FEDERAL UNIVERSITY OF ITAJUBÁ - UNIFEI
POSTGRADUATE PROGRAM IN
COMPUTER SCIENCE AND TECHNOLOGY

Application of Real-ESRGAN in Improving IR Sensor Images for Use in SAR Operations.

Vinícius Henrique Geraldo Correa

Itajubá, September 13, 2024

FEDERAL UNIVERSITY OF ITAJUBÁ - UNIFEI
POSTGRADUATE PROGRAM IN
COMPUTER SCIENCE AND TECHNOLOGY

Vinícius Henrique Geraldo Correa

Application of Real-ESRGAN in Improving IR Sensor Images for Use in SAR Operations.

Thesis submitted to the Postgraduate Program in Computer Science and Technology (PPGCTC) as part of the requirements to obtain the title of Master of Science in Computer Science and Technology.

Concentration Area: Computer Mathematics

Supervisor: Prof. Dr. Alexandre Carlos Brandão Ramos

September 13, 2024
Itajubá

FEDERAL UNIVERSITY OF ITAJUBÁ - UNIFEI
POSTGRADUATE PROGRAM IN
COMPUTER SCIENCE AND TECHNOLOGY

Application of Real-ESRGAN in Improving IR Sensor Images for Use in SAR Operations.

Vinícius Henrique Geraldo Correa

Thesis approved by the examining board in September 11, 2024, giving the author the **Master's degree in Computer Science and Technology.**

Examination Board:
Dr. José Renato Garcia Braga
Prof. Dr. Carlos Henrique da Silveira

Itajubá
2024

Vinícius Henrique Geraldo Correa
Application of Real-ESRGAN in Improving IR Sensor Images for Use in SAR Operations/ Vinícius Henrique Geraldo Correa. – Itajubá, September 13, 2024-
87 p. : il.

Supervisor: Prof. Dr. Alexandre Carlos Brandão Ramos

Thesis (Master of Science)
Federal University of Itajubá - UNIFEI
Postgraduate Program in Computer Science and Technology, September 13, 2024.

1. IR Sensor. 2. UAV. 3. Real-ESRGAN. 4. SAR. II. Federal University of Itajubá. III. Faculdade de xxx. IV. Application of ESRGAN in Improving IR Sensor Images for Use in SAR Operations

Acknowledgements

I want to express my sincere gratitude to God for the strength, wisdom, and health, which have always been the essential pillars in the dedication and success of all my life and career achievements. I am deeply grateful to my family, whose love, unconditional support, and constant encouragement have motivated me to move forward. In particular, I thank my father, Ademilson Geraldo Correa, for his support and help in times of ups and downs, and my late mother, Edna Aparecida Fernandes, for shaping the person I am today. I would also like to thank my brothers, Diogo and Victor, who always believed in me and supported me at every stage of my academic training. Your words of encouragement and love were fundamental for me to get here.

I thank my advisor, Prof. Dr. Alexandre Carlos Brandao Ramos, for the research idea and acceptance in guiding this incredible learning journey and for his patience and valuable advice.

I also express my gratitude to the institutions that provided the necessary support to carry out this research. I want to thank Unifei for the exemplary structure and organization, and highly qualified teachers, who offered to help develop this work.

Finally, I would like to thank everyone who, directly or indirectly, contributed to the realization of this project. Every gesture of support, every word of encouragement and every contribution were fundamental to completing this thesis. To everyone, my most sincere thank you.

This dissertation is the result of a collective effort, and I am immensely grateful to have had the help and support of so many special people throughout this academic journey.

" A mind that opens to a new idea will never return to its original size"

(Albert Einstein)

Resumo

A utilização de Veículos Aéreos Não Tripulados (VANTs) em operações de busca e salvamento tem crescido significativamente, principalmente devido à redução de custos e ao menor risco associado. No entanto, a eficácia desses veículos está intimamente ligada à qualidade dos sensores utilizados para captura e identificação de alvos, tornando a investigação desses equipamentos uma área crucial.

Este estudo apresenta uma revisão sistemática da literatura sobre a aplicação de Redes Adversariais Generativas (GANs) em imagens geradas por VANTs com foco em busca e resgate. Além disso, introduzimos uma metodologia que utiliza a ferramenta Real-ESRGAN para aprimorar imagens obtidas por VANTs durante missões de busca e salvamento, com foco em sensores que operam na faixa infravermelha. Os resultados da aplicação dessa técnica em nosso conjunto de dados, combinados com a validação utilizando a ferramenta YOLOv8, revelam melhorias significativas na qualidade das imagens. Isso sugere que a abordagem proposta pode ter aplicações valiosas no pós-processamento e na identificação de alvos humanos durante operações de busca e resgate.

Palavras-chaves: Visão computacional; Processamento digital de imagens; Busca e salvamento; Redes Generativas Adversariais.

Abstract

The use of Unmanned Aerial Vehicles (UAVs) in search and rescue operations has grown significantly, primarily due to reduced costs and lower associated risks. However, the effectiveness of these vehicles is closely linked to the quality of the sensors used for target capture and identification, making the investigation of these devices a crucial area of research.

This study presents a systematic review of the literature on the application of Generative Adversarial Networks (GANs) in UAV-generated images, with a focus on search and rescue. Additionally, we introduce a methodology that uses the Real-ESRGAN tool to enhance images obtained by UAVs during search and rescue missions, specifically targeting sensors that operate in the infrared spectrum. The results of applying this technique to our dataset show significant improvements in image quality, suggesting that this approach may have valuable applications in post-processing and in the identification of human targets in search and rescue operations.

Key-words: Computer vision; Digital image processing; Search and rescue; Generative adversarial networks.

List of Figures

Figure 1 – U-Net model with Spectral Normalization 20
Figure 2 – Generator Network for Real-ESRGAN 20
Figure 3 – Wavelength Spectrum Span 33
Figure 4 – Frames captured from three videos recorded by an infrared camera . . 34
Figure 5 – Frames after enhancement by Real-ESRGAN. 36
Figure 6 – Example illustrating the absence of noticeable contrast in some images. 37
Figure 7 – Example illustrating the lack of sharpness in certain regions of images after super-resolution. 37
Figure 8 – Artifacts after Super-resolution process. 37
Figure 9 – Example of a validation batch in Test 2. 38
Figure 10 – Training Loss Graphs for Tests 1 and 2. 42
Figure 11 – Precision and Recall across 300 epochs for Tests 1 and 2. 43
Figure 12 – Validation loss values for Tests 1 and 2. 44
Figure 13 – mAP50 and mAP50-95 metrics for both Tests. 45

List of Tables

Table 3 – Inclusion criteria. 29
Table 4 – Tests using YOLOv8 applied to the infrared datasets 32
Table 5 – Tests using YOLOv8 applied to the infrared datasets 35
Table 6 – Comparison of metrics from Tests 1 and 2 with results from other studies. 40
Table 7 – Comparison of metrics against Bicubic Interpolation. 40
Table 8 – Comparison of metrics against Bilinear Interpolation. 40
Table 9 – Comparison of metrics against Lanczos Interpolation. 41
Table 10 – Comparison of metrics against Nearest Neighbour Interpolation. 41
Table 11 – Comparison of metrics against Real-ESRGAN super-resolution. 41

List of abbreviations and acronyms

BSRGAN	Blind Super-Resolution Generative Adversarial Network	19
CDC	Component divide-and-conquer	19
DAN	Deep Alternating Network	19
ESRGAN	Enhanced Super Resolution Generative Adversarial Networks	18
FAB	*Força Aérea Brasileira*	15
GAN	*Generative adversarial Networks*	17
HR	High-resolution	15
LADD	*Lacmus Drone Dataset*	23
LR	Low-resolution	15
PICOC	*Population; Intervention; Comparison; Outcome; Context*	27
PRISMA	*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*	17
Real-ESRGAN	Real Enhanced Super Resolution Generative Adversarial Networks	18
RealSR	Real-World Super-Resolution	19
RRDB	Residual-in-Residual Dense Blocks	19
SAR	*Search and Rescue*	15
SLR	*Systematic Literature Review*	17
SN	Spectral Normalization	19
UAV	*Unmanned Aerial Vehicle(s)*	15
YOLO	*You Only Look Once*	23

List of symbols

$D(\mathbf{x})$	Discriminator's output for real data sample $\mathbf{x}$.	17
$V(D, G)$	Objective function.	17
$\log$	Natural logarithm function.	18
$\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}$	Expectation over real data distribution $p_{\text{data}}(\mathbf{x})$.	17
$\mathbf{x}$	Real data sample.	12
$\mathbf{z}$	Random noise vector.	12
$\max_D$	Maximizing with respect to the parameters of the discriminator.	17
$\min_G$	Minimizing with respect to the parameters of the generator.	17
$1 - D(G(\mathbf{z}))$	Arithmetic operation used in the loss function.	18
$G(\mathbf{z})$	Generator's output (fake data) given noise vector $\mathbf{z}$.	17
$\mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}$	Expectation over noise distribution $p_{\mathbf{z}}(\mathbf{z})$.	18
$p_{\mathbf{z}}(\mathbf{z})$	Distribution of noise vector.	12
$p_{\text{data}}(\mathbf{x})$	Distribution of real data samples.	12

Contents

1	**INTRODUCTION**	**15**
1.1	**Super-resolution**	**15**
1.2	**Systematic Literature Review**	**17**
1.3	**Generative adversarial Networks**	**17**
1.4	**Real-ESRGAN**	**18**
1.5	**OpenCV**	**20**
1.6	**YOLOv8**	**22**
1.7	**Motivation**	**22**
1.8	**Scope of the research**	**23**
1.9	**Related Work**	**23**
1.10	**Goals**	**24**
1.11	**Work Structure**	**25**
2	**METHODOLOGY**	**26**
2.1	**Systematic Literature Review**	**26**
2.1.1	Research Questions	26
2.1.2	PICOC framework	27
2.1.3	Search string and databases	27
2.1.4	First Step of Analysis - Abstract Analysis	28
2.1.5	Second Step of Analysis - Quality Assessment	28
2.1.6	Third step of Analysis - Metric Compilation	29
2.1.7	Fourth step of Analysis - Clustering and Article Finalization	32
3	**DEVELOPMENT**	**33**
3.1	**Target Detection Implementation**	**33**
3.1.1	Datasets Creation	33
3.1.2	Using Real-ESRGAN to produce SR images for SRGAN training	33
3.1.3	YOLOv8	33
4	**RESULTS**	**36**
4.1	**Real-ESRGAN**	**36**
4.2	**YOLOv8**	**38**
5	**ANALYSIS AND DISCUSSION**	**46**
6	**CONCLUSION**	**48**
6.1	**Validation Limits**	**48**

6.2 Main contributions 49
6.3 Future works 50
6.4 Final considerations 50

APPENDIX 51

APPENDIX A – SLR MANUSCRIPT 52

APPENDIX B – 1ST ROUND OF THE MANUSCRIPT REVIEW . 75

ANNEX 83

ANNEX A – CERTIFICATE OF ACCEPTANCE OF SLR PUBLICATION 84

BIBLIOGRAPHY 86

1 Introduction

SAR (*Search and Rescue*) is a critically important activity for preserving human life, safety, and societal comfort, primarily focusing on locating survivors and victims of accidents and natural disasters [1, 2, 3, 4, 5, 6, 7]. Depending on the search area, weather conditions, terrain, and other factors, this operation is conducted with the assistance of aerial vehicles manned by specialized teams such as firefighters, the Brazilian Air Force a.k.a FAB (*Força Aérea Brasileira*), and other agencies. Today, with advancements in automation, robotics, and sensors, UAV (*Unmanned Aerial Vehicle(s)*) have become significant allies in these operations. The use of drones offers advantages including cost reduction in equipment and personnel, while also playing crucial roles in expansive research areas within computer vision and digital image processing, given that photographic cameras are the primary sensors employed on these vehicles.

In a UAV-assisted search and rescue operation, the initial task involves enhancing sensor images to improve the identification of targets such as individuals, animals, or specific objects of interest. The effectiveness of this identification is significantly influenced by sensor quality, with factors like noise, light spectrum, blur, and object distance potentially degrading the performance of detection algorithms used for automated or manual target localization.

1.1 Super-resolution

To address the sensors' quality challenges, super-resolution techniques can be employed. Super-resolution is a method that enhances image resolution beyond its original quality by using algorithms to generate HR (High-resolution) images from LR (Low-resolution) inputs. This process involves reconstructing or estimating finer details in the image, thus improving clarity and effectiveness in target identification despite the limitations of the initial sensor quality.

Some classic examples of algorithms, as those presented on [8] focused on image super-resolution include Nearest neighbor, Bilinear, Quadratic, Bicubic, and Lanczos interpolations. These methods are used to enhance image quality by applying filters based on neighboring pixels to improve the images. For instance:

Nearest Neighbor Interpolation: This is the simplest and most direct method, where the value of the target pixel is simply the value of the nearest pixel in the original image. While it is fast and easy to implement, it tends to result in images with hard edges and blocky artifacts, as it does not consider surrounding pixels for smoothing. It is more suitable for situations where speed is critical and image quality is less of a priority.

Bilinear Interpolation: This method is simpler and faster compared to bicubic interpolation. It uses the 4 nearest pixels (a 2x2 grid) to estimate the value of the target pixel. Bilinear interpolation calculates a weighted average of these 4 pixels, taking into account the distance of the target pixel from each neighboring pixel. Although it is less sophisticated than bicubic interpolation, bilinear still provides a significant improvement over the basic nearest neighbor method, especially for moderate resolution increases.

Quadratic Interpolation: This method is slightly more complex than bilinear interpolation and is based on a quadratic polynomial. It uses a second-degree polynomial formula to estimate the pixel value, taking into account a larger set of neighboring pixels (generally a larger area than bilinear but smaller than bicubic). Quadratic interpolation can provide improved image quality compared to bilinear interpolation, but it still does not achieve the smoothness and precision of bicubic interpolation.

Bicubic Interpolation: This method considers the 16 closest pixels (a 4x4 grid) around the target pixel. Bicubic interpolation is more advanced than bilinear and quadratic methods because it uses a cubic polynomial to compute the value of the interpolated pixel. This results in smoother transitions between pixels and better image quality with fewer visual artifacts such as blurring or blockiness. It is often used in applications requiring high image quality, such as photo editors and printing software.

Lanczos Interpolation: this method uses a windowed sinc function, defined by a parameter called the "Lanczos kernel," to weigh surrounding pixels when computing new pixel values. The kernel's size, typically determined by the number of lobes (often 2 or 3), controls how many neighboring pixels influence the interpolation. This approach effectively balances sharpness and smoothness, making it particularly useful for maintaining image quality during scaling operations.

These algorithms are used to enhance image quality and reduce noise in images captured by various devices, including CT scans, X-rays, CCTV cameras, and smartphones, and they can also be applied to low-quality images produced by UAVs. UAVs, due to their high-speed movement, can introduce artifacts such as motion blur, further affecting image clarity. Interpolation techniques, originally designed for resizing images, also help mitigate noise and improve resolution by enhancing overall image detail and reducing distortions.

While classical image processing algorithms operate on pixel-level data within the image itself with kernels, generation algorithms can create new data from a trained latent space. This capability enables segments of low-quality images generated by drones to be artificially reconstructed, potentially enhancing contrast, resolution, and consequently improving the overall quality of images captured by these sensors.

1.2 Systematic Literature Review

A SLR (*Systematic Literature Review*) is a rigorous scientific method to synthesize evidence from multiple studies on a specific topic. Unlike common reviews, which may be more informal and less structured, a systematic review follows a predefined and detailed protocol to ensure objectivity and minimize biases.

The PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) protocol is a widely used guideline for planning, conducting, and reporting systematic reviews. It includes steps such as defining the research question, systematically searching for relevant studies, rigorously selecting articles, extracting data, and synthesizing results in a transparent and replicable manner.

The main difference between a systematic review and a common review lies in the robust methodology and systematic approach to identify, assess, and integrate all available evidence impartially. This enables researchers to conduct a comprehensive and reliable analysis of the available information on a specific topic, contributing to evidence-based clinical, policy, or research decisions grounded in solid and up-to-date evidence.

The first stage in the development of this work involved a systematic literature review using the PRISMA methodology. This SLR aimed to gather data on the application of GAN algorithms used in images captured by drones, to provide an overview of how these generative algorithms can be conceptualized and applied in search and rescue operations for people.

1.3 Generative adversarial Networks

The GAN (*Generative adversarial Networks*) are convolutional artificial neural networks designed to generate new images by training on a latent space. Developed by Ian Goodfellow in 2014 as an enhancement over autoencoders [9, 10, 11], they incorporate principles from zero-sum game theory. This framework involves two competing neural networks in a minimax-like framework: the generator, which creates new images, and the discriminator, which evaluates these generated images against real ones. The process of training lies on the minimax loss function described in Equation 1.1, where $V(D, G)$ is the objective function, represented by the right part of Equatiion 1.1 which we're minimizing ($\min_G$) with respect to the Generator or maximazing ($\max_D$) with respect of the discriminator. $\mathbf{x}$ and $\mathbf{z}$ denote the real data sample and random noise vector, respectively; $p_{\text{data}}(\mathbf{x})$ represents the distribution of real data samples($\mathbf{x}$); $p_{\mathbf{z}}(\mathbf{z})$ is the distribution of noise vector($\mathbf{z}$); $D(\mathbf{x})$ denotes the discriminator's output for a real data sample ($\mathbf{x}$), while $G(\mathbf{z})$ the generator's output, which generates fake data given a noise vector ($\mathbf{z}$); $\mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}$ indicates the expectation over the distribution of real data, while

$\mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}$ is the expectation over the distribution of the noise vector; log represents the natural logarithm function and $1 - D(G(\mathbf{z}))$ signifies the arithmetic operation used in the loss function.

In summary, the generator aims to minimize the loss function described, while the discriminator aims to maximize it.

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))] \tag{1.1}$$

The adversarial process requires the generator to continuously improve its outputs to better deceive the discriminator, with the training constrained by Nash's equilibrium. This makes defining an optimal stopping condition for the algorithm challenging. As the generator becomes more proficient at creating realistic images, it becomes increasingly difficult for the discriminator to distinguish them from real images, achieving the desired outcome of the algorithm. The result is a generator capable of producing artificial images that are indistinguishable from real-world images. GANs have various applications, such as colorizing black-and-white photographs, generating text, enhancing images through super-resolution, and more.

The Gans face several critical issues that can hinder their effectiveness. The vanishing gradient problem can disrupt generator training by diminishing the gradient flow, impairing learning and slowing convergence. Model collapse presents another challenge, where the generator may repetitively produce identical outputs, limiting diversity and reducing the quality of generated samples. Moreover, failure to converge can occur, reducing the discriminator's feedback effectiveness and preventing the generator from achieving optimal performance. These issues underscore the ongoing challenges in GANs' stability and effectiveness in generating high-quality, diverse outputs.

1.4 Real-ESRGAN

The tests and development in this study are based on applying the work of [12], utilizing their pre-trained model to enhance infrared images for subsequent detection of individuals in search and rescue operations. The primary contribution of these authors lies in introducing the Real-ESRGAN (Real Enhanced Super Resolution Generative Adversarial Networks) model for blind image super-resolution. Real-ESRGAN is designed to remove artifacts and enhance details in real-world images, using purely synthetic data during training. This innovative approach aims to improve the visual quality of images, surpassing traditional methods of blind super-resolution. Additionally, Real-ESRGAN demonstrates superior performance in artifact removal and restoration of textural details compared to previous approaches such as ESRGAN (Enhanced Super Resolution Generative Adversarial Networks)[13], DAN (Deep Alternating Network)[14], CDC (Component divide-and-conquer)[15], RealSR (Real-World Super-Resolution)[16], and BSRGAN (Blind Super-Resolution Generative Adversarial Network)[17].

In the context of Real-ESRGAN, the architecture of the Generative Adversarial Networks (GANs) is designed to effectively enhance image resolution. For the generator—the component responsible for producing high-resolution images from low-resolution inputs—the authors use a deep neural network that incorporates several Residual-in-Residual Dense Blocks RRDB (Residual-in-Residual Dense Blocks). These blocks are essential for capturing intricate details and generating high-quality images.

The training of Real-ESRGAN involves comparing high-resolution images with their low-resolution, degraded counterparts. This comparison helps the model learn how to reconstruct the high-quality details from the lower-quality inputs. Real-ESRGAN's training process is more extensive than ESRGAN because it uses a training dataset that covers a broader spectrum of image degradations. This makes the model's task of distinguishing between the original high-resolution images and their degraded versions more complex.

