. The present section presents interpretations of recurring terms in related literature, generally perceived by the research-community:

- Any change that is introduced into an original input of DNN model, resulting in an inaccurate prediction, is commonly referred to as an **adversarial perturbation**. Frequently, it exhibits a configuration akin to that of minimal & subtle noise.
- An **adversarial example** or image refers to an image that has been intentionally altered to yield an erroneous prediction by a model. Typically, the process involves introducing disruptive noise to an authentic image in order to calculate it. The antithesis of an adversarial example is frequently characterized by a unique image or original visual representation.
- The term **adversarial medium** is used to describe an object that contains an adversarial perturbation. The presence of an adversarial medium is a prerequisite for executing an attack in the physical domain.
- The phenomenon of utilizing adversarial samples to launch an attack on DNN-based models is commonly referred to as **adversarial attacks**.
- The concept of a **digital attack** assumes that the attacker possesses adequate knowledge of the actual digital input of the model. The attacker in the digital domain has the ability to manipulate the pixel values of target image.
- The availability of the digitized version of the model's input is not always a prerequisite for the occurrence of **physical attacks**. The perpetuation of these attacks

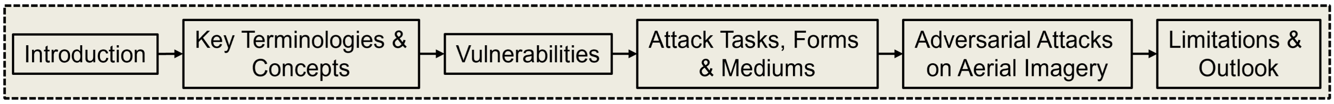
Introduction
Key Terminologies & Concepts
Vulnerabilities
Attack Tasks, Forms & Mediums
Adversarial Attacks on Aerial Imagery
Limitations & Outlook

FIGURE 3. Outlook of survey's framework.

is predominantly executed by means of modifying the intended image via tangible means, such as affixing adhesive labels or positioning patches onto or in the vicinity of the intended image.

- The term **white-box** adversarial attacks pertains to a situation wherein an attacker possesses full awareness of the target model's structure, parameters, and training data. In the given context, the attacker is capable of obtaining unrestricted access to all internal information about the model and subsequently utilizing it to construct adversarial examples.
- **Black-box** attacks pertain to a situation in which the attacker possesses prohibited or negligible accessibility to the intricacies of the target model, which includes its structure, settings, or data used for training. Within this context, the attacker is limited to engaging with the target model solely through the provision of input samples and subsequent observation of the outputs that correspond to those samples. The methodology of black-box attacks frequently encompasses transfer attacks, which entail the training of a substitute model on a distinct dataset by the attacker, who subsequently utilizes it as a proxy to produce adversarial examples. It is commonly assumed that the transferability of adversarial examples to the target model is facilitated by the shared output-input association between the substitute model and the target model. Black-box attacks are typically considered more challenging due to the attacker's limited knowledge regarding the target model.
- **Grey-box** adversarial attacks fall in the intermediate position between white-box and black-box attacks. The grey-box scenario refers to a situation where the adversary possesses only partial information or limited access to the target model and its parameters, which is insufficient compared to the comprehensive knowledge that is available in a white-box attack. Incomplete knowledge may encompass details such as the structure of the model, while excluding specific model parameters or training data. Grey-box attacks involve situations where the attacker possesses a surrogate model or a substitute model that provides an approximation of the target model's behavior. The surrogate model is trained on a distinct dataset and exhibits resemblances with the target model. The utilization of surrogate models by the attacker enables the creation of adversarial examples that can be transferred to the target model, thereby taking advantage of shared characteristics in the way they function.
- Within the framework of adversarial attacks, the term **target image** denotes a particular image that the attacker endeavors to alter or mislead the target model into an incorrect prediction. The selection of the target image is a deliberate act by the attacker, often driven by a particular aim or purpose. The process of choosing the target image holds significant importance in the context of adversarial attacks, as it plays a pivotal role in defining the precise objective of the attacker's manipulations.
- Within the discipline of adversarial attacks, the term **target model** pertains to a particular machine learning model or system that an attacker endeavors to manipulate or deceive. The commonly held assumption is that the target model is a carefully trained model which has been implemented for practical purposes. The aim of the attacker is to circumvent or deceive the model's protective measures and capitalize on its vulnerabilities to influence the model's predictive process.
- **Target label** denotes the intended misclassification assigned to a malicious entity's sample. Stated differently, it denotes the classification that the attacker aims for the targeted model to anticipate. The aforementioned notion assumes particular significance in the context of targeted attacks, wherein the objective is to influence the model's output towards a specific category or identification as per the attacker's preference.
- The main objective of an attacker in **targeted attacks** is to influence the model's prediction by causing it to recognize a preconceived target class or label, regardless of the authentic label associated with the input data. The attacker employs perturbations on the input data, exploiting the model's vulnerabilities to manipulate its decision-making mechanism.
- **Untargeted attacks** are a type of attacks where the attacker intentionally selects adversarial instances to cause the target model to produce inaccurate predictions that deviate from the actual data. The primary aim of the adversary is to increase the prediction error of the model, rather than focusing on a particular target label.
- The utilization of **adversarial training** is a technique implemented to enhance the resilience of machine learning models in the face of adversarial attacks. During the training phase, the integration of adversarial examples is implemented to improve the model's resilience against adversarial attacks.
- The domain of **adversarial defense** is currently a subject of active research, given the persistent emergence of novel attack strategies by attackers and the corresponding efforts by defenders to enhance the resilience

of their models. The objective of adversarial defense is to develop machine learning models that exhibit not only high accuracy and efficiency but also robustness against adversarial attacks, thereby guaranteeing their dependability and safety in practical scenarios.

- **Physical world attacks**, also known as real-world attacks, refer to attacks that occur in the tangible domain rather than the digital or virtual domain.

III. VULNERABILITIES OF DEEP NEURAL NETWORKS

Deep Neural Networks (DNNs) have become an omnipresent tool in many industries, ranging from healthcare to finance to transportation. These models are designed to learn complex patterns and make accurate predictions on new data, which has led to their widespread adoption. DNNs are not impervious to attacks, though, and new studies have revealed that they are susceptible to a variety of adversarial attacks [81]. Adversarial attacks are intended to cause the model to anticipate incorrectly, frequently by introducing undetectable perturbation to the input data [26]. Understanding these vulnerabilities is crucial to building more robust models and ensuring the security of DNN-based systems. In this context, this section will discuss some of the most common vulnerabilities of DNNs to adversarial attacks [82], and highlight some of the techniques being used to exploit these vulnerabilities.

- Adversarial attacks often use **gradient-based optimization** techniques to find the most effective perturbations. These methods identify the ways that the model is most susceptible to input changes by using the variations of the DNN's loss function. Examples of methods that take benefit of such weakness include the Projected Gradient Descent (PGD) [35] and the Fast Gradient Sign Method (FGSM) [26].
- DNNs are trained on large datasets with millions of examples. These datasets often contain **non-robust features** that are highly correlated with the class labels, but not necessarily indicative of the underlying data distribution. Adversarial attacks take advantage of these non-robust features to create perturbations that cause the DNN to make incorrect predictions. For example, in [23], the authors show that DNNs trained on ImageNet dataset are vulnerable to adversarial attacks that exploit non-robust features such as high-frequency components.
- **High model complexity** DNNs are more susceptible to overfitting to training data, which can reduce their resistance to adversarial attacks. This overfitting is exploited by adversarial attacks to produce perturbations that lead to inaccurate predictions from the model. For instance, in [31], the authors demonstrate that DNNs with a high complexity are more susceptible to adversarial attacks than models with a lower complexity.
- DNNs are **sensitive to small perturbations** in the input data. Adversarial attacks take advantage of this vulnerability by adding small, imperceptible perturbations to the input data, which can cause the model to misclassify the input. For example, in [83], the authors show that even small perturbations can cause DNNs to misclassify images.
- To target another model, it is often sufficient to use adversarial instances originally created for a different model. This is a potential weakness in DNNs and is referred to as **transferability**. The authors of [28], for instance, demonstrate that adversarial examples created for one model can be used to effectively target other models.
- DNNs are often trained on **limited amounts of data**, which can make them less robust to adversarial attacks. For example, in [40], the authors show that DNNs trained on a small subset of the MNIST dataset are more vulnerable to adversarial attacks than DNNs trained on the full dataset.
- Some components of DNNs, such as activation functions, are **non-differentiable**. This can make it difficult to compute gradients and find effective adversarial examples. For example, in [84], the authors show that DNNs with ReLU activation functions are more vulnerable to adversarial attacks than DNNs with sigmoid activation functions.
- While **adversarial training** can make DNNs more resistant to adversarial attacks, it is not a solution. Adversarial training can result in overfitting and may not generalize effectively to new adversarial examples. For instance, the authors of [27] demonstrate that adversarial training can enhance the robustness of DNNs, but it is still possible to generate adversarial examples that deceive these models.
- DNNs are typically trained on a **limited input domain**, such as images or text. Adversarial attacks that exploit properties of the input domain, such as the spatial structure of images, may be particularly effective. For example, in [85], the authors show that adversarial attacks on image recognition systems can be significantly more effective than attacks on text-based systems.
- DNNs can also be vulnerable to adversarial attacks that compromise the **privacy** of users. For example, in [86], the authors show that DNNs used for facial recognition can be vulnerable to attacks that extract sensitive information from facial images. This can include gender, age, and other demographic information.

IV. DOMAIN OF ADVERSARIAL ATTACKS: A PERSPECTIVE ANALYSIS

Since its inception as proposed by [23], scholars predominantly concentrated on safeguarding DNN-based models against adversarial attacks. Yet, the advent of recent research has led to the proposal of diverse methods for carrying out these attacks in both the digital and physical worlds. Most digital attack strategies utilize methods that introduce small perturbations into the input data, resulting in inaccurate predictions generated by the model. The physical adversarial attack is distinguished by its focus on the established

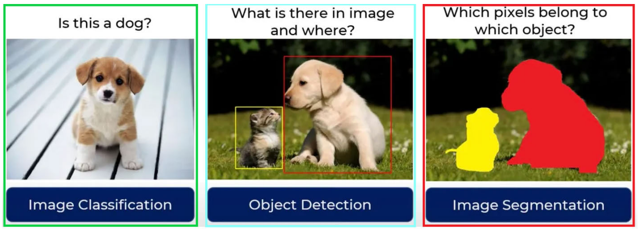
Is this a dog?
What is there in image and where?
Which pixels belong to which object?
Image Classification
Object Detection
Image Segmentation

FIGURE 4. Visual segregation of classification, detection and segmentation tasks [87].

DNN algorithms in the real world, which presents a more formidable undertaking owing to the intricate physical environment. In order to be deemed acceptable in real world, physical attacks must effectively acquire three fundamental targets: resilience, imperception and effectiveness. In a dynamic context, the adversarial medium is required to sustain its attack capability, which encompasses resistance to various factors such as cross-models, cross-scenes, and measurable limits. The technique employed must possess effectiveness in impeding the victim model's performance and should be simple for implementation within the physical environment. The medium in question must possess a level of difficulty in its detection or remain imperceptible to the naked human eye. Table 1 and Table. 2 provide a comprehensive overview of the most impactful contributions and advancements in digital attack techniques (categorized by methodologies) and physical attack techniques (categorized by adversarial medium), respectively.

A. ADVERSARIAL ATTACKS IN THE REALM OF COMPUTER VISION

Adversarial attacks have been found to be applicable to various domains within computer vision [60] including but not limited to image classification, object detection, and semantic segmentation, which fall within the purview of this review article:

- The task of categorizing an image into a specific class is a conventional problem in the field of computer vision, commonly referred to as **image classification**. Given an input x as image and a proficiently trained classifier f, the output is represented by $f(x) = y_{cls}$ where y_{cls} denotes the anticipated category of the input image. The categorization of visual information through image classification is a crucial element in the field of computer vision, providing a foundational framework for machines to comprehend and classify visual data. Deep learning models possess the ability to acquire complex visual features from unprocessed pixel data, leading to the development of image classification systems that are both highly precise and adaptable.
- **Object detection** refers to the procedure of identifying the location of an object within an image while simultaneously categorizing it. Given an input image x, a proficient deep neural network (DNN) object detection model denoted as f produces an output denoted as $f(x) = y_{object}$ which comprises three distinct components: objectness, classification, and boundary. Object detection has undergone significant advancements over time, with the incorporation of deep neural networks that enable the acquisition of resilient representations and precise localization of objects. As a result, object detection has become a crucial undertaking in both computer vision research and practical applications.
- The primary aim of **semantic segmentation** is to categorize each individual pixel into a pre-established group of classes. Given an image as input denoted by x and a deep neural network based semantic-segmentation model that has been properly trained, the output is represented as f(x) = y_{segt} where, y_{segt} refers to the output that defines the various components, including case categorization along with semantic-segmentation masking. Every visual element identified by mask incorporates unique color scheme. In contrast to the process of image classification, which involves assigning a solitary label to the entirety of an image, semantic segmentation offers a more comprehensive comprehension of an image by assigning a label to each individual pixel based on its corresponding class or category. The process of semantic segmentation furnishes machines with comprehensive and intricate data regarding the scene, thereby enabling them to comprehend the spatial arrangement and interconnections among diverse objects or regions.

