Kütahya Dumlupınar University
Institute of Graduate Studies

Journal of Scientific
Reports-A
E-ISSN: 2687-6167
Number 54, September 2023

RESEARCH ARTICLE

EVALUATING THE ROBUSTNESS OF YOLO OBJECT DETECTION ALGORITHM IN TERMS OF DETECTING OBJECTS IN NOISY ENVIRONMENT

Halit BAKIR[1*], Rezan BAKIR[2]

[1]Sivas University of Science and Technology, Faculty of Engineering and Natural Sciences, Department of Computer Engineering , Sivas, Turkey, halit.bakir@sivas.edu.tr, ORCID: 0000-0003-3327-2822
[2]Sivas University of Science and Technology, Faculty of Engineering and Natural Sciences, Department of Computer Engineering , Sivas, Turkey, rezan.bakir@sivas.edu.tr, ORCID: 0000-0002-4373-2231

Receive Date: 27.02.2023 *Accepted Date: 24.04.2023*

ABSTRACT

Our daily lives are impacted by object detection in many ways, such as automobile driving, traffic control, medical fields, etc. Over the past few years, deep learning techniques have been widely used for object detection. Several powerful models have been developed over the past decade for this purpose. The YOLO architecture is one of the most important cutting-edge approaches to object detection. Researchers have used YOLO in their object detection tasks and obtained promising results. Since the YOLO algorithm can be used as an object detector in critical domains, it should provide a quite high accuracy both in noisy and noise-free environments. Consequently, in this study, we aim to carry out an experimental study to test the robustness of the YOLO v5 object detection algorithm when applied to noisy environments. To this end, four case studies have been conducted to evaluate this algorithm's ability to detect objects in noisy images. Specifically, four datasets have been created by injecting an original quality image dataset with different ratios of Gaussian noise. The YOLO v5 algorithm has been trained and tested using the original high-quality dataset. Then, the trained YOLO algorithm has been tested using the created noisy image datasets to monitor the changes in its performance in proportion to the injected Gaussian noise ratio. To our knowledge, this type of performance evaluation study did not conduct before in the literature. Furthermore, there are no such noisy image datasets have been shared before for conducting these types of studies. The obtained results showed that the YOLO algorithm failed to handle the noisy images efficiently besides degrading its performance in proportion to noise rates.

Keywords: *Deep learning, Image processing, YOLO, Object detection, Gaussian noise.*

1. INTRODUCTION

Technology has a life cycle just like humans. Every new technology is born, develops, and becomes a raw material for another technology. Deep learning and image processing have become the raw material of every technology today. The use of image processing and deep learning techniques has gained importance in many fields such as medicine, the defense industry, astronomy, geology, etc. One area that has attained great progress in the last years is object detection. Object detection is a computer vision technique that generally utilizes machine learning or deep learning techniques for determining the location and scale of all object instances in images or videos. Nowadays, object detection has entered numerous fields, and its applications have varied, from identity detection and self-driving cars to security and medical uses. Object detection is considered a challenging problem due to various potential reasons such as the limited amount of annotated data, class imbalance, and so on.

Evaluating object detection algorithms in a noisy environment presents unique challenges and considerations. In a noisy environment, various sources of noise, such as sensor limitations, low lighting conditions, or environmental interferences, can significantly impact the performance of object detection algorithms. When evaluating these algorithms in such conditions, it becomes crucial to assess their robustness and reliability in accurately detecting and localizing objects amidst the noise. Evaluating object detection algorithms in a noisy environment typically involves analyzing their performance metrics, such as detection accuracy, localization precision, and false positive rates, while considering the specific noise characteristics and their potential effects on algorithm performance. Furthermore, researchers often employ specialized datasets or introduce synthetic noise to simulate realistic scenarios and assess the algorithms' ability to handle noise-induced challenges. The evaluation process helps in understanding the algorithm's performance limitations, guiding improvements, and facilitating the development of more noise-robust object detection systems.

This study focuses on evaluating the widely recognized YOLO object detection algorithm in various environments with different levels of noise. Specifically, our assessment involves examining the algorithm's performance in detecting objects within images that have been subjected to progressively increasing amounts of Gaussian noise.

1.1. Motivation and Contribution

Many successful deep learning architectures such as Yolo, VGG-19, ResNet, Inception, Xception, and MobileNet are used for detecting and recognizing objects in images. However, distortions may occur while acquiring images because of various sources. In other words, noises like electrical interference, poor lighting, and gaussian noise may lessen the quality of the image. In such situations, deep learning architectures that detect or classify objects rapidly and accurately in clear images may fail (i.e. in noisy images). On the other hand, as it's known, the information obtained in some critical fields such as the defense industry, radar systems, medicine, etc. should have zero or close to zero error rates. And since the images acquired in such domains do not always have a high quality due to the environmental conditions. Therefore, there is an urgent need for adaptive models that can be used in more than one domain and that can minimize the error rate when detecting objects in both noisy and noise-free images. With the advent of the YOLO algorithm, several applications have employed YOLO for object detection and recognition in various fields and the obtained results were encouraging, but in most cases, this algorithm has been trained and tested using high-quality images. This motivates us to conduct a specific study to evaluate YOLO robustness in detecting objects in noisy environments. A dataset consisting of 40 classes was used in this study to evaluate the object detection performance of the YOLO algorithm. YOLO was first evaluated using the original dataset's high-quality images. Then four datasets have been created by injecting noise gradually into the original images. To the best of our knowledge, this type of dataset has not been publicly shared before in the literature. YOLO's detection performance is evaluated using the suggested and constructed datasets. The literature has not previously addressed this type of performance evaluation study to the best of our knowledge. Furthermore, doing such a study can help researchers realize the maximum rate of distortion at which the model can produce results whether it is good or bad results.

1.2. Research question

Verifying whether YOLO is good model for general object detection or not, and whether YOLO algorithm could be used as end-to-end model for conducting object detection task in critical systems.

The rest of the paper is organized as follows. Section 2 includes the related studies. Section 3 describes the used methodologies. Section 4 presents experimental results. Section 5 includes conclusion and future works.

2. RELATED WORKS

Due to the rapid technological change during the last years there has been a rapid and successful expansion of computer vision research. One area that has attained great advancement is object detection. Object detection is a fundamental task required by most computer vision systems. Researchers in the last years have made a great effort to make considerable progress in various directions in order to conduct a robust object detection algorithms and approaches. For example, some of the studies adopted machine learning methods [1] while others tried to develop new representations and models for specific computer vision problems or tried to develop efficient existing solutions [1–7]. As an instance, in [1] study, a new algorithm was proposed for visually salient object detection, then it was utilized to extract salient objects to be used for training the machine learning-based object detection part of the proposed system.

A lot of researchers recently have utilized deep learning algorithms in the domain of computer vision, especially in image classification and object detection. For example, Bakır et.al in their study[8], proposed CNN and ANN based approaches for diagnosing and detecting lung diseases. In another study [9], a ResNet deep learning architecture was employed to classify the malaria parasite effectively. Authors in [10] study proposed several deep learning architectures such as VGG-16, ResNet, and Inception v3 in order to extract features to be used in detecting cataract disease from retinal fundus images. On the other hand, in [11] study, an approach utilizing deep learning was introduced to detect objects by analyzing images captured by an unmanned aerial vehicle's (UAV) onboard camera during an autonomous flight trajectory. Subsequently, an algorithm was devised to autonomously guide the UAV to land in close proximity to the detected object.