The authors employ a neural network architecture known as U-Net combined with Spectral Normalization SN (Spectral Normalization) to manage this increased complexity, as depicted in Figure 1. U-Net is particularly effective for image-to-image tasks due to its encoder-decoder structure with skip connections, which helps preserve details and improves the overall output quality. Spectral Normalization is a technique used to stabilize the training of GANs by controlling the spectral norm of weight matrices in the network, thus improving the model's stability and performance. Since ESRGAN is a computationally heavy network, the authors use a technique called pixel-unshuffle [18] before feeding the inputs into the main ESRGAN architecture on the generator network, Figure 2 shows the generator network.

The pixel-unshuffle is the inverse operation of pixel-shuffle. This step reduces the spatial size of the input images while enlarging the channel size. By performing most of the calculations on a smaller resolution space, this approach helps reduce GPU memory and computational resource consumption, making the training process more efficient.

Additionally, the Real-ESRGAN model was trained using synthetic images. This approach has several advantages. Synthetic images, which are artificially generated, allow for the creation of diverse and controlled degradation scenarios that might be rare or difficult to capture in real-world images. Training on these synthetic datasets helps the model generalize better, leading to improved performance when applied to real-world images.

Pre-trained models offer significant advantages in machine learning and computer

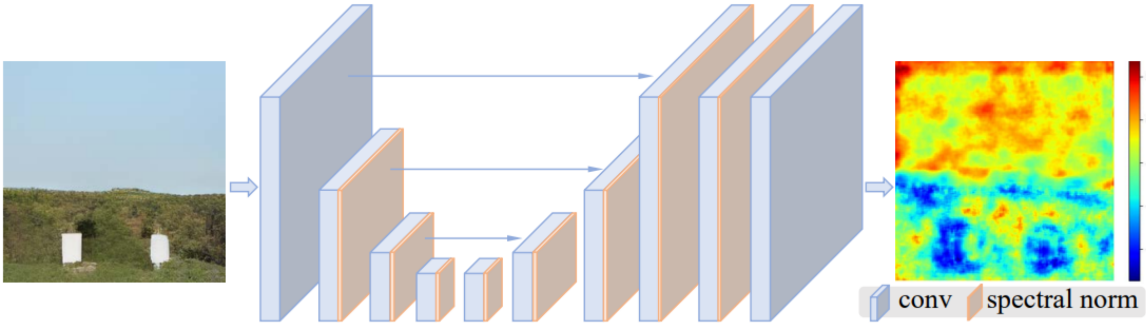


Figure 1 – U-Net model with Spectral Normalization proposed by the authors on Real-ESRGAN. **Source:** [12]

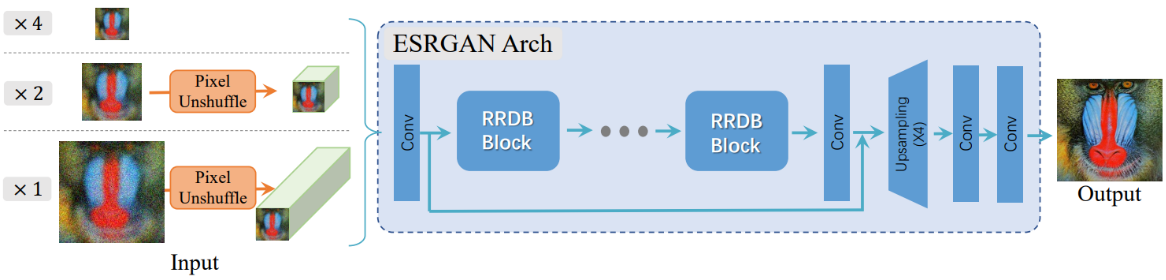


Figure 2 – The Generator Network in Real-ESRGAN takes LR images and transforms them into high-resolution HR images. This process includes a technique called "Pixel-unshuffle" which rearranges the pixels to make the image smaller in terms of its spatial dimensions but richer in channels. **Source:** [12]

vision tasks. They serve as starting points for training specific models, saving time and computational resources, especially useful with limited training datasets. These models capture general and abstract features from large datasets, improving their ability to generalize across different datasets and tasks. Utilizing pre-trained models can reduce the risk of overfitting, as they have already learned useful representations from diverse datasets. They enable rapid prototyping and validation of ideas in new projects, accelerating the development cycle. In many cases, pre-trained models outperform models trained from scratch, particularly in computer vision tasks where pre-training on large image datasets captures valuable visual information. Another important aspect to consider is that training a GAN network is computationally intensive and time-consuming. Therefore, using pre-trained models can significantly accelerate the process.

1.5 OpenCV

OpenCV, an acronym for Open Source Computer Vision Library[1], is an open-source library widely used in the field of computer vision and digital image processing.

[1] https://opencv.org/

Initially developed in C++, the library also has support for the Python language, which facilitates its use by developers and researchers in different areas of computer science. With OpenCV, we can load images and represent them in matrices, which can be interpreted by OpenCV itself or by the Matplotlib library, facilitating the use of linear algebra techniques. By the way, in 2005, for the first time in history, an autonomous vehicle called Stanley, which used OpenCV[2], traveled 132 miles in the Mojave Desert, winning the DARPA Grand Challenge, awarded a prize of 2 million dollars.

Computer vision is a branch of artificial intelligence focused on enabling computers to understand and interpret images and videos, similar to how humans do. OpenCV is a powerful library that offers many tools and functions to help you work on computer vision projects. Some of its key features include image pre-processing, applying filters, and post-processing.

Image preprocessing involves preparing images for subsequent analysis or processing. This stage may include converting color images to grayscale, normalizing brightness and contrast, and removing noise. OpenCV provides functions that streamline these operations, making the process quick and efficient.

The application of filters is another essential functionality that can be performed using OpenCV. Filters are used to enhance specific features of images such as edges, contours, textures, and can also be employed to address certain image degradations. We can create custom filters and apply them manually to images, expanding upon the convolution process mathematically performed. This allows for research, replication, and development of new techniques and frameworks for image processing.

Post-processing, in turn, involves enhancing images or extracting relevant information after the initial processing. This can include object segmentation, face detection, or identifying points of interest in an image.

In the context of artificial intelligence, OpenCV is often used in conjunction with other machine learning libraries such as TensorFlow and PyTorch. During the training of neural networks, it is common for images to need modification or adjustment to enhance the model's performance. OpenCV enables these modifications to be efficiently executed, simplifying the process of creating and improving AI models.

In our specific project, we utilize OpenCV to extract images from videos captured by infrared sensors, which will form the basis of our dataset. These images will undergo enhancement using Real-ESRGAN and subsequent testing with YOLOv8.

[2] https://docs.opencv.org/3.4/d0/de3/tutorial_py_intro.html

1.6 YOLOv8

YOLO (You Only Look Once) is a highly versatile computer vision tool capable of identifying, classifying, tracking, and segmenting objects in images. The application of YOLO for object detection involves the following steps:

1. **Dataset Preparation:** To train a YOLO network to detect objects, preparation of training and validation images along with their annotations is necessary. Annotating an image involves mapping the location of objects of interest, and these annotations must be exported in a format accepted by YOLO.

2. **Training and Evaluation:** After dataset preparation, the YOLO network is trained using the annotated images. Post-training, the model can be evaluated using various metrics to assess its performance.

3. **Integration with SAR:** In the context of Search and Rescue (SAR), the application of YOLO typically follows the enhancement of images by GAN algorithms, such as Real-ESRGAN with pre-trained models. This preprocessing step aims to improve image quality before YOLO performs object detection, enhancing the accuracy and reliability of target identification in SAR operations.

Version 8 enhances previous versions by improving its backbone network, feature fusion techniques, and overall architecture, thereby achieving superior real-time object detection capabilities with enhanced speed and accuracy.

1.7 Motivation

The use of low-cost equipment in UAVs, especially sensors, presents a trade-off that requires careful analysis in their application for search and rescue purposes.

On one hand, there is the advantage of affordability, but on the other hand, there is the issue of compromised efficiency in target identification due to the low quality of generated images, caused by factors such as blur, stability, resolution, noise, lens distortions, sharpness, among others. Additional factors like terrain topography, variations in relief, and object distance can also negatively affect image quality, introducing distortions and noise that hinder accurate data interpretation. These limitations can lead to misinterpretations, impairing target detection accuracy.

Given the significant potential of GANs for improving image quality, correcting imperfections, and enhancing resolution—particularly in the context of images captured by drones—we reviewed the literature and identified Real-ESRGAN [12] as the most suitable solution for our objectives. This algorithm is specifically designed to enhance image quality and works effectively in both visible and infrared spectra. The Real-ESRGAN models and code are available on GitHub[3], which simplifies its implementation and adaptation to our specific needs.

Therefore, aiming to mitigate the negative effects of low-cost sensors and adverse environmental conditions, we propose the study and utilization of Real-ESRGAN to evaluate the enhancement of images captured by UAVs. This, in turn, should improve the accuracy and reliability of target detection processes.

1.8 Scope of the research

Initially, we conducted a systematic literature review aimed at gathering data and information on the use of Generative Adversarial Networks (GANs) in the context of target detection in images generated by Unmanned Aerial Vehicles (UAVs). This review allowed us to identify the approaches and motivations behind the application of such algorithms in object detection from UAV-generated images, as well as the metrics employed to evaluate the outcomes.

Subsequently, we constructed two datasets: the first comprised images extracted from videos captured by an infrared camera, and the second consisted of the same images from the first dataset enhanced using Real-ESRGAN.

Following the previous stage, annotations were added to the dataset images to train YOLO (*You Only Look Once*)(version 8) for detecting people, aiming to compare the results between normal and enhanced images.

Next, we conducted a comparison of the high-resolution images with classical super-resolution algorithms to evaluate their performance using similarity-oriented metrics.

Finally, we discussed the potential implications of the results obtained, such as the feasibility of using GANs for enhancement and analysis of UAV images in real-world applications. We also suggested future research directions in this area, including the integration of additional sensors and the development of more advanced algorithms for multispectral image processing.

1.9 Related Work

Regarding Search and rescue, The most prominent work related in detecting people is the LADD (*Lacmus Drone Dataset*) [4]. This dataset was voluntarily created by

[3] https://github.com/xinntao/Real-ESRGAN

[4] https://www.kaggle.com/datasets/mersico/lacmus-drone-dataset-ladd-v40

organizations such as Owl and LisaAlert. It consists of 1365 diverse images captured by drones at a height of 40-50 meters, depicting people in various poses. The dataset also includes annotations in a format compatible with YOLO, which facilitates the training of automatic person detection in images captured by UAVs. However, the dataset is in the visible light spectrum.

Regarding detecting targets in infrared scenarious, [19] presents the ITD-YOLOv8 model, designed for detecting infrared targets using UAVs. The model utilizes the HIT-UAV dataset, comprising 2898 infrared images with classes including people, bicycles, and vehicles. The key contributions include an improved YOLOv8 backbone network for accurate target detection in complex environments and a lightweight convolutional operation for model efficiency. The model's ability to detect diverse targets in various scenarios enhances its applicability in search and rescue missions, where quick and precise target identification is crucial for successful operations.

[20] introduces the YOLOv8-EGP algorithm for infrared road object detection, which enhances the YOLOv8 model by improving accuracy in detecting small targets in infrared images. The study utilized the FLIR_ADAS_v2 dataset, consisting of 10,467 infrared images, and focused on detecting six specific classes: person, bike, car, bus, light, and sign. The YOLOv8-EGP model showed significant improvements in accuracy compared to the original model, making it suitable for real-world applications such as vehicle-assisted driving, nighttime road recognition, and intelligent transportation. The enhanced model's capabilities in detecting small targets contribute to its effectiveness in search and rescue operations.

[21] continues their work from [20] and presents an improved infrared road object detection algorithm based on an attention mechanism in the YOLOv8 model. The main contributions include incorporating the CPCA attention module and the CGBD downsampling module to enhance model accuracy and performance. They use the same FLIR_ADAS_v2 dataset. The enhanced model showed a 1.4% increase in mean average precision (mAP) compared to YOLOv8s, demonstrating improved precision and recall. This advancement in infrared object detection can benefit search and rescue operations by accurately detecting and tracking objects in challenging environments, enhancing safety and efficiency in such scenarios.

1.10 Goals

- Conduct a systematic literature review using the Prisma method to investigate how GAN algorithms are used to enhance low-quality images for subsequent target detection.

- Assess the practicality of using the Real-ESRGAN algorithm, with pre-trained models, to enhance images captured by infrared sensors.

- Apply YOLOv8 to the enhanced images and compare the results with those from other studies on target detection.

1.11 Work Structure

This study is structured as follows: Section 2 outlines the methodology employed in the systematic literature review, along with the tools utilized throughout the processes. Section 3 provides a detailed account of the step-by-step procedures used to test and validate the super-resolution application on our dataset. It includes a demonstration of how the Real-ESRGAN algorithm was implemented on images from our dataset using pre-trained models and how we use YOLOv8. Section 4 presents the primary outcomes derived from the tests conducted with the Real-ESRGAN and YOLOv8 algorithms. In Section 5, we analyze and discuss the findings presented in Section 4. Finally, Section 6 concludes our study with insights, discussions on validation limitations, and proposals for future research based on our contributions.

2 Methodology

2.1 Systematic Literature Review

In the first step of our research, we conducted a systematic literature review. This review aimed to analyze various GAN algorithms utilized for image enhancement, edge detection, or target identification, evaluating their metrics, objectives, and potential applications in search and rescue operations. The SLR plan followed the principles of the PRISMA methodology, a key reference for constructing reviews of this scope. The review began in mid-2023 and was completed in the early months of 2024.

2.1.1 Research Questions

Given the focus on search and rescue applications, we centered our research around the following research question, which was replicated in the RSL article: **How can GAN algorithms help detect edges or objects in images generated by UAVs?**

To assist in structuring the studies, we also formulated the following sub-questions:

1. How can GANs be addressed on SAR operations?
2. What benefits are gained from using a pre-trained model rather than training one from scratch?
3. Which metrics are most suitable for validating these algorithms?

Sub-question 1 above guides the main objective of the research, aimed at examining how edge and object detections can be addressed in search and rescue targets, specifically focusing on vulnerable individuals such as people or animals in distress situations.

Sub-question 2 above arose from the observation that Real-ESRGAN was able to enhance our images using a pre-trained model, which was not specifically related to those images. In other words, a different dataset could also be useful for improving images unrelated to it.

Sub-question 3 stemmed from the need to observe the metrics used in validating GANs, considering that these algorithms converge based on Nash equilibrium [9]. Therefore, analyzing their performance becomes challenging to achieve.

2.1.2 PICOC framework

In the context of a systematic literature review, PICOC (*Population; Intervention; Comparison; Outcome; Context*) is an acronym representing criteria for study selection. These criteria help to clearly define and specify the relevant articles for the review, ensuring that the research is conducted in a structured and comprehensive manner. For our study, we formulated the following PICOC criteria to guide the selection of studies:

- Population: Our population consisted of studies applying GAN algorithms to images generated by UAVs.
- Intervention: We seek algorithms and techniques for enhancing images generated by UAVs, specifically for detecting people.
- Comparison: We did not utilize this item, as our focus is not on comparing studies but rather on conducting a review of articles using GAN algorithms for object and edge detection.
- Outcome: Our outcome focused on validating the most effective solutions employing GANs in enhancing target and edge detection, especially when applied to people and animals.
- Context: Our context centers on publications using GANs for analysis in UAV images, particularly focusing on object detection, edge detection, and object classification.

2.1.3 Search string and databases

To maximize the retrieval of relevant articles addressing the main research question and sub-questions, we formulated the following search string: **("edge detection" OR "object detection") AND (uav OR drones) AND (gan OR "genera tive adversarial networks")**. We chose Scopus and IEEE Xplore as they are significant databases known for listing high-impact articles. After exporting the articles, we inputted their information into the Parsifal tool (https://parsif.al/) to facilitate the review process.

The search string returned 69 raw results, comprising 42 from the SCOPUS database and 27 from IEEE Xplore. After removing duplicates and secondary studies, 54 articles (31 from SCOPUS and 23 from IEEE Xplore) were submitted for abstract analysis. Our aim during this phase was to eliminate false positives, secondary studies, and those not related to the PICOC framework used in our research.

2.1.4 First Step of Analysis - Abstract Analysis

We conducted abstract analysis to eliminate false positives, secondary studies, or those unrelated to the theme of the systematic literature review. After this analysis, the number of articles under review decreased to 23 studies from the SCOPUS database and 16 from IEEE Xplore.

2.1.5 Second Step of Analysis - Quality Assessment

In this stage of quality assessment, we used the inclusion criteria defined in Table 3 to assign scores to the articles read in greater depth. Each inclusion criterion (IC) serves a specific purpose in selecting studies that use GANs on UAV images. Here's an explanation of how each criterion contributes:

- **IC1** helps in identifying studies where GANs are applied specifically to enhance edge detection or object detection in UAV images. It ensures that the focus is on the use of GANs for improving the clarity and accuracy of object boundaries or entire objects in the images captured by UAVs.
- **IC2** is relevant because using pre-trained models in GANs can significantly affect the performance and efficiency of the algorithm. Studies that employ pre-trained models often benefit from transfer learning, where knowledge gained from a large dataset in a different domain can be leveraged to improve performance on UAV images without requiring extensive training on UAV-specific data.
- **IC3** ensures that studies provide transparency regarding the evaluation metrics used to assess the performance of the GAN model applied to UAV images. Metrics such as accuracy, precision, recall, and F1-score are essential for understanding how well the GAN-enhanced images align with the intended objectives (e.g., detection accuracy of objects).
- **IC4** focuses on studies that specifically target images captured in the visible light spectrum by UAVs.
- **IC5** This criterion focuses on studies that specifically investigate images captured in the infrared spectrum by UAVs. It considers that the human body, which is the primary target for detection algorithms, emits heat in this spectral range. This characteristic can aid UAVs in search operations.
- **IC6** targets studies specifically applied to detect people or animals in UAVs images as the main objective of the study is for search and rescue.
- **IC7** was established to identify studies that incorporate any version of YOLO as part of their methodology for detecting objects in UAV images.

Table 3 – Inclusion criteria.

Code	Criteria
IC1	Does the GAN algorithm aim to assist in detecting edges or objects?
IC2	Does the authors employ a pre-trained model?
IC3	Does the paper provide the metrics used for the applied model?
IC4	Is the solution proposed in the study aimed at images within the visible light spectrum?
IC5	Does the solution presented in the study target images within the infrared spectrum?
IC6	Is the SAR algorithm designed to detect people or animals?
IC7	Does the study utilize any version of YOLO in its development?

The scores regarding the inclusion criteria were assigned as follows: if the article fully met the criteria, it received a score of 1.0; if it partially met the criteria, it received a score of 0.5; if the study did not meet the criteria or there was no mention of it, it received a score of 0.0. Although it would be possible to compile a table with these data, there is no reason to classify the articles as better or worse based on their scores. This is because the review yielded various types of results regarding the application of GANs, each with its particularities.

2.1.6 Third step of Analysis - Metric Compilation

Once quality scores were assigned, articles most aligned with the study objectives were organized into tables to qualitatively evaluate the metrics used. It was observed that the studies employed the following metrics:

Precision (P): Precision indicates the proportion of correctly predicted positive instances (true positives) relative to all instances predicted as positive (true positives + false positives). A high precision value signifies that the model has a low false positive rate, meaning that when it predicts an object, it is likely to be correct.

Recall (R): Recall, also known as sensitivity or true positive rate, evaluates the model's ability to identify and capture all relevant instances of the target objects within the dataset. It calculates the ratio of correctly predicted positive instances (true positives) to all actual positive instances in the dataset (true positives + false negatives). A high recall value indicates that the model can effectively detect most of the relevant objects present in the dataset.

The Structural Dissimilarity Index Measure (DSSIM) is a metric that addresses the structural information in images during the optimization process of a neural network model. DSSIM is used to improve the estimation of both the luminance and chrominance pixels of images, particularly in tasks such as image translation or synthesis. By incorporating DSSIM loss into the optimization process, the model becomes more aware of the structural information in the images, leading to better-shaped objects in the generated outputs. DSSIM defines a region (window size) in the images where it predicts the luminance and chrominance, thereby enhancing the overall quality of the generated images by encouraging spatial smoothness. The use of DSSIM helps the model produce more accurate and visually appealing results by focusing on preserving the structural details of the images.

The Peak Signal-to-Noise Ratio (PSNR) is a widely used metric for assessing the quality and similarity of images and video. It measures the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation.

The Structural Similarity Index (SSIM) is a metric used to assess the similarity between two images. SSIM takes into account three components of similarity: luminance, contrast, and structure. It measures how well the structure of the images is preserved when comparing a reference image to a target image. A higher SSIM value indicates a higher degree of similarity between the two images in terms of their structure and content. SSIM is commonly used in image processing and computer vision tasks to evaluate the quality of image restoration, compression, and transformation algorithms. It provides a more comprehensive assessment of image similarity compared to simple pixel-wise metrics like Mean Squared Error (MSE).

