Recent Advances in Science and Engineering
Web page info: https://rase.yildiz.edu.tr
DOI: 10.14744/rase.2023.0004

Research Article

Improving object detection of UAV images with Real-ESRGAN

Settar YİĞİT*, Tülay YILDIRIM

Faculty of Electrical Electronic Engineering, Yildiz Technical University, Istanbul, Türkiye

ARTICLE INFO

Article history
Received: November 13, 2023
Revised: November 23, 2023
Accepted: December 07, 2023

Keywords:
Super-resolution, ESRGAN, YOLO, object-detection, UAV.

ABSTRACT

Unmanned Aerial Vehicles (UAVs), commonly known as drones, in recent years continue to gain popularity in various fields, ranging from the entertainment sector to the service industry. Their application areas are expanding every day, making them increasingly versatile. They are widely used in the defense industry, contributing to the have a voice of the countries that possess them. Given this context, object detection, tracking, and other customized tasks carried out using imagery obtained from UAVs have become significantly important. However, the images obtained from UAVs are generally low resolution and quality, as they need to be captured from a safe flight distance. This situation is a disadvantage for object detection applications. To reduce this disadvantage, various Super Resolution techniques have been developed. In this paper, the focus is on the critical importance of improvements in this field, especially within the defense sector, by utilizing ESRGAN and YOLO together to enhance the resolution of images captured from UAVs. The primary objective of this study is to enhance the efficacy of object detection by simultaneously augmenting the number of detected objects and improving the accuracy of the detection process. This research presents a comparative analysis of the outcomes achieved through two distinct approaches. Firstly, object detection is executed utilizing a pre-trained YOLO-V7 model on a LR image extracted from the VisDrone Dataset. Subsequently, the same YOLO-V7 model is deployed, but object detection is carried out on the SR version of the same LR image obtained from the ESRGAN network. The findings from this investigation unequivocally demonstrate that conducting object detection on the SR image not only results in a notable increase in the quantity of detected objects but also leads to a significant enhancement in the overall accuracy of the detection process.

Cite this article as: Yiğit S, Yıldırım T. Improving object detection of UAV images with Real-ESRGAN. Recent Adv Sci Eng 2023;3:2:33–39.

INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have rapidly evolved into a versatile and increasingly popular technology in recent years. UAVs find applications across a wide spectrum, especially in the defense industry with active use, encompassing areas such as border security, target identification, electronic warfare, among others. Beyond defense, they are extensively employed in fields like agricultural surveillance, environmental monitoring, energy

***Corresponding author.**
*E-mail address: settarygt@gmail.com

sector inspections, law enforcement operations, aerial photography and even entertainment drone flights. This diversity stands as one of the fundamental factors contributing to the growing popularity of UAVs. Furthermore, the advancement of technology and the reduction in costs have made UAVs more accessible, leading to a broader adoption of their advantages by a greater number of individuals and organizations.

One of the common applications of UAVs is object detection. In these applications, models consisting of YOLO (You Only Look Once) [1] and its derivatives are popularly used. YOLO is a deep learning algorithm and model used in the field of object detection. YOLO aims to detect and classify objects in an image by examining the entire image in a single pass. Unlike some other object detection methods, YOLO can detect and classify multiple objects simultaneously without requiring multiple processing steps. In other words, it can recognize objects in an image all at once.

However, since the images obtained from UAVs are generally of low-resolution (LR), it directly impacts the accuracy of object detection applications in a negative way. To address this drawback, various super-resolution (SR) models have been developed [2–7]. In this article primarily focused on models derived from Generative Adversarial Networks (GAN) [3, 4, 8]. GAN networks consist of two comparative structures. The first one, the Generator (G), aims to create a new image from simple noise data. On the other hand, the Discriminator (D), aims to distinguish whether the images generated by the G are real or fake [9]. In super-resolution applications, the G tries to obtain high-resolution (HR) from a LR image, while the D tries to distinguish whether this image is real HR or generated from LR. In object detection, training with SR images has been observed to results 8% better in quality compared to training with LR images [10]. These GAN models are Super-Resolution GAN (SRGAN), Enhancement SRGAN (ESRGAN) and Real-ESRGAN, chronologically. However, due to the normalization algorithms they employ, SRGAN and ESRGAN architectures significantly lag behind Real-ESRGAN in improving object detection [11]. Therefore, in this article focused on the Real-ESRGAN architecture. Figure 1 illustrates the general architecture of how object detection is performed from the LR image at the proposed method in this article.