The object detection algorithms are classified into two main categories: Single-stage models and multi-stage models. For example, in the two-stage models like R-CNN [12], the initial model is employed to identify object regions, while the subsequent model is utilized to classify the objects and further enhance the precision of their localization. As a result, learn the localization and classification stage separately makes this method relatively slow. SPP-Net [13] and Fast R-CNN [14] presented the

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

concept of region-wise feature extraction. Proposing to use spatial pyramid pooling (SPP), SPP-net can generate a fixed-length representation regardless of image size/scale [13]. On the other hand, Fast R-CNN utilizes various methods (such as using region of interest (RoI) pooling layer) to improve training and testing speed while boosting detection accuracy. Later, in Faster R-CNN [15] approach, a Region Proposal Network (RPN) was introduced to generate region proposals with minimal computational overhead. This was achieved by sharing convolutional features from the full image with the detection network.

In the single-stage approach, a fixed number of predictions is made on the grid which means the model can directly predict object bounding boxes for an image. The most famous single-stage models are You Only Look Once (YOLO) [16],YOLO v2 [17] YOLO v3 [18] , YOLO v4 [19] , and SSD: Single Shot MultiBox Detector [20], Broadly speaking, research in the field of object detection can be categorized into two primary streams: conventional methods for object detection and detection methods based on deep learning techniques. Traditional object detection methods are built on handmade features and shallow trainable architecture, such as in [21–31]. However, these types of methods can easily fail to handle complex combinations of low-quality images beside being inaccurate, relatively slow with low performance on unfamiliar dataset [32]. With the rapid development in deep learning, more robust architectures have been introduced to address the problems present in traditional methods such as [31, 33–35].

There are few studies on object detection or classification in a noisy environment. As an example, Nayan, Al-Akhir, et al in their article [36] proposed a new low-cost technique for error-free object recognition in noisy images. In the study, a comprehensive experimental evaluation with conventional detectors retrained on noisy images is presented, taking advantage of the Single Shot MultiBox Detector SSD. In another study, Kushagra Yadav et al. [37] proposed a new method to reduce the effect of noise on the image object detection task. The proposed method consisted of two stages. In the first stage, Residual Dense Network (RDN) was used to reduce noise from the low-quality image, and in the second stage, the standard Single-Shot Multibox object detector was adopted to complete the object detection process. Furthermore, the proposed model is tested using Gaussian noise images. and the proposed approach is evaluated in the Pascal Visual Object Classes comparison. Moreover, Milyaev, S., and I. Laptev [38] proposed a new, low-cost method for image noise removal by considering object detection in noisy images. The proposed method is based on combining the standard Deformable Parts Model and Regions with Convolutional Neural Network object detectors. The proposed method is compared with other noise removal techniques as well as with standard detectors retrained on noisy images. In Elena Medvedeva's work [39], an improved method is proposed to detect moving objects in distorted images caused by white Gaussian noise. Including two stages, the proposed method represents the video sequence through a three-dimensional discrete Markov process. The first stage is the filtering of three-dimensional non-linear images, which allows objects' contours to be preserved. The second stage involves identifying objects of interest based on their boundaries and luminosity.

On the other hand, for detecting objects in noisy environments, J. F. Que et al. [40] proposed a Yolov3-based method. In particular, the YOLO v3 algorithm is used to create an LSS object detection system that can adapt to environmental noise. Multiple experiments were performed on both noisy and

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

noiseless datasets, and it is stated that the proposed method improves object detection accuracy in noisy environments. Geonsoo Lee et al. [41] proposed a Feature Enhancement Network (FEN) to deal with noise in small object detection. The authors presented a self-monitoring approach to training SEN without any labels. They also noted that the proposed approach can be seamlessly combined with a variety of off-the-shelf object detectors. Maheep Singh et al. [42] propose a new Distinctive Object Detection (SOD) technique in noisy environments using a convolutional neural network (CNN). De-noising the image is achieved using CNN, which uses coordinate descent to modulate the signal. Gaussian noise and white noise were tested in the study and the performance of the proposed V-SIN technique was evaluated on two publicly available image datasets with different evaluation metrics. Aditya Gautam and Mantosh Biswas [43] adopted Whale Optimization Algorithm (WOA) for edge detection in Gaussian noise images. According to them, experimental results showed that the proposed technique outperformed conventional edge detectors and the considered technique.

Furthermore, the YOLO is accepted as one of the most important and accurate object detection algorithms that has been used in a wide range of applications. For example, in [44] a YOLOv5 was utilized to detect the bacterial spot disease in the bell pepper plants from the symptoms seen on the leaves. As we can see from related works, the suggested performance evaluation study has not been previously addressed in the literature which encourages us to conduct such a study to provide insight into YOLOv5's performance in detecting objects in noisy environments. In [45], YOLO-SA which is a YOLO landslide detection model is proposed in order to improve the speed, accuracy, and parameters of landslide detection models. In [46], channel pruned YOLO v4 has been adopted for apple flower real-time detection. In [47], YOLO-Tomato, a YOLO based model has been proposed for handling the challenges of fruit detection. In [48], YOLO-face, a YOLO v3 based model has been proposed for improving the performance of face detection.

When we looked at the literature it can be concluded that the YOLO algorithm has accepted as an accurate algorithm which can be used in multiple domains such as agriculture and biomedical. The images that have been used in most of the literature works is high-quality images and the YOLO algorithm has been used as an end-to-end model without any pre-processing phase. So, this work aims at investigating the robustness of the YOLO algorithm in detecting objects in low-quality or noisy images. This type of investigation works is very important to shed the light on the acceptable amount of distortion or noise to save the performance of these types of models. To this end, we proposed to inject the Gaussian noise gradually into an image dataset and monitor the performance of this algorithm i.e. YOLO algorithm. The results showed that when the noise amount was 25% the performance of the YOLO algorithm did not be affected so much and it was still at an acceptable rate. On the other hand, the performance of YOLO started dropping significantly. The performance of YOLO became very bad when the amount of injected Gaussian noise reached 100%.

3. MATERIAL AND METHOD

In general, digital images became an important part of modern systems including airplanes, aircraft, autonomic systems, and medical systems. The images collected in such systems are not high quality in most cases. This is related to the fact that the quality of images degrades due to the existence of noise. Image noise is a random variation of brightness or color information. The noise can occur due to

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

different reasons such as electricity, heat, and sensor illumination levels. The noise in images can be manifest in different formats. we will briefly discuss the most important of it in the following section.

3.1. Types of Noise

Noise in images refers to unwanted random variations or distortions that affect the visual quality of a picture. It can appear as graininess, speckles, or artifacts, and is primarily caused by factors such as sensor limitations, low-light conditions, or compression algorithms. Noise can reduce the clarity, sharpness, and overall fidelity of an image, impacting its visual appeal and potentially hindering the interpretation of important details. To enhance image quality, various techniques like denoising algorithms and post-processing methods are employed to minimize or remove noise while preserving the essential information and maintaining a balance between noise reduction and image sharpness. There are several types of image noise: Gaussian Noise, Impulse Noise, Salt and Pepper Noise, Speckle Noise, and Poisson Noise.

3.1.1. Gaussian noise

Gaussian noise or Random noise (also called electronic noise) is statistical noise having values distributed in a normal Gaussian. The noise is created by adding a Gaussian function to the image. The type of noise is very similar to nature's noise types, thus this type of noise has been adopted in the present study to evaluate the robustness of the YOLO algorithm. The noise's values are Gaussian-distributed. In the case of a Gaussian random variable Z, the probability density function $\mathcal{P}$ can be expressed as in Eq. 1.

$$\mathrm{PG}(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{1}$$

Where Z represents the grey level, μ the mean grey value and σ its standard deviation.

3.1.2. Impulse noise

Impulse noise is a type of random noise that affects digital images. It manifests as isolated pixels with significantly higher or lower intensity values compared to their neighboring pixels, resembling white and black specks. Impulse noise can occur due to various factors, such as transmission errors in digital communication, faults in image sensors, or data corruption during storage or processing. It can degrade image quality, introduce visual artifacts, and adversely impact subsequent image analysis tasks. Denoising techniques specifically designed to handle impulse noise are commonly employed to mitigate its effects and restore the clarity and integrity of the image. There are three main types of impulse noise. Salt Noise, Pepper Noise, Salt and Pepper Noise. Salt noise can be generated by adding random bright values to the image. Pepper noise can be generated by adding random dark values to the image. Bright values and dark values are produced by adding 255-pixel values and zero-pixel values, respectively.

