Adversarial Defense in Aerial Detection

Yuwei Chen, Shiyong Chu
Aviation Industry Development Research Center of China
No.14 Xiao Guan Dong Li, Chaoyang District, Beijing, China
catcornic@gmail.com, csy3191dl@163.com

Abstract

The excellent performance of artificial intelligence algorithms in target detection greatly improves the efficiency of detection. However, this alternative to human processing of image information faces many challenges, one of which is adversarial examples (AE). For aerial detection, it is a function widely used in many fields to obtain detection pictures of optical, infrared, and synthetic aperture radars (SAR) from high altitudes to identify ground targets. But in the current research results, optical sensors, infrared sensors, and SAR will be attacked by adversarial patches and perturbation. When these attacks exist, it is risky to let intelligent algorithms perform aerial detection. This paper will focus on the characteristics of each detection mode and propose Adaptive Defense Pipeline (ADP) in addition to improving algorithm robustness through training. According to different weather conditions, the ADP sets the weight coefficients of the detection results of multiple sensors to synthesize the detection results, and on this basis, the second confirmation is added. At the same time, we compare the traditional aerial detection results of a single sensor with the weighted results using ADP and verify that the proposed method could indeed improve the efficiency of aerial detection using artificial intelligence algorithms in an adversarial environment.

1. Introduction

With the further development of computer science, AI algorithm has greatly improved the efficiency of target detection with their excellent and fast computing performance. For aerial detection, it is obviously a necessary means to improve efficiency. In this field, [8] [11] [29] [1] and [30] have all explored the application methods of AI in aerial detection, which could greatly improve the efficiency of target object recognition in aerial detection.

However, with the discovery of adversarial examples [9], the "Achilles heel" of AI has appeared in front of the world, and the weak noise interference of AI will become a key issue affecting the security of AI algorithms. In this field, [2], [31], [27], [15], [32], [39], [12], conducted research on the generation of adversarial examples, adversarial patches, and perturbation. At the same time, [38], [19], [10], [16], [21], [36], [20], [26], [16], [3] began to conduct experiments on AI attacks in digital environments or real-world environments.

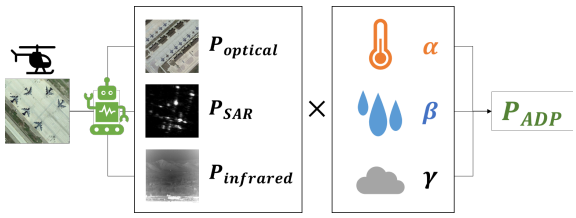

Figure 1. In the Adaptive Defense Pipeline, a variety of sensors are used to detect targets. According to the temperature, rainfall and cloud thickness at that time, weight coefficients are set to generate aerial detection results in the region, which could be used to prevent adversarial attacks to a certain extent.

Therefore, the risk faced by AI algorithms in aerial detection is obvious. The detector with the intelligent algorithm is likely to fail to output correct results. In this field, researchers have carried out research on the attack of AI algorithms in aerial detection, including optical detection attack([22]), infrared imaging attack([7]), SAR imaging attack([13]), *etc*. Each of these attacks will result in a series of irreparable losses, making the decision to apply AI algorithms a difficult trade-off for aerial detection activities.

In order to defend against adversarial attacks, many research teams have proposed defense strategies. [37] proposed the Progressive Diversified Augmentation (PDA) method, which improves the robustness of the AI algorithm by progressively injecting diverse adversarial noises during training. And [17] proposed Adversarial Noise Propagation (ANP), which could be easily combined with other adversarial training methods to further improve the robustness of the model by utilizing the potential of hidden layers. [25] promoted the development of algorithm robustness by studying the relationship between the architecture of the al-

gorithm and the robustness. In addition, [18], [24], [28] are also studying the methods or means of training to improve the robustness of the algorithm in the face of adversarial attacks.

For the aerial detection field, most of which are based on training to improve the robustness of the algorithm. [34] proposed a new adaptive training strategy to train more robust intelligent algorithms in this way. [7] proposed adversarial training for infrared detection imaging, and improved the accuracy of the detection algorithm to a certain extent.