TABLE 1. A summary of representative publications on physical adversarial attacks (Categorized by adversarial medium).

Physical Adversarial Attacks	
Adversarial Attack Medium	**Reference Works**
Patch	[46], [85], [88]–[94]
Sticker	[95]–[105]
Makeup	[51]
3D Printed Object	[106]
Clothing	[45], [104], [107]–[111]
Eye Glasses	[112]
Image	[27], [113]–[117]
Light	[118]–[122]
Bulb	[43]
Camera	[123], [124]

Figure 4 visually illustrates the distinction between aforesaid CV tasks.

B. DIGITAL ADVERSARIAL ATTACKS

Digital adversarial attacks refer to a category of attacks on machine learning models that exploit their vulnerabilities to manipulate the output of the model. These attacks entail the introduction of minute perturbations to the input data, which can result in the model producing inaccurate outputs with a significant degree of certainty. Numerous machine learning models have been demonstrated to be vulnerable to adversarial attacks, encompassing those utilized for image recognition, speech recognition, and natural language processing, as evidenced by various studies [23], [26], [31]. Whilst the effects of these attacks on machine learning models can be detrimental, there exist techniques to combat them, including adversarial training [35] and input preliminary processing [84]. Researchers in the fields of machine learning and computer security are currently engaged in efforts to comprehend and forestall adversarial attacks. The ensuing discourse presents details pertaining to frequently employed digital adversarial attacks in CV tasks.

- Szegedy et al. [23] introduced the earliest adversarial attack technique known as **Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS)**, which involved creating adversarial perturbations through the maximization of prediction error of the network. The L-BFGS algorithm aims to detect subtle alterations in images that result in incorrect classification by a neural network. By formulating the underlying issue as an optimization problem, the primary aim is to identify the perturbation that minimizes the discrepancy between the network's output on the modified image and the desired classification. The L-BFGS method employs a restricted-memory technique to estimate the inverse Hessian matrix, which encodes the second-order derivative details of the objective function. The L-BFGS method circumvents the requirement of explicitly calculating and retaining the complete Hessian matrix, which can be both computationally and memory intensive for optimization problems of significant scale, by estimating its inverse. The L-BFGS optimization algorithm integrates the first-order gradient information with the estimate of the second-order inverse Hessian matrix to incrementally update the variables, thereby achieving the minimization of the objective function. During each iteration, the descending direction is calculated utilizing the estimated inverse Hessian, followed by a line search to ascertain a suitable step size. The iterative process of the algorithm persists until it attains convergence, which usually occurs when the gradient reaches a sufficiently diminutive value. However, the optimization process in L-BFGS presents challenges in practical situations as it necessitates the optimization of parameters in a layer-wise manner across numerous layers.
- In order to enhance the efficiency of adversarial attacks, Goodfellow et al. [26] introduced the **Fast Gradient Sign Method (FGSM)**. The concept of adversarial training was motivated by the initial observation made by Szegedy et al. [23] that the robustness of classifiers to adversarial examples can be improved through their training with adversarial images. Nevertheless, the computational cost of resolving the optimization problem for a considerable quantity of images is high. The Fast Gradient Sign Method (FGSM) was proposed as a proficient approach to calculate adversarial perturbations, in order to tackle the aforementioned issue. The Fast Gradient Sign Method (FGSM) computes perturbations by leveraging the gradient of the cost function of the model in relation to the input image. This technique generates perturbations that are bounded by a norm. The authors, employed the Fast Gradient Sign Method (FGSM) to substantiate their linearity hypothesis. This hypothesis posits that the susceptibility of neural networks to adversarial perturbations is a result of their linear conduct in high-dimensional spaces, which is induced by activation functions such as ReLUs. The aforementioned hypothesis presents a divergent perspective from the dominant notion of the era, which posited non-linearity as the primary factor contributing to susceptibility in intricate networks. Fast Gradient Sign Method (FGSM) was further extended through integration of L_2 [140] and L_∞ [34] norms into the produced perturbation. Kurakin et al. The authors of [34] introduced the Iterative Fast Gradient Sign Method (I-FGSM), which is an iterative variant of FGSM. Dong et al., through Momentum Iterative-FGSM [39], subsequently improved the optimization process by incorporating momentum.
- The **Basic Iterative Method (BIM)** [34]represents a notable advancement in the discipline of adversarial attacks and is founded upon the principles of the Fast Gradient Sign Method (FGSM). The iterative process of creating an adversarial image using BIM entails modifying an initial image through the incorporation of scaled gradient direction additions or subtractions. Kurakin et al. demonstrated the effectiveness of BIM by utilizing printed adversarial images in real-world environments to mislead the ImageNet inception model. The advent of BIM has also acted as a stimulant for the progression of physical world attacks. In addition, it has incorporated the premise of targeted attacks, whereby the algorithm boosts the model's confidence on a specific target classification through modification of the iterative process. The iterative process culminates when the the intended number of iterations has been attained. BIM is also mentioned as Iterative Least-likely Class Method (ILCM).
- An untargeted attack called **DeepFool** repeatedly creates minor image perturbations in the direction of the closest decision boundary [41]. The objective of this framework is to identify the least amount of perturbations necessary to manipulate a deep neural network into incorrectly classifying an input sample. The DeepFool algorithm was developed on the premise of decision boundaries within spaces of high dimensionality. The process involves an iterative computation of the distance between the input sample along with the decision boundaries of the neural network, followed by the determination of the optimal direction where the sample needs to be perturbed to successfully cross the boundary. The iterative procedure persists until the sample is erroneously classified or attains a predetermined threshold of iterations.
- The **Projected Gradient Descent (PGD)** method employs an iterative technique to generate perturbations

TABLE 2. A summary of most impactful publications and advance techniques in the field of digital adversarial attacks.

Authors & Reference	Year	Contribution	Methodology	Attack Type
Szegedy et al. [23]	2013	Limited-Memory Broyden- Fletcher-Goldfarb-Shanno (L-BFGS)	Gradient-based Attack	White Box
Goodfellow et al. [26]	2014	Fast Gradient Sign Method (FGSM)	Gradient-based Attack	White Box
Moosavi-Dezfooli et al. [41]	2016	Deepfool	Gradient-based Attack	White Box
Goodfellow et al. [34]	2016	Basic Iterative Method (BIM)	Gradient-based Attack	White Box
Papernot et al. [81]	2016	Jacobian-based Saliency Map Attack (JSMA)	Gradient-based Attack	White Box
Madry et al. [35]	2017	Projected Gradient Descent (PGD)	Gradient-based Attack	White Box
Moosavi-Dezfooli et al. [125]	2017	Universal Adversarial Perturbations (UAP)	Optimization-based Attack	White Box
Carlini, N. et al. [40]	2017	Carlini and Wagner (C&W)	Optimization-based Attack	White Box
Kurakin et al. [27]	2018	Iterative-Fast Gradient Sign Method (I-FGSM)	Gradient-based Attack	White Box
Dong et al. [39]	2018	Momentum Iterative-Fast Gradient SIgn Method (MI-FGSM)	Gradient-based Attack	White Box
Rony et al. [126]	2019	Decoupled Direction and Norm (DDN) attack	Gradient-based Attack	White Box
Yao et al. [127]	2019	Trust Region based Adversarial Attack	Gradient-based Attack	White Box
Dong et al. [128]	2020	Robust Superpixel-Guided Attentional Adversarial Attack	Gradient-based Attack	White Box
Guo et al. [129]	2020	Linear Backpropagation (LinBP)	Gradient-based Attack	White Box
Dong et al. [129]	2020	Greedyfool: Distortion-aware Sparse Adversarial Attack	Gradient-based Attack	White Box
Sriramanan et al. [130]	2020	Guided Adversarial Margin Attack (GAMA)	Gradient-based Attack	White Box
Rahmati et al. [131]	2020	Geometric Decision-based (GeoDA) Attack	Query-based Attack	Black Box
Shi et al. [132]	2020	Customized Adversarial Boundary (CAB) Attack	Query-based Attack	Black Box
Li et al. [133]	2020	Projection & Probability-Driven Black-box (PPBA) Attack	Query-based Attack	Black Box
Li et al. [134]	2020	Query-efficient Boundary-based Blackbox Attack (QEBA)	Query-based Attack	Black Box
Ru et al. [135]	2020	Bayesian Optimisation-based Attack	Query-based Attack	Black Box
Wei et al. [136]	2021	Transferable Adversarial Attack	Query-based Attack	Black Box
Wei et al. [137]	2022	Meaningful Printable Adversarial Attack	Query-based Attack	Black Box
He et al. [138]	2023	Point Cloud Adversarial Perturbation	GAN-based Attack	White Box
Shi et al. [139]	2023	Customized Iteration and Sampling Attack (CISA)	Query-based Attack	Black Box

in incremental stages, while simultaneously ensuring that the modified image remains within a pre-defined epsilon-ball centered around the original image [35]. The PGD algorithm begins by initializing an initial input example, and then iteratively modifies the input through incremental adjustments in the direction of the gradient of a selected loss function. The implementation of a loss function is a method utilized to achieve a particular goal, which may involve either optimizing the estimated error of the model or reducing the level of confidence linked with a particular category. After each iteration, the modified input is mapped onto a group of viable solutions. The customary characterization of this combination entails a sphere of radius epsilon centered at the initial input. The objective of this cartographic representation is to ensure that any disturbances remain confined within the specified boundaries. The projection process functions to limit the magnitude of disturbances, ensuring that they do not exceed a predetermined threshold and preserving the visual similarity between the original and modified versions.

- The **Carlini and Wagner (C&W)** attack is a method that aims to detect the most minimal perturbation that can cause an image to be misclassified as a specific target category. The attack strategy is expressed in the form of an optimization problem, wherein an objective function is established to encompass both the objective of misclassification and a component that gauges the perceptibility of the perturbations [40]. The utilization of the misclassification component fosters the adversarial example to be categorized as either the intended target class or an alternative class, other than its actual class. The inclusion of a perturbation component serves to impose a penalty on significant perturbations, thereby preserving the imperceptibility of the resulting adversarial example. The C&W attack approach presents the optimization problem in a manner that permits adaptability in the specification of the distance measure between the initial and modified instances. Various distance metrics, including the L_0, L_2, or L_∞ norm, can be selected. The selection of a distance metric has a significant impact on both the properties of the resulting adversarial examples and the computational intricacy of the optimization procedure. The C&W attack strategy utilizes optimization methods, including gradient descent and binary search, to progressively modify perturbations until the intended misclassification and perceptibility standards are achieved. The attack methodology alters the objective function and optimization procedure to achieve a suitable equilibrium between the effectiveness and perceptibility of the attack.

- The **One-Pixel** targeted attack, involve the manipulation of a single pixel within an input image, resulting in an incorrect classification outcome [141]. The optimization procedure commonly entails the selection of the target class for misclassification, the development of an objective function that quantifies the extent of the misclassification error, and the imposition of constraints on the positioning and strength of the pixels. Constraints may be imposed on the pixel location and color value thresholds to ensure that the perturbation remains undetected. It is crucial to understand that the effectiveness of the One Pixel Attack technique may depend on various factors, including the complexity of the model, the dataset used, and the choice of optimization algorithm. In addition, countermeasures that target minimizing the impact of adversarial attacks, such as input transformations or adversarial training, possess the capability to alleviate the impacts of the One Pixel Attack.
- The objective of **Universal Adversarial Perturbations** is to generate perturbations that are independent of the image content and can deceive an intended model on any given image. The transfer-ability of these perturbations across diverse images renders them efficacious in deceiving multiple models. The perturbations adhere to a constraint whereby the likelihood of the target model incorrectly classifying an image that has undergone perturbation is equal to or exceeds a predetermined scalar value δ. Additionally, the perturbation must remain within a specific norm-bound η. An iterative algorithm is utilized to calculate the perturbations, which displace data points from their respective class regions. This process results in the accumulation of perturbations that alter the labels, while adhering to a predetermined norm-bound.

