OBJECT DETECTION MODELS SENSITIVITY & ROBUSTNESS TO SATELLITE-BASED ADVERSARIAL ATTACKS

Jade Eva Guisiano, Domenico Barretta, Éric Moulines, Thomas Lauvaux, Jérémie Sublime

OBJECT DETECTION MODELS SENSITIVITY & ROBUSTNESS TO SATELLITE-BASED ADVERSARIAL ATTACKS

Jade Eva Guisiano[1,2,3,5], Domenico Barretta[4], Éric Moulines[2], Thomas Lauvaux[5] and Jérémie Sublime[1]

1. ISEP School of Digital Engineers, 2. École Polytechnique, 3. United Nations Environment Program,
4. University of Campania "Luigi Vanvitelli", 5. University of Reims-Champagne Ardenne

ABSTRACT

The use of object detection algorithms for the analysis of satellite imagery is increasing in various fields, including environment and defense, as they enable the automatic detection, recognition and localization of targets. Satellite images often exhibit significant variations, including differences in resolution and noise levels between different satellites. Additional distortions can be caused by factors such as the position of the satellite and the specific area being scanned, resulting in changes in tangential distortion, brightness and saturation. Depending on the severity, these variations can affect the visual clarity of objects in the images and thus impair the effectiveness of object detection algorithms. This study therefore investigates the effects of such fluctuations on the performance of 3 categories of object recognition algorithms - YOLO, FASTER-RCNN and RT-DETR - by applying the principle of adversarial attacks to the inference phase of the algorithms. This experiment makes it possible to uncover the weaknesses of the algorithms and then provides information on how these models could be improved to be more robust to variations in satellite imagery. The case study presented is based on the automatic detection of 3 types of oil and gas infrastructure: compressor, tank and well in the Permian Basin (USA).

Index Terms— Object Detection, Deep learning, Remote Sensing, Adversarial Attacks, Oil & Gas

1. INTRODUCTION

Object detection algorithms are pivotal in enabling the automated detection and recognition of specific objects within images. These algorithms are particularly useful for tasks such as automatic counting and the tracking of objects over time. In the realm of remote sensing, they have a broad spectrum of applications, ranging from environmental monitoring [1, 2, 3] to defense purposes [4, 5, 6]. Object detection algorithms can be broadly classified into 3 main categories: one-stage, two-stage, and transformer encoder-decoder architectures. Additionally, they are versatile in their application, being compatible with various learning methods, including supervised, semi-supervised, and self-supervised learning approaches. In this study, we focus on supervised learning, with particular emphasis on the algorithms YOLO (one-stage) [7], FASTER-RCNN (two-stage) [8] and RT-DETR (transformer) [9]. In supervised learning, a database of images labeled with objects is created so that these algorithms can learn to recognize and to retrieve specific objects. However, a notable challenge with this approach, especially in the context of remote sensing, is the limited representational diversity of the objects in the training database. For example, satellite images exhibit a wide range of variations due to different satellite sensor types (e.g. resolution, noise) and environmental factors related to the position of the satellite and the sampled area (e.g. tangential distortion, brightness and saturation) [10]. These variations can significantly affect the algorithm's ability to generalize across different representations of the same object [11, 12].

The aim of this study is to evaluate the robustness of YOLO, FASTER-RCNN, and RT-DETR against five specific types of variations commonly found in satellite imagery. To achieve this, we used the concept of adversarial attack during the inference phase as described in previous studies [13, 14, 15, 16]. In this method, the accuracy of both the pre-trained and fine-tuned algorithms is evaluated when they are presented with modified images (counterexamples) containing different types of induced perturbations. For each of the five types of variations, we systematically generated stepwise negative examples and then tested the performance of the algorithms for each sub-variation. The case study presented in this paper focuses on the automatic detection of 3 types of oil and gas infrastructure in the Permian Basin, USA – compressor, tank, and well – as shown in Figure 1.