However, we believe that training and development are endless and require a lot of time and resources. Therefore, here we discuss how to utilize the AI algorithm in the existing aerial detection sensors without considering algorithm iterative optimization and carry out the effective defense with minimal cost, so as to give full play to the working efficiency of the AI algorithm in aerial detection. First, we update the traditional single sensor to detect the same target with multiple sensors. Then we analyze the detection performance characteristics of optical sensors, infrared sensors, and SAR sensors under different weather conditions, design the weight coefficient according to these, and form the final detection results based on the detection results of various sensors. Finally, considering the risk that the misjudgment of the result may bring to other subsequent activities, the threshold of manual confirmation is set to ensure that the weighted detection results could reduce the labor cost and misjudgment cost to the greatest extent. In summary, our main contributions are as follows:

- We discuss the vulnerability of the AI algorithm in current aerial detection and identification and the risks it brings.
- We designed the Adaptive Defense Pipeline (ADP) based on the weighted fusion of multiple detection modes and carried out simulation experiments in commercial software to prove the advantages of ADP.

2. Related Work

At the beginning of the application research of AI algorithm, Dudgeon and Lacoss [6] put forward the process of automatic target detection: The sensor collects the signal data, then filters the background noise, and then screens out the non-target information through recognition processing, and finally forms the target list. After the appearance of AE, many research teams in the field have carried out research on adversarial attacks and defenses in the field of aerial detection.

2.1. Adversarial Attacks in Aerial Detection

Among the attacks against different types of sensors, Edwards and Rawat [7] studied the infrared detection attack, Li et al. [13] explored the influence of adversarial attacks on SAR image recognition, Lu et al. [22], Xu and Ghamisi [33] and den Hollander et al. [5] explored the AE of optical remote sensing images. Chen et al. [4] researched the adversarial attack of remote sensing images and concluded that the average probability of the DNN algorithm being fooled in SAR is 76.01%, and the average probability of optical data set being fooled is 60.28%. Furthermore, Lian et al. [14] proposed an adaptive-patch-based physical attack (AP-PA) framework to enhance the attack adaptability of the defensive patch in a complex and real physical environment. At the same time, for different detectors, Lian et al. [14] have done experiments to prove that it could maintain high attack efficiency.

2.2. Adversarial Defenses in Aerial Detection

From the perspective of training, Xu et al. [34] put forward the defense mode of aerial detection intelligent algorithm against adversarial attacks, and directly used the generated AE to carry out training, to improve the robustness of the intelligent algorithm in the face of adversarial attacks. Edwards and Rawat [7] used adversarial training to enhance the defense of AE in the field of aerial infrared detection. Raja et al. [23] also proposed to improve the robustness of unmanned aerial vehicles (UAV) to AE through training. It could be seen that the current research focuses on improving the robustness of algorithms through training, but there is not much research on how to improve the robustness of existing intelligent algorithms to AE. Xue et al. [35] located the AE position in the aerial detection image, removed the poisoning part and redrawn the image, and then transmitted it to the subsequent algorithm for identification. In this way, the robustness of detection was improved. However, there is not much research on the robustness of other types of sensors except optical sensors.

3. Adversarial Attacks on Different Sensors

In aerial detection, infrared sensors, SAR sensors, optical sensors, *etc*. are often used to collect imagery, and then the target is classified and identified by using the characteristic values of the target object. All types of sensors could be attacked in adversarial environments.

3.1. Optical Sensor

Optical sensor based on visible light detection is the most common means in aerial detection, and it is also a major topic of the adversarial attack research field. Compared with other imaging methods, optical images could provide more details of objects. However, because the imaging principle of the optical sensor is greatly influenced by illumination, optical detection needs sufficient illumination and high visibility to play its maximum effectiveness. At the same time, for optical sensors, the imaging angle will also affect

the extraction of target features, thus affecting the detection efficiency of targets.

A considerable number of adversarial attacks could affect the optical sensor, such as sticking AE on or near the target object, which could attack the identification module of the optical sensor. Under the natural environment suitable for optical detection, adversarial information could also be the better image in the sensor. Therefore, there is a high probability that the optical sensor will be attacked by AE.