C. PHYSICAL ADVERSARIAL ATTACKS

The creation of physical adversarial attacks involves the conversion of digital adversarial attacks into physical forms. The act of deceiving an object identification system in the physical world is commonly achieved through the presentation of a digital adversarial example, either through printing or projection. Physical adversarial attacks involve the creation of an adversarial image through digital adversarial attack techniques, which are subsequently reproduced to produce an adversarial medium that can be utilized in the real world. Generating physical adversarial attacks poses several challenges that must be addressed to achieve effectiveness. Several methodologies have been developed to address them:

- The challenge of accurately capturing subtle value differences between neighboring pixels poses a significant obstacle in the field of adversarial image generation. The integration of the total variation loss concept [106] into the objective function has been proposed by researchers as a means of addressing this challenge. The regularization term of **total variation loss** is utilized to measure the degree of variation in the pixel intensities of an image. The optimization process is enhanced by integrating total variation, resulting in the production of adversarial images that demonstrate seamless pixel transitions in their visual appearance. The utilization of this regularization technique is of utmost importance in generating authentic and visually credible adversarial examples, thereby amplifying their efficacy in practical situations.
- In the domain of physical adversarial attacks, a notable barrier emerges as a result of the possible divergence of hues present in the printed adversarial patches when compared to their intended manifestation in the physical environment. In order to address this issue, Sharif et al. in their work [112] incorporated **non-printability score** into the objective function to evaluate the printability of colors in the adversarial patch within the physical domain. The assessment of the non-printability score pertains to the accuracy of color reproduction during the printing procedure, which is a crucial aspect in the development of a potent physical adversarial attack. The aforementioned scoring mechanism functions as a means of quantifying the precision of color replication, thereby guaranteeing the development of robust and dependable physical adversarial attacks.
- The **Expectation over Transformations (EoT)** [142] concept is frequently utilized to bolster the resilience of adversarial attacks against a range of real-world variations. EoT involves the utilization of a adversarial dataset that undergoes various transformations such as rotations, translations, and scale changes. These transformations are applied to simulate the inherent variability that is encountered in the physical world, and the attack is optimized using this dataset. Through exposure to such variations, the attack proficiency is enhanced.

D. ADVERSARIAL MEDIUM IN CONTEXT OF PHYSICAL ADVERSARIAL ATTACKS

The term adversarial medium pertains to the physical material or medium utilized for conducting physical adversarial attacks. The medium utilized in physical adversarial attacks may encompass a diverse range of media or objects, including but not limited to printed images, stickers, 3D-printed objects, or physical changes created on the external appearance of an object. These media are specifically engineered or adapted to capitalize on the susceptibilities of machine learning systems during their interactions with the real environment. Physical adversarial mediums encompass a range of techniques that can be employed to deceive computer vision systems. These may include the use of printed images with meticulously designed patterns [27], the placement of stickers on objects in a strategic manner to induce misclassification [95], or the physical alteration of traffic signals or stop signs to mislead autonomous vehicles [85].

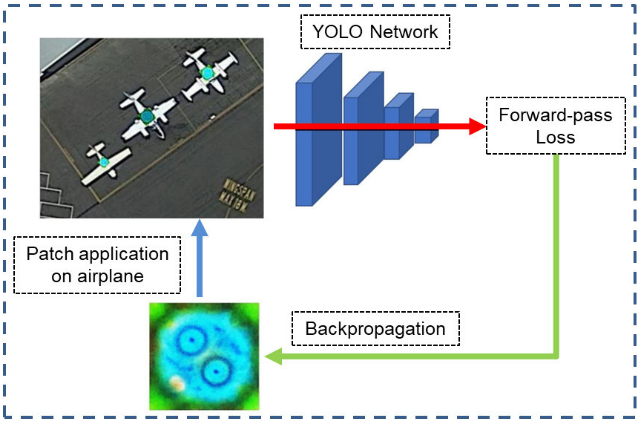

FIGURE 5. Illustration of patch training methodology proposed by [143].

Stickers [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], and [105] or **patches** [46], [85], [88], [89], [90], [91], [92], [93], [94] are placed on objects or clothing to obscure or distort their features, making them difficult for recognition systems to identify. These stickers can be designed to contain adversarial perturbations that can cause recognition systems to misclassify the object. Clothing [45], [107], [108] is designed with patterns or designs that can fool recognition systems. These patterns can be used to create false positives or negatives in the recognition system. **3D printed objects** based attacks [106] are designed to contain adversarial perturbations that can fool object recognition systems. These objects can be designed to look similar to real objects, but contain subtle changes that can cause recognition systems to misclassify them. Similarly in **image-based attacks** [27], digital images are manipulated to produce physical effects when printed or displayed. For example, certain patterns or colors can be used to create optical illusions or distortions that can make it difficult for recognition systems to identify objects or people. **Eyeglasses** [112] and **makeup** [51] have also been used as a medium to perform physical attacks on face recognition systems. **Light** [118], [119], [120], [121], [122], **bulbs** [43] and direct manipulation of sensor [123], [124] have also been used as adversarial mediums in real-world scenarios.

V. ADVERSARIAL ATTACKS ON AERIAL IMAGERY

The rapid advancement of machine learning models employed in the examination of aerial imagery has enabled the investigation of new territories in significant fields such as urban planning [144], environmental surveillance [145], emergency management [146], and other emerging areas [147]. The vulnerability of these models to adversarial attacks poses a significant challenge to the reliability and accuracy of their outcomes. Adversarial attacks targeting aerial imagery employ advanced techniques. The aforementioned techniques involve the incorporation of subtle perturbations into input data with the express purpose of misleading the underlying algorithms. The possible consequences of such attacks can be considered severe, leading to erroneous classification, inaccurate recognition of entities, or compromised decision-making in circumstances where safety is of utmost importance. Therefore, there is an urgent requirement for a thorough examination and robust protective strategies to effectively mitigate this emerging hazard.

The challenges posed in the aerial dimensions of physical world attacks are inherently demanding and require careful consideration. The challenges encompass concerns pertaining to atmospheric impacts, illumination, sensor resolution, fluctuations in lighting conditions, diverse fields of views and distances. The aforementioned factors introduce intricacies that have a substantial influence on the efficacy of adversarial attacks within the aerial domain. As a result, the majority of extant research in this domain has primarily concentrated on visual tasks conducted on the ground, wherein the difficulties pertaining to the aerial dimension are comparatively less pronounced. The study of vision tasks that are conducted on the ground has been the subject of extensive research. As a result, models that are highly resilient and effective defense mechanisms have been created to counteract adversarial attacks.

This section provides insights into research work related to adversarial attacks on aerial imagery primarily focusing on three CV tasks such as classification, segmentation, and object detection. Table 3 presents an overview of research conducted in the areas of detection, classification, and segmentation within the field of computer vision. The table is organized chronologically, allowing researchers to categorize the research based on various factors such as the computer vision tasks involved, the targeted model or network, the attack setting, the adversarial medium, and the datasets evaluated.

A. CLASSIFICATION AND SEMANTIC-SEGMENTATION IN CONTEXT OF AERIAL IMAGERY ADVERSARIAL ATTACKS

Czaja et al. [149] presented a novel technique in classification of aerial images, wherein they generated adversarial instances to classify satellite images. Their approach involved targeting smaller regions of the image for inducing an incorrect prediction by the classifier, which was successfully demonstrated. The utilization of a deep convolutional neural network (CNN) for object detection and the generation of adversarial examples through a variant of the fast gradient sign method (FGSM) is employed by the authors. Adversarial examples are produced through the manipulation of input images, resulting in the misclassification of objects depicted within said images. The authors employ the Functional Map of the World (fMoW) dataset [150] to assess the efficacy of the adversarial examples. This dataset comprises high-resolution satellite images that are accompanied by ground truth object labels. The study conducted by the authors demonstrates the efficacy of their adversarial examples in misleading the object detection CNN, resulting in the misclassification of objects within the fMoW dataset. The study conducted by

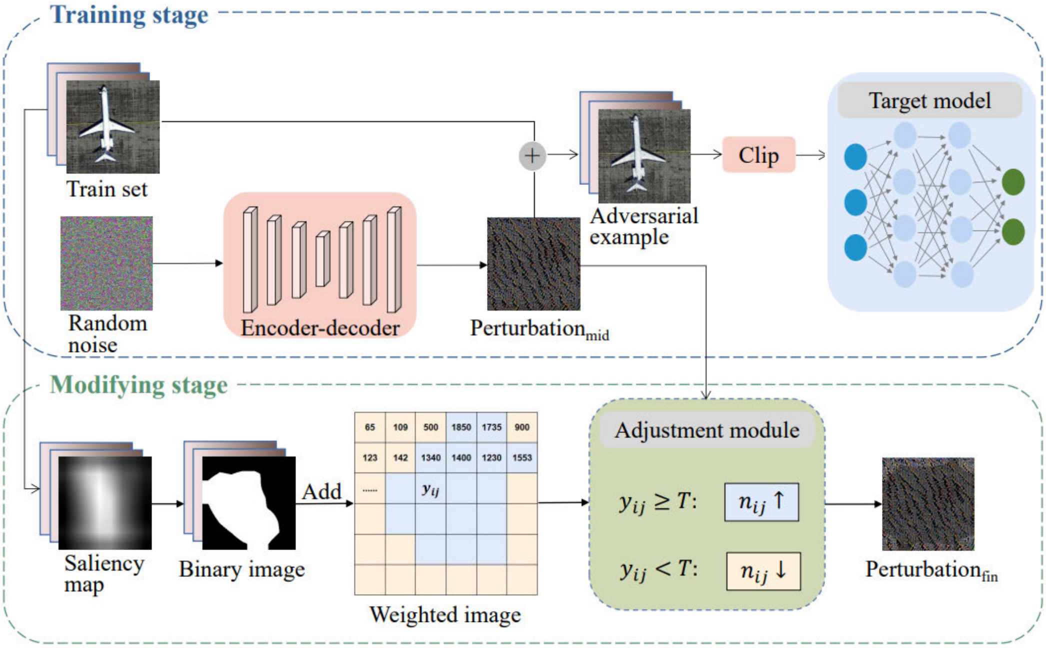

FIGURE 6. Overview of the concept of generic universal adversarial perturbation. The symbol '↑' denotes a rise in the magnitude of n_{ij}, while the symbol '↓' denotes a decline in the magnitude of n_{ij} [148].

Chen et al. [152] centered on investigating the performance of Remote Sensing Image (RSI) recognition models when subjected to adversarial examples. The FGSM [26] and BIM [34] algorithms were employed to launch attacks on various converged RSI recognition models across diverse datasets. The study conducted experiments on various convolutional neural network models, including ResNet50 [177] and InceptionV1 [178], which were trained on diverse datasets. The classifiers were successfully manipulated to exhibit erroneous predictions. Authors further discussed the classification problem in aerial imagery, particularly in the domains of military and autonomous driving, posing a significant security threat. The potential ramifications can be quite severe. The authors also explain that within the military domain, an attacker has the capability to produce adversarial perturbations aimed at a specific target, such as an aircraft, and conceal the object by means of a physical obstruction.

Yin et al. [148] conducted a study on various classifier models, namely VGG-16 & 19 [179], ResNet34 [177], and ResNet101 [180], that were trained on PatternNet dataset [170]. The authors demonstrated that the Universal Adversarial Perturbation (UAP) framework has the potential to cause misclassification in these models. The authors presented a novel approach to deceive classifiers by enhancing the universal adversarial perturbation framework using an encoder-decoder network fusion. The flow chart depicted in Figure 6 illustrate the process of generating UAP. During the training phase, the input is a random noise z that follows a normal distribution with mean 0 and variance 1, denoted as $N(0, 1)$. The generator incorporates multi-layer convolution, pooling, and upsampling procedures, thereby ensuring improved extraction of high-dimensional features. The process of generating an adversarial example involves the introduction of perturbation to a given clean example, followed by the application of a clipping operation. Subsequently, the designated model is employed to make a prediction. In the context of a target model denoted as $C_\theta(x)$, which is characterized by a parameter θ, it is possible to accurately classify a given clean example x when the output of the model, $C_\theta(x)$, matches the correct label c associated with the example in question. In the event that perturbation δ is introduced to the example, the model will inaccurately classify the example provided that $C(x + \delta)$ does not equal c. The objective of the UAP is to identify a perturbation v that satisfies the formula $C(x + v) \neq c$ for a significant number of clean examples. The intent is to identify a perturbation v that results in the misclassification of a majority of the examples.

Similarly, Xu et al. [159] conducted a comparative analysis of various targeted classifiers and semantic segmentation models. To achieve this, they employed black box attacks, namely Mixup and Mixcut attacks, and identified vulnerabilities that were shared across multiple networks. The study presents a novel approach for generating ubiquitous adversarial perturbations and evaluates its effectiveness on diverse datasets and models, with the aim of enhancing the dependability of remote sensing image recognition systems. Mixup-Attack approach, as proposed, is illustrated in Figure 11. The central idea pertains to generating a mixup image by employing a linear synthesis of training images derived from disparate categories. The aforementioned composite image is subsequently utilized to carry out adversarial attacks at the level of features on the input image. Upon

TABLE 3. Publications summary on aerial imagery in context of adversarial attacks & respective scenarios.