The Inception Score (IS) is a metric used to evaluate the quality and diversity of generated images in image generation tasks, particularly in the context of Generative Adversarial Networks (GANs). IS measures the quality of generated images based on two aspects: how realistic the images look (measured by the Inception-v3 network's classification accuracy) and how diverse the generated images are (measured by the entropy of the class distributions). A higher IS indicates that the generated images are both visually realistic and diverse in terms of different classes or categories. IS is a popular metric for assessing the performance of GAN models in generating high-quality and diverse images.

The Mean Absolute Error (MAE) is a metric commonly used to measure the average magnitude of errors between predicted and actual values in a dataset. In the context of image analysis, MAE can be used to evaluate the accuracy of pixel-wise predictions or image-to-image translations by calculating the absolute differences between corresponding pixels in the generated and ground truth images and then averaging these differences across all pixels. A lower MAE value indicates a smaller average error between the predicted and actual pixel values, reflecting a higher level of accuracy in the image prediction or translation task. MAE is a widely used metric in various machine learning and image processing applications to quantify the overall prediction error.

The Segmentation Score (S-Score) is a metric used to evaluate the quality of segmented images, particularly in the context of image translation tasks where segmentation plays a crucial role. The S-Score assesses how well the shape of a segmented region is preserved or transformed after the image translation process. It is calculated by comparing the segmented regions of the generated images with the ground truth segmented regions of the real images. A higher S-Score indicates that the shape of the segmented regions is well-maintained or accurately transformed during the image translation process, reflecting the effectiveness of the segmentation task in preserving the structural information of the objects in the images. S-Score is a valuable metric for evaluating the performance of segmentation models in image translation and related tasks.

The Dice Similarity Coefficient (DSC) is a metric commonly used in image segmentation tasks to evaluate the similarity between two sets of segmented images. It measures the spatial overlap between the segmented regions of two images and is calculated as the ratio of twice the intersection of the segmented regions to the sum of the pixels in both segmented regions. A higher DSC value indicates a greater overlap and similarity between the segmented regions of the two images, reflecting the accuracy of the segmentation task. DSC is a widely used metric in medical image analysis, computer vision, and other fields where image segmentation is a critical component of the analysis.

The Fréchet Inception Distance (FID) is a metric commonly used to evaluate the quality of generated images in image generation tasks, particularly in the context of Generative Adversarial Networks (GANs). FID measures the similarity between real and generated images by comparing statistics extracted from a pre-trained deep neural network (typically Inception-v3) on both sets of images. A lower FID score indicates that the generated images are more similar to the real images in terms of visual quality and diversity. It is a popular metric for assessing the performance of GAN models in generating realistic and diverse images.

Intersection over Union (IoU): IoU measures the overlap between the predicted bounding box and the ground truth bounding box. It is calculated as the intersection area divided by the union area of the two boxes.

Average Precision (AP): AP is a metric that considers precision and recall across different thresholds. It is commonly used to evaluate the performance of object detection models, especially in the context of precision-recall curves.

Mean Average Precision (mAP): mAP is the average of AP values calculated for multiple classes or categories in a multi-class object detection task. It provides an overall performance measure for the detector across all classes.

Accuracy: Accuracy is a general metric that measures the overall correctness of the predictions made by the model. In object detection, accuracy can be calculated based on the correct detection of objects within images.

False Positive Rate (FPR) and False Negative Rate (FNR): These metrics measure the rate of false positives and false negatives generated by the detector, respectively.

Receiver Operating Characteristic (ROC) Curve: The ROC curve is a graphical representation of the true positive rate against the false positive rate at various threshold settings. It is useful for evaluating the performance of binary classifiers.

2.1.7 Fourth step of Analysis - Clustering and Article Finalization

The fourth round of analysis occurred during the correction and revision phase of the SLR article, following the initial peer review. Articles were grouped into clusters based on shared objectives (Table 4). In total, five clusters were created, named A through E. Cluster A includes studies focusing on GANs for data augmentation, super-resolution, and motion prediction. Cluster B comprises studies utilizing GANs for deblurring, augmentation, and super-resolution. Cluster C consists of studies employing GANs for small object detection, fusion, and anomaly detection. Cluster D encompasses studies using GANs for image translation and fusion. Cluster E includes studies focusing on GANs for weather correction, adverse condition handling, and deblurring. This clustering aids researchers in focusing their research efforts on potential solutions applicable to search and rescue applications. After completing the systematic literature review, we proceeded to the stage of developing and analyzing the Real-ESRGAN algorithm with pre-trained models on our dataset. These steps will be discussed next.

Appendix A provides a detailed overview of all the stages conducted in the systematic literature review submitted to Drones Journal [1]. Appendix B displays the first round of corrections made to the manuscript. You can check this published SLR on [22].

Table 4 – Tests using YOLOv8 applied to the infrared datasets

Cluster	**nº grouped studies**
A: Data augmentation and enhancement	7
B: Super-resolution and deblurring	5
C: Anomaly and small-object detection	6
D: Image translation and fusion	3
E: Adverse condition handling	4

[1] EISSN: 2504-446X

3 Development

3.1 Target Detection Implementation

3.1.1 Datasets Creation

After drafting and submitting the SLR, we constructed three datasets to test with YOLOv8 for human detection. These datasets were created by extracting frames from three videos recorded by an infrared camera, depicting human targets in various poses and scenarios such as grassy areas, sidewalks, nearby trees, houses, and streets. The wavelength range of these images spans from 7 to 14 micrometers in the infrared spectrum (Figure 3). 500 random images of size 640x512 pixels were extracted from the first and second videos to form Dataset I, which was used for YOLOv8 training. Subsequently, 100 random frames were selected from all three videos to create Dataset II to validate the YOLOv8 training. Figure 4 displays four samples from the three videos.

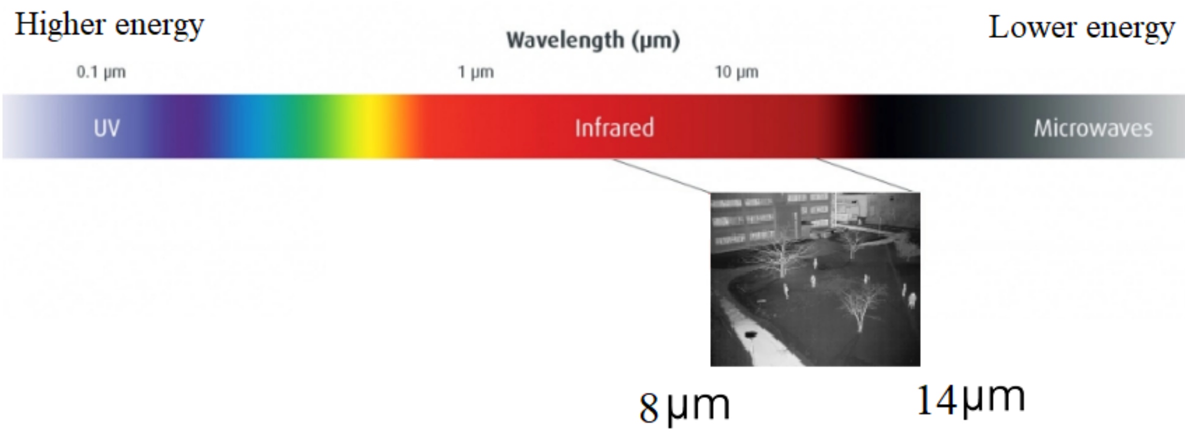


Figure 3 – The range of wavelengths covered by the samples in our dataset.

3.1.2 Using Real-ESRGAN to produce SR images for SRGAN training

Dataset III, also consisting of 100 random images from the three videos, was employed for the second test on YOLOv8, but the images were first processed by Real-ESRGAN with 4x upscaling using the RealESRGAN_x4plus pre-trained model.

3.1.3 YOLOv8

To utilize YOLOv8, we annotated the datasets using the CVAT[1] tool and extracted annotations in the format accepted by YOLO. We annotated only the "human" class,

[1] https://www.cvat.ai/

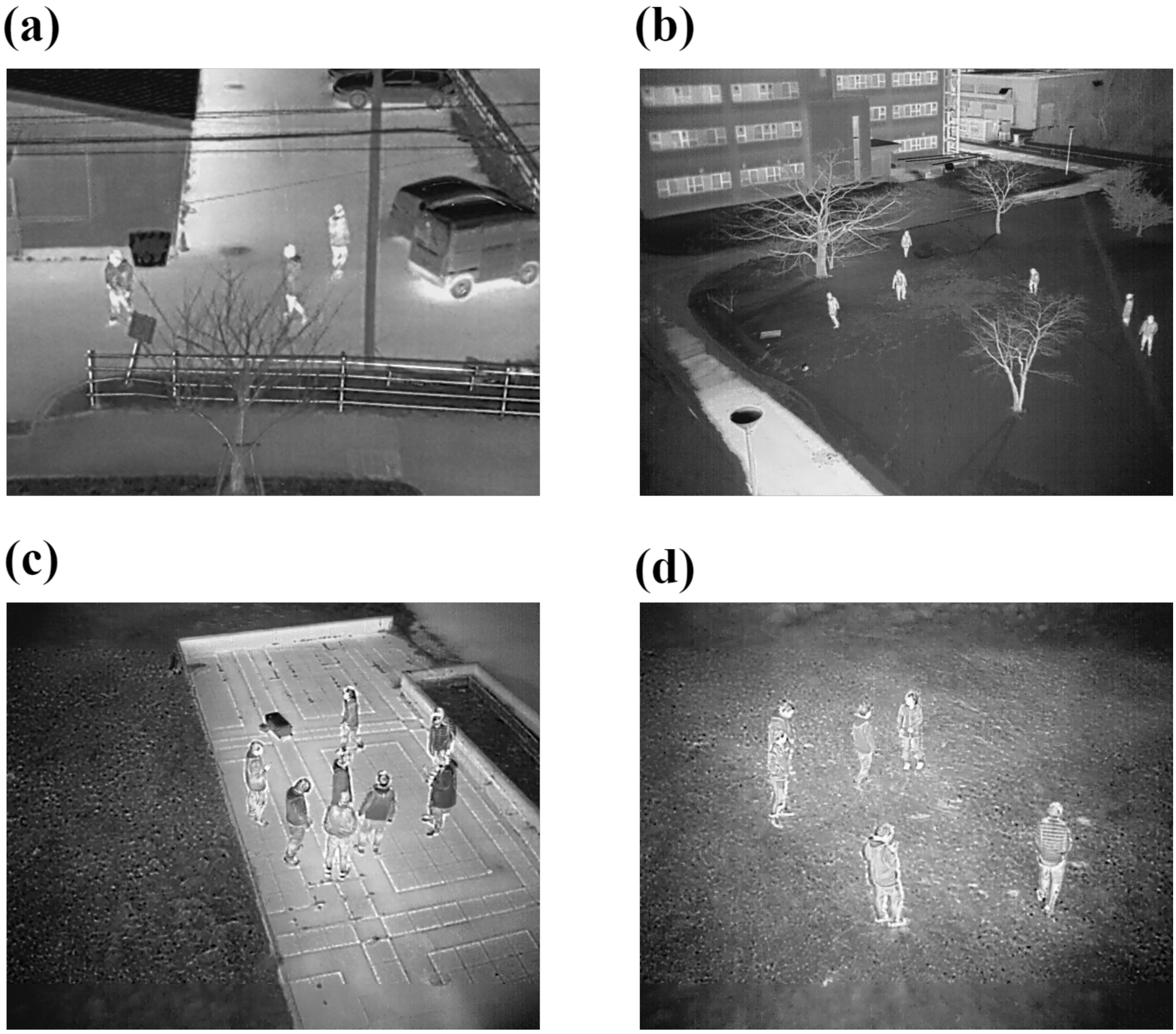


Figure 4 – Frames captured from three videos recorded by the infrared camera. Plot (a) corresponds to the first video, plot (b) to the second video, and plots (c) and (d) to the third video.

as the primary objective was to detect humans in various scenarios. Subsequently, we employed Google Colab with L4 GPU acceleration for the training process of YOLOv8. The data was uploaded to Google Drive and referenced by the notebook running YOLOv8 on Colab. The default parameters were used on YOLO training.

In summary, two tests were conducted using the datasets: the first (Test 1) validated training using the normal dataset with original images (Dataset II), while the second (Test 2) validated training using the normal dataset with enhanced images (Dataset III). Both tests utilized 300 epochs with 16 images per batch. Table 5 summarizes the YOLOv8 schema applied in this study.

Table 5 – Tests using YOLOv8 applied to the infrared datasets

Test nº	Train dataset	Validation dataset	Epochs	Batch size
1	I	II	300	16
2	I	III	300	16

4 Results

4.1 Real-ESRGAN

The main idea of this study was to assess the application of the pre-trained Real-ESRGAN algorithm for enhancing images in the infrared spectrum. Figure 5 illustrates a comparison between regular images from the created dataset and their enhanced version using the super-resolution tool. We observed significant improvements in the delineation of trees, building windows, and people after the process of enhancement by the algorithm.

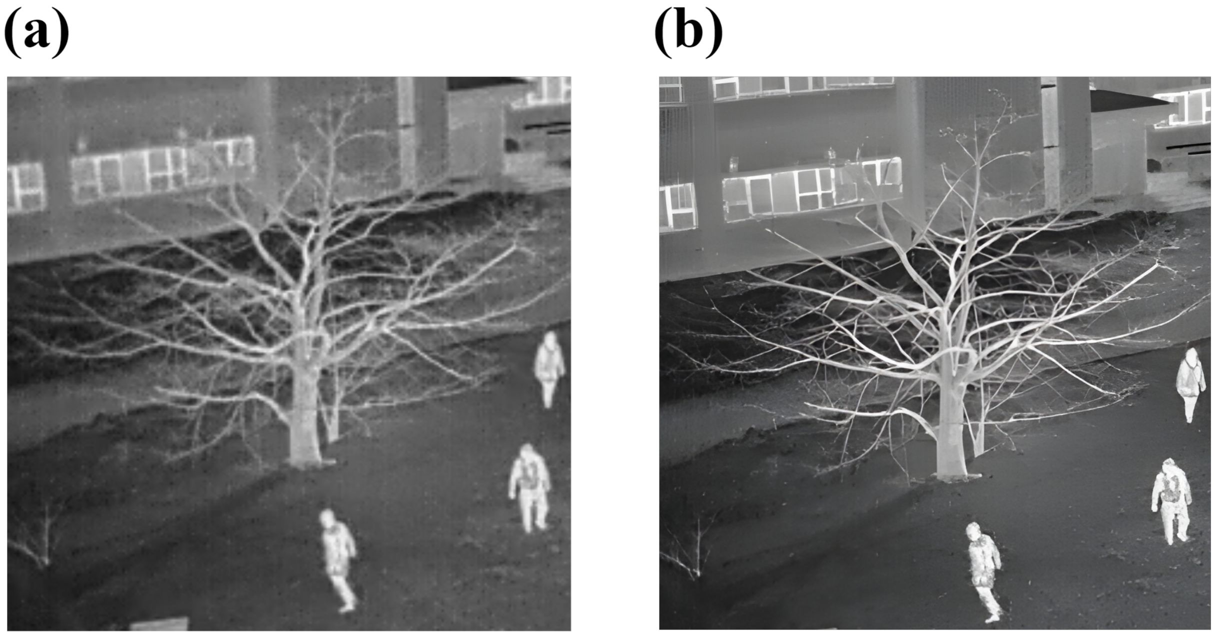


Figure 5 – Example of satisfactory contrast enhancement in image resolution by Real-ESRGAN. (a) Image before enhancement. (b) Image after the process.

Despite noticeable improvements in image clarity, we observed minimal contrast differences in some frames despite the increased scaling, as depicted in Figure 6, where we observed low sharpness and contrast in the outlines and segments of people and the ground after applying the algorithm, this reveals some limitations of Real-ESRGAN in certain types of images.

It is also worth noting that the algorithm did not succeed in enhancing sharpness and contrast in specific regions of the images, as observed in Figure 7, in the red square of Figure 7(b), we can observe the absence of improvements after applying the super-resolution algorithm.

Furthermore, artifacts were observed in certain areas of the images, as shown in Figure 8, where a grid pattern can be observed in the upper right region of Figure 8(b).

(a) (b)

Figure 6 – Example of minimal contrast improvement in an image after super-resolution enhancement. (a) Image before enhancement. (b) Same image after enhancement.

(a) (b)

Figure 7 – Regions where the algorithm was not effective in improvement. (a) Image before algorithm application. (b) Region (red rectangle) showing absence of noticeable contrast after enhancement.

(a) (b)

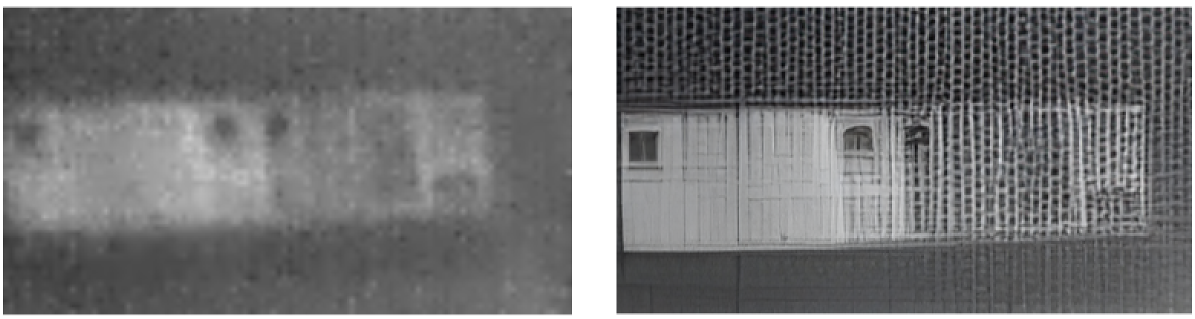


Figure 8 – Artifacts observed in a region of an image after applying the enhancement algorithm. (a) Original image depicting a window. (b) Observed artifacts.

4.2 YOLOv8

We used the standard batch size of 16 images for training and validation in YOLOv8, as seen in Figure 9. We trained one model at a time.

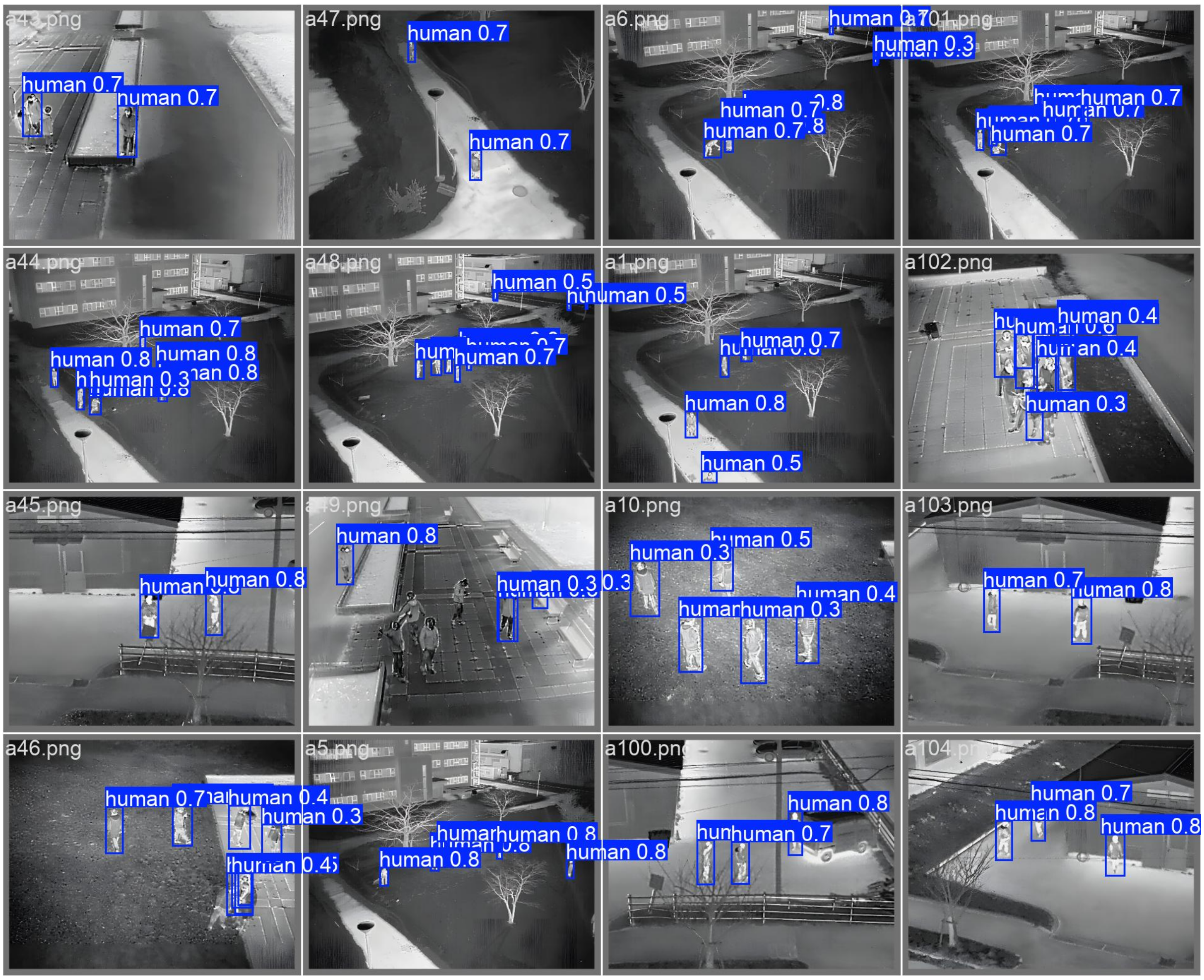


Figure 9 – Example of a batch during validation in the YOLOv8 process for Test 2.