The rest of this paper is organized as follows. Section **Literature Review** provides the related works. Section **Methodology** presents the proposed system in detail and experimental results. Finally, the concluding remarks are shown in **Conclusion**.

LITERATURE REVIEW

This article investigates the impact of Super Resolution models on object detection. In this section, we discuss existing methods related to our work.

The problem of enhancing the resolutions of LR images has been a long-standing issue, and various solutions have been proposed over the years. In 2017, the SRGAN architecture was introduced for this problem [4]. GAN based methods, such as SRGAN and ESRGAN, demonstrated remarkable performance in improving LR images. In 2018, the SRGAN architecture was further developed, leading to the proposal of the ESRGAN architecture. In this architecture, a structure called Residual Dense Block (RDB) was introduced along with an improved loss function. RDB made learning easier and allowed for the creation of deeper models. Additionally, the Batch Normalization Block in the SRGAN architecture was removed, reducing computational complexity and enabling faster training [8]. In 2021, the Real-ESRGAN architecture focused on a U-Net based Discriminator. The proposed U-Net based architecture provides detailed per-pixel feedback to the generator while maintaining the global coherence of synthesized images by offering global image feedback as well [12]. It has been suggested that using pure synthetic data during the training of the Real-ESRGAN model offers better visual performance [3].

However, despite the realistic appearance of the generated HR images, especially high-frequency details (such as image edges) suffer from degradation. Some studies have suggested that edge information plays a crucial role in object detection and that images generated by GAN networks have a negative impact in this regard. Therefore, preserving this feature enhances object detection accuracy [10]. When combined with edge detection, a 2% increase in object detection accuracy has been observed [13]. Deep learning-based object detection architectures are broadly categorized into two main types: One-stage detectors and Two-stage detectors. While One-stage detectors are

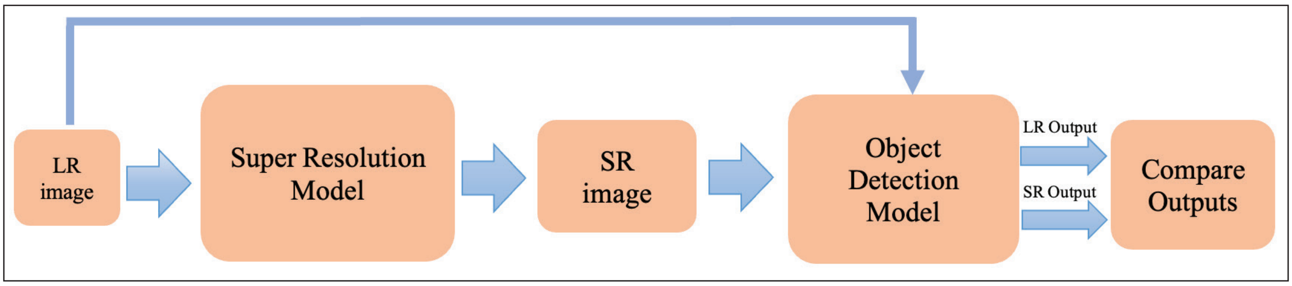

Figure 1. General Architecture of Object Detection from LR Image and SR Image.

efficient, they have lower accuracy rates compared to Two-stage detectors.

Xing et al., [14], worked on object detection with the UAV datasets which they created. In their studies, the PReLu function was proposed instead of the generally used ReLu activation function. It has also been suggested to remove the normalization block as in other studies. Through the utilization of SR images, the mean Average Precision (mAP) increased from 64.82% to 68.38, while the missing rate decreased from 16.98% to 14.23%

The impact of SR images on object detection was demonstrated in a paper [15]. In the utilized architecture, an Edge Enhancement network was employed between the GAN block and the Detection block. The study used the COWC dataset and the one-stage detector SSD for object detection. When trained and tested with LR data, an accuracy rate of 61.9% was achieved. However, when ESRGAN was used to enhance the resolution and subsequently train and test with SR data, the accuracy increased to 85.8%. Finally, when an Edge Enhancement network was used between ESRGAN and SSD, resulting in edge-enhanced super-resolution GAN (EESRGAN), an accuracy rate of 86% was attained [10]. In another similar study conducted in by Zou et al. [13], a four-stage object detection architecture was presented. In this approach, the ESRGAN block was initially used to increase resolution. Subsequently, an Edge Enhancement block was employed to extract edge information. Then image segmentation was applied, and the processed image was fed into the object detection block (YOLOv3). Using the COCO test-dev dataset, they achieved an accuracy rate of 42.2%, a significant improvement compared to the 31% accuracy rate of YOLOv3 without the enhancement stages.