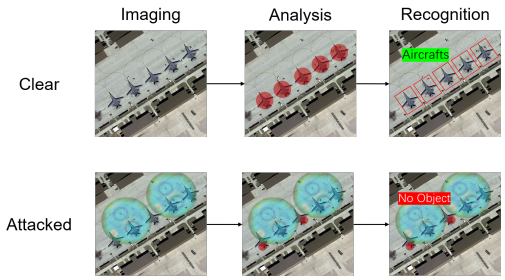

Figure 2. Taking the optical imaging of aerial detection as an example, when the image is clean, targets could be successfully identified through the imaging of the sensor and the feature analysis. When AE appears in the image, the algorithm will be confused, which will eventually lead to the failure of target detection.

3.2. Infrared Sensor

The infrared sensors could work under all-day conditions, but the detection ability of the image will decrease when it meets the weather conditions of fog and haze. Theoretically, any object whose temperature is above absolute zero will radiate infrared rays. Theoretically, any object with a temperature above absolute zero will radiate infrared rays. The higher the temperature, the shorter the wavelength, and the larger the volume, the stronger the radiation.

For the adversarial attacks, according to the research[7], because the infrared image is a single channel image, the attack methods for RGB could also attack the infrared image. Therefore, the attackers could post or place the AE suitable for attacking RGB images on or near the unit to be concealed, and then transmit the attack information to the intelligent algorithm when the infrared sensor detects the area, and then complete the attack.

3.3. SAR Sensor

SAR is a microwave active sensor, which could penetrate clouds, rain, snow, and smoke, and could detect all day and long distances. It has a penetrating ability to disguise a certain amount of cover, and at the same time, it has an excellent performance in detecting surface texture characteristics and artificial metal objects. When the target object is discovered, the type of the object is judged by the scattering mechanism of each component.

For the SAR sensor itself, in the high-resolution SAR image, the complex background will interfere with the accurate identification of the target in the SAR image, and a large number of background highlight scattering points are distributed around the target, which makes it difficult to accurately locate and identify. Therefore, before the SAR sensor performs object and target detection, it will first perform false alarm elimination activities.

However, the filtering of clutter could not effectively filter adversarial information such as perturbation on the target. When the target object changes its reflexivity to carry perturbation or adversarial information by updating the material or changing the surface texture, the image formed by the SAR sensor will disturb the subsequent target detection and cause the target identification to fail.

Table 1. Characteristics of various types of aerial detection sensors

Type	Temperature	Rainfall	Cloud
Optical	Almost irrelevant	Rainfall↑ Accuracy↓	Cloud↑ Accuracy↓
Infrared	Temperature↑ Sensitivity↑	Rainfall↑ Accuracy↓	Cloud↑ Accuracy↓
SAR	Almost irrelevant	Almost irrelevant	Almost irrelevant

4. Adaptive Defense Pipeline

In order to improve the robustness of intelligent algorithms against adversarial attacks in aerial detection, most research teams start with the continuous training of algorithms. In addition to this method, we design ADP with the current algorithm unchanged.

4.1. Using Multiple Sensors

By clarifying the detection target area and the target type that might be involved, aerial detection could set the relevant detection function parameters in advance. In a single detection mission, a variety of sensors are used to detect the target, and the detection results of various sensors are obtained.

4.2. Setting Weight Coefficients

Different meteorological characteristics, such as temperature, rainfall, and cloud thickness, will affect the sensor. The weight coefficient is set according to the current meteorological setting of the detection area. Set α as the weight coefficient of the optical sensor's imaging results, set β as the weight coefficient of the infrared sensor's imaging results, and γ as that of SAR. The sum of these three coeffi-