In the first part, the selected object detection algorithms and their pre-training/fine-tuning are presented. Then the satellite variations and their adversarial example are defined and described in details. Finally, the results of the experiments are discussed.

(a) Tank (b) Compressor (c) Well

Fig. 1. Base test images of O&G infrastructures.

2. OBJECT DETECTION ALGORITHMS

Object recognition algorithms are used to automatically identify and to localize multiple instances of specific object classes in images and videos. For effective training, supervised models require a database of labeled images. Labeling in this context involves marking objects of interest with bounding boxes and assigning appropriate labels to them. During the training phase, the algorithm learns both the localization and the recognition of the target objects. This learning process can be implemented with different algorithm architectures, including single-stage, two-stage and transformer models (encoder-decoder).

Models architectures: The You Only Look Once (YOLO) architecture, a one-stage object recognition model, was originally introduced by Redmon et al. [7]. Among the various iterations, YOLO version 8 (v8) is notable for its advanced precision and recent development [17]. YOLO v8 includes five pre-trained models (denoted as n, s, m, l, x), each trained on the COCO 2017 dataset [18]. Remarkable results from previous research [19], which are particularly relevant to this study, show that the model m shows superior performances compared to its counterparts.

FASTER-RCNN, which is categorized as a two-stage algorithm, is presented in [8]. This algorithm provides 3 different backbone architectures, all trained with the COCO 2017 database: the Feature Pyramid Network (FPN), the ResNet conv4 backbone with a conv5 head (C4), and the ResNet conv5 backbone with dilations in conv5 (DC5). Previous work by [19] - in a similar environment to our study - shows that the FPN architecture in combination with ResNet101 is the most effective configuration.

RT-DETR [9], an extended version of DETR [20], differs from one- and two-stage detectors by considering object recognition as a direct set prediction problem within a unified architecture. RT-DETR is based on a transformer-encoder-decoder framework and provides two pre-trained models based on the COCO 2017 database: large and extra-large. Empirical studies have shown that the large model outperforms the extra-large model in terms of performance.

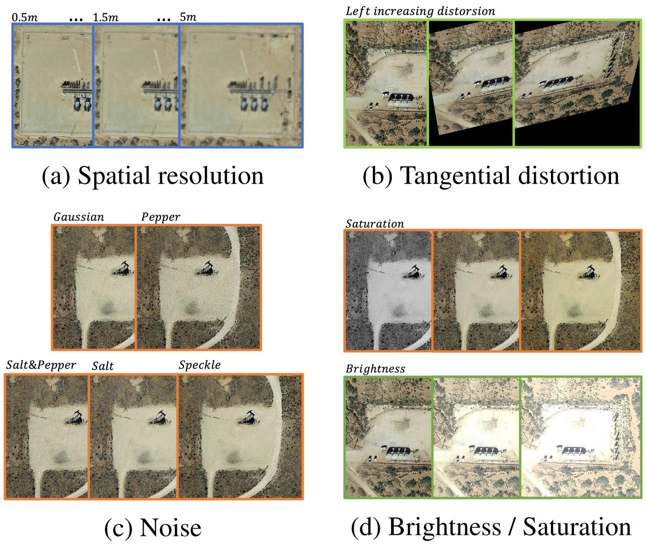
Fine-tuning: In this study, we compare 3 categories of object detection models: single-stage (YOLO v8, model ”m”), two-stage (FASTER-RCNN, model ”R101-FPN”) and transformer-encoder-decoder (RT-DETR, model ”l”). Each model was fine-tuned for specific object detection in oil and gas infrastructure after initial pre-training in the COCO database, with a focus on compressors, wells and tanks. The study utilized the specially curated OG database, which includes 930 of high-resolution aerial imagery from the Permian Basin, USA, with a total of 1951 annotated instances. Fine-tuning was performed using an NVIDIA GeForce RTX 3090 GPU. For the implementation, YOLO v8 and RT-DETR used Ultralytics, while FASTER-RCNN used Detectron2. Consistent learning rates and epochs were set for all models, with batch sizes adjusted accordingly. The performance metrics showed that YOLO v8 led with a mean average precision (mAP) of 92.6, RT-DETR with 86.4 and FASTER-RCNN with 48.8.