Akhyar et al. [16] conducted a study on the steel industry's significant problem of defect detection and proposed a solution that involved the integration of GAN and a one-stage detector (SSD). They used the Severstal Steel Dataset for testing. When testing with LR images using another one-stage detector (YOLO-X), they achieved a 65% accuracy rate. However, when testing with SR images generated by GAN networks and using SSD for object detection, they reached an accuracy rate of 80.4%. In a different application involving pothole detection, Salaudeen and Çelebi et al. [17] used ESRGAN to enhance SR images and EfficientDet for object detection. They achieved a 32% accuracy rate with SR images, while the accuracy rate with LR images lagged significantly at 10.6%. In 2023, Lv et al. [18] used Real-ESRGAN in combination with SSD for missing bolt sub-detection, resulting in a 9.59% increase in accuracy compared to without GAN. Zhang et al., [19] used similar method for detection Rice Leaf disease. Also, Chen et al., [20] proposed a method for helmet wearing detection. Maqsood et al., [21] achieving the wheat stripe rust classification using the GAN methods.

METHODOLOGY

The proposed architecture aims to enhance the resolution of images taken from UAVs, allowing for the detection of objects that would normally go unnoticed. With this goal in mind, the architecture is presented in two main sections. Initially, the LR image is processed within the GAN block, where a pretrained Real-ESRGAN model is employed to generate an SR image. Subsequently, these SR images are individually fed into the Object Detection block, where the YOLOv7 architecture is utilized for object detection. The results obtained from this process are then compared with the results of directly using YOLOv7 on LR images, and the success rate of the architecture is determined.

Real ESRGAN, presented in Figure 2, is an improved version of ESRGAN. One of the key differences is the shift from the classical "first-order" degradation model to a "high-order" degradation model, as previous versions were unable to restore images with unknown and complex degradations. This is because when we take photos with our cellphones, they may exhibit various degradations such as camera blur, sensor noise, sharpening artifacts, JPEG compression, etc. To address this complexity and improve training stability, a new degradation model has been proposed. Additionally, changes have been made in the D structure, incorporating U-Net design and Spectral Normalization, while no changes have been made in the G structure.

However, the architecture used in this work is designed to demonstrate the only impact of SR images obtained from LR images on object detection. As stated in the introduction, LR images are first processed using the Real-ESRGAN structure to obtain SR images. The used Real-ESRGAN [3] architecture is a pre-trained model trained with the DIV2K

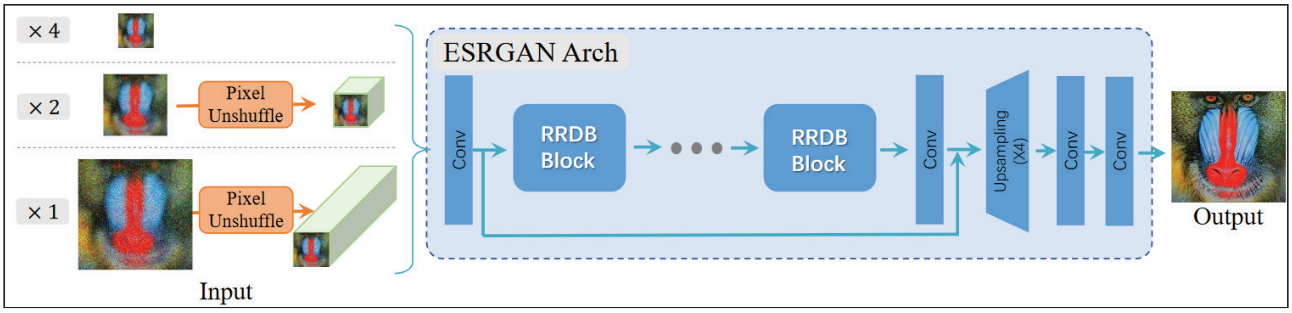

Figure 2. Architecture of Real-ESRGAN [3].