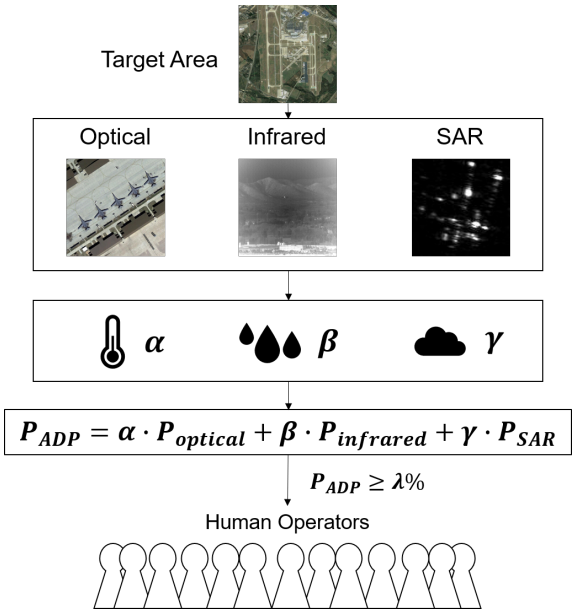

Figure 3. Schematic diagram of ADP. After the detection target range is determined, the weights are set according to weather, time, and other factors, and then all kinds of sensors detect the area at the same time. The results of detection and identification are fused according to the weights. When the overall target probability is higher than a certain value, it will be transferred to human-in-the-loop processing.

cients should be 1, that is:

$$\alpha + \beta + \gamma = 1 \tag{1}$$

The weight coefficients could adaptive change according to each meteorological feature. For example, when the temperature rises, β should increase; When rainfall increases, α and β should decrease; And α and β should decrease as cloud thickness increases. The weight coefficient of this aerial exploration mission could be calculated by combining the percentage increase and decrease of all items and the equation.

4.3. Fusing Results and Confirming

After the target detection result is given by a single detector, the fusion result is generated by weighting according to the weight coefficient set in the previous method4.2. Set the result of the optical sensor to identify a certain target here as $P_{optical}$, the result of the infrared sensor as $P_{infrared}$, and the result of the SAR sensor as P_{SAR}, the final detection result's confidence P_{ADP} follows the equation1.

$$P_{ADP} = \alpha \times P_{optical} + \beta \times P_{infrared} + \gamma \times P_{SAR} \tag{2}$$

After the fusion result P_{ADP} comes out, according to the determination of the leader or commander, and considering the risk and cost that could be borne by misjudgment, a threshold λ is set. When $P_{ADP} \geq \lambda$, it enters the confirmation part of human-in-the-loop; When $P_{ADP} < \lambda$, it could be considered that there are no potential targets in this area, and the aerial detection mission in this area is over.

After entering the Human-in-the-loop confirmation, the human operators would confirm the potential targets again. Human operators could judge and confirm the detected targets by using their own experience and according to the filtered typical images. At this time, human operators do not need to face massive image information but only need to examine the images filtered by AI.

5. Experiments

5.1. Experiment Settings

Using "Command Modern Operation" software, we define the Defensive Side and the Offensive Side. In the experiment, the Offensive Side needs to detect a certain area of the Defensive Side.

We set the temperature to a random integer between 0 degrees Celsius and 25 degrees Celsius. Set the rainfall scale as a random integer from 0 to 40, with 0 indicating clear weather and 40 indicating Heavy Storm. Set the cloud thickness rating as a random integer from 0 to 10, with 0 indicating that the sky is clear and 10 indicating that the sky is full of thick clouds.

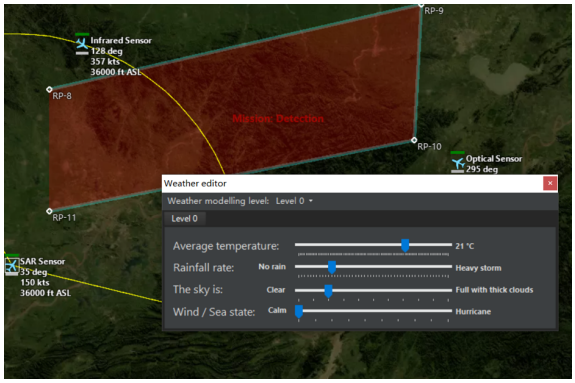

Figure 4. Temperature, rainfall, cloud thickness, and other levels are set in the software to control the experimental conditions of aerial detection.