3. SATELLITE-BASED ADVERSARIAL ATTACKS

While the 3 models fine-tuned with satellite imagery from the OG database — in particular YOLO v8 and RT-DETR — show promising accuracy, it is crucial to assess their robustness for practical remote sensing. Our study primarily targets the detection and recognition of three oil and gas infrastructures: Compressors, wells and tanks in a variety of satellite images. However, these images are subject to a number of variations that affect their visual representation. Technological differences between satellite sensors can lead to variations in spatial resolution and noise levels. In addition, systematic errors can occur due to the Earth’s rotation, geometric distortions caused by topography shifts, variations in satellite altitude and attitude, and instrument anomalies. Furthermore, environmental factors such as the nature of the terrain (e.g. deserts) can influence image attributes such as luminosity and saturation.

To assess the robustness of our algorithms against variations in the satellite images, we implemented adversarial attacks [13, 14, 15, 16]. In this method, the precision of the 3 pre-trained and fine-tuned algorithms is evaluated in response to various perturbations generated as negative examples on the input images. Such a process is crucial to identify weaknesses in the models and suggest areas for refinement. For example, a notable drop in precision or a complete failure to recognise an object when confronted with certain adversarial examples could indicate a lack of resilience to that specific variation. We generated adversarial examples for each type of ”base” image that mimic variations in satellite imagery, such as changes in resolution, tangential distortion, noise, brightness, and saturation.

Spatial resolution variations. Different satellites, equipped with different imaging systems and specifications, have different spatial resolutions. These differences result from factors such as sensor types, technological advances and the intended functions of the satellites. To mimic these resolution differences, as shown in Figure 2a, we applied average pool-



(a) Spatial resolution (b) Tangential distortion

(c) Noise (d) Brightness / Saturation

Fig. 2. Adversarial examples.

ing with a square window (kernel size of 5 and stride of 3). In this way, ten images with resolutions ranging from $0.5m$ to $5m$ were generated, which are shown in Figure 1. This range mirrors the resolution spectrum of actual satellite images and thus allows a thorough evaluation of the performance of the object detection algorithms at different spatial resolutions.

Tangential Distortion. Tangential scale distortions, common in satellite imagery, result from the compression of image features, especially those further from the nadir point. Factors contributing to this distortion include sensor optics, the scanning system's motion, Earth's curvature and rotation, and terrain relief variations. To emulate varying levels of this distortion as depicted in Figure 2b, we employed a left affine perspective transformation using the OPENCV library. This technique involved a transformation matrix that maps 3 points from the original image to their new positions in the distorted image. We progressively adjusted the value of the left element in the matrix's third line from 50 down to 0, in steps of 5. This method enabled a controlled increase in leftward distortion across the images, providing a framework to evaluate the algorithms' effectiveness under different degrees of tangential scale distortions.

Noise.

As shown in Figure 2c, the different on-board sensors of each satellite can cause a range of noise types. These include additive (Gaussian), multiplicative (speckle) and impulsive (salt and pepper) noise. To replicate Gaussian noise, we added a normally distributed random value to each pixel. For Salt and Pepper noise, we randomly altered pixels to extreme values (0 for dark, 1 for bright). Speckle noise, being multiplicative, was created using the formula out $=$ image $+ n \times$ image, with N as uniform noise defined by a specific mean and variance. In this study, we simulated these 3 noise types, as well as the isolated effects of salt (bright) and pepper (dark) values. Such a methodology enables a thorough assessment of the object detection algorithms' resilience against various noise types commonly found in satellite images.