YOLOv8 facilitates the extraction of metric data and results obtained during the process, which simplifies the analysis and processing of the outcomes.

The Box loss serves as the optimization criterion in object detection models, specifically YOLO, aiming to minimize the disparity between predicted and ground truth bounding box coordinates (x, y, width, height) of objects in images. Cls loss addresses the classification aspect by minimizing the difference between predicted class probabilities within bounding boxes and the actual class labels of objects, ensuring accurate class prediction. DFL loss refers to techniques that dynamically adjust feature representations during training in object detection tasks, potentially enhancing model adaptability to varying data characteristics. Figure 10 illustrates the comparison of these losses across epochs for Test 1 and Test 2. The downward trend in all three losses indicates improvement over the 300 epochs. The identical results between tests are attributed to both using the same dataset

(I) for training YOLO as depicted early in Table 5.

Regarding precision and recall, our results—illustrated in Figure 11—show significant differences between the two tests, with Test 2 outperforming Test 1.

After 300 epochs, we noted significant differences in the validation loss graphs, as shown in Figure 12. Test 2 achieved a lower Box loss of 1.85, compared to 2.06 in Test 1, and a lower Cls loss of 1.08, compared to 2.02 in Test 1. Additionally, Test 2 recorded a DFL loss of 1.52, whereas Test 1 recorded 1.75.

mAP50 evaluates how accurately object detection models localize and classify objects using an IoU threshold of 0.5. It computes the average precision across all classes, indicating better performance with higher scores. mAP50-95 expands this assessment across IoU thresholds from 0.5 to 0.95, offering a broader evaluation of the model's ability to detect objects with different degrees of overlap with ground truth, providing a comprehensive measure of its performance. Higher mAP scores (both mAP50 and mAP50-95) indicate better overall performance of the object detection model in accurately identifying and localizing objects in images. In our results, as shown in Figure 13, we also observed significant differences in the behavior of these metrics across 300 epochs, all converging to an acceptable value of mAP scores. At the end of 300 epochs, Test 1 achieved a mAP50 score of 48.7%, while Test 2 achieved 83.4%. For mAP50-95, Test 1 scored 0.23, whereas Test 2 scored 0.44%.

Table 6 – Comparison of metrics from Tests 1 and 2 with results from other studies.

Study	**Precision(%)**	**Recall(%)**	**Map50%**	**F1-Score(%)**	Epochs
[19]	90.3	**88.6**	**93.5**	**89.4**	300
[20]	85.6	74.0	82.9	79.3	300
[21]	84.2	70.2	78.2	76.5	300
Test 1	55.6	57.7	48.7	56.6	300
Test 2	**92.9**	71.6	83.4	80.9	300

Table 6 presents a comparison of metrics between Tests 1 and 2 alongside other studies. It's important to note that [19], [20], and [21] were trained on datasets different from ours, which introduces bias into the comparison. Therefore, this table should be viewed merely as a reference for the results obtained by those studies. We observed significant differences between the results of the two tests: Test 2 exhibited a 67.1% higher precision compared to Test 1, with a recall of 24.1% higher. The Map50 in Test 2 was 71.2% higher than in Test 1, and the F1-score in Test 2 was 42.9% better.

To evaluate how Real-ESRGAN compares to other super-resolution methods, we use PSNR and SSIM metrics. These metrics help us compare the original image with its enhanced version to determine their similarity. We compare Real-ESRGAN against several interpolation techniques, including Nearest Neighbor, Bicubic, Bilinear, and Lanczos, all implemented using OpenCV and Python. Since PSNR and SSIM require images to be the same size, we scaled all images to 640x512 pixels with a 4x factor to ensure accurate comparisons. The results of these comparisons are shown in Tables 7, 8, 9, 10, and 11.

Table 7 – Comparison of metrics against Bicubic Interpolation.

Algorithm	**PSNR**$_{dB}$	**SSIM**
Bicubic	∞	1.00
Lanczos	47.97	0.99
Bilinear	41.60	0.98
NN	36.49	0.92
Real-ESRGAN	32.73	0.79

Table 8 – Comparison of metrics against Bilinear Interpolation.

Algorithm	**PSNR**$_{dB}$	**SSIM**
Bilinear	∞	1.00
Bicubic	41.60	0.98
Lanczos	40.50	0.97
NN	36.55	0.92
Real-ESRGAN	32.91	0.80

Table 9 – Comparison of metrics against Lanczos Interpolation.

Algorithm	**PSNR**$_{dB}$	**SSIM**
Lanczos	∞	1.00
Bicubic	47.97	0.99
Bilinear	40.50	0.97
NN	36.37	0.91
Real-ESRGAN	32.70	0.79

Table 10 – Comparison of metrics against Nearest Neighbour Interpolation.

Algorithm	**PSNR**$_{dB}$	**SSIM**
NN	∞	1.00
Bilinear	36.55	0.92
Bicubic	36.49	0.92
Lanczos	36.37	0.91
Real-ESRGAN	32.43	0.75

Table 11 – Comparison of metrics against Real-ESRGAN super-resolution.

Algorithm	**PSNR**$_{dB}$	**SSIM**
Real-ESRGAN	∞	1.00
Bilinear	32.91	0.80
Bicubic	32.73	0.79
Lanczos	32.70	0.79
NN	32.43	0.79

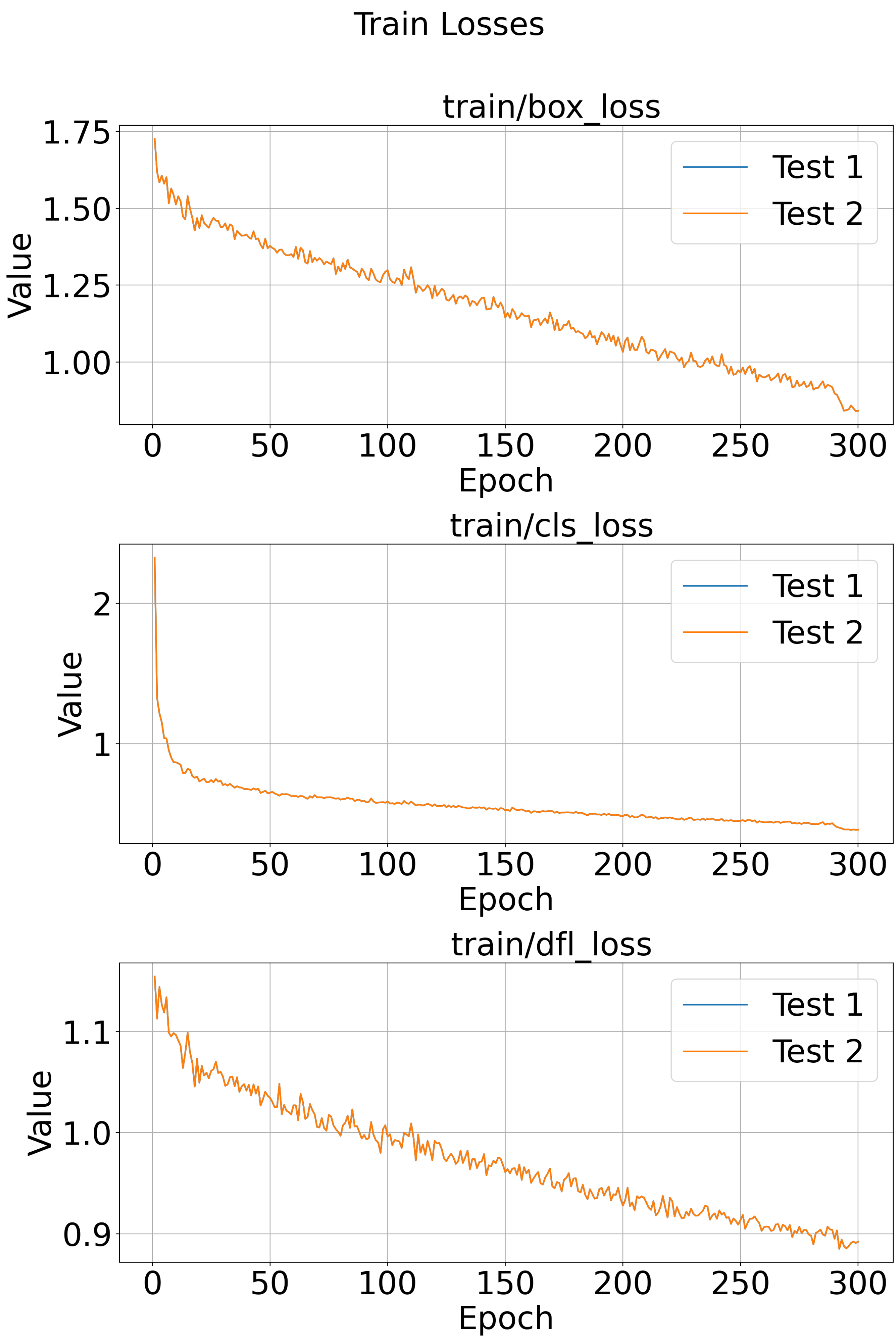


Figure 10 – Comparison of training loss graphs for the two tests over 300 epochs.

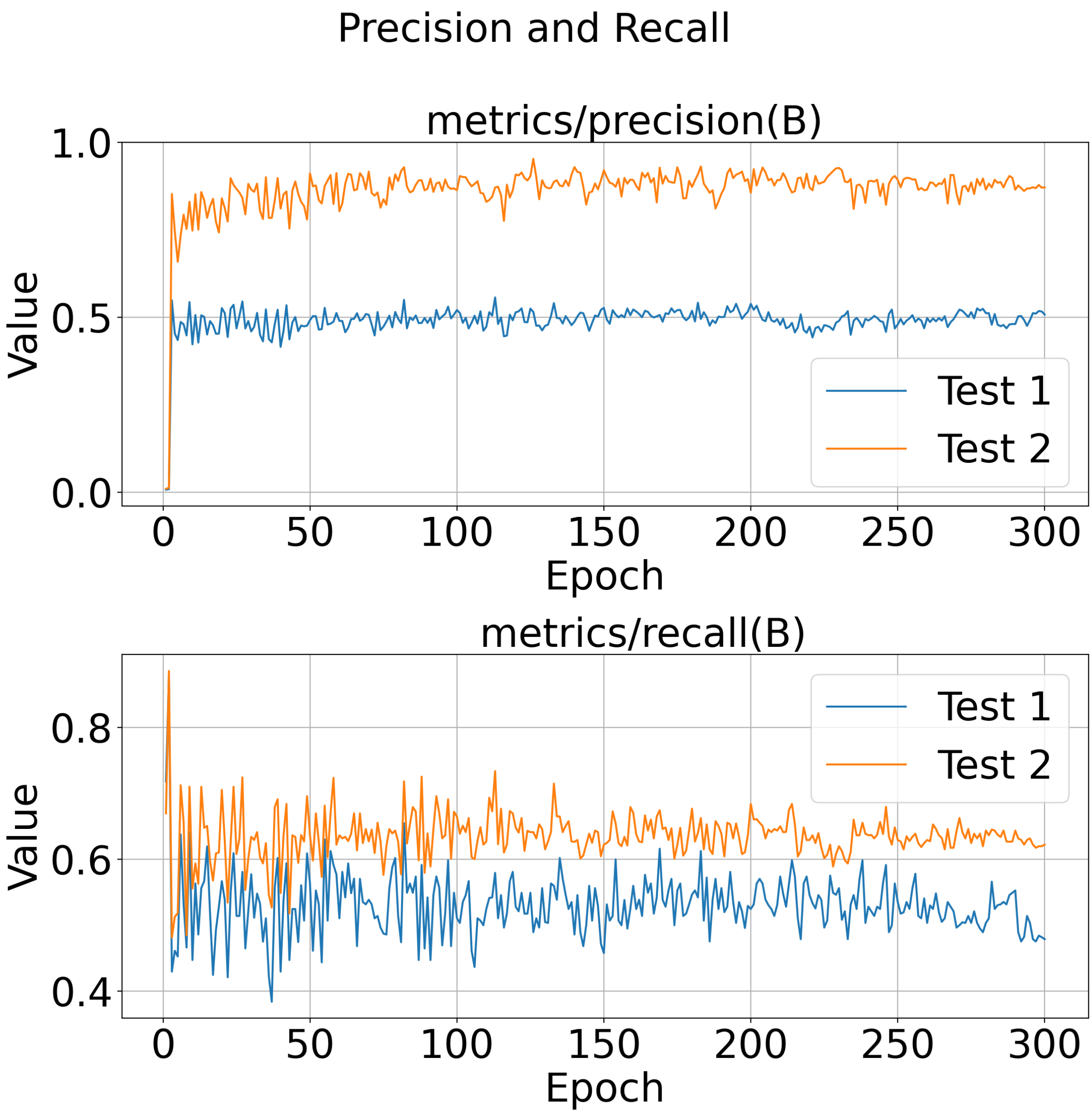


Figure 11 – Comparison of Precision and Recall across 300 epochs for Tests 1 and 2.

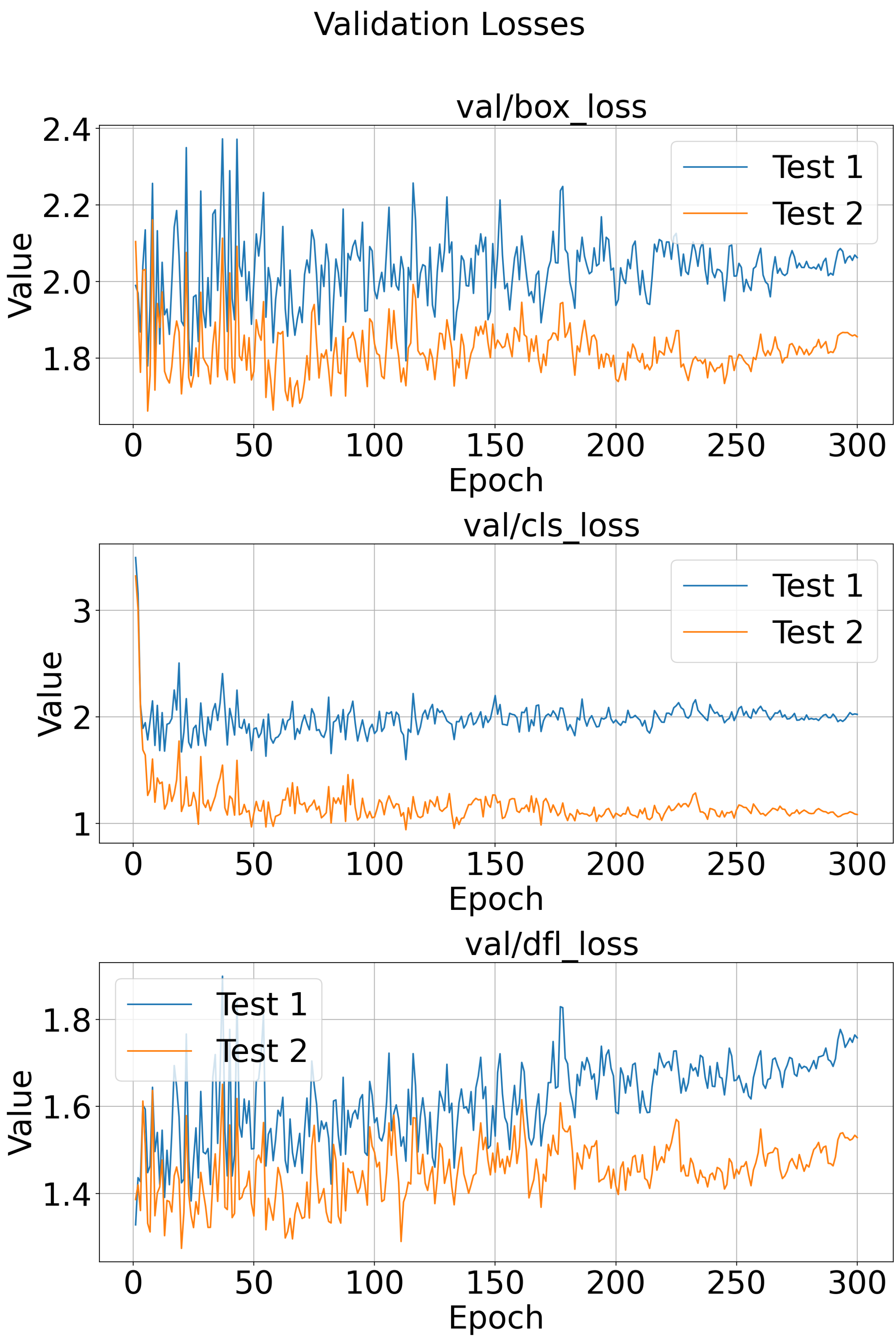


Figure 12 – Validation loss values for Tests 1 and 2 throughout 300 epochs.

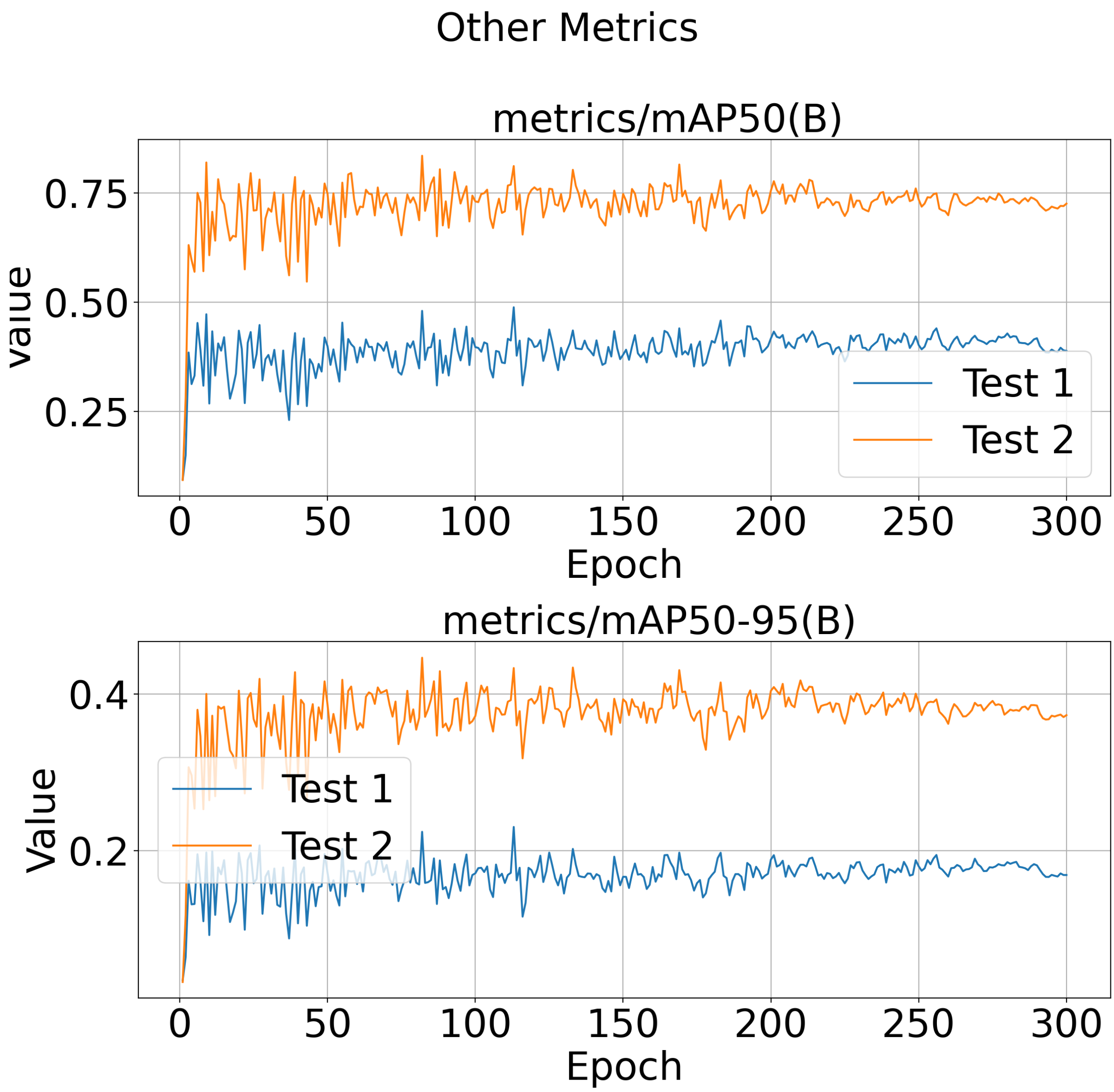


Figure 13 – mAP50 and mAP50-95 for both Tests.