On this basis, when the temperature rises 1 degree from 0 degrees, the probability of successful detection of the infrared sensor increases by 4%, and the weight increases by 4%. When the rainfall level increases by 1 level, the successful probability of the infrared sensor and optical sensor decreases by 2% and the weight decreases by 2%. When the cloud thickness level increases by 1 level, the successful probability of the infrared sensor and optical sensor decreases by 10% and the weight decreases by 10%.

At the same time, on the basis of using the software's

own general airborne optical sensor, infrared sensor, and SAR, set the detection success rate coefficient is between 30% and 60% random number under adversarial attacks. Let the operator believe there is a target in the area when the probability exceeds 45%.

We conducted 10 experiments to compare the average probability of success of a single type of sensor in aerial detection with the average probability of success of the ADP under the same meteorological conditions and adversarial attacks.

5.2. Experiments on Single Type Sensor

In such experiments, only a single type of sensor is considered to carry out target detection in the Defensive Side's area.

For the same target, the aircraft carries an optical sensor, infrared sensor, and SAR sensor to the target area for detection. According to the characteristics of various sensors under weather conditions, the probability of success is modified.

The successful detection probability interval of the optical sensor under adversarial conditions is $P_{optical}$, which is set to a random value in the range of $(40 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1)$ to $(60 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1)$. The successful detection probability interval of the infrared sensor is $P_{infrared}$, which is set to a random value in the range of $(40 + C_{Temp} \times 0.04 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1)$ to $(60+ C_{Temp} \times 0.04 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1)$, and the successful detection probability interval of SAR sensor is P_{SAR}, which is set to a random value in the range of 40 to 60.

Among them, C_{Temp} represents the temperature, C_{Rain} represents the rainfall level, and C_{Cloud} represents the cloud thickness level.

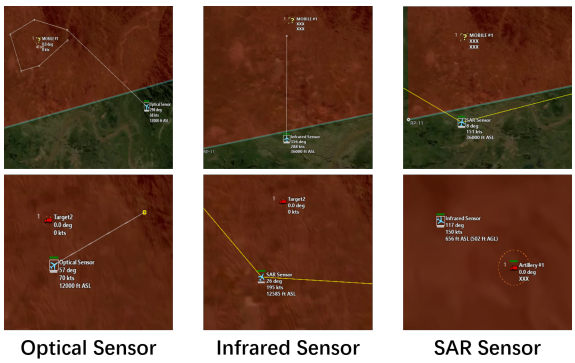

Figure 5. Under different weather conditions, we use optical sensors, infrared sensors, and SAR to detect the same target on the Defense Side.

Each type of sensor conducts 10 detection missions on the same target under 10 random weather conditions. The detection results are shown in the table 2.

Table 2. Single sensor experimental results

Conditions (Temperature/Rainfall/Cloud)	$P_{optical}$	$P_{infrared}$	P_{SAR}
21 / 19 / 6	41%	39%	55%
20 / 15 / 3	44%	48%	56%
21 / 26 / 10	49%	55%	60%
15 / 19 / 8	53%	50%	51%
22 / 4 / 1	49%	45%	47%
9 / 35 / 6	50%	49%	53%
3 / 13 / 5	41%	39%	56%
23 / 4 / 10	41%	53%	43%
11 / 37 / 10	43%	40%	43%
21 / 11 / 2	50%	44%	45%

5.3. Experiments with ADP

Based on the experiments 5.2 of detecting the target with a single sensor, the weight coefficient is set according to the weather conditions.

Table 3. Adjustment of weight coefficients under weather conditions

Sensor type	Optical	Infrared	SAR
Temperature	1	$1 + C_{Temp} \times 0.04$	1
Rainfall	$1 - C_{Rain} \times 0.02$	$1 - C_{Rain} \times 0.02$	1
Cloud	$1 - C_{Cloud} \times 0.1$	$1 - C_{Cloud} \times 0.1$	1
Total	$3 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1$	$3 + C_{Temp} \times 0.04 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1$	3

Therefore, according to the table 3, the calculation formula of each weighting coefficient is as follows:

$$Total = 9 + C_{Temp} \times 0.04 - C_{Rain} \times 0.04 - C_{Cloud} \times 0.2 \tag{3}$$

$$\alpha = \frac{3 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1}{Total} \tag{4}$$

$$\beta = \frac{3 + C_{Temp} \times 0.04 - C_{Rain} \times 0.02 - C_{Cloud} \times 0.1}{Total} \tag{5}$$

$$\gamma = \frac{3}{Total} \tag{6}$$

The final weighted probability of successful detection is P_{ADP}, which could be calculated by equation 2.