3.1.3. Salt and pepper noise

This type of noise is a combination of salt and pepper noise. It can be injected into an image by adding both random bright and random dark values all over the image. The Salt & Pepper noise values can

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

range from 0 to 255. Pepper noise tends to have an intensity value close to 0, while salt noise tends to have an intensity value close to 255. See in Eq. 2.

$$\eta(x, y) = \begin{cases} 0, & \text{pepper noise} \\ 255, & \text{salt noise} \end{cases} \tag{2}$$

3.1.4. Poisson noise

Poisson noise, also known as photon noise or shot noise, is a type of statistical noise that commonly occurs in digital imaging systems, particularly in low-light conditions. It is caused by the inherent randomness associated with the detection of light particles (photons) by an image sensor. Poisson noise follows a Poisson distribution and is characterized by variations in the number of photons detected at each pixel, resulting in random fluctuations in pixel intensity. In images, Poisson noise appears as a granular pattern with slight intensity variations across the scene. It is more pronounced in darker regions where fewer photons are detected. Denoising techniques designed to handle Poisson noise typically involve statistical modeling and estimation to reduce the noise while preserving image details and avoiding excessive smoothing. These techniques are commonly used in applications such as astrophotography, medical imaging, and scientific imaging where low-light conditions are prevalent. A nonlinear response of image detectors and recorders causes this type of noise, known as quantum noise or shot noise. For a random variable $X \geq 0$ the Poisson noise can be dictated using in Eq. 3.

$$PY|X\,(y \mid x) = 1/y!\,(ax + y)e^{y-(ax+y)}, x \geq 0, y = 0,1, \ldots \tag{3}$$

$\alpha > 0$: scaling factor, and $\lambda \geq 0$: the dark current parameter.

As a result, input random variable X is transformed into output random variable Y, which is indicated by Y as in in Eq. 4.

$$Y = P(aX + y) \tag{4}$$

3.1.5. Speckle noise

Speckle noise is a type of granular noise that commonly affects images acquired through coherent imaging systems such as ultrasound, synthetic aperture radar (SAR), and laser imaging. It arises from the interference patterns created by the constructive and destructive interference of coherent waves within the imaging system. Speckle noise appears as a grainy pattern with random variations in intensity, resulting in a speckled or textured appearance in images. It can obscure fine details, reduce contrast, and degrade the overall quality of the image. Denoising techniques for speckle noise often involve the use of filters, statistical models, or advanced algorithms specifically designed to reduce the noise while preserving important image structures and details. These techniques play a crucial role in enhancing the visual quality and interpretability of images acquired through coherent imaging systems. Speckle noise is a multiplicative noise that takes place in low-level luminance images such as Magnetic Resonance Image (MRI) images. This type of noise can be produced by multiplying random pixel values with different pixels of an image. Speckle noise can be modeled as in in Eq. 5.

$$g(m,n) = f(m,n)u(m,n) + \eta(m,n) \quad (5)$$

Where $g(m,n)$ point to a corrupted image matrix at the spatial position(m,n); $u(m,n)$ and $\eta(m,n)$ stand for the multiplicative and additive component of the noise, respectively; and f is the original image. Figure 1, illustrate an image injected with different types of noises.

Denoising techniques are utilized to reduce or eliminate noise from images, enhancing their visual quality and improving the accuracy of subsequent analysis. Various approaches are employed to tackle noise, such as spatial filtering, statistical methods, and machine learning algorithms. Spatial filtering methods, including median filtering and Gaussian filtering, work by replacing each pixel's value with a filtered value based on its neighboring pixels. Statistical methods, such as wavelet denoising or total variation denoising, exploit the statistical properties of noise to remove its presence while preserving image details. Machine learning-based techniques employ deep neural networks trained on large datasets to learn the noise patterns and perform denoising effectively. These techniques play a vital role in restoring image fidelity and enhancing the overall visual appeal of images in various domains, including photography, medical imaging, and computer vision. We intend to assess the efficacy of these denoising techniques in our forthcoming research endeavors, as their evaluation falls beyond the scope of the present paper.

3.2. Object Detection

Object detection is a fundamental task in computer vision that involves identifying and localizing objects of interest within an image or a video sequence. It plays a crucial role in various applications, such as autonomous driving, surveillance, and augmented reality. Object detection algorithms aim to automatically detect and classify objects in images, often using deep learning techniques. One popular approach is the region-based convolutional neural network (R-CNN) family of algorithms, which generate region proposals and then classify them using a convolutional neural network (CNN). Another widely used algorithm is the You Only Look Once (YOLO) model, which divides the input image into a grid and predicts bounding boxes and class probabilities directly from the grid cells. Other notable algorithms include Single Shot MultiBox Detector (SSD) and Faster R-CNN, which improve on speed and accuracy. These algorithms have significantly advanced the field of object detection, enabling robust and efficient detection of objects in various real-world scenarios. So, in this study, we will evaluate the performance of the well-known YOLOv5 in terms of detecting objects in noisy environments. Therefore, we will briefly talk about this algorithm in the next sub-section.

3.2.1. YOLOv5

Proposed by Redmond et. al YOLO (You Only Look Once) is one of the most common real-time object detection algorithms. The algorithm depends on dividing images into a grid system. Each cell in the grid is responsible for detecting objects within itself. The YOLO model solves object detection as a regression problem instead of a classification problem by directly predicting the image pixels as objects and its bounding box attributes [49]. Therfore, the YOLO algorithm uses bounding box regression to predict the center, height, width, and class of each object. Moreover, this algorithm uses the Intersection over union (IOU) concept to select the bounding boxes that fit the objects in the image as perfectly as possible, in this way the algorithm can prevent detecting the object more than one time.

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

After the invention of the YOLO algorithm, a multiple version of this algorithm has been proposed and developed such as YOLOv2 [17] and YOLOv3 [18],and YOLO v4 [19]. Each version of YOLO has been proposed for solving specific problem in the previous versions and improving the detection accuracy of the original YOLO algorithm. For example, in YOLO v2, it has been proposed to use batch normalization operation in the YOLO algorithm, which improves the performance of YOLO and solves the problem of detection of small objects. Also, in YOLO v3, it has been proposed to use the logic of residual neural networks (especially skip connections) in YOLO algorithm, which improved the performance of YOLO significantly compared with YOLO v2. In 2020, Glenn Jocher introduced YOLOv5 using the Pytorch framework. The algorithm is pre-trained on the MS COCO dataset. YOLOv5 is considered one of the authorized cutting-edge models with outstanding support and is convenient to use in production. All the versions of YOLO algorithm have been adopted in diverse applications mainly due to their faster inferences, high detection accuracy, and better generalization besides being open source.

Since YOLO is one of the state-of-the-art object detection algorithms used in a huge number of previously deep learning-based science studies, and since this algorithm has been introduced as one of the most advanced object detection algorithms that can be used in multiple domains and can achieve an outperforming performance, in this work we proposed to test the robustness of YOLO algorithms with low-quality images and see how its performance can be affected if the environment contains some type of noise.