5 Analysis and discussion

The study aimed to explore the application of deep learning, specifically Generative Adversarial Networks, in enhancing low-quality images captured by drones for target identification, particularly in search and rescue operations. Drones offer cost-effective solutions in such scenarios, yet their imagery often suffers from inherent quality issues. Traditional image processing methods struggle to generate new data effectively, prompting the investigation into GANs for super-resolution, a technique capable of synthesizing high-quality images from low-resolution inputs based on trained latent spaces.

To further explore and validate the use of GANs to address SAR operations, a systematic literature review was conducted, confirming the efficacy of GAN-based approaches in improving drone-captured images for object detection and edge delineation. This systematic review facilitated the clustering of various studies, enabling targeted research into specific gaps in each application scenario of GANs for object detection by UAVs, facilitating future work.

Initially focusing on the Real-ESRGAN algorithm, suited for real-world images, the study hypothesized its potential utility in drone imagery enhancement. Surprisingly, even when using a pre-trained model unrelated to the specific dataset, satisfactory results were achieved. This observation underscored the robustness of Real-ESRGAN across varied datasets, highlighting its adaptability and effectiveness in diverse applications.

Despite significant advancements in image clarity through Real-ESRGAN, challenges persisted in achieving uniform contrast enhancement across all frames, as evidenced in Figure 6. Moreover, certain regions of the images exhibited persistent issues with sharpness and contrast, indicating limitations in the algorithm's performance under specific conditions (Figure 7). Additionally, artifacts were occasionally observed in localized areas of the enhanced images (Figure 8), suggesting areas for further algorithm refinement.

In implementing the YOLOv8 model, standard procedures were adhered to, with a batch size of 16 used for both training and validation. Both tests exhibited similar loss patterns during training due to the same training dataset(Dataset I) applied across experiments. However, after 300 epochs, Test 2 showed higher box, class, and DFL losses compared to Test 1. Despite this, Test 2 achieved superior precision, recall, mAP50, and F1-score relative to Test 1. These results are promising when compared to other studies detailed in Table 6. It is important to note that our study employed a relatively small dataset with just one class (human) for YOLOv8 process, compared to other studies.

In comparing Real-ESRGAN with traditional interpolation methods, it was found that Real-ESRGAN had lower PSNR and SSIM values. This might initially suggest that

Real-ESRGAN performs worse. However, it's important to consider that interpolation methods apply filters that make only minor adjustments to the image, leading to small differences from the original. On the other hand, Real-ESRGAN significantly modifies pixel values to enhance the image, which can introduce artifacts and cause more noticeable deviations from the original. Therefore, the reduced PSNR and SSIM values reflect the extent of these changes rather than indicating a fundamental flaw in the algorithm.

In conclusion, the integration of Real-ESRGAN for image enhancement and YOLOv8 for object detection showcases great promising advancements for detecting human targets in drone imagery for search and rescue operations.

6 Conclusion

The application of Real-ESRGAN on infrared spectrum images in this study demonstrated improvements in sharpness, enhancing the accuracy of object identification algorithms. The systematic literature review conducted and published highlighted the primary applications of GANs in the context of target identification in drone-generated images, showcasing their use across various target types such as animals, fires, trees, humans, among others. It also emphasized the key metrics used to analyze the results.

6.1 Validation Limits

While the results presented demonstrate the significant potential of using pre-trained Real-ESRGAN models for enhancing images followed by YOLOv8 application in target detection, several validation limitations need to be addressed.

Regarding the dataset, it was relatively small compared to more comprehensive studies, which may limit the generalizability of the results to different drone image capture conditions.

Considering the targets, we used YOLOv8 to detect only one target, which is "human". It is important to evaluate how the algorithm performs when multiple classes are used in the images after image super-resolution.

Regarding image annotation procedures for the YOLOv8 application, annotations were manually performed, and the accuracy of these annotations directly influences the performance of detection algorithms. Inaccurate or inconsistent annotations may introduce biases in the results.

Concerning the application of Real-ESRGAN in our study, the pre-trained models used may not have been specifically optimized for images captured by drones in the infrared spectrum. This could affect the algorithm's ability to handle specific nuances of these images.

Regarding metrics, although standard metrics such as precision, recall, and F1-score were employed to evaluate results, they may not fully capture the complexity of object detection tasks in dynamic and challenging environments.

The generalizability of the results may be specific to the experimental context and conditions applied. Generalizing to different search and rescue scenarios may require additional adaptations in algorithms and datasets.

These limitations underscore the importance of future research to enhance the robustness and applicability of image enhancement and object detection algorithms in drone images, particularly in critical search and rescue scenarios where precision and reliability are paramount.

Additionally, artifacts in images were inherent to the algorithm itself, as acknowledged by the authors, affecting certain regions of the enhanced images. Unlike cartoon images that tend to be predictable, real images captured by drones can exhibit random issues such as terrain variation, lighting variations, blurring, and weather factors, compromising the captured data. Therefore, careful consideration is essential during image capture to mitigate these issues.

Moreover, the generation algorithms to which GANs belong tend to consume significant computational resources and time, necessitating further investigation for real-time target detection applications. For instance, applying Real-ESRGAN to 1000 images using an RX570 GPU took approximately 90 minutes to complete the enhancement process.

6.2 Main contributions

In summary, this study's primary contribution stems from the observation and validation that pre-trained models in a GAN network can effectively enhance images captured by UAVs, particularly in the infrared spectrum, for subsequent target identification, primarily focusing on people and animals in search and rescue operations. This finding emerged from tests using the Real-ESRGAN algorithm alongside a pre-trained model to enhance images within our infrared dataset. Importantly, the images used in our tests were initially unrelated to those used in the pre-trained models proposed by the authors, suggesting the potential development of a comprehensive super-dataset capable of enhancing various image classes.

This work opens avenues for future researchers to apply Super-ESRGAN to UAV images across diverse spectra for super-resolution, followed by algorithms like YOLO for target identification in search and rescue operations.

The systematic literature review solidified studies utilizing GAN algorithms for improving target and edge detection in UAV-generated images, along with identifying useful validation metrics across studies. By clustering studies in the systematic literature review, researchers can identify critical gaps and foundational studies, providing insights for developing algorithms and techniques to address and mitigate these gaps, as well as identifying the most useful metrics for validating methodologies for this purpose.

Tests conducted with YOLOv8 demonstrated its versatility and efficiency in detecting targets using a reduced dataset and limited training epochs. While differences in results between tests using normal and enhanced images were minor, they highlight the potential for future investigations with enhanced resources and more refined methodologies.

6.3 Future works

In our study, artifacts in the enhanced images stemming from the Real-ESRGAN algorithm were noted, as anticipated by the authors regarding potential issues in certain regions of the improved images. Future research directions could focus on mitigating these artifacts effectively. Additionally, while Real-ESRGAN proved highly effective in enhancing images within our dataset, its real-time application during operations was not addressed in this study. This area presents a compelling avenue for investigation, especially with the advancements in real-time operating systems and embedded technologies like FPGA, Beagle-bone Black, Raspberry Pi, ESP32, among others, which support Python and its libraries, including OpenCV—the primary tool used in this study.

Moreover, OpenCV's capability to handle image representation through pixel matrices provides insights for developing robust applications. This includes extending training epochs and dataset sizes, leveraging parallel and/or distributed processing libraries like OpenMP, Vulkan, among others, to assess algorithm performance across various GPUs or clusters. Furthermore, this study paves the way for image enhancement in other capture spectra and diverse applications such as contour detection, area-of-interest identification, anomaly detection in parts, among others, fostering interdisciplinary collaborations with engineering disciplines such as mechanical and aeronautical engineering offered by Unifei.

6.4 Final considerations

The dissertation described in this work was highly enriching, challenging, and rewarding, providing a comprehensive understanding of all stages involved in a scientific project, from its planning to its publication. This highlights the quality and excellence of the graduate program in Computer Science and Technology at Unifei, in engaging the development and consolidation of new scientific knowledge that significantly contributes to societal well-being.

Appendix

APPENDIX A – SLR manuscript

Article

Applications of GANs to aid target detection in SAR operations: a systematic literature review.

Vinícius Correa [1,2,*], Alexandre Ramos [1]

[1] Institute of Mathematics and Computing, Federal University of Itajubá, Itajubá 37500-903, Brazil; correa@unifei.edu.br; ramos@unifei.edu.br
[2] Institute of Technological Sciences, Federal University of Itajubá, Itabira 35903-087, Brazil
* Correspondence: correa@unifei.edu.br

Abstract: Research on unmanned autonomous vehicles (UAV) for search and rescue missions is widespread due to its cost-effectiveness and enhancement of security and flexibility in operations. However, a significant challenge arises from the quality of sensors, terrain variability, noise, and the size of targets in the images and videos taken by them. Generative Adversarial Networks (GANs), introduced by Ian Goodfellow among their variations, can offer excellent solutions for improving the quality of sensors regarding super-resolution, noise removal and other image processing issues. To identify new insights and guidance on how to apply the GANs to detect living beings in SAR operations, a PRISMA-oriented systematic literature review was conducted to analyze primary studies that explore the usage of GANs for edge or object detection in images captured by drones. The results demonstrate the utilization of GAN algorithms in the realm of image enhancement for object detection, along with the metrics employed for tool validation. These findings provide insights on how to apply or modify them to aid in target identification during search stages.

Keywords: GAN; UAV; SAR.

1. Introduction

Search and rescue is a highly significant field in saving lives in perilous environments, environmental disasters, and accidents involving both people and animals worldwide [1–9]. The use of Unmanned Aerial Vehicles (UAVs) proves to be particularly valuable in these operations[10,11], especially in hard-to-reach areas, as this technology offers conveniences such as reduced operation costs, agility, safety, remote operation, and the use of sensors calibrated across various light spectra, among others. Concerning these sensors, one of the primary challenges to address is the quality of the camera, as the steps of mapping, remote sensing, target class identification, visual odometry [12], and UAV positioning rely on the analysis and interpretation of data captured by these sensors. Thus, processing the images generated by UAVs through computer vision algorithms aided by deep learning techniques is a highly important area of investigation in the deployment of these small vehicles.

The issues presented by sensors include motion blur, generated by the discrepancy between the velocity and instability of the UAV during image capture, the quality of the terrain where the image was captured, the distance from the targets (the farther away, the more difficult the identification process), video noise, and artifacts generated by camera quality, among others.

Some studies have been analyzing and proposing algorithms to address those quality issues in the sensors. [13] propose a method for removing non-uniform motion blur from multiple blurry images by addressing images blurred by unknown, spatially varying motion blur kernels caused by different relative motions between the camera and the scene. [14] proposes a novel motion deblurring framework that addresses challenges in image deblurring, particularly in handling complex real-world blur scenarios and avoiding over- and under-estimation of blur, which can lead to restored images remaining blurred or

Citation: Correa, V.; Ramos, A. Title. *Journal Not Specified* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

introducing unwanted distortion. They use BSDNEt to disentangle blur features from blurry images, modeling motion blur, synthesizing blurry images based on extracted blur features, and demonstrating generalization and adaptability in handling different blur types. Regarding noise, [15] review with some techniques, using PSNR and SSIM among other metrics to evaluate the studies under analysis.

[16] aims for brightness and contrast enhancement, and [17] handle the resolution enhancement on the images.

The Generative Adversary networks [18](GANs) based solutions has shown interesting results to address image enhancement, as some variations of this algorithm can generate high quality new images from degraded ones [18–26]. [27] conduct a review upon these GAN based techniques addressed on super-resolution, also using PSNR and SSIM to evaluate and compare results among important datasets, including people and animal data, but they do not address the SAR context on their study, neither target detection.

Considering the applicability of GAN networks for image enhancement, we conducted a systematic literature review focused on the utilization of GAN algorithms for improvements on target detection in images captured by UAVs, aiming to gain insights into the techniques and metrics employed in this task and potential adaptations for search and rescue applications.

2. Research Method

2.1. Research definition

Our research was divided in 3 parts as depicted in Figure 1. To initiate our research endeavor, we formulated the following search question.

How can GAN algorithms help detect edges or objects in images generated by UAVs?

This question served as a guiding beacon, assisting in delineating the scope of our study analysis.

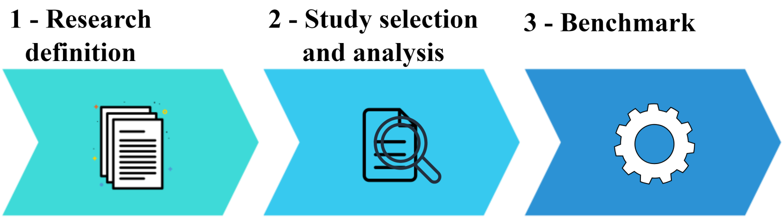


Figure 1. Research steps utilized on our study. Initially, we framed our investigation by formulating a primary research question along with three subsidiary questions that were aligned with the practical applications of our findings. Subsequently, we executed several essential steps, including duplicate removal, abstract analysis, quality assessment, and study selection. These processes were pivotal in ensuring the integrity and rigor of our research. Finally, in terms of our findings, we introduced and adhered to a standardized benchmark approach on our data. This benchmark framework was designed to analyze an application of super-resolution on some selected images by state-of-art GAN model along with its pre-trained default model.

To offer additional guidance for our analysis, particularly concerning metrics and the utilization of pre-trained models within the studies under scrutiny, we have put forth the following supplementary questions:

1. How can GANs be addressed on SAR operations?
2. What benefits are gained from using a pre-trained model rather than training one from scratch?
3. Which metrics are most suitable for validating these algorithms?

To study the role of Generative Adversarial Networks (GANs) in aiding the detection of individuals and wildlife in search and rescue operations, we formulated a generic search string for edge and object detection in images generated by Unmanned Aerial Vehicles (UAVs), intending to capture the broadest possible range of results in the application area of GANs for target detection in UAV-generated images. Below is the chosen search string tailored for this purpose.

("edge detection" OR "object detection") AND (uav OR drones) AND (gan OR "generative adversarial networks")

Assisted by Parsifal [28], the subsequent phase unfolded, characterized by a methodical and systematic methodology. Figure 2 depicts the sequential progression followed during this stage.

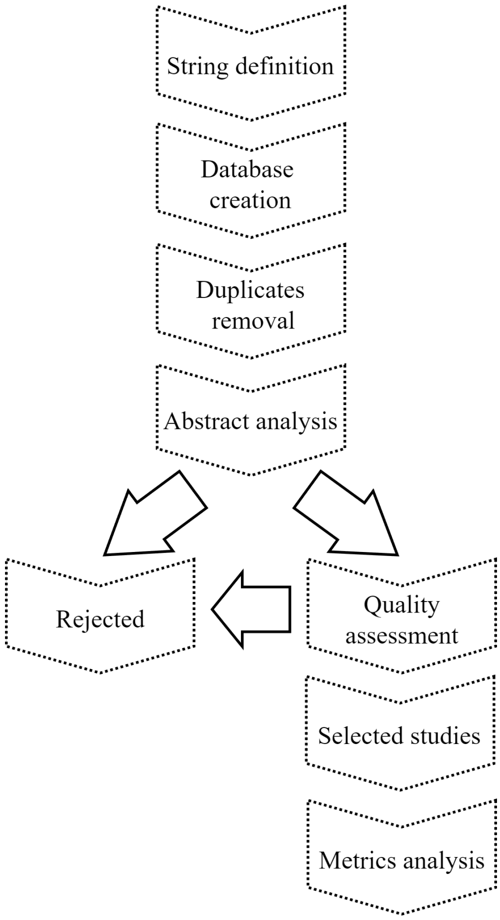


Figure 2. Flowchart illustrating the step-by-step process followed on studies analysis. Studies were scrutiniz

Figure 3 summarizes the implementation of the PICOC framework in this review.

Aspect	Scope
Population	Studies that use GAN algorithms applied to images generated by UAVs.
Intervention	Algorithms and image enhancement techniques for detecting people in search and rescue operations carried out by UAVs.
Comparison	N/A As this is not a comparative study but rather a review of articles utilizing GAN algorithms.
Outcome	Validation of the best GAN solutions for improving edge or target detection. Detection of objects and targets, specifically focused on the search and localization of people.
Context	Publications centered on the utilization of Generative Adversarial Network (GAN) algorithms for image analysis from Unmanned Aerial Vehicles (UAVs), particularly emphasizing edge detection, object detection, and classification tasks.

Figure 3. PICOC framework applied to this work.

The search string yielded 42 results from Scopus and 27 from IEEE Xplore databases. An outline of the study selection and quality assessment process is illustrated in Figure 4

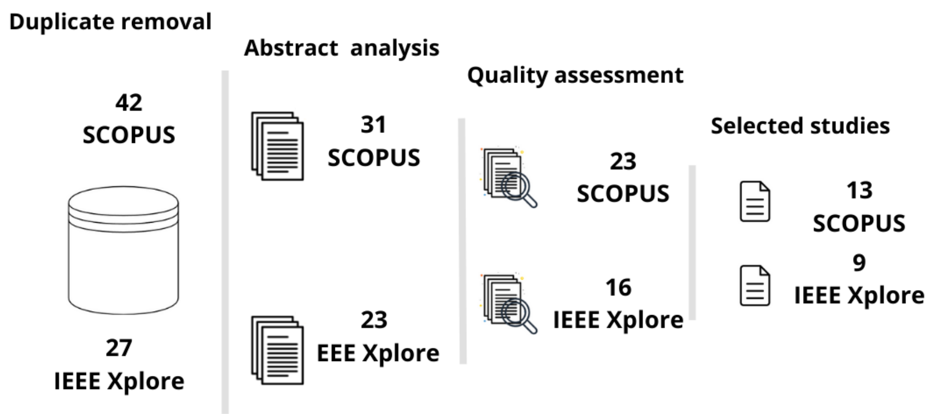


Figure 4. Study selection process. After removing duplicates and secondary studies, 31 articles from the Scopus database and 23 from IEEE Xplore were subjected to abstract analysis. Subsequently, 23 from Scopus and 16 from IEEE Xplore were selected for quality assessment, while the remainder were rejected. Finally, 13 studies from Scopus and 9 from IEEE Xplore were selected for evaluation of the metrics presented in validating the tools used for enhancing and detecting objects in images generated by UAVs.

2.2. Abstract analysis

After removing duplicate entries, we assessed the abstracts of the papers to make preliminary selections, concentrating on the posed inquiries. Duplicate entries and secondary works were excluded during the initial screening process. Following this, we examined the abstracts of the articles to finalize our initial selections. Table 1 outlines the criteria utilized for rejecting papers at this stage.

Table 1. Exclusion criteria.

Code	Criteria
EX1	False positives
EX2	Secondary studies
EX3	Not directly related

Five documents obtained from the data collected from the SCOPUS database were excluded because they only contained informational notes from conference guides, and did not presented any significant studies. The three following articles were removed regarding EX1, criteria: [29] Addresses the application of GAN in protecting private images, given the growing volume of images collected on IOT devices; [30] addresses concerns regarding security and reliability in deep learning models aimed at industrial cyber-physical systems ICPSs; [31] addresses a two-stage insulator defect detection method based on Mask R-CNN(Masked Convolutional Neural Networks) focusing on the use of unmanned aerial vehicles (UAVs) in the inspection of electrical power systems.

[32] were removed by EX2, It discusses several machine learning techniques, including GANs in vehicle detection by UAVs.

The following studies were removed regarding EX3: [33] discusses the automatic landing of UAVs in unknown environments based on the perception of 3D environments. However, the algorithm uses Random Forest and not GAN; [34] augments the GAN network training dataset by adding transformed images with increased realism, using the PTL technique to deal with the degradation difference between the real and virtual training images. Edge or object detection is not mentioned in the article, just improvement of the dataset; [35] detail investigations into the detection of electromagnetic interference through spoofing in UAV GPS systems.

2.3. Quality assessment

The incorporation of GAN is a requisite in our dataset; studies lacking its integration were excluded during the quality assessment process. Our evaluation of quality was conducted under the inclusion criteria outlined in Table 2.

Table 2. Inclusion criteria.