When the detection result rate P_{ADP} is more than 45%, enter the human-in-the-loop confirming step. It is assumed that after entering the human-in-the-loop step, the efficiency of final aerial detection could be maximized.

Based on the test results of a single sensor, the weight setting and the final weighted detection results are shown in the table 4.

Table 4. Experimental results after using ADP

Conditions (Temperature/Rainfall/Cloud)	α	β	γ	P_{ADP}
21 / 19 / 6	0.256	0.363	0.381	45.60%
20 / 15 / 3	0.279	0.372	0.349	49.67%
21 / 26 / 10	0.218	0.341	0.441	55.90%
15 / 19 / 8	0.252	0.334	0.414	51.17%
22 / 4 / 1	0.296	0.389	0.315	46.82%
9 / 35 / 6	0.251	0.305	0.444	51.03%
3 / 13 / 5	0.295	0.310	0.395	46.30%
23 / 4 / 10	0.247	0.366	0.387	46.16%
11 / 37 / 10	0.211	0.285	0.503	42.14%
21 / 11 / 2	0.287	0.380	0.333	46.05%

5.4. Analysis of Results

If the final threshold is set at 45%, it could be clearly seen from the table that in 10 experiments, the optical sensor found the target 5 times, the infrared sensor found the target 6 times, and the SAR sensor found the target 8 times. However, weighted by the ADP algorithm, targets in this region could be found 9 times under the same meteorological conditions, indicating that ADP could guarantee the robustness of intelligent detection results even under adversarial attacks to a certain extent.

6. Conclusion

In this study, we explored the ADP besides improving the robustness of the model algorithm through training. Considering the influence of meteorological factors such as temperature, rainfall, and cloud thickness on various types of sensors during aerial detection, we set the weight coefficients of optical sensors, infrared sensors, and SAR, and fused them to generate the final detection results. According to the fusion results of target detection information, the human operator makes the secondary confirmation, which ensures the robustness of target detection under adversarial attack to a certain extent.

We believe that if we could continue to train algorithms on the basis of ADP, we could further improve the robustness of intelligent algorithms for aerial detection and provide better solutions for intelligent aerial detection.

7. Broader Impacts

We explore the defense method of aerial detection when the current intelligent algorithm faces adversarial attacks, in order to find an emergency option outside the training algorithm, that is, how to improve the robustness in the face of adversarial attacks without time and resources to train the algorithm. We believe that this is a new possible choice besides the training algorithm, so as to reduce the influence of adversarial attacks on the application of intelligent algorithms and promote the AI algorithms to provide services for humans more widely. In the meantime, we will continue to explore the addition of training algorithms on the basis of ADP to bring more contributions to robustness.

References

[1] Sourav Kumar Bhoi, Kalyan Kumar Jena, Sanjaya Kumar Panda, Hoang Viet Long, Raghvendra Kumar, P Subbulakshmi, and Haifa Bin Jebreen. An internet of things assisted unmanned aerial vehicle based artificial intelligence model for rice pest detection. *Microprocessors and Microsystems*, 80:103607, 2021. 1

[2] Tom B Brown, Dandelion Mané, Aurko Roy, Martín Abadi, and Justin Gilmer. Adversarial patch. *arXiv preprint arXiv:1712.09665*, 2017. 1

[3] Nicholas Carlini and David Wagner. Audio adversarial examples: Targeted attacks on speech-to-text. In *2018 IEEE security and privacy workshops (SPW)*, pages 1–7. IEEE, 2018. 1

[4] Li Chen, Zewei Xu, Qi Li, Jian Peng, Shaowen Wang, and Haifeng Li. An empirical study of adversarial examples on remote sensing image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7419–7433, 2021. 2