Authors & Reference	Year	Task	Targeted Network / Architecture	Attack Configuration	Attack Scenario	Evaluated Dataset	Evaluation Metric
Czaja et al. [149]	2018	Classification	CNN-I	White-box	Digital-patch attack	fMoW [150]	Attack Success Rate, Total Error Rate
Hollander et al. [143]	2020	Detection	Yolo-V2	White-box	Digital-patch attack	DOTA [151]	Average Precision
Chen, L. et al. [152]	2019	Classification	ResNet-50, Inception-V1	White-box	Digital adversarial attack	NWPU [153], UCM [154], CLRS [155]	Fooling Rate
Lu et al. [156]	2021	Detection	Yolo-V3, Yolo-V5, FasterR-CNN	White-box	Digital-patch attack	NWPU-VHR10 [157], DOTA [151], RSOD [158]	Average Precision
Xu et al. [159]	2022	•Classification •Semantic segmentation	**•Classification** (AlexNet, VGG16, InceptionV3, ResNet-18, ResNet-101, DenseNet-121, DenseNet-201, RegNet-X400MF, RegNet-X16GF) **•Semantic Segmentation** (FCN32s, FCN8s, DeepLabV2, UNet, Seg-Net, PSP-Net, SQ-Net, Link-Net, FRRNetA)	Black-box	Digital adversarial attack	**•Classification** (AID [160], UCM [154]) **•Semantic segmentation** (Zurich-Summer [161], Vaihingen [162])	Attack Success Rate
Du et al. [163]	2022	Detection	Yolo-V3	White-box	Physical-patch attack	COWC(M) [164], COCO [165]	Average Objectness Reduction Rate, Objectness Score Ratio
Zhang et al. [166]	2022	Detection	Yolo-V3, Yolo-V5	White-box	Physical-patch attack	VisDrone2019 [167]	Attack Success Rate
Van et al. [168]	2022	Detection	Yolo-V3, Yolt-V4	White-box	Digital-patch attack	VisDrone2019 [167]	Mean-F1
Wise et al. [169]	2022	Detection	FasterR-CNN	White-box	Digital-patch attack	COCO [165]	Mean Average Precision
Yin et al. [148]	2022	Classification	VGG16, VGG19, ResNet-34, ResNet-101	White-box	Digital-patch attack	Pattern-Net [170]	Attack Success Rate
Lian et al. [171]	2022	Detection	YOLO-v2, YOLO-v3, YOLO-v5s, YOLO-v5n, YOLO-v5m, YOLO-v5x, YOLO-v5l, FasterR-CNN, SSD, Swin-Transformer	White-box	Digital-patch attack & Physical-patch Attack on image-based scaled scenario	RSOD [158], DOTA [151]	Average Precision
Sun et al. [172]	2023	Detection	FCOS, FasterR-CNN, YOLO-v4, RetinaNet	White-Box	Digital-patch attack	DIOR [173], DOTA [151]	Mean Average Precision
Lian et al. [174]	2023	Detection	YOLO-v2, YOLO-v3, YOLO-v5, SSD, FasterR-CNN , Swin-Transformer, CascadeR-CNN, RetinaNet, MaskR-CNN, FoveaBox, Free-Anchor, FSAF, RepPoints, TOOD, ATSS, Varifocal-Net	White-Box	Physical-Contextual background attack	RSOD [158], DOTA [151]	Average Precision
Liu et al. [175]	2023	Detection	ROI-transformer, Gliding Vertex	White-Box	Digital adversarial attack	DOTA [151]	Mean Average Precision, Vanishing Rate
Tang et al. [176]	2023	Detection	Yolo-V2, Yolo-V3, Yolo-V4	White-Box	Digital-patch attack	NWPU-VHR10 [157], DOTA [151], RSOD [158]	Attack Success Rate

encountering a surrogate model that is parameterized by θ_s, authors move on with extracting the superficial features of both the original input image x alongside the mixup image $\tilde{x}$. The function $\mathcal{L}_{mix}$, which pertains to the mixing of images, is formulated by minimizing the KL divergence that exists among the features of the mixup image and also the input image. The application of cross-entropy loss $\mathcal{L}_{ce}$ is utilized to facilitate the attack, as the limitation of $\mathcal{L}_{mix}$ does not incorporate the network's predictions. The calculation of $\mathcal{L}_{ce}$ involves the comparison between the predicted logits pertaining to the target image and the corresponding true label. Therefore, the all-encompassing objective function $\mathcal{L}$ is the combination of $\mathcal{L}_{ce}$ and $\mathcal{L}_{mix}$, with suitable weighting. The creation of universal adversarial examples can be accomplished by integrating the gradients, also known as adversarial perturbation, of the all-encompassing objective function L into the original unmodified image. The mixup attack methodology also integrates the momentum mechanism to augment the consistency of update directions throughout the iterations.

B. OBJECT DETECTION IN CONTEXT OF AERIAL IMAGERY ADVERSARIAL ATTACKS

Hollander et al. [143] studied and employed adversarial attacks to deceive the Yolo-V2 object detection model [181] that had been trained on the DOTA dataset [151]. This was achieved by strategically placing patches of varying sizes on specific images, thereby rendering them camouflaged and

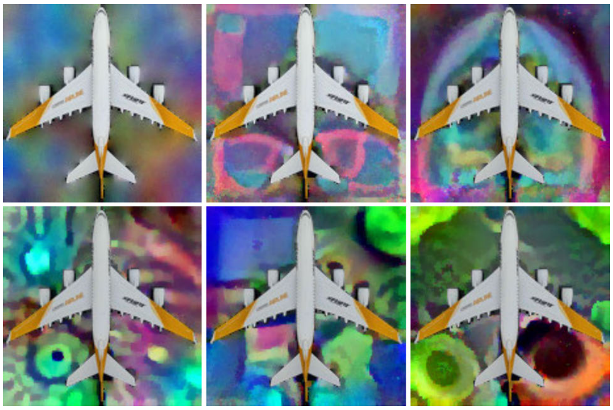

FIGURE 7. Depiction of contextual adversarial patches generated through CBA to conceal aircraft from object detectors [174].

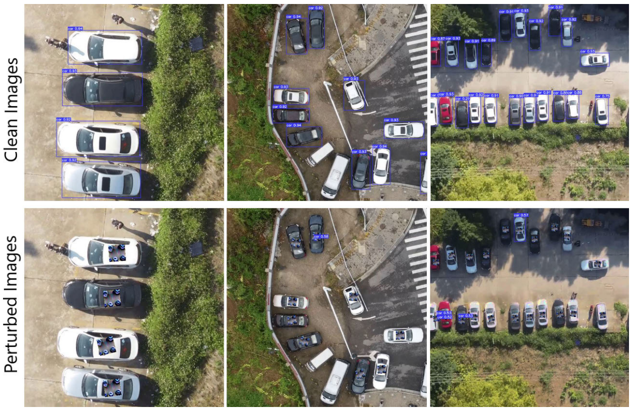

FIGURE 8. Clean & perturbed images after application of adversarial patch [166].

undetectable from aerial surveillance. The training approach utilized by authors is based on the methodology proposed by Thys et al. [46], as depicted in Figure 5. During each iteration, a batch of images that feature airplanes was utilized for patch training. The ongoing iteration of the adversarial patch, which commences with a randomly generated patch, was applied to the airplanes that were originally annotated. Prior to being placed on the planes, the patches went through a series of transformations, including scaling, rotation, noise corruption, and contrast stretching, in order to simulate real-life capturing circumstances. The patches also underwent through an arbitrary rotation of 360 degrees due to the absence of a pre-established position in aerial imagery. The outcomes were the trained patches that predominantly exhibit circular symmetrical designs. The YOLOv2 [181] neural network was utilized for the purpose of training, however, weights remained unchanged throughout the backpropagation process. The patches were optimized through loss formula i.e., $L = \alpha L_{nps} + \beta L_{tv} + L_{obj}$. The L_{nps} guarantees the production of printable hues within the patch. The utilization of L_{tv} serves to hinder the emergence of a noise pattern in the patch. The term L_{obj} denotes the highest score of objectness that an image can attain in the YOLO output. Consequently, this term will diminish the level of confidence in the detections made in the image. In every iteration that followed, the patch was modified. The patch attack was found to be effective in reducing the effectiveness of the model that was targeted. The study revealed a correlation between the magnitude of the patch and its ability to deceive object detection systems. Lu et al. [156] addressed the problem of physical attacks resulting from differences in size and scale of the target image. Their study successfully showcased the effectiveness of adversarial patches in attacking targets of varying sizes. The aforementioned models, namely Yolo-V3 [182], Yolo-V5 [183], and FasterR-CNN [184], were subjected to critical evaluation by the authors across various benchmark datasets. Their study introduced a novel attack strategy, termed PatchNoobj, which aimed to enable the adversarial patch to accommodate the scale variation of an aircraft and effectively obscure it from the perception of an object detector. Figure 9 depicts the framework architecture of PatchNoobj. The PatchNoobj architecture comprises of two distinct components, namely a patch applicator & a detector. The task of affixing adversarial patches to airplanes of varying dimensions is assigned to the patch applicator, whereas the detector employs a comprehensive object detection methodology and is bound for the iterative improvement of the adversarial patches through the loss function. The authors initially established the target-ground truth of the aircraft that required the attachment of the adversarial patch, as well as the untarget-ground truth of the object not requiring such attachment. Afterwards image was fed to the detector. Adversarial patch scaling was determined by utilizing the target ground-truth, and the mask was created to identify the precise location for attaching the adversarial patch. The untargeted ground-truth was utilized in the computation of loss for optimizing the adversarial patch. The two aforementioned ground-truths bear resemblance to the ground-truth pertaining to the bounding-box in object detection, and all three share the common format of $[x, y, w, h]$. Subsequently, the image is fed into the patch applicator, and a predetermined adversarial patch of fixed dimensions is initialized randomly. The process involved determining the scaling of the adversarial patch, generating a mask based on the target ground truth, and subsequently affixing the scaled adversarial patch onto the aircraft within the image in accordance with the mask. Finally, adversarial examples containing adversarial patches were introduced into the detector. The loss between the detector's outcome and the untargeted ground-truth was then computed using a loss function. The adversarial patch was subsequently updated iteratively by loss optimization.

Van et al. [168] underscored the susceptibility of adversarial patches, specifically with regard to the detection of said patches by object detectors. The employment opportunity also presented a notion for fabricating semi-transparent patches that were predominantly detected by object recognition systems. The researchers conducted experiments on

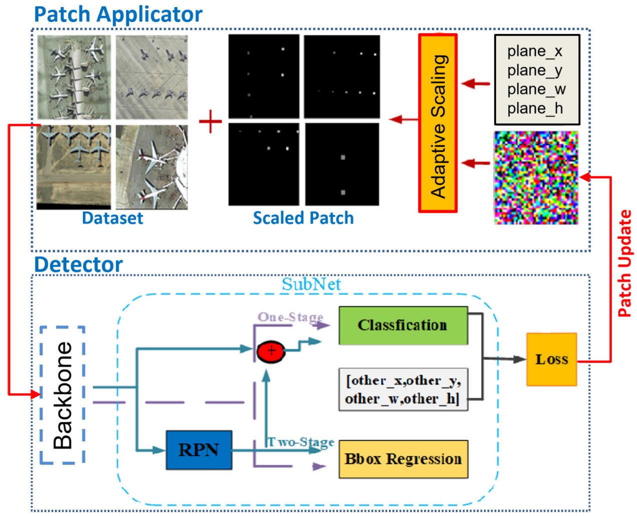

FIGURE 9. Depiction of PatchNoobj Framework proposed by [156].

YoloV3 [182]and YoloV4 [185] models that were trained using the Visdrone2019 dataset [167].

Wise et al. [169] introduced an innovative strategy for producing camouflage patches that possess the ability to effectively conceal substantial ground assets while remaining imperceptible. A novel approach was suggested for generating unnoticeable patches, which involves augmenting the object detection loss while simultaneously reducing color perception. The experiments employed a FasterR-CNN [184] model that had been trained on the COCO-2017 dataset [165]. In accordance with the methodology outlined in [186], a patch tailored to each image was created. This was achieved by optimizing the patch to maximize prediction loss, while simultaneously constraining it within the RGB color space to ensure its suitability for printing.

Nevertheless, all methodologies cited in the literature [143], [156], [168], [169] have been validated exclusively in virtual settings or by means of simulated environments, without conducting empirical trials in authentic real-world contexts. Du et al. [163] conducted a groundbreaking demonstration of physical world adversarial attacks, wherein they deceived a Yolo-V3 [182] that had been trained on COCO [165] and COWC-M [164] datasets by affixing on&off patches in physical scenarios. The authors utilized the method proposed by Thys et al. [46] for optimizing adversarial patches was adapted for use in aerial scenes. The study also exhibited the effectiveness of patch attacks in various weather conditions; nevertheless, it failed to address the issue of weather conditions. Additionally, a novel metric *Average Objectness Reduction Rate (AORR)* was introduced for evaluating performance in both digital and physical domains for directly measuring the effect on objectness score. A higher AORR can be achieved through the implementation of a more efficient attack strategy.

Zhang et al. [166] were able to effectively deceive object detection models [182], [183], which are designed to identify multiple objects from diverse perspectives and elevations in authentic settings. The authors introduced a novel approach for targeting object detectors through the development of a composite optimization problem that incorporates both detection and object loss considerations. The authors first conducted optimization of adversarial patches, which were subsequently subjected to scaling and rotation prior to being applied onto clean images. This process resulted in the creation of perturbed images through the use of an applicator function. The iterative optimization process of the adversarial patch involves updates through the gradient ascent algorithm, with the aim of minimizing the loss function. This loss function is composed of four distinct components, namely object-loss, detection-loss, non-printability-score (NPS) loss, and total-variation (TV) loss. The application of the apply function is observed to result in the transfer of clean images to adversarial examples, as depicted in Figure 8. The dimensions of the patch correspond with those of the car-roof, and the orientation of the patch with respect to the car undergoes variation subsequent to a randomly generated patch rotation.