3.3. Used and Created Datasets

Military Aircraft detection dataset has been utilized in this study. The dataset consists of images related to 43 different types of aircraft. Particularly, the dataset contains objects related to 43 different classes her object has been labelled with a bounding box in PASCAL VOC format. The original dataset contains 10658 different labelled images. The original dataset can be founded and downloaded from Kaggle repository [50]. In order to evaluate the robustness of the YOLO algorithm against Gaussian noise we propose to inject this noise gradually into the images in the original dataset and monitor how the performance of YOLO will be affected. The Gaussian noise has been added into the images using the formulas Eq.6 and Eq.7.

$$\mathrm{noise} = \mathrm{np.random.normal}(\mathrm{loc} = 0, \mathrm{scale} = 1, \mathrm{size} = \mathrm{image.shape}) \tag{6}$$

$$\mathrm{noisyImage} = \mathrm{image} + \mathrm{noise} * \mathrm{noiseRate} \tag{7}$$

In particular, in the first case study, we proposed creating a Gaussian noise matrix, multiplying it by 0.3 (noiseRate=0.3), and adding it to the original images' pixels. In the second case study, we proposed creating a Gaussian matrix, multiplying it by 0.5 (noiseRate=0.5), and adding it to the original images' pixels. In the third case study, we proposed creating a Gaussian matrix, multiplying it by 0.8 (noiseRate=0.8), and adding it to the original images' pixels. In the fourth case study, we proposed creating a Gaussian matrix, multiplying it by 1 (noiseRate=1), and adding it to the original images' pixels. Accordingly, in this way, we can control the proportion of noise that will be added to images, and we can create four different noisy image datasets from the original dataset namely 30%-Gaussian-Dataset, 50%-Gaussian-Dataset, 80%-Gaussian-Dataset, and 100%-Gaussian-Dataset. The

JSR A
Journal of Scientific Reports

created datasets will be available on demand. To the best of our knowledge, this is the first time that these types of datasets are constructed and made available for future works. Algorithm 1, illustrates the Pseudocode of the Python script used for injecting the noise and creating the proposed noisy datasets. Figure 2 illustrate an example of image injected with different proportions of gaussian noise.

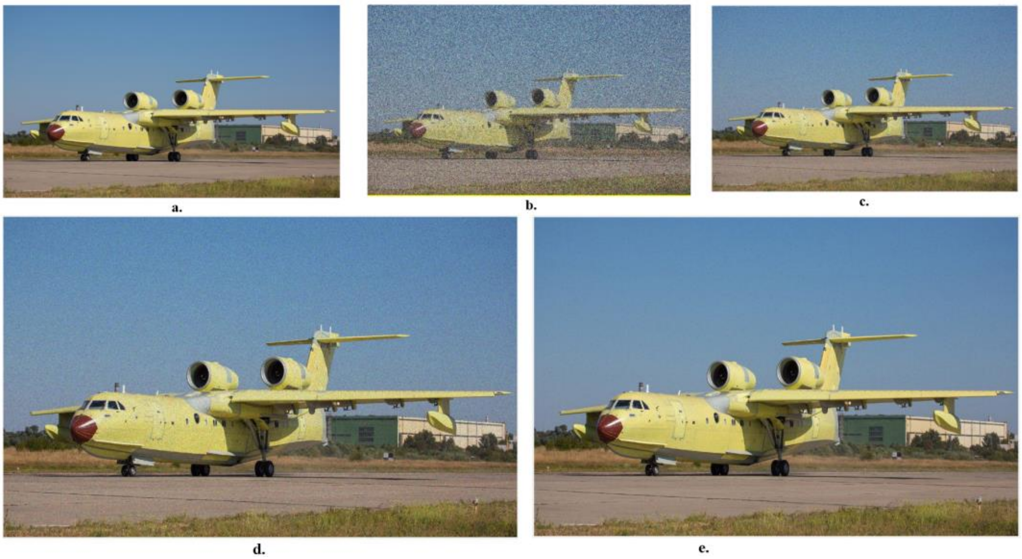


Figure 1. Example images injected with different types of noise.
a. Orignal image, **b.** Gaussian noisy image, **c.** Salt&paper noisy image, **d.** Speckle noisy image, **e.** Poisson noisy image.

3.4. Proposed Test Bed

In order to test the YOLO model in detecting noisy images we have to create a noisy image dataset with different noise rates. To this end, as explained in the previous section, we injected images with different Gaussian noise proportions. Then we applied YOLO to detect objects within created image datasets besides the original dataset to evaluate the overall performance. Figure 3 shows the block diagram of the conducted experiment. As can be seen in Figure 3 the test bed mainly consists of two phases. In the first phase, the original dataset was split into a training dataset utilized to train the YOLO model and a testing dataset used to evaluate the detection performance of the trained YOLO model. Thus, in the first phase, we trained and tested the YOLO v5 algorithm in order to evaluate its detection accuracy when the used image dataset contains high-quality images. Afterward, we saved the weights trained by the algorithm to be used to detect objects from the same images but after injecting some level of noise into them. Particularly, the second phase is composed of four different case studies as follows:

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

3.4.1. Case study 1

In this case study, we proposed to inject a very small amount of gaussian noise into the images. Particularly, we proposed to inject only 30% of gaussian noise into each image in the dataset. After that, we used the trained YOLO algorithm in order to investigate if YOLO algorithm can detect the object from the images with the same accuracy achieved on the original image dataset.

3.4.2. Case study 2

In this case study, we increased the amount of injected noise slightly to 50% and tested the performance of the trained YOLO algorithm in terms of detecting objects in these mild noise-contained images.

Algorithm 1. Algorithm used for constructing noisy image dataset.

```
Input: Image_dataset
Output: Noisy_image_dataset

For img in Image_dataset:
Image = read_Image_file(img)
Image = reScale_image_pizels(image)
#creating noise matrix with same shape of the original image
noise = np.random.normal(loc = 0, scale = 1, size = Image.shape)
noisy_image = np.clip((img + noise * noise_proportion),0,1)
Noisy_image_dataset ← image_saver(noisy_image)
```

Table 1. dataset description.

Number of classes	Number of objects	Number of images
43	17145	10658

3.4.3. Case study 3

In this case study, we increased the amount of noise to 80%, and again tried to detect the objects from the images using the trained YOLO algorithm.

3.4.4. Case study 4

In this case study, the injected amount of gaussian noise has been increased to 100%, and the trained YOLO algorithm has been applied to detect the objects from these fully obfuscated images.

In brief, in the second phase, we injected the original images with different gaussian noise rates to create four different datasets. Then we applied YOLO trained model to the created noisy image datasets.

JSR A
Journal of Scientific Reports

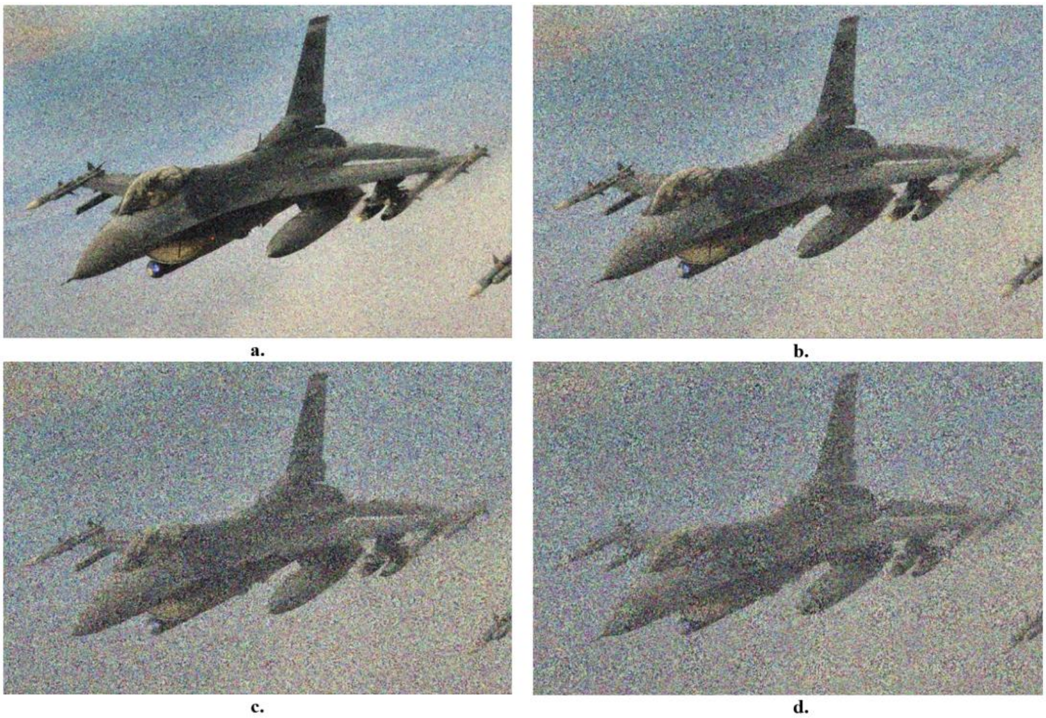


Figure 2. Example of an image injected with different proportions of noise.
a. Gaussian 30%, **b.** Gaussian 50%, **c.** Gaussian 80%, **d.** Gaussian 100%.