[5] Richard den Hollander, Ajaya Adhikari, Ioannis Tolios, Michael van Bekkum, Anneloes Bal, Stijn Hendriks, Maarten Kruithof, Dennis Gross, Nils Jansen, Guillermo Perez, et al. Adversarial patch camouflage against aerial detection. In *Artificial Intelligence and Machine Learning in Defense Applications II*, volume 11543, pages 77–86. SPIE, 2020. 2

[6] Dan E Dudgeon and Richard T Lacoss. An overview of automatic target recognition. *The Lincoln Laboratory Journal*, 6(1):3–10, 1993. 2

[7] DeMarcus Edwards and Danda B Rawat. Study of adversarial machine learning with infrared examples for surveillance applications. *Electronics*, 9(8):1284, 2020. 1, 2, 3

[8] Luis F Gonzalez, Glen A Montes, Eduard Puig, Sandra Johnson, Kerrie Mengersen, and Kevin J Gaston. Unmanned aerial vehicles (uavs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors*, 16(1): 97, 2016. 1

[9] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014. 1

[10] Qian Huang, Isay Katsman, Horace He, Zeqi Gu, Serge Belongie, and Ser-Nam Lim. Enhancing adversarial example transferability with an intermediate level attack. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4733–4742, 2019. 1

[11] Lihua Jian, Zhen Li, Xiaomin Yang, Wei Wu, Awais Ahmad, and Gwanggil Jeon. Combining unmanned aerial vehicles with artificial-intelligence technology for traffic-congestion recognition: electronic eyes in the skies to spot clogged roads. *IEEE Consumer Electronics Magazine*, 8(3):81–86, 2019. 1

[12] Alexey Kurakin, Ian J Goodfellow, and Samy Bengio. Adversarial examples in the physical world. In *Artificial intelligence safety and security*, pages 99–112. Chapman and Hall/CRC, 2018. 1

[13] Haifeng Li, Haikuo Huang, Li Chen, Jian Peng, Haozhe Huang, Zhenqi Cui, Xiaoming Mei, and Guohua Wu. Adversarial examples for cnn-based sar image classification: An experience study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:1333–1347, 2020. 1, 2

[14] Jiawei Lian, Shaohui Mei, Shun Zhang, and Mingyang Ma. Benchmarking adversarial patch against aerial detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. 2

[15] Aishan Liu, Xianglong Liu, Jiaxin Fan, Yuqing Ma, Anlan Zhang, Huiyuan Xie, and Dacheng Tao. Perceptual-sensitive gan for generating adversarial patches. In *33rd AAAI Conference on Artificial Intelligence*, 2019. 1

[16] Aishan Liu, Tairan Huang, Xianglong Liu, Yitao Xu, Yuqing Ma, Xinyun Chen, Stephen Maybank, and Dacheng Tao. Spatiotemporal attacks for embodied agents. In *European Conference on Computer Vision*, 2020. 1

[17] Aishan Liu, Xianglong Liu, Hang Yu, Chongzhi Zhang, Qiang Liu, and Dacheng Tao. Training robust deep neural networks via adversarial noise propagation. *IEEE Transactions on Image Processing*, 2021. 1

[18] Aishan Liu, Xianglong Liu, Hang Yu, Chongzhi Zhang, Qiang Liu, and Dacheng Tao. Training robust deep neural networks via adversarial noise propagation. *IEEE Transactions on Image Processing*, 30:5769–5781, 2021. 2

[19] Aishan Liu, Jun Guo, Jiakai Wang, Siyuan Liang, Renshuai Tao, Wenbo Zhou, Cong Liu, Xianglong Liu, and Dacheng Tao. X-adv: Physical adversarial object attacks against x-ray prohibited item detection. *arXiv preprint arXiv:2302.09491*, 2023. 1

[20] Shunchang Liu, Jiakai Wang, Aishan Liu, Yingwei Li, Yijie Gao, Xianglong Liu, and Dacheng Tao. Harnessing perceptual adversarial patches for crowd counting. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, pages 2055–2069, 2022. 1

[21] Yanpei Liu, Xinyun Chen, Chang Liu, and Dawn Song. Delving into transferable adversarial examples and black-box attacks. *arXiv preprint arXiv:1611.02770*, 2016. 1