Code	Criteria
IC1	Does the GAN algorithm aim to assist in detecting edges or objects?
IC2	Does the authors employ a pre-trained model?
IC3	Does the paper provide the metrics used for the applied model?
IC4	Is the solution proposed in the study aimed at images within the visible light spectrum?
IC5	Does the solution presented in the study target images within the infrared spectrum?
IC6	Is the SAR algorithm designed to detect people or animals?
IC7	Does the study utilize any version of YOLO in its development?

In IC1, we evaluate whether the primary emphasis of the GAN solution proposed in the study is on detecting edges or objects, recognizing that numerous works may target alternative applications such as UAV landing or autonomous navigation. Given that our benchmark relies on pre-trained GAN models, IC2 examines whether the study utilizes a pre-trained model or if the authors train one from scratch using a specific dataset. A pivotal aspect of evaluating the study involves scrutinizing the metrics employed to assess the algorithm; thus, IC3 analyzes if the paper provides these metrics for the GAN. IC4 and

IC5 enable us to ascertain if the GAN solution presented in the study is trained and/or applied to images in the visible light or infrared spectra, aligning with our benchmark's focus on these two spectra. In IC6, we verify if the object or edge detection aims to identify people or animals, given our benchmark's concentration on search and rescue operations. As YOLO stands out as one of the most commonly used tools for real-time object detection, in IC7, we verify if the paper employs any version of YOLO or a complementary tool. Each Inclusion Criterion (IC) is assigned a score of 0 if the paper fails to meet the criteria or does not mention it in the study, 0.5 if it partially meets the criteria, and 1 if it fully meets the criteria.

3. Results

Using the quality assessment framework, our objective is to meticulously review the studies to determine if they achieve the stated objectives of employing GANs for target detection oriented for SAR operations aid. Table 3 presents the scoring outcomes for each study. Initially, our analysis was directed towards verifying the Inclusion Criteria (ICs). Subsequently, we scrutinized studies with a focus on detecting people or animals, followed by those targeting other objects, then those addressing images in the infrared spectrum, further examining studies utilizing some version of YOLO, and finally those dealing with pre-trained models.

The following studies were excluded during the quality assessment stage; therefore, they are not included in this results: [36] was inaccessible via our institutional tools and network; consequently, it was excluded from our review. Studies [37] and [38] were deemed false positives as they did not mention the utilization of GANs in their paper. Additionally, it was observed that they were authored by the same individuals and focused on similar research themes. [39] and [40] were rejected because the primary focus of the research was on creating datasets using GANs, which lies outside the scope of our systematic review.

Figure 5(a) shows a bar chart relating the number of papers per score rate, and Figure 5(b) relates the number of studies per year and type of publication.

From the articles focused on the detection of humans and animals, we obtained the following outcomes:
[53] propose a weight-GAN sub-network to enhance the local features of small targets and introduce sample balance strategies to optimize the imbalance among training samples, especially between positive and negative samples, and easy and hard samples, a technique for object detection free to address issues of images generated by drone movement instability and tiny object size, which can hinder identification, lighting problems, rain, fog, among others. The study reported improvements in detection performance compared to other methods, such as achieving a 5.46% improvement over Large Scale Images, a 3.91% improvement over SRGAN, a 3.59% improvement over ESRGAN, and a 1.23% improvement over Perceptual GANs. This work would be an excellent reference for addressing issues related to images taken from medium or high altitude in SAR operations. The authors use accuracy as metric, comparing its value with the SRGAN, ESRGAN, and Perceptual GAN models. Other metrics presented for evaluating the work include AP (average precision) and AR (average recall).

[43] uses Faster-RCNN, but for detecting stingrays. The work proposes the application of a GLO model (a variation of GANs where the discriminator is removed and learns to map images to noise vectors by minimizing the reconstruction loss) to increase the dataset to improve object detection algorithms. The used model (C-GLO) learns to generate synthetic foreground objects (stingrays) given background patches using a single network, without relying on a pre-trained model for this specific task. In other words, the article utilizes a modified GAN network to expand the dataset of stingray images in oceans, considering the scarcity of such images, which complicates the training of classification algorithms. Thus, the dataset was augmented through C-GLO, and the data were analyzed by Faster-RCNN.

They utilize the AP metric to assess the performance of RCNN applied to images of various latent code dimensions.

[60] proposes a model for generating pedestrian silhouette maps, used for their recognition; however, the application of GANs is not addressed. It was rejected because it lacks the use of GANs in the development.

[61] was also rejected due to its failure to incorporate GAN usage in development, despite utilizing YOLOv3 for target detection in UAV images.

considering studies oriented to object detection, [46] aims to address object detection challenges in aerial images captured by UAVs in the visible light spectrum. To enhance object detection in these images, the study proposes a GAN-based super-resolution method. This GAN solution is specifically designed to up-sample images with low-resolution object detection challenges, improving the overall detection accuracy in aerial imagery.

Table 3. Data from quality assessment. "Total" represents the cumulative points assigned to each study based on inclusion criteria. "Base" denotes the source database where the papers were indexed, while "Pub. type" indicates the publication format, distinguishing between journal papers (a) and conference papers (b).

Study	year	IC1	IC2	IC3	IC4	IC5	IC6	IC7	Total	Citations	Base	Pub. type
[41]	2017	0.5	1.0	0.5	1.0	0.0	0.0	0.0	3.0	-	IEEE Xplore	b
[42]	2017	5.0	0.0	1.0	1.0	0.0	0.0	0.0	2.5	-	IEEE Xplore	b
[43]	2018	0.0	0.0	1.0	1.0	0.0	1.0	0.0	3.0	-	IEEE Xplore	b
[44]	2019	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	-	IEEE Xplore	a
[45]	2019	1.0	0.0	1.0	1.0	0.0	0.0	0.0	3.0	-	IEEE Xplore	b
[46]	2019	0.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	33	SCOPUS	a
[47]	2019	1.0	0.0	1.0	1.0	0.0	0.0	0.0	3.0	8	SCOPUS	b
[48]	2019	1.0	0.0	1.0	1.0	0.0	0.0	0.0	3.0	-	IEEE Xplore	b
[49]	2020	0.0	0.0	1.0	0.5	0.0	0.0	1.0	2.5	12	SCOPUS	a
[50]	2020	0.5	1.0	1.0	1.0	0.0	0.0	0.0	3.5	15	SCOPUS	a
[51]	2020	1.0	0.0	1.0	1.0	0.0	0.0	0.0	3.0	7	SCOPUS	a
[52]	2020	0.5	1.0	1.0	1.0	0.0	0.0	0.0	3.5	-	IEEE Xplore	b
[53]	2021	1.0	0.0	1.0	1.0	0.0	1.0	0.0	4.0	5	SCOPUS	a
[54]	2021	1.0	0.0	1.0	1.0	1.0	0.0	1.0	5.0	1	SCOPUS	a
[55]	2021	0.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	23	SCOPUS	a
[56]	2021	1.0	0.0	1.0	1.0	0.0	0.0	0.0	3.0	0	SCOPUS	a
[57]	2021	0.0	0.0	1.0	1.0	0.0	0.0	1.0	3.0	26	SCOPUS	a
[58]	2021	0.5	0.0	1.0	0.0	0.0	0.0	1.0	2.5	10	SCOPUS	a
[59]	2021	0.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	6	SCOPUS	a
[60]	2021	1.0	0.0	0.0	1.0	0.0	1.0	0.0	3.0	-	IEEE Xplore	b
[61]	2021	0.0	0.0	0.0	0.5	0.0	1.0	1.0	2.5	-	IEEE Xplore	b
[62]	2022	1.0	0.0	1.0	1.0	1.0	0.0	1.0	5.0	1	SCOPUS	b
[63]	2022	1.0	0.0	1.0	1.0	0.0	0.0	1.0	4.0	23	SCOPUS	a
[64]	2022	0.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	0	SCOPUS	b
[65]	2022	1.0	0.0	1.0	1.0	0.0	0.0	1.0	4.0	8	SCOPUS	a
[66]	2022	0.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	8	SCOPUS	a
[67]	2022	1.0	0.0	1.0	1.0	1.0	0.0	0.0	3.0	8	SCOPUS	a
[68]	2022	1.0	0.0	1.0	1.0	0.0	0.0	0.0	3.0	1	SCOPUS	a
[69]	2022	1.0	0.0	1.0	0.0	0.0	0.0	0.0	2.0	-	IEEE Xplore	b
[70]	2022	1.0	1.0	1.0	0.0	0.0	0.0	0.0	3.0	-	IEEE Xplore	b
[71]	2022	0.0	0.0	1.0	0.0	1.0	0.0	0.0	2.0	-	IEEE Xplore	b
[72]	2022	0.0	0.0	1.0	1.0	0.0	0.0	0.0	2.0	-	IEEE Xplore	b
[73]	2023	1.0	0.0	1.0	1.0	0.0	0.0	1.0	4.0	1	SCOPUS	b
[74]	2023	0.0	0.0	1.0	1.0	0.0	0.0	1.0	3.0	2	SCOPUS	a
[75]	2023	0.5	0.0	1.0	0.0	1.0	0.0	0.0	2.5	-	IEEE Xplore	b
[76]	2023	1.0	0.0	1.0	1.0	1.0	0.0	1.0	5.0	-	IEEE Xplore	b
[77]	2023	1.0	0.0	1.0	0.5	0.0	0.0	1.0	3.5	-	IEEE Xplore	b

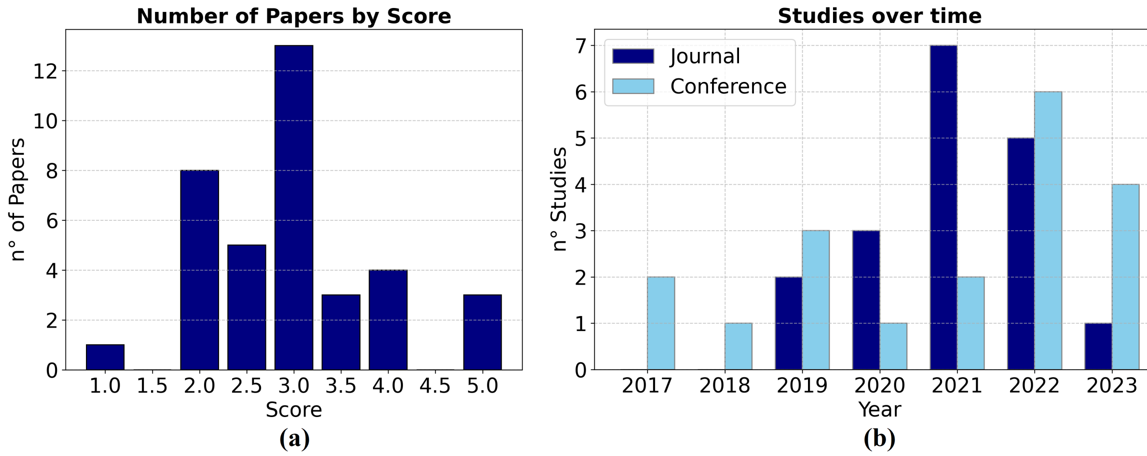


Figure 5. (a) Studies divided by quality assessment score. (b) Quantity of studies under analysis over the publication years, classified by journal or conference publication.

[73] utilized a GAN-based real-time data augmentation algorithm to enhance the training data for UAV vehicle detection tasks, specifically focusing on improving the accuracy of detecting vehicles, pedestrians, and bicycles in UAV images. By incorporating the GAN approach, along with enhancements like using FocalLoss and redesigning the target detection head combination, the study achieved a 4% increase in detection accuracy over the original YOLOv5 model.

[66] introduces a novel two-branch Generative Adversarial Network architecture designed for detecting and localizing anomalies in RGB aerial video streams captured by UAVs at low altitudes. The primary purpose of the GAN-based method is to enhance anomaly detection and localization in challenging operational scenarios, such as identifying small dangerous objects like improvised explosive devices (IEDs) or traps in various environments. The GAN architecture consists of two branches: a detector branch and a localizer branch. The detector branch focuses on determining whether a given video frame depicts a normal scene or contains anomalies, while the localizer branch is responsible for producing attention maps that highlight abnormal elements within the frames when anomalies are detected. In the context of search and rescue operations, the GAN-based method can be instrumental in identifying and localizing anomalies or potential threats in real-time aerial video streams. For example, in search and rescue missions, the system could help in detecting hazardous objects, locating missing persons, or identifying obstacles in disaster-affected areas. By leveraging the GAN's capabilities for anomaly detection and localization, search and rescue teams can enhance their situational awareness and response effectiveness in critical scenarios.

[56] proposes a novel end-to-end multi-task GAN architecture to address the challenge of small object detection in aerial images. The GAN framework combines super-resolution (SR) and object detection tasks to generate super-resolved versions of input images, enhancing the discriminative detection of small objects. The generator in the architecture consists of an SR network with additional components such as a gradient guidance network (GGN) and an edge-enhancement network (EEN) to mitigate structural distortions and improve image quality. In the discriminator part of the GAN, a faster region-based convolutional neural network (FRCNN) is integrated for object detection. Unlike traditional GANs that estimate the realness of super-resolved samples using a single scalar, realness distribution is used as a measure of realness. This distribution provides more insights for the generator by considering multiple criteria rather than a single perspective, leading to improved detection accuracy.

[67] introduces a novel detection network, Region Super-Resolution Generative Adversarial Network (RSRGAN), to enhance the detection of small infrared targets. The GAN

component of RSRGAN focuses on super-resolution enhancement of infrared images, improving the clarity and resolution of small targets like birds and leaves. This enhancement aids in accurate target detection, particularly in challenging scenarios. In the context of search and rescue operations, the application of RSRGAN could be beneficial for identifying small targets in infrared imagery with greater precision. By enhancing the resolution of images containing potential targets, such as individuals in distress or objects in need of rescue, RSRGAN could assist search and rescue teams in quickly and accurately locating targets in various environmental conditions. The improved detection capabilities offered by RSRGAN could enhance the efficiency and effectiveness of search and rescue missions, ultimately contributing to saving lives and optimizing rescue efforts.

[47] concentrates on enhancing small object detection in UAV aerial imagery captured by optical cameras mounted on Unmanned Aerial Systems (UAVs). The proposed GAN solution, known as classification-oriented super-resolution generative adversarial networks (CSRGAN), aims to improve the classification results of tiny objects and enhance detection performance by recovering discriminative features from original small objects. In the context of search and rescue operations, the application of CSRGAN could be beneficial for identifying and locating small objects, such as individuals or objects, in aerial images. By enhancing the resolution and classification-oriented features of these small objects, CSRGAN could assist in improving the efficiency and accuracy of search and rescue missions conducted using UAVs. This technology could aid in quickly identifying and pinpointing targets in large areas, ultimately enhancing the effectiveness of search and rescue operations.

[59] focuses on LighterGAN, an unsupervised illumination enhancement GAN model designed to improve the quality of images captured in low illumination conditions using urban UAV aerial photography. The primary goal of LighterGAN is to enhance image visibility and quality in urban environments affected by low illumination and light pollution, making them more suitable for various applications in urban remote sensing and computer vision algorithms. In the context of search and rescue operations, the application of LighterGAN could be highly beneficial. When conducting search and rescue missions, especially in low light or nighttime conditions, having clear and enhanced images from UAV aerial photography can significantly aid in locating individuals or objects in need of assistance. By using LighterGAN to enhance images captured by UAVs in low illumination scenarios, search and rescue teams can improve their visibility, identify potential targets more effectively, and enhance overall situational awareness during critical operations.

[42] explores the use of GANs to enhance image quality through a super-resolution deblurring algorithm. The GAN-based approach aims to improve the clarity of images affected by motion blur, particularly in scenarios like UAV (Unmanned Aerial Vehicle) image acquisition. By incorporating defocused fuzzy kernels and multi-direction motion fuzzy kernels into the training samples, the algorithm effectively mitigates blur and enhances image data captured by UAVs.

[48] introduces a novel approach utilizing a GAN to address the challenge of small object detection in aerial images captured by drones or Unmanned Aerial Vehicles (UAVs). By leveraging the capabilities of GAN technology, the research focuses on enhancing the resolution of low-quality images depicting small objects, thereby facilitating more accurate object detection algorithms.

[44] utilizes a Generative Adversarial Networks GAN solution to augment typical easily confused negative samples in the pretraining stage of a saliency-enhanced multi-domain convolutional neural network (SEMD) for remote sensing target tracking in UAV aerial videos. The GAN's purpose is to enhance the network's ability to distinguish between targets and the background in challenging scenarios by generating additional training samples. In SAR operations, the study can assist in distinguishing between targets and the background.

[45] introduces a Generative Adversarial Network named VeGAN, trained to generate synthetic images of vehicles from a top-down aerial perspective for semantic segmentation

tasks. By leveraging the GAN for content-based augmentation of training data, the study aims to enhance the accuracy of a semantic segmentation network in detecting cars in aerial images. We can take the study as a basis for training the identification of other targets.

[55]The study aimed to enhance maize plant detection and counting using deep learning algorithms applied to high-resolution RGB images captured by UAVs. To address the challenge of low-quality images affecting detection accuracy, the study proposed a GAN-based super-resolution method. This method aimed to improve results on native low-resolution datasets compared to traditional upsampling techniques. This study was rejected because it focused on agricultural purposes rather than SAR.

[57] The study utilizes a Generative Adversarial Network (GAN), specifically the CycleGAN model, for domain adaptation in bale detection for precision agriculture. The primary objective is to enhance the performance of the YOLOv3 object detection model in accurately identifying bales of biomass in various environmental conditions. The GAN is employed to transfer styles between images with diverse illuminations, hues, and styles, enabling the YOLOv3 model to effectively detect bales under different scenarios. We also reject it because of its non-SAR purpose.

[68] The study utilizes a conditional generative adversarial network (cGAN) for the automated extraction and clustering of peach tree crowns based on UAV images in a peach orchard. The primary focus is on monitoring and quantitatively characterizing the peach tree crowns using remote sensing imagery. It was also rejected because it doesn't focus on agriculture.

[69] proposes a novel approach using the Pix2Pix GAN architecture for Unmanned Aerial Vehicle (UAV) detection. The GAN is applied to detect UAVs in images captured by optical sensors, aiming to enhance the efficiency of UAV detection systems. By utilizing the GAN framework, the study focuses on improving the accuracy and effectiveness of identifying UAVs in various scenarios, including adverse weather conditions. We reject this study because it's aimed on air defense by identifying UAVs in the air by some sensors on the ground.

[64] It employs GAN networks to enhance transmission line images. The article doesn't mention YOLO; it uses a dataset from scratch. Therefore, we can conclude that the GAN network was used for super-resolution. We cannot classify it as a study focused on SAR. Hence, we reject the study at this stage.

Considering studies with images in the infrared spectrum, [54] employ GANs to facilitate the translation of color images to thermal images, specifically aiming to enhance the performance of color-thermal ReID (Re-identification). This translation process involves converting probe images captured in the visible range to the infrared range. By utilizing the GAN framework for color-to-thermal image translation, the study aims to improve the effectiveness of object recognition and re-identification tasks in cross-modality scenarios, such as detecting objects in thermal images and matching them with corresponding objects in color images. Yolo and any other object detector was mentioned, the study utilizes various metrics for evaluating the ThermalReID framework and modern baselines. For the object detection task, they use Intersection over Union (IoU) and mean Average Precision (mAP) metrics. In the ReID task, they employ Cumulative Matching Characteristic (CMC) curves and normalized Area-Under-Curve (nAUC) for evaluation purposes.

In [62] The primary objective of utilizing GANs is to address the challenge posed by the differing characteristics of thermal and RGB images, such as varying dimensions and pixel representations. By employing GANs, the study aims to generate thermal images that are compatible with RGB images, ensuring a harmonious fusion of data from both modalities.

The StawGAN in [75] is used to enhance the translation of night-time thermal infrared images into daytime color images. The StawGAN model is specifically designed to improve the quality of target generation in the daytime color domain based on the input thermal infrared images. By leveraging the GAN architecture, which comprises a generator and a discriminator network, the StawGAN model aims to produce more realistic and wellshaped objects in the target domain, thereby enhancing the overall image translation process.

[76] employs GAN as a sophisticated image processing technique to enhance the quality of input images for UAV target detection tasks. The primary objective of integrating GAN technology into the research framework is to elevate the accuracy and reliability of the target detection process, particularly in the context of detecting UAVs. By harnessing the capabilities of GANs as image fusion technology, the study focuses on amalgamating images captured from diverse modalities, such as those obtained from both the infrared and visible light spectrums. This fusion process is crucial as it enriches the visual information available for identifying and pinpointing UAV targets within the imagery. Essentially, the GAN functions as a tool to generate fused images by adapting and refining the structures of both the generator and discriminator components within the network architecture. Through this innovative approach, the research aims to enhance the precision and robustness of the target detection mechanism embedded within the YOLOv5 model. By leveraging the power of GAN-based image fusion, the study endeavors to optimize the focus and clarity of the target detection process, ultimately leading to improved performance in identifying UAV targets within complex visual environments.