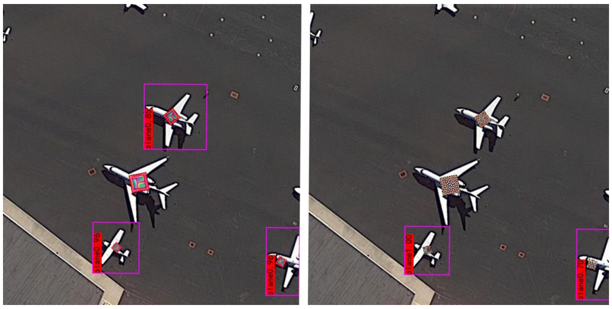

FIGURE 10. Illustration of results after application of different adversarial patches on DOTA dataset [176].

Lian et al. [171] proposed a novel attack algorithm termed as adaptive-patch-based physical attack (AP-PA) that is designed to produce adversarial patches with the objective of concealing objects from aerial detection systems. Experimentation was performed on multiple detection models including single stage [182], [183], [187], single shot [188], two stage [184] and transformer based detectors [189]. DOTA [151] dataset was used to train object detctors whereas RSOD [158] dataset was utilized for patch optimization. The iterative process of the algorithm commences with a randomly generated patch, which is subsequently modified to achieve an adversarial outcome. The process involves the application of a modified patch onto an image, followed by the utilization of a detector to identify objects within the modified image. The computational procedure computes detection loss and subsequently modifies the patch through the utilization of gradient descent in order to minimize the loss. The aforementioned procedure is iteratively executed until the objective function is minimized and patch is obtained. Objective function is the sum of three losses i.e., non-printability score, total variation and objectiveness score, as discussed earlier in IV-C.

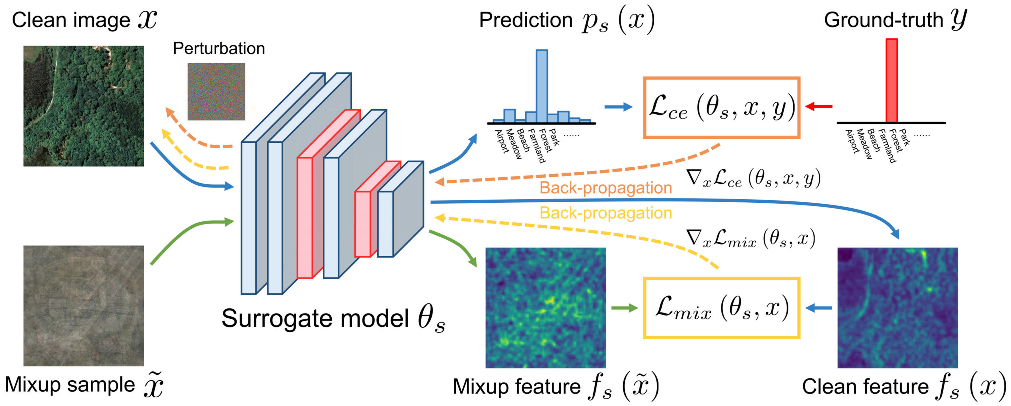
Clean image x
Perturbation
Prediction $p_s(x)$
Ground-truth y
Airport
Meadow
Beach
Farmland
Forest
Park
......
$\mathcal{L}_{ce}(\theta_s, x, y)$
Back-propagation
$\nabla_x \mathcal{L}_{ce}(\theta_s, x, y)$
Back-propagation
$\nabla_x \mathcal{L}_{mix}(\theta_s, x)$
Surrogate model θ_s
$\mathcal{L}_{mix}(\theta_s, x)$
Mixup sample $\tilde{x}$
Mixup feature $f_s(\tilde{x})$
Clean feature $f_s(x)$

FIGURE 11. Representative flowchart of mixup-attack to conduct a black-box adversarial attack on a remote sensing scene classification task. The surrogate model depicts the convolutional layers and pooling layers as light-blue and pink blocks, respectively [159].

Lian et al. [174] introduced a novel approach *contextual background attack (CBA)* with the aim of deceiving aerial detectors in real-world scenarios through background adversarial patches. Experimentation was conducted on multiple models [3] and DOTA [151] datset, whereas patches were optimized on RSOD [158] dataset. The optimization of patches was undertaken to ensure adequate coverage of significant contextual background areas, thereby enhancing the detection process. The target was effectively concealed within the image through masking with with patches, thereby rendering them unnoticeable. The optimization also ensures that the adversarial patches produced exhibit a high degree of similarity to the surrounding context, thereby augmenting their efficacy in misleading the detectors. Figure 7 depicts various contextual adversarial patches generated by CBA to attack object detectors.

Liu et al. [175] proposed Sensitive Pixel Localization C&W (SPL-C&W) vanishing attack algorithm which modifies the original image through the optimization of patches within the image. This process generates an adversarial example that can effectively deceive the object detector, resulting in the misclassification of objects present in the image. The approach involves the utilization of the backpropagation in conjunction with proposed attack sensitivity mapping algorithm for the purpose of calculating the perturbations and subsequently updating the perturbed image on the basis of the sensitive pixels. RIO-transformers [190] and Gliding vertex [191] detection models and DOTA [151] dataset were utilized during the study.

Tang et al. [176] proposed an attack technique based on the works of [46] and [143]. However, a novel patch optimization loss function was introduced. Rather than optimizing adversarial patch directly through detection confidence score, authors proposed to use intermediate outputs as loss function. This proposal was based on the assumption that intermediate outputs having better degree of freedom than final confidence score, thereby enabling a more comprehensive representation of the inputs' variability. Yolo-family object detction models [182], [183], [185] and NWPU-VHR10 [157], DOTA [151] and RSOD [158] datasets were utilized during the study. Results of adversarial patch application on DOTA dataset are illustrated in Figure 10.

VI. CONSTRAINTS AND POTENTIAL SOLUTIONS

This section will undertake an individual exploration of the constraints encountered in research field of adversarial attacks in context of aerial imagery, as well as potential avenues for their resolution.

A. HOMOGENEITY IN ATTACK SCENARIOS

The constraint of insufficient variety in attack scenarios pertains to the inclination of attacks on aerial imagery to primarily focus on particular computer vision tasks, such as object identification, detection or semantic segmentation. Although the aforementioned tasks hold significance, certain crucial tasks pertaining to the analysis of aerial imagery, such as anomaly detection, target identification, or scene understanding, have not been given due consideration in conjunction with adversarial attacks. The current state of aerial imagery research indicates a restricted range of attack scenarios, thereby limiting the diversity of such scenarios.

In order to overcome this constraint, upcoming studies need to work towards examining and scrutinizing adversarial attacks on varied aerial imaging assignments that extend beyond common ones. Through the broadening of attack scenarios, researchers may gain valuable insights into the existing vulnerability and construct resilient defense tactics for the purposes of anomaly detection, target identification, scene comprehension, and other vital applications within the discipline of aerial imagery analysis. The incorporation of a wider viewpoint is expected to enhance the general effectiveness of the defense mechanism against such attacks.

B. CONSTRAINTS IN PHYSICAL WORLD ATTACK SCENARIOS

Physical world attacks pose a severe challenge since they may take advantage of vulnerabilities which are challenging to predict or counteract solely through software-driven safeguards. The success of these attacks is contingent upon the exploitation of the physical characteristics of the surrounding environment, as well as the constraints imposed by the sensors or cameras integrated into aerial imaging systems. Unfortunately, studies often impose limitations on the investigation and understanding of real-world attacks on aerial imaging systems. Most research efforts focus on digital attacks, where tampering occurs within the digital version of aerial imagery. As a result, there could be insufficient emphasis on the effectiveness of actual physical attacks and the corresponding strategies employed to mitigate them.

Reducing the vulnerability to physical attacks requires a comprehensive approach that combines both concrete and abstract measures of protection. It is imperative that research efforts encompass the thorough examination and evaluation of practical measures that have been implemented in real-world scenarios, and their associated impacts on aerial perception systems. Future research may involve examining techniques for detecting and mitigating physical attacks, developing robust sensor fusion algorithms that can function optimally in challenging conditions, and exploring novel strategies for protecting the security and reliability of the aerial imaging systems.

C. LACK OF ATTACK GENERALIZATION AMONG DISSIMILAR SENSORS

The development of attack strategies that can be effectively migrated across diverse imaging devices poses a significant challenge. The effectiveness of adversarial attacks that are tailored for a particular aerial sensor, like RGB sensors, may not demonstrate resilience when employed on other sensor modalities that possess distinct resolutions or types. for instance, different sensors record distinct scene elements. Thermal imaging systems detect thermal signatures, RGB imaging systems detect visible light, and multi-spectral sensors detect a broad spectrum of wavelengths. Based on sensor modalities, noise patterns, and sensitivities, adversarial perturbations affect sensor inputs in a manner that varies.

The assessment of the resilience of aerial imagery systems in practical situations where various sensor modalities may be utilized necessitates a comprehension of the transference and efficacy of attacks across diverse sensor types. This facilitates the creation of all-encompassing defense mechanisms that take into consideration the variety of sensors employed in airborne platforms. It is imperative for researchers to investigate the development of attack designs that can effectively exploit shared vulnerabilities across various sensors, or alternatively, to devise approaches to attack that are customized to the distinctive features associated with each sensor modality.

D. COMPUTATIONAL OVERHEAD OF ATTACK IN REAL-TIME SCENARIOS

The computational overhead associated with real-time applications is a crucial factor to consider. In particular, the prompt evaluation of aerial imagery becomes of utmost importance for time-sensitive scenarios, such as monitoring or emergency management. Numerous adversarial attack techniques entail substantial computational costs, rendering them unfeasible for real-time installation situations.

To overcome this constraint, it is imperative to devise computationally efficient adversarial attack techniques that are customized for applications that operate in real-time. It is vital for researchers to investigate methods that can lower the computational cost of current attack methodologies while maintaining their efficacy. The process of generating attacks can be expedited by employing approximation techniques or algorithmic optimizations.

VII. RECOMMENDATIONS AND FUTURE DIRECTIONS

This section provides a discussion, brief synopsis of recommendations, and potential future directions.

A. OPTIMIZATION OF RANGE-INDEPENDENT ADVERSARIAL PATCHES

One of the primary challenges currently encountered in this field pertains to semantic segmentation and object detection models. When contemplating object detection as a metaphor, changes in the distance involving the sensor and the target could significantly influence the efficacy of the adversarial perturbation. In situations where the adversarial object is in sensor's close proximity, the target may appear magnified, resulting in distinct adversarial features that can effectively deceive the targeted model. In contrast, in cases where the adversarial features manifest as indistinct patterns or the targeted object is situated at a considerable distance from the sensor, the attack is unsuccessful. Therefore, it is crucial to investigate the techniques utilized in producing and enhancing adversarial perturbations that are capable of enduring diverse target sizes and ranges.

B. PATCH IMPERCEPTIBILITY

The objective of employing physical adversarial attacks as a means of protecting the assets is to substantially improve the effectiveness of the attack, while simultaneously minimizing the degree of perceptibility. The visual characteristics of the adversarial patch are a crucial aspect to take into account, given the susceptibility of the human visual system to novel stimuli, which may increase the likelihood of a successful attack. Therefore, it is considered worthwhile to conduct additional research aimed at examining the creation of an adversarial patch that is both more realistic and less noticeable, while taking into account the limitation of perturbation magnitude.

C. ADVERSARIAL ATTACK'S TRANSFERABILITY

The concept of transferability is integral to the phenomenon of adversarial perturbation, as it pertains to the capacity of perturbations to mislead models that are both familiar and unfamiliar. Nevertheless, the present physical attacks utilized in this area of research have faced impediments. The question of transfer-ability can be extended by means of ensemble techniques, which involve the concurrent implementation of multiple models. Furthermore, the exploration of the application of vision transformers for the purpose of attaining model transfer-ability is a promising avenue for future research.

D. ROBUSTNESS TO REAL-WORLD SCENARIOS

The most significant challenge in aerial imagery is the complex physical conditions, particularly the environmental factors in real-world situations. The physical environments that are considered complex encompass a variety of factors, such as weather phenomena, illumination levels, distortion effects, range limitations, opacity, and distortions that may arise from the image sensor system. Contemporary methodologies frequently overlook a significant subset of these variables and fail to account for real-world environmental factors. Thus, the optimization of adversarial mediums to accommodate physical conditions continues to be a significant obstacle to potential research avenues.

E. ADVERSARIAL ATTACKS ON UNLABELLED DATA

The research area concerning adversarial attacks on unlabelled data has become a multifaceted field that has garnered significant attention in recent years. Anticipated future research initiatives in this field are expected to enhance concerns related to adversarial attacks. The study aims to yield supplementary advantages to the community by enhancing attack strategies, strengthening the resilience of machine learning models, evaluating the influence of adversarial attacks on real-world scenarios, and scrutinizing their ethical concerns.