[22] Mingming Lu, Qi Li, Li Chen, and Haifeng Li. Scale-adaptive adversarial patch attack for remote sensing image aircraft detection. *Remote Sensing*, 13(20):4078, 2021. 1, 2

[23] Ashok Raja, Laurent Njilla, and Jiawei Yuan. Adversarial attacks and defenses toward ai-assisted uav infrastructure inspection. *IEEE Internet of Things Journal*, 9(23):23379–23389, 2022. 2

[24] Shiyu Tang, Ruihao Gong, Yan Wang, Aishan Liu, Jiakai Wang, Xinyun Chen, Fengwei Yu, Xianglong Liu, Dawn Song, Alan Yuille, et al. Robustart: Benchmarking robustness on architecture design and training techniques. *arXiv preprint arXiv:2109.05211*, 2021. 2

[25] Shiyu Tang, Siyuan Liang, Ruihao Gong, Aishan Liu, Xianglong Liu, and Dacheng Tao. Exploring the relationship between architecture and adversarially robust generalization. *arXiv preprint arXiv:2209.14105*, 2022. 1

[26] Florian Tramer, Nicholas Carlini, Wieland Brendel, and Aleksander Madry. On adaptive attacks to adversarial example defenses. *Advances in neural information processing systems*, 33:1633–1645, 2020. 1

[27] Jiakai Wang, Aishan Liu, Zixin Yin, Shunchang Liu, Shiyu Tang, and Xianglong Liu. Dual attention suppression attack: Generate adversarial camouflage in physical world. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 1

[28] Jiakai Wang, Zixin Yin, Pengfei Hu, Aishan Liu, Renshuai Tao, Haotong Qin, Xianglong Liu, and Dacheng Tao. Defensive patches for robust recognition in the physical world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2456–2465, 2022. 2

[29] Jinwang Wang, Wen Yang, Haowen Guo, Ruixiang Zhang, and Gui-Song Xia. Tiny object detection in aerial images. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3791–3798. IEEE, 2021. 1

[30] Jason Whelan, Abdulaziz Almehmadi, and Khalil El-Khatib. Artificial intelligence for intrusion detection systems in unmanned aerial vehicles. *Computers and Electrical Engineering*, 99:107784, 2022. 1

[31] Chaowei Xiao, Bo Li, Jun-Yan Zhu, Warren He, Mingyan Liu, and Dawn Song. Generating adversarial examples with adversarial networks. *arXiv preprint arXiv:1801.02610*, 2018. 1

[32] Chaowei Xiao, Jun-Yan Zhu, Bo Li, Warren He, Mingyan Liu, and Dawn Song. Spatially transformed adversarial examples. *arXiv preprint arXiv:1801.02612*, 2018. 1

[33] Yonghao Xu and Pedram Ghamisi. Universal adversarial examples in remote sensing: Methodology and benchmark. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–15, 2022. 2

[34] Yonghao Xu, Bo Du, and Liangpei Zhang. Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: Attacks and defenses. *IEEE Transactions on Geoscience and Remote Sensing*, 59 (2):1604–1617, 2020. 2

[35] Wei Xue, Zhiming Chen, Weiwei Tian, Yunhua Wu, and Bing Hua. A cascade defense method for multidomain adversarial attacks under remote sensing detection. *Remote Sensing*, 14(15):3559, 2022. 2

[36] Hiromu Yakura and Jun Sakuma. Robust audio adversarial example for a physical attack. *arXiv preprint arXiv:1810.11793*, 2018. 1

[37] Hang Yu, Aishan Liu, Gengchao Li, Jichen Yang, and Chongzhi Zhang. Progressive diversified augmentation for general robustness of dnns: A unified approach. *IEEE Transactions on Image Processing*, 2021. 1

[38] Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li. Adversarial examples: Attacks and defenses for deep learning. *IEEE transactions on neural networks and learning systems*, 30 (9):2805–2824, 2019. 1

[39] Zhengli Zhao, Dheeru Dua, and Sameer Singh. Generating natural adversarial examples. *arXiv preprint arXiv:1710.11342*, 2017. 1