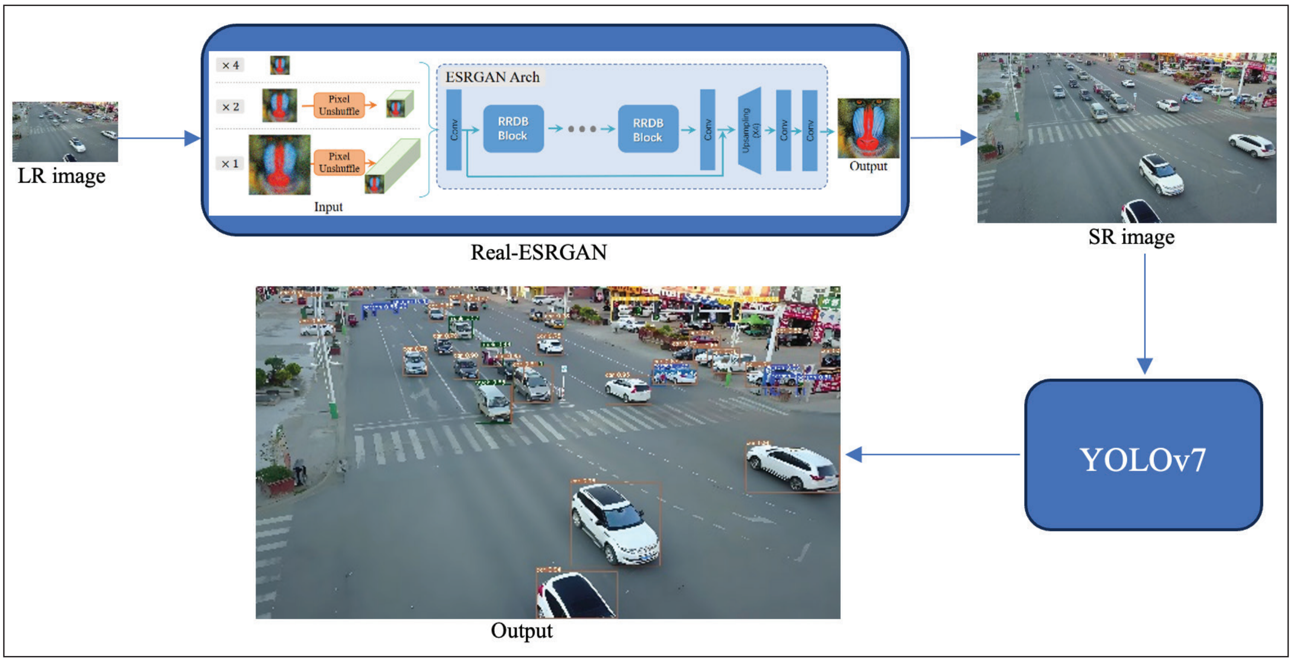

Figure 3. General Architecture of Proposed Model.

[22], Flickr2K [23], and OutdoorSceneTraining [24] datasets [3]. The resulting SR images are then input into the YOLOv7 model for object detection. YOLOv7 [25] is a significant iteration of the YOLO object detection model, widely recognized in the field of computer vision. YOLOv7 is designed to enhance the speed and accuracy of object detection while reducing model complexity. This version builds upon the success of its predecessors by optimizing the network architecture and introducing various improvements. YOLOv7 utilizes a streamlined model design, reducing computational requirements and enabling real-time performance on a variety of platforms. It incorporates advancements in anchor clustering, network scaling, and feature pyramid networks to enhance object detection precision. YOLOv7 is well-suited for a range of applications, including autonomous driving, surveillance, and image analysis, where efficient and accurate object detection is crucial. In the general architecture showed in Figure 3, the LR image from the VisDrone Dataset is input into the Real-ESRGAN block to obtain a higher resolution SR image. Object detection was performed on the resulting SR image using the pre-trained YOLOv7. The results are provided as output.