F. STANDARDIZING EVALUATION METRICS

The prevailing assessment criteria primarily centers on the measures of precision and robustness. The current metrics under use evaluate the efficacy of machine learning models in accurately classifying images and their capacity to withstand adversarial manipulation. Nevertheless, in certain contexts that pertain to aerial imagery, there could exist alternative metrics that are more pertinent and illuminating in encapsulating the tangible consequences of adversarial attacks. The development of standardized evaluation metrics that surpass the conventional measures of accuracy and robustness can enable researchers to acquire a more profound comprehension of the practical implications and outcomes. The utilization of standard metrics can aid in the development of aerial imagery models that are more robust and dependable, customized to meet the distinct demands of the application.

VIII. CONCLUSION

The research article undertakes a pioneering and comprehensive analysis of adversarial attacks on aerial imagery, leveraging current research in the field. The present investigation delves into the diverse spectrum of adversarial attack methodologies employed in both digital and physical domains. The article centers its focus on the noteworthy successes observed in adversarial attacks pertaining to three extensively employed tasks, namely classification, detection, and segmentation. The current research endeavors to shed light on instances of adversarial attacks prevalent in the applied discipline of aerial imagery. The overarching goal of this review is to invigorate further investigations by furnishing researchers by foundational understanding and valuable resource enabling to devise more assertive attacks. The aforementioned evaluation underscores the pressing need for deeper research initiatives aimed at devising more robust defensive tactics against these specific attack modalities. In light of the ongoing advancements witnessed in aerial imaging technology, it is of paramount importance to accord priority to the protection of data integrity and confidentiality. The significance of this technology assumes particular criticality due to its multifarious benefits in the critical domain.

ACKNOWLEDGMENT

The authors wish to extend their heartfelt gratitude to the editors and reviewers for their invaluable feedback and recommendations, which have substantially enhanced the caliber of this article. Additionally, they wish to express their gratitude for the technical and administrative assistance rendered by the School of Avionics and Electrical Engineering, College of Aeronautical Engineering, National University of Science and Technology, in the course of this study. They also express gratitude toward their colleagues and collaborators for their valuable insights and contributions to the present study. Ultimately, they also express their gratitude to their families for their consistent and steadfast backing and motivation throughout their scholarly pursuits.

REFERENCES

[1] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Exp. Syst. Appl.*, vol. 169, May 2021, Art. no. 114417.

[2] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Inf. Fusion*, vol. 73, pp. 22–71, Sep. 2021.

[3] Y. Bazi, L. Bashmal, M. M. A. Rahhal, R. A. Dayil, and N. A. Ajlan, "Vision transformers for remote sensing image classification," *Remote Sens.*, vol. 13, no. 3, p. 516, Feb. 2021.

[4] W. Boulila, M. Sellami, M. Driss, M. Al-Sarem, M. Safaei, and F. A. Ghaleb, "RS-DCNN: A novel distributed convolutional-neural-networks based-approach for big remote-sensing image classification," *Comput. Electron. Agricult.*, vol. 182, Mar. 2021, Art. no. 106014.

[5] S. K. Abid, N. Sulaiman, N. P. N. Mahmud, U. Nazir, and N. A. Adnan, "A review on the application of remote sensing and geographic information system in flood crisis management," in *Proc. Conf. Broad Exposure Sci. Technol.*, 2022, pp. 1–10.

[6] M. Steven and J. A. Clark, *Applications of Remote Sensing in Agriculture*. Amsterdam, The Netherlands: Elsevier, 2013.

[7] J. Wasowski and F. Bovenga, "Remote sensing of landslide motion with emphasis on satellite multi-temporal interferometry applications: An overview," in *Landslide Hazards, Risks, and Disasters*, 2022, pp. 365–438. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/B9780128184646000068

[8] K. Dick, L. Russell, Y. Souley Dosso, F. Kwamena, and J. R. Green, "Deep learning for critical infrastructure resilience," *J. Infrastructure Syst.*, vol. 25, no. 2, Jun. 2019, Art. no. 05019003.

[9] B. Aksoy, M. Melikşahözmen, M. Eylence, S. A. Inan, and K. Eyyubova, "Deep learning-based air defense system for unmanned aerial vehicles," in *Smart Applications With Advanced Machine Learning and Human-Centred Problem Design*. Cham, Switzerland: Springer, 2023, pp. 69–83.

[10] C. Yuan, Z. Liu, and Y. Zhang, "Aerial images-based forest fire detection for firefighting using optical remote sensing techniques and unmanned aerial vehicles," *J. Intell. Robotic Syst.*, vol. 88, nos. 2–4, pp. 635–654, Dec. 2017.

[11] N. Aafaq, A. Mian, W. Liu, S. Z. Gilani, and M. Shah, "Video description: A survey of methods, datasets, and evaluation metrics," *ACM Comput. Surveys*, vol. 52, no. 6, pp. 1–37, Nov. 2020.

[12] A.-R. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech, Lang., Process.*, vol. 20, no. 1, pp. 14–22, Jan. 2012.

[13] 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.

[14] H. Qin, Z. Cai, M. Zhang, Y. Ding, H. Zhao, S. Yi, X. Liu, and H. Su, "BiPointNet: Binary neural network for point clouds," 2020, *arXiv:2010.05501*.

[15] N. Aafaq, A. Mian, W. Liu, N. Akhtar, and M. Shah, "Cross-domain modality fusion for dense video captioning," *IEEE Trans. Artif. Intell.*, vol. 3, no. 5, pp. 763–777, Oct. 2022.

[16] N. Aafaq, N. Akhtar, W. Liu, S. Z. Gilani, and A. Mian, "Spatio-temporal dynamics and semantic attribute enriched visual encoding for video captioning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12479–12488.

[17] N. Aafaq, A. S. Mian, N. Akhtar, W. Liu, and M. Shah, "Dense video captioning with early linguistic information fusion," *IEEE Trans. Multimedia*, vol. 25, pp. 2309–2322, 2022.

[18] L. H. Shehab, O. M. Fahmy, S. M. Gasser, and M. S. El-Mahallawy, "An efficient brain tumor image segmentation based on deep residual networks (ResNets)," *J. King Saud Univ. Eng. Sci.*, vol. 33, no. 6, pp. 404–412, Sep. 2021.

[19] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, and W. Enbeyle, "Deep neural networks for medical image segmentation," *J. Healthcare Eng.*, vol. 2022, Mar. 2022, Art. no. 9580991.

[20] H. Ji, Z. Gao, T. Mei, and B. Ramesh, "Vehicle detection in remote sensing images leveraging on simultaneous super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 676–680, Apr. 2020.

[21] A. Albert, J. Kaur, and M. C. Gonzalez, "Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 1357–1366.

[22] M. Pritt and G. Chern, "Satellite image classification with deep learning," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2017, pp. 1–7.

[23] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," 2013, *arXiv:1312.6199*.

[24] N. Aafaq, N. Akhtar, W. Liu, M. Shah, and A. Mian, "Language model agnostic gray-box adversarial attack on image captioning," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 626–638, 2023.

[25] W. Brendel, J. Rauber, and M. Bethge, "Decision-based adversarial attacks: Reliable attacks against black-box machine learning models," 2017, *arXiv:1712.04248*.

[26] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, *arXiv:1412.6572*.

[27] A. Kurakin, I. J. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," in *Artificial Intelligence Safety and Security*. Boca Raton, FL, USA: Chapman & Hall/CRC, 2018, pp. 99–112.

[28] N. Papernot, P. McDaniel, and I. Goodfellow, "Transferability in machine learning: From phenomena to black-box attacks using adversarial samples," 2016, *arXiv:1605.07277*.

[29] N. Carlini and D. Wagner, "MagNet and 'efficient defenses against adversarial attacks' are not robust to adversarial examples," 2017, *arXiv:1711.08478*.

[30] S. Gu and L. Rigazio, "Towards deep neural network architectures robust to adversarial examples," 2014, *arXiv:1412.5068*.

[31] N. Papernot, P. McDaniel, X. Wu, S. Jha, and A. Swami, "Distillation as a defense to adversarial perturbations against deep neural networks," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2016, pp. 582–597.

[32] B. Wang, J. Gao, and Y. Qi, "A theoretical framework for robustness of (deep) classifiers against adversarial examples," 2016, *arXiv:1612.00334*.

[33] H. Kannan, A. Kurakin, and I. Goodfellow, "Adversarial logit pairing," 2018, *arXiv:1803.06373*.

[34] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," 2016, *arXiv:1611.01236*.

[35] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," 2017, *arXiv:1706.06083*.

[36] N. Pitropakis, E. Panaousis, T. Giannetsos, E. Anastasiadis, and G. Loukas, "A taxonomy and survey of attacks against machine learning," *Comput. Sci. Rev.*, vol. 34, Nov. 2019, Art. no. 100199.

[37] B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Šrndić, P. Laskov, G. Giacinto, and F. Roli, "Evasion attacks against machine learning at test time," in *Proc. Eur. Conf. ECML PKDD*. Prague, Czech Republic: Springer, Sep. 2013, pp. 387–402.

[38] B. Biggio, B. Nelson, and P. Laskov, "Poisoning attacks against support vector machines," 2012, *arXiv:1206.6389*.

[39] Y. Dong, F. Liao, T. Pang, H. Su, J. Zhu, X. Hu, and J. Li, "Boosting adversarial attacks with momentum," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9185–9193.

[40] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 39–57.

[41] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "DeepFool: A simple and accurate method to fool deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2574–2582.

[42] X. Wei, B. Pu, J. Lu, and B. Wu, "Physical adversarial attacks and defenses in computer vision: A survey," 2022, *arXiv:2211.01671*.

[43] X. Zhu, X. Li, J. Li, Z. Wang, and X. Hu, "Fooling thermal infrared pedestrian detectors in real world using small bulbs," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 4, 2021, pp. 3616–3624.

[44] X. Lei, C. Lu, Z. Jiang, Z. Gong, X. Cai, and L. Lu, "Using frequency attention to make adversarial patch powerful against person detector," 2022, *arXiv:2205.04638*.

[45] K. Xu, G. Zhang, S. Liu, Q. Fan, M. Sun, H. Chen, P.-Y. Chen, Y. Wang, and X. Lin, "Adversarial T-shirt! Evading person detectors in a physical world," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 665–681.

[46] S. Thys, W. V. Ranst, and T. Goedemé, "Fooling automated surveillance cameras: Adversarial patches to attack person detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 49–55.

[47] V. Cherepanova, M. Goldblum, H. Foley, S. Duan, J. Dickerson, G. Taylor, and T. Goldstein, "LowKey: Leveraging adversarial attacks to protect social media users from facial recognition," 2021, *arXiv:2101.07922*.

[48] Y. Dong, H. Su, B. Wu, Z. Li, W. Liu, T. Zhang, and J. Zhu, "Efficient decision-based black-box adversarial attacks on face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7706–7714.

[49] C. Bisogni, L. Cascone, J.-L. Dugelay, and C. Pero, "Adversarial attacks through architectures and spectra in face recognition," *Pattern Recognit. Lett.*, vol. 147, pp. 55–62, Jul. 2021.

[50] Y. Xu, K. Raja, R. Ramachandra, and C. Busch, "Adversarial attacks on face recognition systems," in *Handbook of Digital Face Manipulation and Detection*. Cham, Switzerland: Springer, 2022, pp. 139–161.

[51] B. Yin, W. Wang, T. Yao, J. Guo, Z. Kong, S. Ding, J. Li, and C. Liu, "Adv-makeup: A new imperceptible and transferable attack on face recognition," 2021, *arXiv:2105.03162*.

[52] J. Jeong, S. Kwon, M.-P. Hong, J. Kwak, and T. Shon, "Adversarial attack-based security vulnerability verification using deep learning library for multimedia video surveillance," *Multimedia Tools Appl.*, vol. 79, nos. 23–24, pp. 16077–16091, Jun. 2020.

[53] D. Edwards and D. B. Rawat, "Study of adversarial machine learning with infrared examples for surveillance applications," *Electronics*, vol. 9, no. 8, p. 1284, Aug. 2020.

[54] Y. Zheng, Y. Lu, and S. Velipasalar, "An effective adversarial attack on person re-identification in video surveillance via dispersion reduction," *IEEE Access*, vol. 8, pp. 183891–183902, 2020.

[55] K. N. Kumar, C. Vishnu, R. Mitra, and C. K. Mohan, "Black-box adversarial attacks in autonomous vehicle technology," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2020, pp. 1–7.

[56] Z. Xiong, H. Xu, W. Li, and Z. Cai, "Multi-source adversarial sample attack on autonomous vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2822–2835, Mar. 2021.

[57] Y. Deng, X. Zheng, T. Zhang, C. Chen, G. Lou, and M. Kim, "An analysis of adversarial attacks and defenses on autonomous driving models," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2020, pp. 1–10.

[58] J. I. Choi and Q. Tian, "Adversarial attack and defense of Yolo detectors in autonomous driving scenarios," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2022, pp. 1011–1017.

[59] Y. Li, X. Xu, J. Xiao, S. Li, and H. T. Shen, "Adaptive square attack: Fooling autonomous cars with adversarial traffic signs," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6337–6347, Apr. 2021.