[71] focuses on utilizing a Conditional Generative Adversarial Network (CGAN), specifically the Pix2Pix model, to generate depth images from monocular infrared images captured by a camera. This application of CGAN aims to enhance collision avoidance during drone flights at night by providing crucial depth information for safe navigation. The research emphasizes the use of CGAN for converting infrared images into depth images, enabling the drone to determine distances to surrounding objects and make informed decisions to avoid collisions during autonomous flight operations in low-light conditions. This study can be leveraged in drone group operations, but in terms of ground object identification, it is not applicable. Therefore, we reject the study.

[51] The study proposes a novel approach for insulator object detection in aerial images captured by drones by utilizing a Wasserstein-Generative Adversarial Network (WGAN) for image deblurring. The primary purpose of the GAN solution is to enhance the clarity of insulator images that may be affected by factors such as weather conditions, data processing, camera quality, and environmental surroundings, leading to blurry images. By training the GAN on visible light spectrum images, the study aims to improve the detection rate of insulators in aerial images, particularly in scenarios where traditional object detection algorithms may struggle due to image blurriness. It was rejected because it is not oriented towards search and rescue.

While some studies utilized Faster-RCNN [43], [50] and [48] or custom object detection solutions [67], the majority of the selected ones employed some version of YOLO, with the most common being versions 3 and 5, as depicted in Figure 6(a).

[65] aimed to enhance wildfire detection by GANs to produce synthetic wildfire images. These synthetic images were utilized to address data scarcity issues and enhance the model's detection capabilities. Additionally, Weakly Supervised Learning (WSOL) was applied for object localization and annotation, automating the labeling task and mitigating data shortage issues. The annotated data generated through WSOL was then used to train an improved YOLOv5-based detection network, enhancing the accuracy of the wildfire detection model. The integrated use of GANs for image generation, WSOL for annotation, and YOLOv5 for detection aimed to enhance the model's performance and automate the wildfire detection process. This study could also aid in Search and Rescue operations, as the presence of fire in an area may indicate potential areas of interest during search efforts.

[74] is centered on image deblurring in the context of aerial remote sensing to enhance object detection performance. It introduces the Adaptive Multi-Scale Fusion Blind Deblurred Generative Adversarial Network (AMD-GAN) to address image blurring challenges in aerial imagery. The AMD-GAN leverages multi-scale fusion guided by image blurring levels to improve deblurring accuracy and preserve texture details. In the study, the AMD-GAN is applied to deblur aerial remote sensing images, particularly in the visible

light spectrum, to enhance object detection tasks. The YOLOv5 model is utilized for object detection experiments on both blurred and deblurred images. The results demonstrate that deblurring with the AMD-GAN significantly improves object detection indices, as evidenced by increased Mean Average Precision (MAP) values and enhanced detection performance compared to using blurred images directly with YOLOv5.

[77] engages on enhancing small object detection in drone imagery through the use of a Collaborative Filtering Mechanism (CFM) based on a Cycle Generative Adversarial Network (CycleGAN). The purpose of the GAN in the study is to improve object detection performance by enhancing small object features in drone imagery. The CFM, integrated into the YOLO-V5s model, filters out irrelevant features during the feature extraction process to enhance object detection. By applying the CFM module to YOLO-V5s and evaluating its performance on the VisDrone dataset, the study demonstrates significant improvements in detection performance, highlighting the effectiveness of the GAN-based approach in enhancing object detection capabilities in drone imagery

[63] aims to develop a portable and high-accuracy system for detecting and tracking pavement cracks to ensure road integrity. To address the limited availability of pavement crack images for training, a GAN called PCGAN is introduced. PCGAN generates realistic crack images to augment the dataset for improved detection accuracy using an improved YOLO v3 algorithm. The YOLO-MF model, a modified version of YOLO v3 with acceleration and median flow algorithms, is employed for crack detection and tracking. This integrated system enhances the efficiency and accuracy of pavement crack detection and monitoring for infrastructure maintenance, We reject this study because it lacks relation to SAR operation.

[49] focuses on addressing the challenges of motion deblurring and marker detection for autonomous drone landing using a deep learning-based approach. To achieve this, the study proposes a two-phase framework that combines a slimmed version of the DeblurGAN model for motion deblurring with the YOLOv2 detector for object detection. The purpose of the DeblurGAN model is to enhance the quality of images affected by motion blur, making it easier for the YOLOv2 detector to accurately detect markers in drone landing scenarios. By training a variant of the YOLO detector on synthesized datasets, the study aims to improve marker detection performance in the context of autonomous drone landing. Overall, the study leverages the DeblurGAN model for motion deblurring and the YOLOv2 detector for object detection to enhance the accuracy and robustness of marker detection in autonomous drone landing applications. We reject it as its focus is on landing assistance rather than search and rescue.

[57] utilizes a GAN solution, specifically the CycleGAN model, for domain adaptation in the context of bale detection in precision agriculture. The primary objective is to enhance the performance of the YOLOv3 object detection model for accurately detecting bales of biomass in various environmental conditions. The GAN is employed to transfer styles between images with diverse illuminations, hues, and styles, enabling the YOLOv3 model to be more robust and effective in detecting bales under different scenarios. By training the YOLOv3 model with images processed through the CycleGAN for domain adaptation, the study aims to improve the accuracy and efficiency of bale detection, ultimately contributing to advancements in agricultural automation and efficiency. The study was rejected because its focus is more aligned with the application of UAVs in agriculture

[58] introduces InsulatorGAN, a novel model based on conditional Generative Adversarial Nets (GAN), designed for insulator detection in high-voltage transmission line inspection using unmanned aerial vehicles (UAVs). The primary purpose of InsulatorGAN is to generate high-resolution and realistic insulator-detection images from aerial images captured by drones, addressing limitations in existing object detection models due to dataset scale and parameters. In the study, the authors leverage the YOLOv3 neural network model for real-time insulator detection under varying image resolutions and lighting conditions, focusing on identifying ice, water, and snow on insulators. This application of YOLOv3 demonstrates the integration of advanced neural network models within

the context of insulator detection tasks. While the study does not explicitly mention the use of pre-trained models or training from scratch for InsulatorGAN, the emphasis is on enhancing the quality and resolution of generated insulator images through the proposed GAN framework. By combining GAN technology with YOLOv3 for insulator detection, the study aims to advance the precision and efficiency of detecting insulators in transmission lines using UAV inspection, contributing to the field of computer vision and smart grid technologies. We decline it due to its lack of emphasis on search and rescue operations.

Figure 6(a) compiles the YOLO versions and the number of studies that utilize each one of them.

Considering the use of pre-trained models and weights, [50] aims to predict individual motion and view changes of objects in UAV videos for multiple object tracking. To achieve this, the study proposes a novel network architecture that includes a social LSTM network for individual motion prediction and a Siamese network for global motion analysis. Additionally, a GAN is introduced to generate more accurate motion predictions by incorporating global motion information and objects' positions from the last frame. The GAN is specifically utilized to enhance the final motion prediction by leveraging the individual motion predictions and view changes extracted by the Siamese network. It plays a crucial role in generating refined motion predictions based on the combined information from the individual and global motion analysis components of the network. Furthermore, the Siamese network is initialized with parameters pre-trained on ImageNet and fine-tuned for the task at hand. This pre-training step helps the Siamese network learn relevant features from a large dataset like ImageNet, which can then be fine-tuned to extract changing information in the scene related to the movement of UAVs in the context of the study.

[72] The study focuses on using generative adversarial networks (GANs) to enhance object detection performance under adverse weather conditions by restoring images affected by weather corruptions. Specifically, the Weather-RainGAN and Weather-NightGAN models were developed to address challenges related to weather-corrupted images, such as rain streaks and night scenes, to improve object detection accuracy for various classes like cars, buses, trucks, motorcycles, persons, and bicycles in driving scenes captured in adverse weather conditions. The study can provide valuable insights in SAR scenarios in snow-covered regions or other severe weather conditions.

[70] It introduces a GAN for a specific purpose, although the exact application domain is not explicitly mentioned in the provided excerpts. The GAN is crafted to achieve a particular objective within the context of the research, potentially linked to tasks in image processing or computer vision. Furthermore, the study incorporates the use of a pre-trained model, which serves a specific purpose in developing or enhancing the proposed GAN solution. The application of the pre-trained model within the study likely aims to leverage existing knowledge or features to improve the performance or capabilities of the GAN in its intended application domain. This study, despite its focus on human-face recognition, is deemed unnecessary for SAR operations, as the UAV is anticipated to operate at high altitudes where facial images of potential individuals would not be readily identifiable. Therefore, we reject this study.

[41] introduces a dual-hop generative adversarial network (DH-GAN) to recognize roads and intersections from aerial images automatically. The DH-GAN is designed to segment roads and intersections at the pixel level from RGB imagery. The first level of the DH-GAN focuses on detecting roads, while the second level is dedicated to identifying intersections. This two-level approach allows for the end-to-end training of the network, with two discriminators ensuring accurate segmentation results. Additionally, the study utilizes a pre-trained model within the DH-GAN architecture to enhance the intersection detection process. By incorporating the pre-trained model, the DH-GAN can effectively extract intersection locations from the road segmentation output. This integration of the pre-trained model enhances the overall performance of the DH-GAN in accurately identifying intersections within the aerial images. We decline it because it's not closely aligned with

the topic, as it focuses on road detection rather than the detection of objects, animals, or people.

[52] The study aims to enhance tracking performance in UAV videos by transferring contextual relations across views. To achieve this, a dual GAN learning mechanism is proposed. The tracking-guided CycleGAN (T-GAN) transfers contextual relations between ground-view and drone-view images, bridging appearance gaps. This process helps adapt to drone views by transferring contextual stable ties. Additionally, an attention GAN (A-GAN) refines these relations from local to global scales using attention maps. The pre-trained model, a Resnet50 model, is fine-tuned to output context operations for the actor-critic agent, which dynamically decides on contextual relations for vehicle tracking under changing appearances across views. Typically, SAR operations are conducted in remote or hard-to-reach areas, making ground-based image capture impractical. Therefore, we reject the study.

This Sistematic Literature Review yielded highly detailed and diverse results, considering that drone images can serve various purposes such as agriculture, automatic landing, face recognition, identification of objects on power lines, and so on. Since our focus was primarily on SAR operations, we rejected papers that were unrelated to this topic. Table 4 and 5displays the selected results along with the metrics used to evaluate the proposals. Here, AP stands for average precision; AR means average recall; AUC (Area under the receiver operating characteristic curve.); ROC (Receiver operating characteristic curve); MAP (Mean average precision); NIQE (Natural image quality evaluator); AG (Average gradient); PIQE (Perception index for quality evaluation), PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index), FID (Fréchet Inception Distance); DSC - (The Dice Similarity Coefficient); S-Score (Segmentation Score); MAE (Mean Absolute Error); IS (Inception Score); SMD (Standard Mean Difference); EAV (Edge-Adaptive Variance); PI (Perceptual Index); IOU (Intersection over Union) and CP_x and CP_y refer to the center position errors in the longitudinal and lateral driving direction. Figure 6(b) shows the most used metrics, along with the number of studies that use them.

Table 4. Classic metrics from selected studies

Study	Accuracy	Precision	AP	Recall	AR	ROC	AUC	MAP	F1-Score
[43]	-	-	✓	-	-	-	-	-	-
[44]	-	✓	-	-	-	-	-	-	-
[45]	-	-	-	-	-	-	-	-	-
[46]	✓	-	✓	-	-	-	-	-	-
[47]	✓	-	-	-	-	-	-	-	-
[48]	-	-	-	-	-	-	-	✓	-
[50]	-	-	-	-	-	-	-	-	✓
[53]	✓	-	✓	-	✓	-	-	-	-
[54]	-	-	-	-	-	✓	-	-	-
[56]	-	-	✓	-	-	-	-	-	✓
[59]	-	-	-	-	-	-	-	-	-
[62]	-	-	✓	-	-	-	-	✓	-
[65]	-	✓	✓	✓	-	-	-	-	✓
[66]	-	-	-	-	-	-	✓	-	-
[67]	-	-	✓	✓	-	-	-	✓	✓
[42]	-	-	-	-	-	-	-	-	-
[72]	-	-	-	-	-	-	-	✓	-
[73]	-	-	-	-	-	-	-	✓	-
[74]	✓	✓	-	✓	-	-	-	✓	-
[75]	-	-	-	-	-	-	-		-
[76]	-	✓	-	✓	-	-	✓	-	-
[77]	-	✓	-	✓	-	-	-	-	-

Table 5. Other metrics from Studies

Study	PSNR	SSIM	Other	Other	Other	Other	Other
[45]	-	-	IOU	CP_x	CP_y	-	-
[48]	✓	✓	PI	-	-	-	-
[56]	✓	✓	AG	NIQE	-	-	-
[59]	-	-	PIQE	-	-	-	-
[65]	-	-	FDR	FNR	-	-	-
[66]	-	✓	-	-	-		-
[42]	-	-	SMD	EAV	-	-	-
[74]	✓	✓	-	-	-	-	-
[75]	✓	✓	DSC	FID	IS	S-Score	MAE
[77]	-	-	MAP 0.5	MAP 0.95	-	-	-

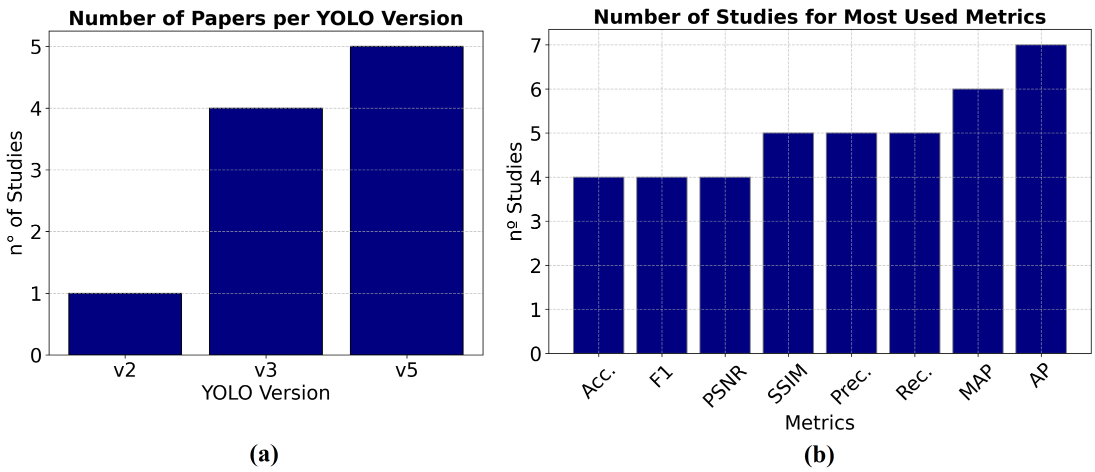


Figure 6. (a) Number of studies regarding the YOLO version used. (b) Studies by the most used metrics.

[76] The study conducted a comprehensive comparison of precision, recall, and Mean Average Precision (MAP) metrics between their proposed improved versions of YOLOv5 and the following models: original YOLOv5; YOLOv5+CBAM and YOLOv5+Image Fusionl. The results demonstrated superior performance of their proposed models in terms of object detection accuracy. This comparative analysis not only highlights the advancements achieved with the enhanced YOLOv5 variants but also underscores the importance of employing classical evaluation metrics in assessing the efficacy of GANs and YOLOv5 in practical applications.

[77] also uses Yolov5 and comparison between different techniques for object detection with mAP 0.5; mAP 0.5:0.95, precision and recall.

[50] presented a comprehensive array of metrics, including Identification Precision (IDP), Identification Recall (IDR), IDF1 score (F1 score), Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), Mostly Tracked targets (MT), Mostly Lost targets (ML), Number of False Positives (FP), Number of False Negatives (FN), Number of ID Switches (IDS), and Number of times a trajectory is Fragmented (FM). The authors utilized diverse datasets and compared object monitoring performance across various techniques, namely Faster-RCNN, R-FCN, SSD, and RDN. This extensive metric evaluation renders this reference an excellent resource for validating metrics applied in target identification in SAR applications.

[74] The study conducts a comparison of Mean Average Precision (MAP), Precision, and Recall in object identification using the methods GT + YOLOV5, Blur + YOLOV5, and AMD-GAN + YOLOV5. It serves as an excellent resource for detection comparisons with YOLOV5. Regarding the GAN utilized, the article employs metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), comparing these

metrics with various methods, including AMD-GAN without NAS MFPN, AMD-GAN with NAS MFPN, AMD-GAN without AMS Fusion, and AMD-GAN with AMS Fusion.

[59] compares the performance of several algorithms, including LighterGAN; EnlightenGAN; CycleGAN; Retinex; LIME (Low-Light Image Enhancement via Illumination Map Estimation), this one shows the highest PIQE score, and DUAL (Dual Illumination Estimation for Robust Exposure Correction), with respect to PIQE metric.

[72] The study employs Average Precision (AP) and mean Average Precision (mAP) metrics to assess the performance of various restoration methods, including Gaussian Gray Denoising Restormer, Gaussian Colour Denoising Restormer, Weather-RainGAN, and Weather-NightGAN. It evaluates these metrics across different object classes such as car, bus, person, and motorcycle. This comparison serves as an excellent resource for applying these metrics in the context of object detection in works involving the application of GANs.

[75] The study utilizes PSNR and SSIM, in addition to FID and IS, to compare image modality translation among the methods pix2pixHD, StarGAN v, 3 PearlGAN, TarGAN, and StawGAN. Furthermore, they compare segmentation performance using metrics such as Dice Similarity Coefficient (DSC), S-Score, and Mean Absolute Error (MAE) specifically for TarGAN and StawGAN. This comprehensive evaluation provides insights into the effectiveness of these methods for image translation and segmentation tasks.

[48] compared PSNR, SSIM and average PI with SRGAN, ESRGAN and their proposal model. They also use the MAP metric to evaluate different methods for object detection such as SSD and Faster R-CNN.

[43] Used Average precision to compared augmentation methods to Faster R-CNN without augmentation, in the context of detection of stingrays.

[45] employed the IOU metric to quantify the degree of overlap between predicted car regions and ground truth car regions in the images. Through the analysis of CP_x and CP_y values, the study has shown how accurately the segmentation network was able to localize and position the detected cars within the images.

4. Benchmark

To finish our analysis, we conduct the third part of the research as a benchmark regarding a GAN model from state-of-art, in order to evaluate the super-resolution feature from a pre-trained model, applied on some samples images from our data. Considering the second research sub-question, What benefits are gained from using a pre-trained model rather than training one from scratch? We selected the Real-ESRGAN algorithm [26]. This study provides an algorithm with pre-trained models that we can utilize to evaluate its performance in the application to our proposed images, which are not related to the model provided by the algorithm, as some are in the infrared spectrum.

The first image was extracted from a video recorded by a sensor in the infrared spectrum, where it is possible to observe the heat signature of three people walking on a lawn near the university. The original image as shown in Figure 7 (a) contained 640X512 pixels, and after processing through the algorithm as depicted in Figure 7 (b), its size was increased to 2560X2040. We cropped important parts of the image for comparison and observed a significant improvement in the contour features of the people, grass, and tree.

The second image, in the visible light spectrum, was captured by a smartphone, depicting some ropes used for guiding a trail-following drone. The original image as shown in Figure 8 (a) had 509X277 pixels, and after processing through the algorithm as depicted in Figure 8 (b), it had 2036X1108 pixels. It is also possible to observe contrast improvements in the yellow, red, and blue lines of the figure.

Figure 9 (a) displays some artifacts found in the region of a window of buildings from Figure 7, while in Figure 9 (b), we have examples of artifacts found in the lawn area of the same figure, indicating some algorithmic flaws. Those artifacts was mention by the authors.