4. EXPERIMENTAL RESULTS

All the experimental studies have been conducted using Python programming language over Google-colab free GPU. Multiple python libraries and frameworks have been used during conducting this study such as Kersas, Tensorflow, Open-CV, etc.

4.1. Evaluation Metrics

To evaluate the robustness of YOLOv5 in different noisy environments we utilized different standard evaluation metrics such as confusion matrix, Recall, Precision, F1-score, and mAP (mean Average Precision), but we only displayed F1-score and mAP in the results since it is enough to summarize the tradeoff of both Recall and Precision metrics and gives us a better idea of the overall accuracy of the model.

4.1.1. Confusion matrix

A confusion matrix is a table that defines how well a classification algorithm performs. Confusion matrixes consist of four components as can be seen in Figure 3.

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

4.1.2. True Positives (TP)
The model predicted a label correctly, which means the actual value matched the predicted value.

4.1.3. True Negatives (TN)
The model does not predict the label correctly; the actual value is negative while the predicted value is positive.

4.1.4. False Positives (FP)
The actual value is negative but the predicted value is positive.

4.1.5. False Negatives (FN)
The actual value is positive but the predicted value is negative.

4.1.6. Accuracy
This metric is calculated using Eq.8.

$$\mathrm{Acc} = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP+TN+PF+FN})} \tag{8}$$

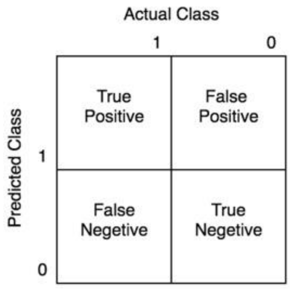


Figure 3. confusion matrix.

4.1.7. Precision
This metric is calculated using Eq.9.

$$\mathrm{P} = \frac{\mathrm{TP}}{(\mathrm{TP+FP})} \tag{9}$$

4.1.8. Recall
This metric is calculated using Eq.10.

$$R = \frac{TP}{(TP+FN)} \tag{10}$$

JSR A
Journal of Scientific Reports

Bakır, H. and Bakır, R., Journal of Scientific Reports-A, Number 54, 1-25, September 2023

4.1.9. F1-score
This metric is calculated using Eq.11.

$$F = 2 * \frac{(R*p)}{(R+p)} \tag{11}$$

4.1.10. mean Average Precision (mAP)
This metric is calculated using Eq.12.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{12}$$

As it can clearly be seen from the results in Figure 4 the YOLO approved its efficiency for detecting objects from high-quality images relevant to the original dataset. Particularly, the average value of mAP reached 73.5% and the average value of F1 score values reached 69%, which denotes a great success in detecting objects in high quality images. After that, we tested the YOLO algorithms using the constructed four noisy image datasets. we saw that the YOLO algorithm obtained good results when tested in a noise-free environment. However, when adding noise to the images, the YOLO model showed a gradually decreasing in performance in a proportion to the amount of injected noise. For example, it can be noted from Figure 5 the big degrading in mAP score, where it decreased from 73.5% when applied to the original dataset to 48.6%, 35.6%,24.4 %, and 23.5 % when applied to detect the objects from the same images after injecting gaussian noise with 30%, 50%, 80%, and 100% proportions respectively. Moreover, Figure 6 illustrates the F1-score curve obtained when the trained YOLO algorithm has been applied for detecting objects from the created four noisy image datasets. We can note from the Figure that the average F1-score value degraded from 69% when the trained YOLO algorithm has been used for detecting objects from the original image dataset to 48%, 34%, 23%, and 21% obtained when the YOLO-trained algorithm adopted for detecting the objects from the same image but after injecting gaussian noise with 30%, 50%, 80%, and 100% proportions respectively. Moreover, Figure 7 illustrates some detection examples obtained by applying YOLO to the original dataset. We can see from the figure that the trained YOLO algorithm could detect all the objects in the image except one object, and almost all the detected labels were true.

Then when we have applied the YOLO algorithm on the 30% Gaussian-Dataset and 50% Gaussian-Dataset as can be seen in Figure 8, we can note from the figure that the number of undetected objects increased gradually based on the amount of the injected Gaussian noise, where the YOLO algorithm cannot detect 12 objects from the 30%-Gaussian images and 22 objects from 50%-Gaussian images. Also, we can note from the figure that almost all the labels of the detected objects have been defined wrongly by the YOLO algorithm.

By increasing the amount of noise injected into the original images the performance of YOLO continued to drop down as we can see from Figure 9, where the number of undetected objects reached 38 and 45 when the injected Gaussian noise proportion was 80% and 100% respectively.

JSR A
Journal of Scientific Reports

5. CONCLUSION AND FUTURE WORKS

Object detection is a major task in most computer vision problems. It is typically working to identify and locate objects within an image. YOLO is one of the state-of-the-art algorithms highly used recently in various object detection tasks. The object detection forms the main task in different critical and real time systems such as aircraft goal tracking systems, radar systems, biomedical systems, and so on. And since the images collected in these systems cannot always be in the same quality as the images used for testing it, therefore, in this study, we made various experiments for testing the performance and the robustness of the YOLO detection algorithm in a noisy environment. To this end, we have trained and tested YOLO v5 algorithm using different noisy image datasets.

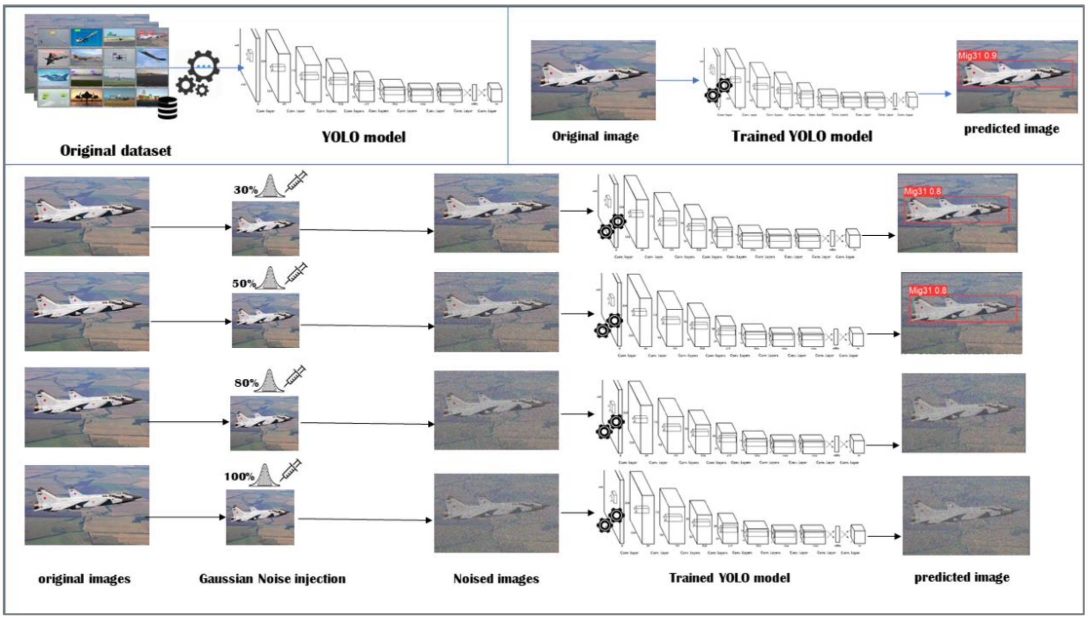


Figure 3. The block diagram of the conducted evaluation experiment.