Brightness & Saturation. Satellite images are influenced by various factors, including atmospheric conditions, optical properties, sensor characteristics and data processing techniques. For example, atmospheric scattering can reduce color intensity, resulting in lower saturation. Min addition, certain geographical areas, such as deserts and polar regions, can influence the image properties due to the high solar reflectance, which leads to increased brightness [21](Figure 2d). To replicate different brightness levels, we used a function from the OPENCV Python library parameterized with $\alpha \in [0, 1]$ and $\beta \in [-127, 127]$. We set α to 1 and varied β from 0 to 100 in steps of ten, with lower β values corresponding to lower brightness. To simulate different degrees of saturation, we used a function from the Pillow Python library with a single parameter. This saturation parameter ranged from 0 (colourless image) to 1.5 and was increased in steps of 0.15. These methodologies enabled us to evaluate the object detection algorithms' robustness against a spectrum of brightness and saturation levels, reflecting the diverse conditions encountered in real-world satellite imagery.

Results:

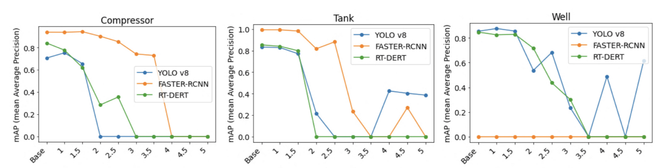


Fig. 3. Impact of tangential resolution variations (from $0.5m$ to $5m$).

Figure 3 illustrates the variations in mean Average Precision (mAP) for each object detection model as a function of spatial resolution changes, focusing on 3 types of infrastructures (compressors, tanks, and wells). We set the mAP to zero in instances where the models fail to detect and recognize the infrastructures. The effectiveness of the algorithms in recognising these objects varies with the resolution. YOLO v8 in particular can detect compressors at resolutions of up to $1.5m$, whereas FASTER-RCNN extends this capability up to $4m$. For tanks and wells, YOLO v8 successfully detects and recognizes them at resolutions as high as $5m$, with mAPs fluctuating between 40% and 60%. It's important to mention that FASTER-RCNN consistently fails to detect wells, even in the base image. Furthermore, we observe that compressors are more challenging to detect at lower resolutions (detected up to $4m$), in contrast to tanks and wells, which are detectable up to $5m$. Figure 4 shows that FASTER-RCNN exhibits remarkable consistency in its performance, with minimal sensitivity to tangential distortions; the mean Average Precision

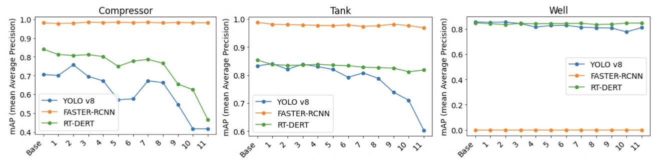


Fig. 4. Impact of tangential distortion variations (left gradual stretching from level 1 to 11).

(mAP) for compressors and tanks remains relatively stable across the different distortion levels. In contrast, YOLO v8 exhibits more significant fluctuations in mAP in response to these distortions. Both YOLO and RT-DETR show a notable decrease in performance at the 5th level of distortion when detecting and recognizing compressors. In the case of tank detection, YOLO is more adversely impacted than RT-DETR, particularly from the 7th level of distortion onward. Interestingly, both YOLO and RT-DETR maintain almost consistent performance in the tank case, unaffected by the levels of distortion presented. Figure 5 shows the varying impact of

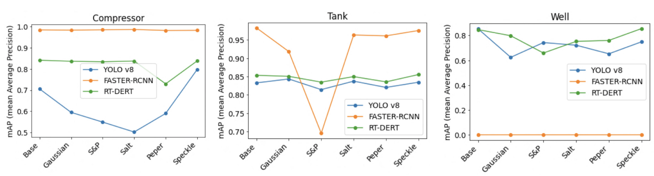


Fig. 5. Impact of noises (gaussian, salt&pepper, salt, pepper and speckle).