[60] D. Wang, W. Yao, T. Jiang, G. Tang, and X. Chen, "A survey on physical adversarial attack in computer vision," 2022, *arXiv:2209.14262*.

[61] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access*, vol. 6, pp. 14410–14430, 2018.

[62] N. Akhtar, A. Mian, N. Kardan, and M. Shah, "Advances in adversarial attacks and defenses in computer vision: A survey," *IEEE Access*, vol. 9, pp. 155161–155196, 2021.

[63] H. Wei, H. Tang, X. Jia, H. Yu, Z. Li, Z. Wang, S. Satoh, and Z. Wang, "Physical adversarial attack meets computer vision: A decade survey," 2022, *arXiv:2209.15179*.

[64] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2805–2824, Sep. 2019.

[65] M. Ozdag, "Adversarial attacks and defenses against deep neural networks: A survey," *Proc. Comput. Sci.*, vol. 140, pp. 152–161, 2018.

[66] Y. Zhou, M. Han, L. Liu, J. He, and X. Gao, "The adversarial attacks threats on computer vision: A survey," in *Proc. IEEE 16th Int. Conf. Mobile Ad Hoc Sensor Syst. Workshops (MASSW)*, Sep. 2019, pp. 25–30.

[67] X. Wei, B. Pu, J. Lu, and B. Wu, "Visual adversarial attacks and defenses in the physical world: A survey," Tech. Rep., 2023. [Online]. Available: https://arxiv.org/abs/2211.01671

[68] R. Reza Wiyatno, A. Xu, O. Dia, and A. de Berker, "Adversarial examples in modern machine learning: A review," 2019, *arXiv:1911.05268*.

[69] A. Chakraborty, M. Alam, V. Dey, A. Chattopadhyay, and D. Mukhopadhyay, "Adversarial attacks and defences: A survey," 2018, *arXiv:1810.00069*.

[70] S. Qiu, Q. Liu, S. Zhou, and C. Wu, "Review of artificial intelligence adversarial attack and defense technologies," *Appl. Sci.*, vol. 9, no. 5, p. 909, Mar. 2019.

[71] X. Huang, D. Kroening, W. Ruan, J. Sharp, Y. Sun, E. Thamo, M. Wu, and X. Yi, "A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability," *Comput. Sci. Rev.*, vol. 37, Aug. 2020, Art. no. 100270.

[72] G. R. Machado, E. Silva, and R. R. Goldschmidt, "Adversarial machine learning in image classification: A survey toward the defender's perspective," *ACM Comput. Surveys*, vol. 55, no. 1, pp. 1–38, Jan. 2023.

[73] Y. Li, M. Cheng, C.-J. Hsieh, and T. C. M. Lee, "A review of adversarial attack and defense for classification methods," *Amer. Statistician*, vol. 76, no. 4, pp. 329–345, Oct. 2022.

[74] A. Serban, E. Poll, and J. Visser, "Adversarial examples on object recognition: A comprehensive survey," *ACM Comput. Surveys*, vol. 53, no. 3, pp. 1–38, May 2021.

[75] Y. Xu, B. Du, and L. Zhang, "Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: Attacks and defenses," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1604–1617, Feb. 2021.

[76] I. Debicha, B. Cochez, T. Kenaza, T. Debatty, J.-M. Dricot, and W. Mees, "Review on the feasibility of adversarial evasion attacks and defenses for network intrusion detection systems," 2023, *arXiv:2303.07003*.

[77] C. Kong, S. Wang, and H. Li, "Digital and physical face attacks: Reviewing and one step further," 2022, *arXiv:2209.14692*.

[78] W. Jin, Y. Li, H. Xu, Y. Wang, S. Ji, C. Aggarwal, and J. Tang, "Adversarial attacks and defenses on graphs: A review, a tool and empirical studies," 2020, *arXiv:2003.00653*.

[79] S. Bhambri, S. Muku, A. Tulasi, and A. B. Buduru, "A survey of black-box adversarial attacks on computer vision models," 2019, *arXiv:1912.01667*.

[80] S. M. K. Abbas Kazmi, N. Aafaq, M. A. Khan, A. Saleem, and Z. Ali, "Adversarial attacks on aerial imagery : The state-of-the-art and perspective," in *Proc. 3rd Int. Conf. Artif. Intell. (ICAI)*, Feb. 2023, pp. 95–102.

[81] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *Proc. IEEE Eur. Symp. Secur. Privacy (EuroSP)*, Mar. 2016, pp. 372–387.

[82] B. Biggio and F. Roli, "Wild patterns: Ten years after the rise of adversarial machine learning," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2018, pp. 2154–2156.

[83] P. Tabacof and E. Valle, "Exploring the space of adversarial images," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 426–433.

[84] W. Xu, D. Evans, and Y. Qi, "Feature squeezing: Detecting adversarial examples in deep neural networks," 2017, *arXiv:1704.01155*.

[85] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning visual classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1625–1634.

[86] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2015, pp. 1322–1333.

[87] B. Dickson. *New Deep Learning Model Brings Image Segmentation to Edge Devices*. Accessed: May 23, 2023. [Online]. Available: https://venturebeat.com/ai/new-deep-learning-model-brings-image-segmentation-to-edge-devices/

[88] T. B. Brown, D. Mané, A. Roy, M. Abadi, and J. Gilmer, "Adversarial patch," 2017, *arXiv:1712.09665*.

[89] A. Liu, "Perceptual-sensitive GAN for generating adversarial patches," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 1028–1035.

[90] Z. Wang, S. Zheng, M. Song, Q. Wang, A. Rahimpour, and H. Qi, "AdvPattern: Physical-world attacks on deep person re-identification via adversarially transformable patterns," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8340–8349.

[91] M. Pautov, G. Melnikov, E. Kaziakhmedov, K. Kireev, and A. Petiushko, "On adversarial patches: Real-world attack on ArcFace-100 face recognition system," in *Proc. Int. Multi-Conf. Eng., Comput. Inf. Sci. (SIBIRCON)*, Oct. 2019, pp. 0391–0396.

[92] A. Liu, J. Wang, X. Liu, B. Cao, C. Zhang, and H. Yu, "Bias-based universal adversarial patch attack for automatic check-out," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 395–410.

[93] F. Nesti, G. Rossolini, S. Nair, A. Biondi, and G. Buttazzo, "Evaluating the robustness of semantic segmentation for autonomous driving against real-world adversarial patch attacks," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 2826–2835.

[94] S. Liu, J. Wang, A. Liu, Y. Li, Y. Gao, X. Liu, and D. Tao, "Harnessing perceptual adversarial patches for crowd counting," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2022, pp. 2055–2069.

[95] Y. Zhang, H. Foroosh, P. David, and B. Gong, "CAMOU: Learning physical vehicle camouflages to adversarially attack detectors in the wild," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–20.

[96] J. Li, F. Schmidt, and Z. Kolter, "Adversarial camera stickers: A physical camera-based attack on deep learning systems," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3896–3904.

[97] S. Komkov and A. Petiushko, "AdvHat: Real-world adversarial attack on ArcFace face ID system," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 819–826.

[98] T. Wu, X. Ning, W. Li, R. Huang, H. Yang, and Y. Wang, "Physical adversarial attack on vehicle detector in the Carla simulator," 2020, *arXiv:2007.16118*.

[99] A. Zolfi, M. Kravchik, Y. Elovici, and A. Shabtai, "The translucent patch: A physical and universal attack on object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15227–15236.

[100] D. Wang, T. Jiang, J. Sun, W. Zhou, Z. Gong, X. Zhang, W. Yao, and X. Chen, "FCA: Learning a 3D full-coverage vehicle camouflage for multi-view physical adversarial attack," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, 2022, pp. 2414–2422.

[101] J. Wang, A. Liu, Z. Yin, S. Liu, S. Tang, and X. Liu, ''Dual attention suppression attack: Generate adversarial camouflage in physical world,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8561–8570.

[102] X. Han, G. Xu, Y. Zhou, X. Yang, J. Li, and T. Zhang, ''Physical backdoor attacks to lane detection systems in autonomous driving,'' in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 2957–2968.

[103] N. Suryanto, Y. Kim, H. Kang, H. T. Larasati, Y. Yun, T.-T.-H. Le, H. Yang, S.-Y. Oh, and H. Kim, ''DTA: Physical camouflage attacks using differentiable transformation network,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 15284–15293.

[104] Z. Hu, S. Huang, X. Zhu, F. Sun, B. Zhang, and X. Hu, ''Adversarial texture for fooling person detectors in the physical world,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13297–13306.

[105] Y. Duan, J. Chen, X. Zhou, J. Zou, Z. He, J. Zhang, W. Zhang, and Z. Pan, ''Learning coated adversarial camouflages for object detectors,'' in *Proc. 31st Int. Joint Conf. Artif. Intell.*, Jul. 2022, pp. 891–897.

[106] A. Athalye, L. Engstrom, A. Ilyas, and K. Kwok, ''Synthesizing robust adversarial examples,'' in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 284–293.

[107] L. Huang, C. Gao, Y. Zhou, C. Xie, A. L. Yuille, C. Zou, and N. Liu, ''Universal physical camouflage attacks on object detectors,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 717–726.

[108] Z. Wu, S.-N. Lim, L. S. Davis, and T. Goldstein, ''Making an invisibility cloak: Real world adversarial attacks on object detectors,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 1–17.

[109] Y.-C.-T. Hu, J.-C. Chen, B.-H. Kung, K.-L. Hua, and D. S. Tan, ''Naturalistic physical adversarial patch for object detectors,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7828–7837.

[110] J. Tan, N. Ji, H. Xie, and X. Xiang, ''Legitimate adversarial patches: Evading human eyes and detection models in the physical world,'' in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 5307–5315.

[111] X. Zhu, Z. Hu, S. Huang, J. Li, and X. Hu, ''Infrared invisible clothing: Hiding from infrared detectors at multiple angles in real world,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13307–13316.

[112] M. Sharif, S. Bhagavatula, L. Bauer, and M. K. Reiter, ''Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition,'' in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 1528–1540.

[113] S.-T. Chen, C. Cornelius, J. Martin, and D. H. P. Chau, ''ShapeShifter: Robust physical adversarial attack on faster R-CNN object detector,'' in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2018, pp. 52–68.

[114] D. Song, K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, F. Tramer, A. Prakash, and T. Kohno, ''Physical adversarial examples for object detectors,'' in *Proc. 12th USENIX Workshop Offensive Technol. (WOOT)*, 2018, pp. 1–15.

[115] R. Duan, X. Ma, Y. Wang, J. Bailey, A. K. Qin, and Y. Yang, ''Adversarial camouflage: Hiding physical-world attacks with natural styles,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 997–1005.

[116] Z. Kong, J. Guo, A. Li, and C. Liu, ''PhysGAN: Generating physical-world-resilient adversarial examples for autonomous driving,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14242–14251.

[117] W. Feng, B. Wu, T. Zhang, Y. Zhang, and Y. Zhang, ''Meta-attack: Class-agnostic and model-agnostic physical adversarial attack,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7767–7776.

[118] D.-L. Nguyen, S. S. Arora, Y. Wu, and H. Yang, ''Adversarial light projection attacks on face recognition systems: A feasibility study,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 814–815.

[119] A. Gnanasambandam, A. M. Sherman, and S. H. Chan, ''Optical adversarial attack,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 92–101.

[120] R. Duan, X. Mao, A. K. Qin, Y. Chen, S. Ye, Y. He, and Y. Yang, ''Adversarial laser beam: Effective physical-world attack to DNNs in a blink,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16057–16066.

[121] G. Lovisotto, H. Turner, I. Sluganovic, M. Strohmeier, and I. Martinovic, ''SLAP: Improving physical adversarial examples with short-lived adversarial perturbations,'' in *Proc. USENIX*, 2021, pp. 1–18.

[122] Y. Zhong, X. Liu, D. Zhai, J. Jiang, and X. Ji, ''Shadows can be dangerous: Stealthy and effective physical-world adversarial attack by natural phenomenon,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 15324–15333.

[123] A. Sayles, A. Hooda, M. Gupta, R. Chatterjee, and E. Fernandes, ''Invisible perturbations: Physical adversarial examples exploiting the rolling shutter effect,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14661–14670.

[124] B. Phan, F. Mannan, and F. Heide, ''Adversarial imaging pipelines,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16046–16056.

[125] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, ''Universal adversarial perturbations,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 86–94.

[126] J. Rony, L. G. Hafemann, L. S. Oliveira, I. Ben Ayed, R. Sabourin, and E. Granger, ''Decoupling direction and norm for efficient gradient-based l2 adversarial attacks and defenses,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4317–4325.

[127] Z. Yao, A. Gholami, P. Xu, K. Keutzer, and M. W. Mahoney, ''Trust region based adversarial attack on neural networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11342–11351.

[128] X. Dong, J. Han, D. Chen, J. Liu, H. Bian, Z. Ma, H. Li, X. Wang, W. Zhang, and N. Yu, ''Robust superpixel-guided attentional adversarial attack,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12892–12901.