JSR A
Journal of Scientific Reports

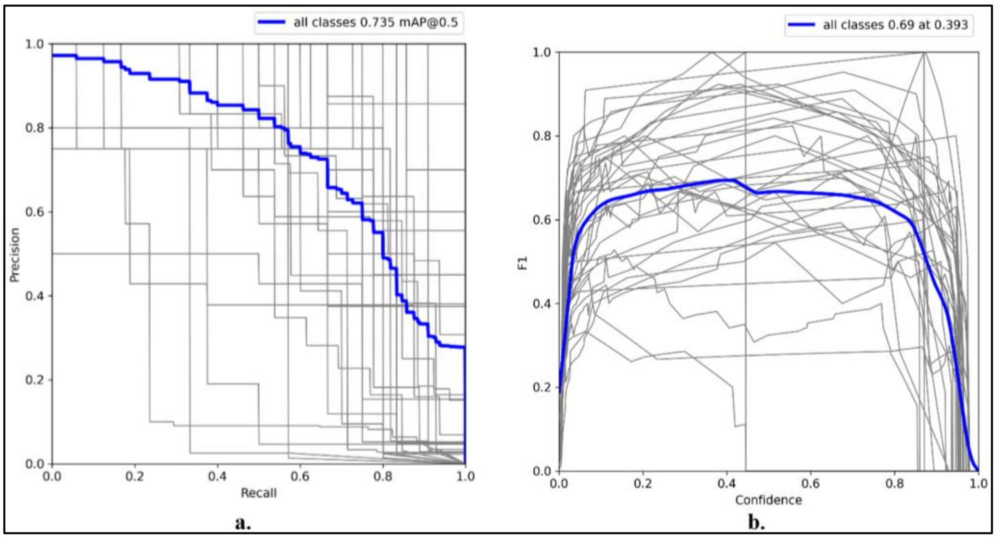


Figure 4. Test results on original dataset. (**a.** mAP, **b.** F1 Score).

Particularly, a test bed has been proposed to test the YOLO algorithm using four different noisy image datasets containing various proportions of gaussian noises. We proposed to inject 30% of gaussian noise into the original image dataset in order to create the first noisy image dataset. After that, we proposed injecting 50% of gaussian noise in order to create the second noisy image dataset. Then, we proposed injecting 75% of gaussian noise into the original image dataset in order to construct the third noisy image dataset, and finally, the fourth noisy image dataset has been constructed by injecting 100% of gaussian noise into the images in the original dataset.

JSR A
Journal of Scientific Reports

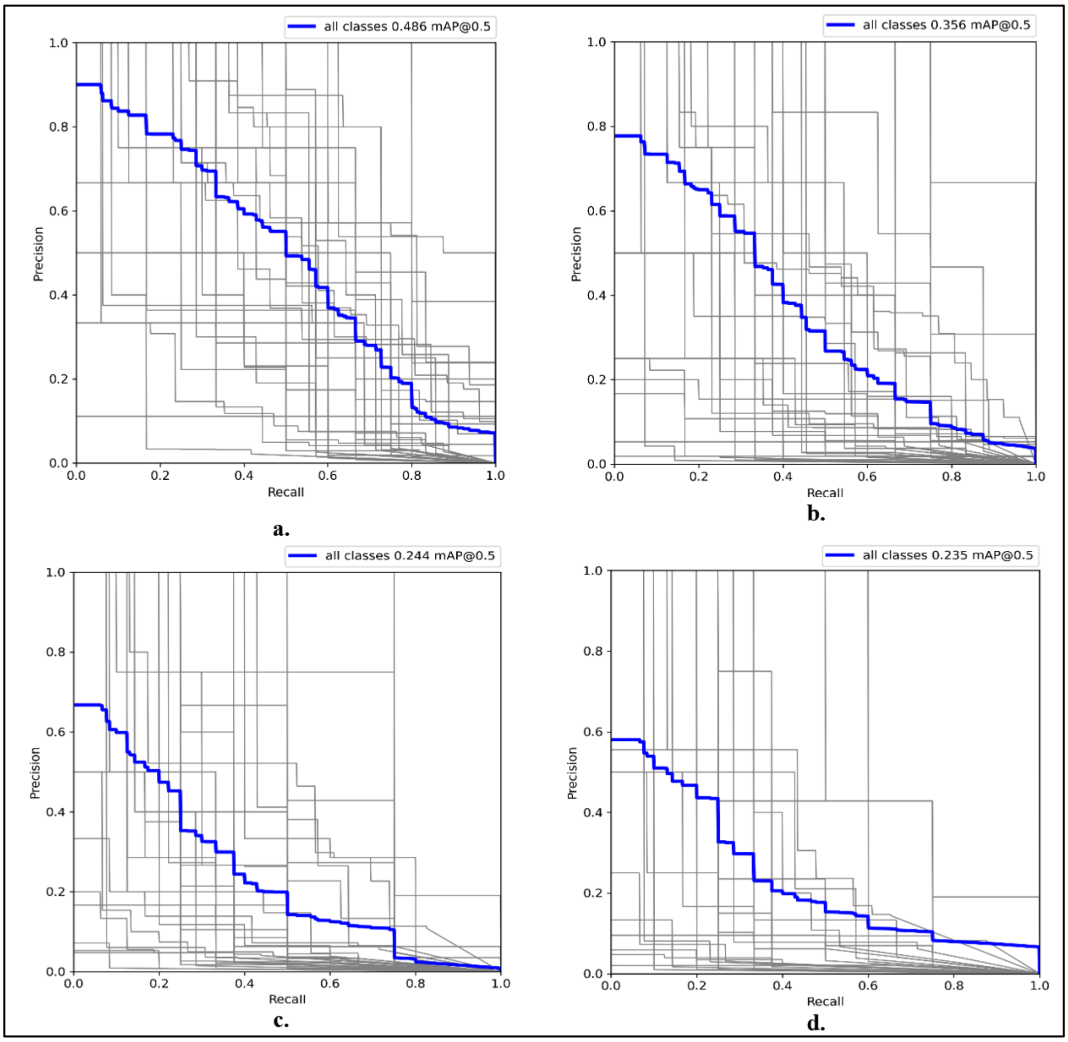


Figure 5. mAP results obtained by applying YOLO on noisy datasets.
a. 30% noise, **b.** 50% noise, **c.** 80% noise, **d.** 100% noise.

The obtained results using four proposed datasets showed that the YOLO model failed to handle noisy images efficiently, while the mAP score reached 73.5% when the trained YOLO v5 algorithm has been tested based on the original image dataset, this score dropped down to 48.6%, 35.6%,24.4 %, and 23.5 % when applied to detect the objects from the same images after injecting gaussian noise with 30%, 50%, 80%, and 100% proportions respectively. Furthermore, the F1-score of the YOLO v5 algorithm was 69% when used for detected objects in the original image dataset compared with 48%, 34%, 23%, and 21% obtained when the YOLO-trained algorithm was adopted for detecting the

JSR A
Journal of Scientific Reports

objects from the same image but after injecting gaussian noise with 30%, 50%, 80%, and 100% proportions respectively. Also, the results showed that the performance of the YOLO algorithm was still at an acceptable rate when the amount of injected noise was 30%. Also, the results showed that when the amount of injected noise reached 50% the performance of YOLO dropped significantly. The results were very bad when the amount of the injected noise reached 100%, where most of the objects could not be detected by YOLO, and the detected objects were detected incorrectly.