Experiments and Results

The LR images used in the proposed model were extracted from the images in the Validation section of the VisDrone dataset [26] include 548 images. The VisDrone dataset is a comprehensive benchmark dataset for visual object tracking, object detection, and single-object tracking tasks. It consists of high-quality aerial video sequences captured by various drone-mounted cameras. The dataset encompasses diverse scenarios, including urban areas, natural landscapes, and congested public events. It features a wide range of challenges such as object occlusion, scale variation, fast motion, and complex object interactions. The VisDrone dataset provides a rich collection of annotated data, including object categories, bounding boxes, object trajectories, and attributes, making it a valuable resource for developing and evaluating computer vision algorithms. Researchers and developers use this dataset to advance the field of visual object tracking and object detection, particularly in aerial and drone-related applications.

In [27], various object detection architectures were compared using three different datasets. In our paper, we will use the comparative results data to perform object detection with the same dataset after increasing its resolution using our own architecture. This will clearly demonstrate the impact of Super Resolution on aerial images object detection. The comparative results obtained are shown in Figure 4. In this paper, the results will be compared by applying the same object detection models again, using the SR versions obtained by employing the ESRGAN model on the same datasets.

The architecture in this paper has not been fully established and the development process continuing. Currently, at this stage of the study, the SR image was generated by employing the pretrained Real-ESRGAN model within an algorithm running on a T4 GPU in the Google Colab environment. The results were obtained as a result of the integration of Real-ESRGAN and YOLOv7. Accordingly, it has been observed that SR image detects more objects and the accuracy rates of the detected objects also increase. The results are shown figures 5-8. In Figures 5 and 7, the object detection results obtained using YOLOv7 with LR images are showed, while Figures 6 and 8 demonstrate the results obtained using SR images.

Model	Backbone	AU-AIR		SDD		VisDrone		FPS
		mAP@0.5	mAP@0.75	mAP@0.5	mAP@0.75	mAP@0.5	mAP@0.75	
CenterNet	ResNet18	72.93	69.63	59.93	41.63	40.63	21.76	**40.02**
CenterNet	ResNet34	61.91	71.43	61.91	47.43	50.65	24.43	32.32
CenterNet	ResNet50	79.33	76.92	66.33	49.92	52.78	25.91	30.19
CenterNet	DLA-43	83.02	79.82	70.42	61.82	53.04	27.76	28.91
CenterNet	ResNet101	85.32	80.13	74.32	66.01	54.90	29.03	17.12
CenterNet	ResNet152	88.72	84.31	79.72	69.31	55.56	29.67	16.07
CenterNet	Res2Net50	90.04	85.49	80.04	70.49	56.43	30.14	15.73
CenterNet	Res2Net101	91.31	**90.03**	**84.31**	72.43	**57.41**	31.65	7.21
CenterNet	hourglass-104	**92.20**	89.40	84.20	**74.40**	57.25	**35.55**	7.19
YoloV1	CNN	60.32	43.98	50.42	39.18	32.41	12.72	17.53
YoloV2	Darknet-19	66.32	50.38	55.32	41.38	34.21	13.22	16.32
YoloV3	Darknet-53	70.32	62.51	58.09	49.51	41.02	18.01	15.23
YoloV4	CSP-Darknet-53	73.03	67.52	68.03	52.52	42.21	20.28	12.62
YoloV5	FPN	80.42	76.92	71.43	60.43	44.32	28.92	12.17
YoloV6	Efficientrep	88.51	81.14	75.61	63.51	50.35	31.45	11.98
YoloV7	RepConN	90.93	88.03	79.43	65.03	53.11	33.35	11.32
SSD	MobileNetV2	72.82	68.37	74.98	43.38	43.32	20.17	32.12
Faster-RCNN	RPN	83.91	76.78	69.3	49.51	47.37	23.65	11.29

Figure 4. Comparison of mAP for three datasets [27].

Figure 4 shows the detection of 15 persons, 18 cars, 1 bus, 4 trucks, and 1 traffic light using LR images with YOLOv7. In Figure 5, object detection with SR images identified 14 persons, 1 bicycle, 29 cars, 1 motorcycle, 1 airplane, 1 bus, 5 trucks, and 3 traffic lights with YOLOv7. The original image has 37 persons, 4 bicycle, 34 cars, 13 van, 1 bus, 16 motor, tricycle 6. This means that while the detection made with LR images achieved a %40,5 mAP (15 of 37) accuracy rate for people and %52,9 mAP (18 of 34) for cars, the detection made with SR images achieved a %37,8 mAP (14 of 37) accuracy rate for people and %85,3 mAP (29 of 34) accuracy for cars.