different noise types on the detection of compressors. Notably, YOLO is significantly affected by salt noise, while RT-DETR is more susceptible to pepper noise. Intriguingly, YOLO demonstrates improved performance with speckle noise in detecting compressors, compared to its performance on the base image. For wells and tanks, RT-DETR exhibits greater resilience to the five types of noise than YOLO, with the notable exception of salt and pepper noise in well detection. On the other hand, FASTER-RCNN generally maintains stable performance, although it shows a heightened sensitivity to salt and pepper noise in the detection of tanks. In Figure

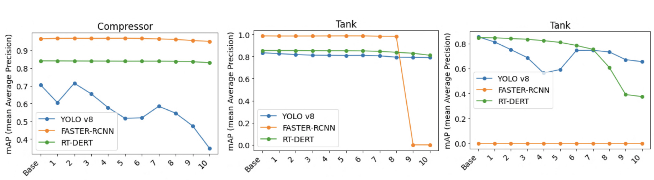


Fig. 6. Impact of brightness.

6, we observe that increasing brightness levels (from 1 to 10) generally leads to a decline in the mean Average Precision (mAP) of the models. This effect is particularly evident in the case of compressors, where YOLO's performance begins to diminish from level 2. However, YOLO is less affected than RT-DETR when detecting wells. For compressors, both FASTER-RCNN and RT-DETR maintain a relatively steady mAP across the range of brightness levels. When detecting and recognising tanks, the performance of FASTER-RCNN drops significantly from level 9, which means that it can no longer perform detection.

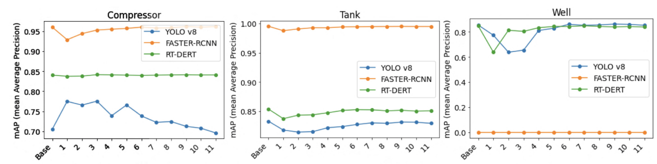


Fig. 7. Impact of saturation.

Figure 7 shows, for the compressor case, that YOLO presents a higher mAP for images with low saturation, which decreases for higher levels of saturation. Conversely, FASTER-RCNN exhibits the opposite trend. Concerning RT-DETR, its mAP value remains almost constant over saturation variations for the compressor case. For the tank case, all 3 models seem to react similarly, with an increasing mAP over saturation augmentation. This observation holds true for the well case between YOLO and RT-DETR.

4. CONCLUSION

This study conducted a systematic assessment of the impact of several satellite image variations—including resolution changes, tangential distortion, noise, brightness, and saturation—on 3 object detection algorithms: YOLO, RT-DETR, and FASTER-RCNN. Our findings indicate that FASTER-RCNN was the least affected by these simulated variations, although it failed to detect any wells. Conversely, YOLO, despite achieving the highest mean Average Precision (mAP) post-training, exhibited the greatest sensitivity. The study underscores that the influence of satellite image variations on algorithm mAPs is highly dependent on the specific object being detected, highlighting the necessity of tailoring algorithm performance to the targeted objects. We identified a hierarchy in the impact of these variations on mAPs: resolution, noise, distortion, brightness, and saturation, in descending order of influence. This ranking offers valuable insights into their relative significance in affecting algorithm performance. By incorporating these impactful variations into the training dataset, our experimental approach seeks to bolster algorithm robustness in practical applications. The comparative analysis of the 3 algorithms sheds light on their individual strengths and weaknesses, providing crucial guidance for choosing the appropriate algorithm for specific tasks under varying satellite imaging conditions.

5. REFERENCES

[1] Jingfan Wang, Lyne Tchapmi, Arvind Ravikumar, Mike Mcguire, Clay Bell, Daniel Zimmerle, Silvio Savarese, and Adam Brandt, "Machine vision for natural gas methane emissions detection using an infrared camera," *Applied Energy*, vol. 257, 10 2019.