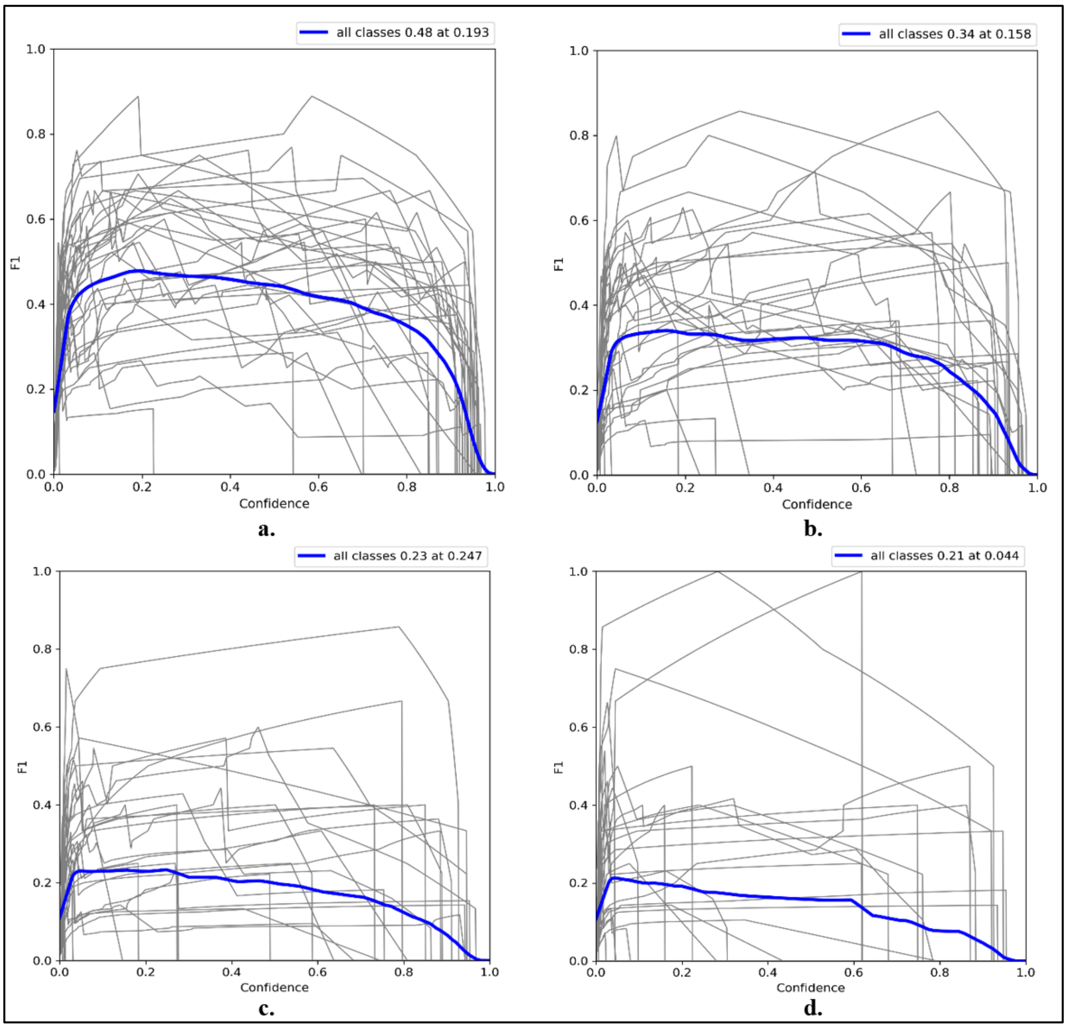
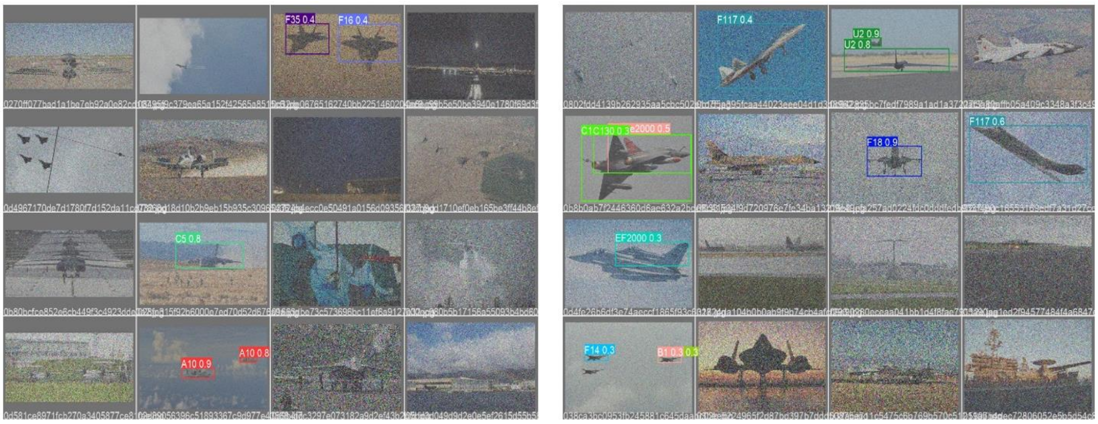


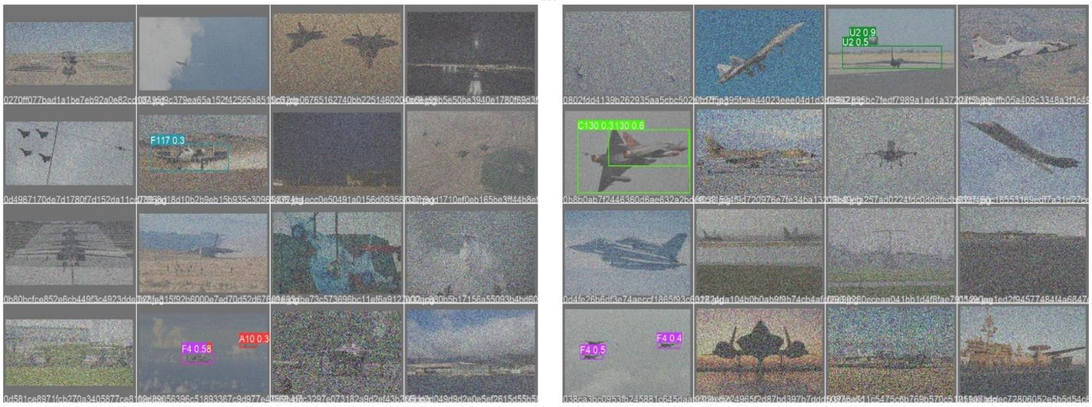
Figure 6. F1 score results obtained by applying YOLO on noisy datasets.
a. 30% noise, **b.** 50% noise, **c.** 80% noise, **d.** 100% noise.

Figure 7. Examples of the YOLO detection on the original images.

Figure 8. Examples of the YOLO detection on the 30%-Gaussian-Dataset and 50% Gaussian-Dataset. **a.** 30% noise, **b.** 50% noise.



a.



b.

Figure 9. Examples of the YOLO detection on the 80% Gaussian-Dataset and 100% Gaussian-Dataset. **a.** 80% noise, **b.** 100% noise.