Figure 5. LR image Object Detection.

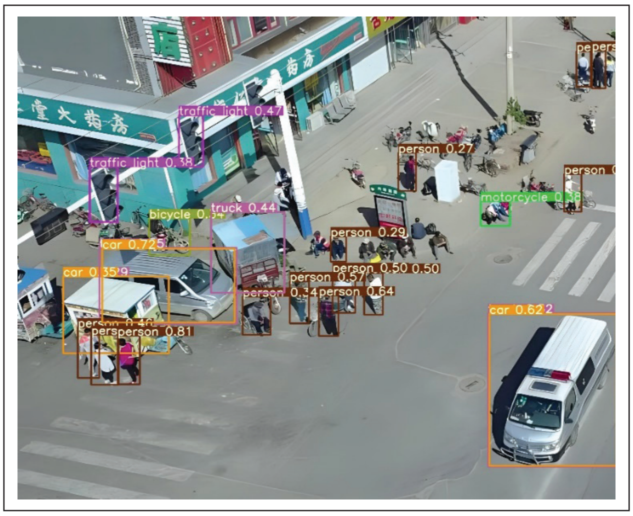

Figure 6. SR image Object Detection.

Figure 7. LR image Object Detection.

Figure 8. SR image Object Detection.

Figure 7 shows the detection of 3 persons, 22 cars, 1 bus, 4 trucks, and 1 traffic light using LR images with YOLOv7. In Figure 8, object detection with SR images identified 7 persons, 28 cars, 1 motorcycle, 1 bus, 4 trucks, and 2 traffic lights with YOLOv7. The original image has 13 persons, 4 tricycle, 40 cars, 3 truck, 1 bus, 4 motor. This means that while the detection made with LR images achieved a %23,1 mAP (3 of 13) accuracy rate for people and %55 mAP (22 of 40) for cars, the detection made with SR images achieved a %53,9 mAP (7 of 13) accuracy rate for people and %70 mAP (28 of 40) accuracy for cars.

CONCLUSION

Object detection in UAV images is one of the popular problems of recent times. Since these images are generally obtained from a safe flying distance, the objects are small and have low resolution. In this article, it is aimed to increase object detection in low resolution images with Real-ESRGAN and YOLO integration. According to the results obtained, more small objects could be detected in high-resolution images and an increase in the accuracy rates of the detected objects was observed. It is envisaged that these results will be further improved with the Edge Enhancement and Image Segmentation blocks obtained in other studies.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The published publication includes all graphics and data collected or developed during the study.

VisDrone Dataset: https://github.com/VisDrone/VisDrone-Dataset

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." arXiv, May 09, 2016. Accessed: Nov. 02, 2023. http://arxiv.org/abs/1506.02640 [CrossRef]

[2] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution." arXiv, Jul. 10, 2017. Accessed: Oct. 23, 2023. [Online]. Available: http://arxiv.org/abs/1707.02921

[3] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data." arXiv, Aug. 17, 2021. Accessed: Oct. 23, 2023. [Online]. Available: http://arxiv.org/abs/2107.10833 [CrossRef]

[4] C. Ledig L. Theis, F. Husz'ar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network." arXiv, May 25, 2017. Accessed: Oct. 23, 2023. http://arxiv.org/abs/1609.04802 [CrossRef]

[5] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang, "Real-World Super-Resolution via Kernel Estimation and Noise Injection," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA: IEEE, pp. 1914–1923, 2020. [CrossRef]

[6] A.-C. Tsai, C.-H. Tsou, and J.-F. Wang, "EESRGAN: Efficient & Effective Super-Resolution Generative Adversarial Network," IETE Technical Review, pp. 1–12, 2023. [CrossRef]

[7] M. Elad, and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," IEEE Transactions on Image Processing, Vol. 6(12), pp. 1646–1658, 1997. [CrossRef]

[8] X. Wang, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks." arXiv, Sep. 17, 2018. Accessed: Oct. 24, 2023. [Online]. Available: http://arxiv.org/abs/1809.00219

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks." arXiv, Jun 10, 2014. Accessed Oct 23, 2023. http://arxiv.org/abs/1406.2661