[2] Gensheng Hu, Pan Yao, Mingzhu Wan, Wenxia Bao, and Weihui Zeng, "Detection and classification of diseased pine trees with different levels of severity from uav remote sensing images," *Ecological Informatics*, vol. 72, pp. 101844, 2022.

[3] Wahidya Nurkarim and Arie Wahyu Wijayanto, "Building footprint extraction and counting on very high-resolution satellite imagery using object detection deep learning framework," *Earth Science Informatics*, vol. 16, 11 2022.

[4] Ryota Yoneyama and Yuichiro Dake, "Vision-based maritime object detection covering far and tiny obstacles," *IFAC-PapersOnLine*, vol. 55, no. 31, pp. 210–215, 2022, 14th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2022.

[5] B Janakiramiaha, Kalyani Gadupudi, Karuna A, Narasimha V, and Marlapalli Krishna, "Military object detection in defence using multi-level capsule networks," 03 2021.

[6] Sikandar Ali, Abdullah, Ali Athar, Maisam Ali, Ali Hussain, and Hee-Cheol Kim, "Computer vision-based military tank recognition using object detection technique: An application of the yolo framework," in *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*, 2023, pp. 1–6.

[7] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," pp. 779–788, 2016.

[8] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[9] Wenyu Lv, Yian Zhao, Shangliang Xu, Jinman Wei, Guanzhong Wang, Cheng Cui, Yuning Du, Qingqing Dang, and Yi Liu, "Detrs beat yolos on real-time object detection," 2023.

[10] S. Rajkumar and Malathi Ganesan, "A comparative analysis on image quality assessment for real time satellite images," *Indian Journal of Science and Technology*, vol. 9, 09 2016.

[11] Nilantha Premakumara, Brian Jalaian, Niranjan Suri, and Hooman Samani, "Improving object detection robustness against natural perturbations through synthetic data augmentation," in *Proceedings of the 2023 Asia Conference on Computer Vision, Image Processing and Pattern Recognition*, New York, NY, USA, 2023, CVIPPR '23, Association for Computing Machinery.

[12] Zijian Zhu, Yichi Zhang, Hai Chen, Yinpeng Dong, Shu Zhao, Wenbo Ding, Jiachen Zhong, and Shibao Zheng, "Understanding the robustness of 3d object detection with bird's-eye-view representations in autonomous driving," 03 2023, CVPR '23.

[13] Haichao Zhang and Jianyu Wang, "Towards adversarially robust object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[14] Debang Li, Junge Zhang, and Kaiqi Huang, "Universal adversarial perturbations against object detection," *Pattern Recognition*, vol. 110, pp. 107584, 2021.

[15] Yifan Zhang, Junhui Hou, and Yixuan Yuan, "A comprehensive study of the robustness for lidar-based 3d object detectors against adversarial attacks," 2023.

[16] Huiming Sun, Lan Fu, Jinlong Li, Qing Guo, Zibo Meng, Tianyun Zhang, Yuewei Lin, and Hongkai Yu, "Defense against adversarial cloud attack on remote sensing salient object detection," 2023.

[17] Juan Terven and Diana Cordova-Esparza, "A comprehensive review of yolo: From yolov1 and beyond," 2023.

[18] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll'a r, and C. Lawrence Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014.

[19] Jade Eva Guisiano, Éric Moulines, Thomas Lauvaux, and Jérémie Sublime, "Oil and gas automatic infrastructure mapping: Leveraging high-resolution satellite imagery through fine-tuning of object detection models," in *Neural Information Processing*, Singapore, 2024, pp. 442–458, Springer Nature Singapore.

[20] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko, "End-to-end object detection with transformers," pp. 213–229, 2020.

[21] Forrest Fankhauser, J. Anthony Tyson, and Jacob Askari, "Satellite optical brightness," *The Astronomical Journal*, vol. 166, no. 2, pp. 59, jul 2023.