It can be concluded from this study that the well-known object detection algorithms can fail in detecting objects in real-life systems. Therefore, in order to make an efficient object detection in noisy environments there is an urgent need to add a co-model that can be utilized to denoising and processing images before handling them by these types of algorithms. In future works, we will try to propose and test some co-models that can be used alongside these types of algorithms in order to improve their performance such that they can detect objects both in noisy and noise-free environments. Several approaches can be utilized to improve the performance of the model such as fine-tuning auto-encoder models in order to obtain as clear version of the image as possible before handling it using the object detection architectures.

ACKNOWLEDGEMENT

This work has been supported by the Scientific Research Projects Coordination Unit of the Sivas University of Science and Technology. Project Number: 2023-GENL-Müh-0007.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] Ramík, D.M., Sabourin, C., Moreno, R., and Madani, K. (2014). A machine learning based intelligent vision system for autonomous object detection and recognition. Applied Intelligence. 40, 358–375.

[2] Nallasivam, M., and Senniappan, V. (2021). Moving human target detection and tracking in video frames. Studies in informatics and control. 30, 119–129.

[3] Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. (2014). Scalable object detection using deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2147-2154).

[4] Han, F., Liu, B., Zhu, J. and Zhang, B. (2019). Algorithm design for edge detection of high-speed moving target image under noisy environment. Sensors, 19(2), p.343.

[5] Razakarivony, S., and Jurie, F. (2016). Vehicle detection in aerial imagery: A small target detection benchmark. J Vis Commun Image Represent. 34, 187–203.

[6] Wang, Z., Du, L., Mao, J., Liu, B., and Yang, D. (2019). SAR target detection based on SSD with data augmentation and transfer learning. IEEE Geoscience and Remote Sensing Letters. 16, 150–154.

[7] Xu, Q., Peng, J., Shen, J., Tang, H., and Pan, G. (2020). Deep CovDenseSNN: A hierarchical event-driven dynamic framework with spiking neurons in noisy environment. Neural Networks. 121, 512–519.

[8] Bakir, H., Oktay, S., and Tabaru, E. (2023). Detection of pneumonia from x-ray images using deep learning techniques. Journal of Scientific Reports-A. 419–440.

[9] Akgül, İ. and and Volkan, K.A.Y.A. (2022). Classification of cells infected with the malaria parasite with ResNet architectures. Journal of Scientific Reports-A, (048), pp.42-54.

[10] Bakir, H. and Yilmaz, Ş. (2022). Using transfer learning technique as a feature extraction phase for diagnosis of cataract disease in the eye. International Journal of Sivas University of Science and Technology, 1(1), pp.17-33.

[11] Tekin, S., Murat, G.O.K., Namdar, M. and Başgümüş, A. (2022). Autonomous guidance system for UAVs with image processing techniques. Journal of Scientific Reports-A, (051), pp.149-159.

[12] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[13] He, K., Zhang, X., Ren, S. and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9), pp.1904-1916.

[14] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[15] Ren, S., He, K., Girshick, R. and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.

[16] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[17] Redmon, J. and Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).

[18] Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[19] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C. (2016). SSD: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

[21] Felzenszwalb, P.F., Girshick, R.B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell. 32, 1627–1645.

[22] Ferrari, V., Jurie, F. and Schmid, C. (2010). From images to shape models for object detection. International journal of computer vision, 87(3), pp.284-303.

[23] Ren, X., and Ramanan, D. (2013). Histograms of sparse codes for object detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 3246–3253

[24] Girshick, R., Felzenszwalb, P. and McAllester, D. (2011). Object detection with grammar models. Advances in neural information processing systems, 24.

[25] Salakhutdinov, R., Torralba, A. and Tenenbaum, J. (2011). Learning to share visual appearance for multiclass object detection. In CVPR 2011 (pp. 1481-1488). IEEE.

[26] Alahi, A., Ortiz, R. and Vandergheynst, P. (2012). Freak: Fast retina keypoint. In 2012 IEEE conference on computer vision and pattern recognition (pp. 510-517). Ieee.

[27] Zhou, X., Yang, C. and Yu, W. (2012). Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE transactions on pattern analysis and machine intelligence, 35(3), pp.597-610.

[28] Zhu, L., Chen, Y., Yuille, A. and Freeman, W. (2010). Latent hierarchical structural learning for object detection. In 2010 IEEE computer society conference on computer vision and pattern recognition (pp. 1062-1069). IEEE.

[29] Felzenszwalb, P.F., Girshick, R.B., and McAllester, D. (2010). Cascade object detection with deformable part models. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 2241–2248

[30] Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., and Li, S. (2013). Salient object detection: A discriminative regional feature integration approach. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 2083–2090

[31] Kim, C., Lee, J., Han, T., and Kim, Y.M. (2018). A hybrid framework combining background subtraction and deep neural networks for rapid person detection. J Big Data. 5.

[32] Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M. and Lee, B. (2022). A survey of modern deep learning based object detection models. Digital Signal Processing, p.103514.

[33] Nobis, F., Geisslinger, M., Weber, M., Betz, J., and Lienkamp, M. (2019). A deep learning-based radar and camera sensor fusion architecture for object detection; A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection.

[34] Elhoseny, M. (2020). Multi-object detection and tracking (modt) machine learning model for real-time video surveillance systems. Circuits Syst Signal Process. 39, 611–630.

[35] Das, S., Pal, S. and Mitra, M. (2016). Real time heart rate detection from ppg signal in noisy environment. In 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI) (pp. 70-73). IEEE.

[36] Nayan, A.-A., Saha, J., Mahmud, K.R., al Azad, A.K., and Kibria, M.G. (2020). Detection of objects from noisy images. In: 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI). pp. 1–6. IEEE

[37] Yadav, K., Mohan, D., and Parihar, A.S. (2021). Image detection in noisy images. In: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). pp. 917–923

[38] Milyaev, S. and Laptev, I. (2017). Towards reliable object detection in noisy images. Pattern Recognition and Image Analysis, 27, pp.713-722.

[39] Medvedeva, E. (2019). Moving object detection in noisy images. In: 2019 8th Mediterranean Conference on Embedded Computing (MECO). pp. 1–4. IEEE

[40] Que, J.F., Peng, H.F., and Xiong, J.Y. (2019). Low altitude, slow speed and small size object detection improvement in noise conditions based on mixed training. In: Journal of Physics: Conference Series. p. 012029. IOP Publishing

[41] Lee, G., Hong, S., and Cho, D. (2021). Self-supervised feature enhancement networks for small object detection in noisy images. IEEE Signal Process Lett. 28, 1026–1030

[42] Singh, M., Govil, M.C., and Pilli, E.S. (2018). V-SIN: visual saliency detection in noisy images using convolutional neural network. In: 2018 Conference on Information and Communication Technology (CICT). pp. 1–6. IEEE

[43] Gautam, A., and Biswas, M. (2018). Whale optimization algorithm based edge detection for noisy image. In: 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS). pp. 1878–1883. IEEE

[44] Mathew, M.P., and Mahesh, T.Y. (2022). Leaf-based disease detection in bell pepper plant using yolo v5. Signal Image Video Process. 1–7.

[45] Cheng, L., Li, J., Duan, P., and Wang, M. (2021). A small attentional YOLO model for landslide detection from satellite remote sensing images. Landslides. 18, 2751–2765.

[46] Wu, D., Lv, S., Jiang, M., and Song, H. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Comput Electron Agric. 178, 105742

[47] Liu, G., Nouaze, J.C., Touko Mbouembe, P.L., and Kim, J.H. (2020). YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. Sensors. 20, 2145

[48] Chen, W., Huang, H., Peng, S., Zhou, C., and Zhang, C. (2021). YOLO-face: a real-time face detector. Vis Comput. 37, 805–813

[49] Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M. and Lee, B. (2022). A survey of modern deep learning based object detection models. Digital Signal Processing, p.103514.

[50] Nakamura, T. (2021). Military Aircraft Detection Dataset. Webpage: https://www.kaggle.com/datasets/a2015003713/militaryaircraftdetectiondataset