[129] Y. Guo, Q. Li, and H. Chen, ''Backpropagating linearly improves transferability of adversarial examples,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 85–95.

[130] G. Sriramanan, S. Addepalli, and A. Baburaj, ''Guided adversarial attack for evaluating and enhancing adversarial defenses,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 20297–20308.

[131] A. Rahmati, S.-M. Moosavi-Dezfooli, P. Frossard, and H. Dai, ''GeoDA: A geometric framework for black-box adversarial attacks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8443–8452.

[132] Y. Shi, Y. Han, and Q. Tian, ''Polishing decision-based adversarial noise with a customized sampling,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1027–1035.

[133] J. Li, R. Ji, H. Liu, J. Liu, B. Zhong, C. Deng, and Q. Tian, ''Projection & probability-driven black-box attack,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 359–368.

[134] H. Li, X. Xu, X. Zhang, S. Yang, and B. Li, ''QEBA: query-efficient boundary-based blackbox attack,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1218–1227.

[135] B. Ru, A. Cobb, A. Blaas, and Y. Gal, ''BayesOpt adversarial attack,'' in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–16.

[136] X. Wei, Y. Guo, J. Yu, H. Yan, and B. Zhang, ''Generating transferable adversarial patch by simultaneously optimizing its position and perturbations,'' in *Proc. ICLR*, 2021, pp. 1–14.

[137] X. Wei, Y. Guo, and J. Yu, ''Adversarial sticker: A stealthy attack method in the physical world,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 2711–2725, Mar. 2023.

[138] F. He, Y. Chen, R. Chen, and W. Nie, ''Point cloud adversarial perturbation generation for adversarial attacks,'' *IEEE Access*, vol. 11, pp. 2767–2774, 2023.

[139] Y. Shi, Y. Han, Q. Hu, Y. Yang, and Q. Tian, ''query-efficient blackbox adversarial attack with customized iteration and sampling,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2226–2245, Feb. 2023.

[140] T. Miyato, S.-I. Maeda, M. Koyama, and S. Ishii, ''Virtual adversarial training: A regularization method for supervised and semi-supervised learning,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1979–1993, Aug. 2019.

[141] J. Su, D. V. Vargas, and K. Sakurai, ''One pixel attack for fooling deep neural networks,'' *IEEE Trans. Evol. Comput.*, vol. 23, no. 5, pp. 828–841, Oct. 2019.

[142] D. Hendrycks and T. Dietterich, ''Benchmarking neural network robustness to common corruptions and perturbations,'' 2019, *arXiv:1903.12261*.

[143] R. den Hollander, ''Adversarial patch camouflage against aerial detection,'' *Proc. SPIE*, vol. 11543, pp. 77–86, Sep. 2020.

[144] V. Chaturvedi and W. T. de Vries, ''Machine learning algorithms for urban land use planning: A review,'' *Urban Sci.*, vol. 5, no. 3, p. 68, Sep. 2021.

[145] A. Singh, F. Ramos, H. D. Whyte, and W. J. Kaiser, ''Modeling and decision making in spatio-temporal processes for environmental surveillance,'' in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 5490–5497.

[146] C. Kyrkou, P. Kolios, T. Theocharides, and M. Polycarpou, ''Machine learning for emergency management: A survey and future outlook,'' *Proc. IEEE*, vol. 111, no. 1, pp. 19–41, Jan. 2022.

[147] N. Aafaq, ''Deep learning for natural language description of videos,'' Tech. Rep., 2021. [Online]. Available: https://research-repository.uwa.edu.au/en/publications/deep-learning-for-natural-language-description-of-videos

[148] Q. Wang, G. Feng, Z. Yin, and B. Luo, ''Universal adversarial perturbation for remote sensing images,'' 2022, *arXiv:2202.10693*.

[149] W. Czaja, N. Fendley, M. Pekala, C. Ratto, and I.-J. Wang, ''Adversarial examples in remote sensing,'' in *Proc. 26th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2018, pp. 408–411.

[150] G. Christie, N. Fendley, J. Wilson, and R. Mukherjee, ''Functional map of the world,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6172–6180.

[151] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, ''DOTA: A large-scale dataset for object detection in aerial images,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3974–3983.

[152] L. Chen, G. Zhu, Q. Li, and H. Li, ''Adversarial example in remote sensing image recognition,'' 2019, *arXiv:1910.13222*.

[153] G. Cheng, J. Han, and X. Lu, ''Remote sensing image scene classification: Benchmark and state of the art,'' *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[154] Y. Yang and S. Newsam, ''Bag-of-visual-words and spatial extensions for land-use classification,'' in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2010, pp. 270–279.

[155] H. Li, H. Jiang, X. Gu, J. Peng, W. Li, L. Hong, and C. Tao, ''CLRS: Continual learning benchmark for remote sensing image scene classification,'' *Sensors*, vol. 20, no. 4, p. 1226, Feb. 2020.

[156] M. Lu, Q. Li, L. Chen, and H. Li, ''Scale-adaptive adversarial patch attack for remote sensing image aircraft detection,'' *Remote Sens.*, vol. 13, no. 20, p. 4078, Oct. 2021.

[157] G. Cheng, J. Han, P. Zhou, and L. Guo, ''Multi-class geospatial object detection and geographic image classification based on collection of part detectors,'' *ISPRS J. Photogramm. Remote Sens.*, vol. 98, pp. 119–132, Dec. 2014.

[158] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, ''Accurate object localization in remote sensing images based on convolutional neural networks,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[159] Y. Xu and P. Ghamisi, ''Universal adversarial examples in remote sensing: Methodology and benchmark,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5619815.

[160] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, ''AID: A benchmark data set for performance evaluation of aerial scene classification,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[161] M. Volpi and V. Ferrari, ''Semantic segmentation of urban scenes by learning local class interactions,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 1–9.

[162] N. Audebert, B. Le Saux, and S. Lefèvre, ''Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images,'' *Remote Sens.*, vol. 9, no. 4, p. 368, Apr. 2017.

[163] A. Du, B. Chen, T.-J. Chin, Y. W. Law, M. Sasdelli, R. Rajasegaran, and D. Campbell, ''Physical adversarial attacks on an aerial imagery object detector,'' in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 3798–3808.

[164] T. N. Mundhenk, G. Konjevod, W. A. Sakla, and K. Boakye, ''A large contextual dataset for classification, detection and counting of cars with deep learning,'' in *Proc. 14th Eur. Conf.* Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 785–800.

[165] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, ''Microsoft COCO: Common objects in context,'' in *Proc. 13th Eur. Conf.* Zurich, Switzerland: Springer, 2014, pp. 740–755.

[166] Y. Zhang, Y. Zhang, J. Qi, K. Bin, H. Wen, X. Tong, and P. Zhong, ''Adversarial patch attack on multi-scale object detection for UAV remote sensing images,'' *Remote Sens.*, vol. 14, no. 21, p. 5298, Oct. 2022.

[167] D. Du, P. Zhu, L. Wen, X. Bian, H. Lin, Q. Hu, T. Peng, J. Zheng, X. Wang, and Y. Zhang, ''VisDrone-DET2019: The vision meets drone object detection in image challenge results,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 213–226.

[168] A. Van Etten, ''The weaknesses of adversarial camouflage in overhead imagery,'' 2022, *arXiv:2207.02963*.

[169] C. Wise and J. Plested, ''Developing imperceptible adversarial patches to camouflage military assets from computer vision enabled technologies,'' 2022, *arXiv:2202.08892*.

[170] W. Zhou, S. Newsam, C. Li, and Z. Shao, ''PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval,'' *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 197–209, Nov. 2018.

[171] J. Lian, S. Mei, S. Zhang, and M. Ma, ''Benchmarking adversarial patch against aerial detection,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5634616.

[172] X. Sun, G. Cheng, L. Pei, H. Li, and J. Han, ''Threatening patch attacks on object detection in optical remote sensing images,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5609210.

[173] G. Cheng, J. Wang, K. Li, X. Xie, C. Lang, Y. Yao, and J. Han, ''Anchor-free oriented proposal generator for object detection,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5625411.

[174] J. Lian, X. Wang, Y. Su, M. Ma, and S. Mei, ''CBA: Contextual background attack against optical aerial detection in the physical world,'' *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5606616.

[175] L. Liu, Z. Xu, D. He, D. Yang, and H. Guo, ''Local pixel attack based on sensitive pixel location for remote sensing images,'' *Electronics*, vol. 12, no. 9, p. 1987, Apr. 2023, doi: 10.3390/ELECTRONICS12091987.

[176] G. Tang, T. Jiang, W. Zhou, C. Li, W. Yao, and Y. Zhao, ''Adversarial patch attacks against aerial imagery object detectors,'' *Neurocomputing*, vol. 537, pp. 128–140, Jun. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231223002989

[177] B. Koonce and B. Koonce, ''Resnet 50,'' in *Convolutional Neural Networks With Swift for Tensorflow: Image Recognition and Dataset Categorization*, 2021, pp. 63–72. [Online]. Available: https://www.oreilly.com/library/view/convolutional-neural-networks/9781484261682/

[178] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, ''Going deeper with convolutions,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[179] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2014, *arXiv:1409.1556*.

[180] Q. Zhang, ''A novel ResNet101 model based on dense dilated convolution for image classification,'' *Social Netw. Appl. Sci.*, vol. 4, no. 1, pp. 1–13, Jan. 2022.

[181] 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.

[182] J. Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, *arXiv:1804.02767*.

[183] G. Jocher, A. Stoken, J. Borovec, A. Chaurasia, L. Changyu, A. Hogan, J. Hajek, L. Diaconu, Y. Kwon, and Y. Defretin, ''Ultralytics/YOLOV5: V5. 0-YOLOV5-p6 1280 models, AWS, supervise. Ly and YouTube integrations,'' Zenodo, Tech. Rep., 2021. [Online]. Available: https://zenodo.org/record/4679653

[184] R. Girshick, ''Fast R-CNN,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[185] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, ''YOLOv4: Optimal speed and accuracy of object detection,'' 2020, *arXiv:2004.10934*.

[186] M. Lee and Z. Kolter, ''On physical adversarial patches for object detection,'' 2019, *arXiv:1906.11897*.

[187] J. Redmon and A. Farhadi, ''YOLO9000: Better, faster, stronger,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[188] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, ''SSD: Single shot multibox detector,'' in *Proc. 14th Eur. Conf.* Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 21–37.

[189] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, ''Swin transformer: Hierarchical vision transformer using shifted windows,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[190] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning RoI transformer for oriented object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2844–2853.

[191] Y. Xu, M. Fu, Q. Wang, Y. Wang, K. Chen, G.-S. Xia, and X. Bai, "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1452–1459, Apr. 2021.

SYED M. KAZAM ABBAS KAZMI received the bachelor's degree in avionics engineering from the National University of Sciences and Technology, Pakistan, in 2008, where he is currently pursuing the master's degree in electrical engineering. His current research interests include deep learning, computer vision, and artificial intelligence.

NAYYER AAFAQ received the B.E. degree (Hons.) in avionics from the College of Aeronautical Engineering (CAE), National University of Sciences and Technology (NUST), Pakistan, in 2007, the M.S. degree (Hons.) in systems engineering from the Queensland University of Technology (QUT), Australia, in 2012, and the Ph.D. degree from the School of Computer Science and Software Engineering (CSSE), The University of Western Australia (UWA). He is currently an Assistant Professor with NUST. His research in computer vision and pattern recognition has published in prestigious venues of the field, including IEEE Computer Vision and Pattern Recognition (CVPR), IEEE Transactions on Multimedia, IEEE Transactions on Artificial Intelligence, IEEE Transactions on Information Forensics and Security, and *ACM Computing Surveys* (ACM CSUR). His current research interests include deep learning, video analysis and intersection of natural language processing (NLP), computer vision (CV), and artificial intelligence. He was a recipient of SIRF Scholarship with UWA. His Ph.D. thesis won Dean's List for Outstanding Thesis Award.

MANSOOR AHMED KHAN received the Ph.D. degree in cryptology from the Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey. He is a Professor and the Head of the Department of Avionics and Electrical Engineering, College of Aeronautical Engineering, National University of Science and Technology, Risalpur. His research publication areas include wireless LAN, cryptographic protocols, cryptography, cyclic redundancy check codes, message authentication, random number generation, deep learning, and telecommunication security.

MOHSIN KHALIL received the B.S. and M.S. degrees in avionics engineering and electrical engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2009 and 2019, respectively. He is currently an Assistant Professor with the National University of Sciences and Technology. His research interests include network performance optimization, 5G networks, deep learning, and green networking.

AMMAR SALEEM received the B.S. degree in avionics engineering from the National University of Science and Technology, Pakistan, in 2007, and the M.S. degree in avionics engineering from Air University Islamabad, Pakistan, in 2013. He is currently pursuing the Ph.D. degree with Sabanci University, Istanbul, Turkey. He is also an Assistant Professor with the College of Aeronautical Engineering, National University of Sciences and Technology. His current research interests include image processing and solving ill-posed inverse problems, especially in context of synthetic aperture radar image reconstruction under the realm of machine learning techniques.