[10] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-object detection in remote sensing ımages with end-to-end edge-enhanced gan and object detector network," Remote Sensing, Vol. 12(9), Article 1432, 2020. [CrossRef]

[11] Z. Zhu, Y. Lei, Y. Qin, C. Zhu, and Y. Zhu, "IRE: Improved Image Super-Resolution Based on Real-ESRGAN," IEEE Access, Vol. 11, pp. 45334–45348, 2023. [CrossRef]

[12] E. Schönfeld, B. Schiele, and A. Khoreva, "A U-Net Based Discriminator for Generative Adversarial Networks." arXiv, Mar. 19, 2021. Accessed: Oct. 24, 2023.: http://arxiv.org/abs/2002.12655

[13] H. Zou, Y. Gao, X. Guo, and M. Zheng, "Small Object Detection Based on Super-Resolution Enhanced Detection Network," in 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI), Kunming, China: IEEE, pp. 99–106, 2021. [CrossRef]

[14] C. Xing, X. Liang, and Z. Bao, "A Small Object Detection Solution by Using Super- Resolution Recovery," in 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China: IEEE, pp. 313–316. 2019. [CrossRef]

[15] Y. Chen, J. Li, Y. Niu, and J. He, "Small Object Detection Networks Based on Classification-Oriented Super-Resolution GAN for UAV Aerial Imagery," in 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China: IEEE, pp. 4610–4615, 2019. [CrossRef]

[16] F. Akhyar, E. N. Furqon, and C.-Y. Lin, "Enhancing Precision with an Ensemble Generative Adversarial Network for Steel Surface Defect Detectors (EnsGAN-SDD)," Sensors, Vol. 22(11), Article 4257, 2022. [CrossRef]

[17] H. Salaudeen and E. Çelebi, "Pothole Detection Using Image Enhancement GAN and Object Detection Network," Electronics, vol. 11(12), Article 1882, 2022. [CrossRef]

[18] H. Lv, A. Zhao, Y. Wang, H. Li, Y. Li, and G. Zou, "Research on Missing Bolt Sub-Detection Based on Real-ESRGAN Super-Resolution Reconstruction," in 2023 4th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China: IEEE,. pp. 520–523, 2023. [CrossRef]

[19] Z. Zhang, Q. Gao, L. Liu, and Y. He, "A High-Quality Rice Leaf Disease Image Data Augmentation Method Based on a Dual GAN," IEEE Access, Vol. 11, pp. 21176–21191, 2023. [CrossRef]

[20] S. Chen, J. Lan, H. Liu, C. Chen, and X. Wang, "Helmet Wearing Detection of Motorcycle Drivers Using Deep Learning Network with Residual Transformer-Spatial Attention," Drones, Vol. 6(12), Article 415, 2022. [CrossRef]

[21] M. H. Maqsood, R. Mumtaz, I. U. Haq, U. Shafi, S. M. H. Zaidi, and M. Hafeez, "Super Resolution Generative Adversarial Network (SRGANs) for Wheat Stripe Rust Classification," Sensors, Vol. 21(23), Article 7903, 2021. [CrossRef]

[22] E. Agustsson, and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA: IEEE, pp. 1122–1131, 2017. [CrossRef]

[23] R. Timofte, E. Agustsson, L. V. Gool, M.-H. Yang, L. Zhang, "NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results," in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA: IEEE, pp. 1110–1121, 2017. [CrossRef]

[24] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform." arXiv, Apr. 09, 2018. Accessed: Nov. 03, 2023. http://arxiv.org/abs/1804.02815

[25] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." arXiv, Jul. 06, 2022. Accessed: Oct. 27, 2023. http://arxiv.org/abs/2207.02696 [CrossRef]

[26] P. Zhu, L. Wen, X. Bian, H. Ling, and Q. Hu, "Vision Meets Drones: A Challenge." arXiv, Apr. 22, 2018. Nov 13, 2023.: http://arxiv.org/abs/1804.07437

[27] Z. Saeed, M. H. Yousaf, R. Ahmed, S. A. Velastin, and S. Viriri, "On-Board Small-Scale Object Detection for Unmanned Aerial Vehicles (UAVs)," Drones, Vol. 7(5), Article 310, 2023. [CrossRef]