Figure 7. Real-ESRGAN on Infrared Camera. (a) Original image. (b) After Super-resolution

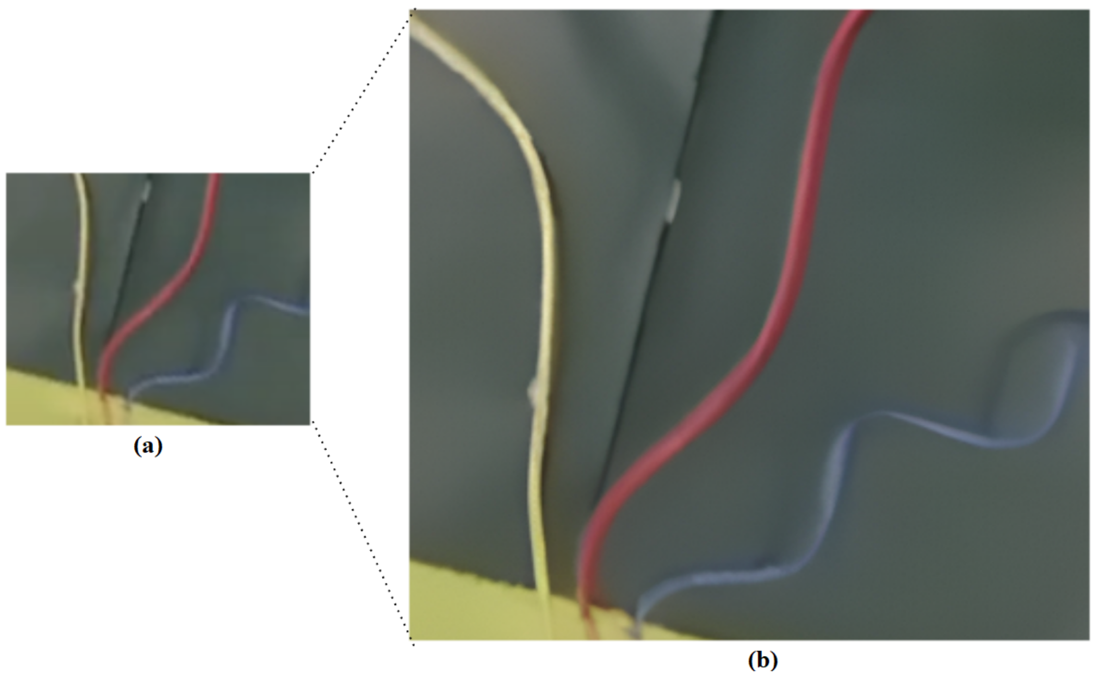
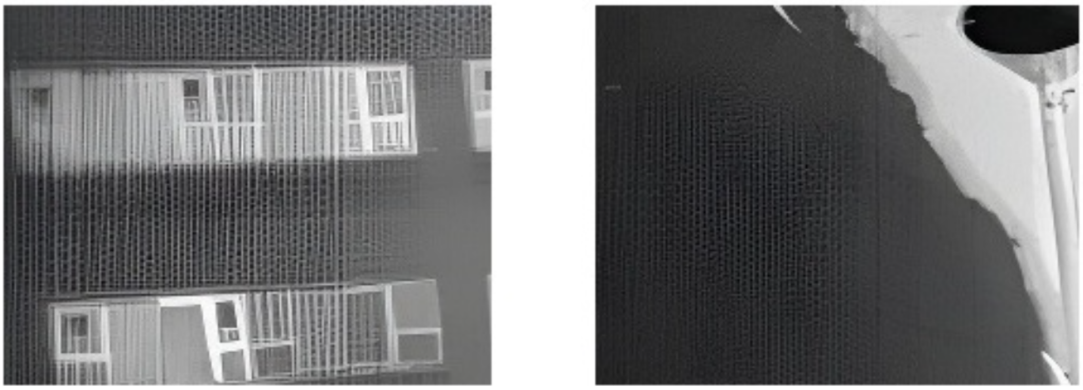


Figure 8. Real-ESRGAN on Visible Light Camera. (a) Original image. (b) After Super-resolution



(a) **(b)**

Figure 9. (a) Artifact 1. (b) Artifact 2

5. Discussion

The search string was devised to retrieve results related to edge and object detection in general, considering these tasks as the initial steps of image preprocessing. Identification of whether the object is an animal, a person, a car, or something else would fall under subsequent algorithmic steps, which may not necessarily executed by a generation algorithm as presented by some studies in the analysis.

Out of the 38 articles analyzed during the quality assessment phase, 19 were conference papers. Additionally, it was observed that the majority of the articles found were recent, with an age of less than 5 years, indicating that research focused on the use of GANs in object detection in images generated for UAVs is recent, as is research focused on the use of GANs in edge detection or object detection. The results show that GAN networks are applied in the image enhancement stage, for subsequent application in edge or target detection, performed by other algorithms such as YOLO or similar ones. Furthermore, the study highlights various types of detection targets in remote sensing literature besides people and animals, such as smoke, insulators, fire, among other targets related to agriculture. From the score analysis, we observe that the results were quite dispersed, with the majority of articles falling within the score intervals of 2 to 4, indicating considerable variability and diferent perspectives in the outcomes.

In light of the article analyses, RSRGAN[67] emerges as the optimal reference point for conducting further investigations into object detection, particularly in the context of adapting detection methods for identifying individuals and animals, as it not only provides super-resolution capabilities but also introduces a proprietary target detection system, accompanied by comparative analyses against alternative detectors.

Regarding the detection algorithms, YOLO predominated in versions 3 and 5, with Faster-RCNN also being used for the same purpose.

In terms of the metrics employed in the studies, we observed the utilization of classical metrics derived from digital image processing. These metrics are used to validate the classification outcomes of images generated through the application of YOLO or other detectors, as well as metrics for assessing noise or image degradation from datasets. The versatility of applications of these various metrics provides us with insights and ideas regarding their potential application for measuring the efficiency of generation algorithms in future works.

Thus, we can address the primary research question as follows: Overall, the studies develop or modify GANs for super-resolution, with the majority employing models trained from scratch using specific datasets. Following this training, the algorithm is applied to the target images, with object identification conducted before and after, aiming to evaluate the comparisons between the GAN algorithms used and their similarities, as well as the subsequent detection stage, where some version of YOLO is predominantly applied.

Consequently, this entire process can be adapted for image enhancement with targets relevant to SAR, aiming to improve UAV sensors for the rapid and real-time identification of these targets.

Addressing the first sub-question: From the studies examined, it is evident that the application of GANs in SAR operations may focus on enhancing sensors for subsequent target detection.

Regarding the second sub-question: A few studies has shown great results regarding pre-trained weights and models. Our framework, developed with Real-ESRGAN with default parameters and pre-trained model, demonstrates that the application of these approach yields satisfactory results in both visible light and infrared spectra, irrespective of the dataset used for pre-training, but it needs further investigation with new and related pre-training models to evaluate more accurately its impacts on practice.

Regarding the third sub-question, our observations revealed the utilization of PSNR and SSIM metrics to validate the results of GAN models. In addition to these traditional metrics, classic CNN metrics were also employed alongside various other metrics equally distributed to validate the detection algorithms and defect analyses, with Average Precision being the metric with the highest number of studies. This give us some light for comparing results and suggests a necessity for further studies aimed at reviewing and developing a framework capable of offering more profound insights into the most effective metrics to guide conclusions regarding the model under analysis.

Regarding Parsif.al tool, during the Quality Assessment phase, it was observed that a duplicated article escaped the automatic removal filter of the tool. Additionally, the data extraction step could automatically retrieve data from the article detail tab, allowing export in CSV format.

6. Conclusions

The systematic literature review has revealed that the application of GAN networks in SAR contexts is currently under development, with a focus on super-resolution for subsequent object detection. Other identified areas of application include the enhancement of training datasets for networks and drone navigation purposes. The results indicate that studies concerning search and rescue operations might be primarily oriented towards image enhancement, dataset expansion, and object identification models for subsequent target identification using classical algorithms such as YOLO, Faster-RCNN and others, with versions 3 and 5 of YOLO being the most prevalent in the evaluated studies. Potential areas for further investigation include real-time applications, target distance in photography, types of search targets, search region quality, and their impacts on photography. Although few results utilize pre-trained models, our benchmark has demonstrated that the utilization of a trained model, regardless of the dataset used, has shown interesting results regarding super-resolution in both the visible light and infrared spectrum. Furthermore, the Super-ESRGAN is a great candidate to be applied on the first stage of image processing for SAR. The validation limitations of this work include the selection of only two research databases (SCOPUS and IEEE), which may overlook other relevant articles not indexed by these databases. Additionally, our benchmark employs pre-trained models on images with targets close to the ground, lacking a comparison with a potential dataset tailored to search and rescue situations where the UAV is scanning at high altitudes. Lastly, the search string focuses on edge and object detection, which could be reconsidered to include animals and people as targets, along with incorporating the term "Search and Rescue" into the search. As future work, we propose studying the behavior of GAN algorithms in real-time, utilizing embedded hardware in UAVs on images captured at medium or high altitudes, to explore possibilities for target detection during real-time rescue operations.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

References

1. McIntosh, S; Brillhart, A; Dow, J; Grissom C. Search and Rescue Activity on Denali, 1990 to 2008.. *Wilderness & Environ. Med.* **2010**, *21(2)*, 103-108.
2. Curran-Sills, G; Karahalios, A; Dow, J; Grissom C. Epidemiological Trends in Search and Rescue Incidents Documented by the Alpine Club of Canada From 1970 to 2005. *Wilderness & Environ. Med.* **2015**, *26(4)*, 536-543.
3. Heggie, T; Heggie, T; Dow, J; Grissom C. Search and Rescue Trends and the Emergency Medical Service Workload in Utah's National Parks. *Wilderness & Environ. Med.* **2015**, *19(3)*, 164-171.
4. Ciesa, M; Grigolato, S; Cavalli, R. Retrospective Study on Search and Rescue Operations in Two Prealps Areas of Italy. *Wilderness & Environ. Med.* **2015**, *26(2)*, 150-158.
5. Freitas, C; Barcellos, C; Asmus, C; Silva, M; Xavier, D. From Samarco in Mariana to Vale in Brumadinho: mining dam disasters and Public Health. *Cad. Saúde Pública..* **2019**, *35*, e00052519.
6. Sássi, C; Carvalho, G; De Castro, L; Junior, C; Nunes, V; Do Nascimento, A; Gonçalves, One decade of environmental disasters in Brazil: the action of veterinary rescue teams. *Front. Public Health.* **2021**, *9*, 624975.
7. Stoddard, M; Pelot R. Historical maritime search and rescue incident data analysis. In *Governance of Arctic Shipping: Rethinking Risk, Human Impacts and Regulation*; Chircop, Aldo; Springer International Publishing, Cham, 2020; pp. 43-62.
8. Wajeeha, N; Torres, R; Gundersen, O; Karlsen, A. The Use of Decision Support in Search and Rescue: A Systematic Literature Review. *ISPRS Int. J. Geo-Inf.* **2023**, *12(5)*, 182.
9. Levine, A; Feinn, R; Foggle, J. Karlsen, A. Search and Rescue in California: The Need for a Centralized Reporting System. *Wilderness & Environ. Med.* **2023**, *34(2)*, 164-171.
10. Prata, I; Almeida, A; de Souza, F; Rosa, P; dos Santos, A. Developing a UAV platform for victim localization on search and rescue operations. In 2022 IEEE 31st International Symposium on Industrial Electronics (ISIE), Anchorage AK, USA, 01-03 June 2022; 721-726.
11. Lyu, M; Zhao, Y; Huang, C; Huang, H. Unmanned aerial vehicles for search and rescue: A survey. *Remote Sens* **2023**, *15(13)*, 3266.
12. Braga, J; Shiguemori, E; Velho, H. Odometria Visual para a Navegação Autônoma de VANT. *Rev. Cereus* **2019**, *11*, 184-194.
13. Cho, S; Matsushita, Y; Lee, S. Removing non-uniform motion blur from images. In 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14-21 October 2007; 1-8.
14. Li, Z; Gao, Z; Yi, H; Fu, Y; Chen, B. Image Deblurring With Image Blurring. *IEEE Trans. Image Process.* **2023**, *32*, 5595 - 5609.
15. Fan, L; Zhang, F; Fan, H; Zhang, C. Brief review of image denoising techniques. *Vis. Comput. Ind. Biomed. Art.* **2019**, *2*, 7.
16. Ibrahim, H; Kong, N. Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consumer Electron.* **2007**, *53(4)*, 1752-1758.
17. Park, S; Park, M; Kang, M. Super-resolution image reconstruction: a technical overview. *IEEE Signal Process. Mag.* **2003**, *20(3)*, 21-36.
18. Goodfellow, I; Pouget-Abadie, J; Mirza, M; Xu, B; Warde-Farley, D; Ozair, S, Bengio, Y. Generative adversarial networks. *Commun. ACM.* **2020**, *63(11)*, 139-144.
19. Ledig, C; Theis, L; Huszár, F; Caballero, J; Cunningham, A; Acosta, A; Shi, W. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21-26 July 2017; 4681-4690.
20. Yuan, Y; Liu, S; Zhang, J; Zhang, Y; Dong, C; Lin, L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; 701–710.
21. Wang, X; Yu, K; Wu, S; Gu, J; Liu, Y; Dong, C; Qiao, Y; Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. *Springer International Publishing.* **2019**, *11133*, 63-79.
22. Bell-Kligler, S; Shocher, A; Irani, M. Blind super-resolution kernel estimation using an internal-gan. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
23. Zhang, K; Gool, L; Timofte, R. Deep unfolding network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; 3217–3226.
24. Zhang, M; Ling, Q. Supervised pixel-wise GAN for face super-resolution. *IEEE Trans. Multimed.* **2020**, *23*, 1938–1950.
25. Zhang, K; Liang, J; Van Gool, L; Timofte, R. Designing a practical degradation model for deep blind image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; 4791–4800.
26. Wang, X; Xie, L; Dong, C; Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; 1905–1914.
27. Wang, X; Sun, L; Chehri, A; Song, Y. A Review of GAN-Based Super-Resolution Reconstruction for Optical Remote Sensing Images. *Remote Sens.* **2023**, *15(20)*, 5062.
28. Parsif.al. Available online: https://parsif.al (Accessed on 25 April 2024).
29. Yu, J.; Xue, H.; Liu, B.; Wang, Y.; Zhu, S.; Ding, M. GAN-Based Differential Private Image Privacy Protection Framework for the Internet of Multimedia Things. *Sensors.* **2021**, *(21)*, 58. https://doi.org/10.3390/s21010058
30. Gipiškis, R; Chiaro, D; Preziosi, M; Prezioso, E; Piccialli, F. The impact of adversarial attacks on interpretable semantic segmentation in cyber–physical systems. *IEEE Syst. J..* **2023**, *17(4)*, 5327 - 5334

31. Hu, M; Xiaoming, J. Two-stage insulator self-explosion defect detection method based on Mask R-CNN. In 2nd IEEE International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Shenyang, China, 17-19 November 2021; 13-18.
32. Bouguettaya, A; Zarzour, H; Kechida, A; Taberkit, A. Vehicle detection from UAV imagery with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33(11)*, 6047-6067.
33. Gan, Z; Xu, H; He, Y; Cao, W; Chen, G. Autonomous landing point retrieval algorithm for uavs based on 3d environment perception. In 2021 IEEE 7th International Conference on Virtual Reality (ICVR), Foshan, China, 20-22 May 2021; 104-108.
34. Shen, Y; Lee, H; Kwon, H; Bhattacharyya, S. Progressive transformation learning for leveraging virtual images in training. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17-24 June 2023; 835-844.
35. Chen, X; Li, T; Liu, H; Huang, Q; Gan, X. A GPS Spoofing Detection Algorithm for UAVs Based on Trust Evaluation. In IEEE 13th International Conference on CYBER Technology in Automation Control and Intelligent Systems (CYBER), Qinhuangdao, China, 11-14 July 2023, 315-319.
36. More, D; Acharya, S; Aryan, S. SRGAN-TQT, an Improved Motion Tracking Technique for UAVs with Super-Resolution Generative Adversarial Network (SRGAN) and Temporal Quad-Tree (TQT). *SAE Tech. Paper.* **2021**, 2022-26-0021.
37. Gong, Y; Liu, Q; Que, L; Jia, C; Huang, J; Liu, Y; Zhou, J. Raodat: An energy-efficient reconfigurable ai-based object detection and tracking processor with online learning. In 2021 IEEE Asian Solid-State Circuits Conference (A-SSCC), Busan, Korea, Republic of,07-10 November 2021; 1-3.
38. Gong, Y; Zhang, T; Guo, H; Liu, Q; Que, L; Jia, C; Zhou, J. An energy-efficient reconfigurable AI-based object detection and tracking processor supporting online object learning. *IEEE Solid-State Circuits Lett.* **2022**, *5*, 78-81.
39. Kostin, A; Gorbachev, V. Dataset Expansion by Generative Adversarial Networks for Detectors Quality Improvement. In CEUR Workshop Proceedings, Saint Petersburg Russia, September 22-25, 2020.
40. Shu, X; Cheng, X; Xu, S; Chen, Y; Ma, T; Zhang, W. How to construct low-altitude aerial image datasets for deep learning. *Math. Biosci. Eng.* **2021**, *18(2)*, 986-999.
41. Costea, D; Marcu, A; Slusanschi, E; Leordeanu, M. Creating roadmaps in aerial images with generative adversarial networks and smoothing-based optimization. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice Italy, 22-29 October 2017, 2100-2109.
42. Tian, B; Yan, W; Wang, W; Su, Q; Liu, Y; Liu, G; Wang, W. Super-Resolution Deblurring Algorithm for Generative Adversarial Networks. In 2017 Second International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin China, 08-10 December 2017, 135-140.
43. Chou, Y; Chen, C; Liu, K; Chen, C. Stingray detection of aerial images using augmented training images generated by a conditional generative model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18-22 June 2018, 1403-1409.
44. Bi, F; Lei, M; Wang, Y; Huang, D. Remote sensing target tracking in UAV aerial video based on saliency enhanced MDnet. *IEEE Access.* **2019**, *7*, 76731 - 76740.
45. Krajewski, R; Moers, T; Eckstein, L. VeGAN: Using GANs for augmentation in latent space to improve the semantic segmentation of vehicles in images from an aerial perspective. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 07-11 January 2019, 1440-1448.
46. Zhou, J; Vong, C; Liu, Q; Wang, Z. Scale adaptive image cropping for UAV object detection. *Neurocomputing.* **2019**, *366*, 305-313.
47. Chen, Y; Li, J; Niu, Y; He, J. VeGAN: 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, 3-5 June 2019, 4610-4615.
48. Xing, C; Liang, X; Bao, Z. VeGAN: Small object detection networks based on classification-oriented super-resolution GAN for UAV aerial imagery. 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, 19-20 October 2019, 313-316.
49. Truong, N; Lee, Y; Owais, M; Nguyen, D; Batchuluun, G; Pham, T; Park, K. SlimDeblurGAN-based motion deblurring and marker detection for autonomous drone landing. *Sensors.***2020**, *20(14)*, 3918.
50. Yu, H; Li, G; Su, L; Zhong, B; Yao, H; Huang, Q. Conditional GAN based individual and global motion fusion for multiple object tracking in UAV videos. *Pattern Recognit. Lett.***2020**, *131*, 219-226.
51. Wang D; Li Y. Insulator object detection based on image deblurring by WGAN. *Dianli Zidonghua Shebei/Electric Power Autom. Equip.***2020**, *40(5)*, 188-194.
52. Song, W; Li, S; Chang, T; Hao, A; Zhao, Q; Qin, H. VeGAN: Cross-view contextual relation transferred network for unsupervised vehicle tracking in drone videos. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass, CO, USA, 01-05 March 2020, 1707-1716.
53. Hou, X; Zhang, K; Xu, J; Huang, W; Yu, X; Xu, H. Object detection in drone imagery via sample balance strategies and local feature enhancement. *Appl. Sci.***2021**, *11(8)*, 3547.
54. Kniaz, V; Moshkantseva, P. Object re-identification using multimodal aerial imagery and conditional adversarial networks. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.***2021**, *54(2)*, 131-136.
55. Velumani, K; Lopez-Lozano, R; Madec, S; Guo, W; Gillet, J; Comar, A; Baret, F. Estimates of maize plant density from UAV RGB images using faster-RCNN detection model: Impact of the spatial resolution *Plant Phenomics.***2021**, *2021*, 9824843.

56. Mu, J; Li, S; Liu, Z; Zhou, Y. Integration of gradient guidance and edge enhancement into super-resolution for small object detection in aerial images. *IET Image Processing*.**15(13)**, *2021*, 3037-3052.
57. Zhao, W; Yamada, W; Li, T; Digman, M; Runge, T. Augmenting crop detection for precision agriculture with deep visual transfer learning—a case study of bale detection. *Remote Sens.***13(1)**, *2021*, 23.
58. Chen, W; Li, Y; Li; Zhao, Z. InsulatorGAN: A transmission line insulator detection model using multi-granularity conditional generative adversarial nets for UAV inspection *Remote Sens.***13(19)**, *2021*, 3971.
59. Wang, J; Yang, Y; Chen, Y; Han, Y. LighterGAN: an illumination enhancement method for urban UAV imagery. *Remote Sens.***13(7)**, *2021*, 1371.
60. Chen, L; Liu, G; Tan, Y; Sun, Z; Ge, H; Duan, F; Zhu, C. VeGAN: Cross-view contextual relation transferred network for unsupervised vehicle tracking in drone videos. A UA-net based Salient Object Detection Method for UAV. In 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), Snowmass, Sanya China, 27-31 December 2021, 1662-1667.
61. Liu, G; Tan, Y; Chen, L; Kuang, W; Li, B; Duan, F; Zhu, C. VeGAN: The development of a UAV target tracking system based on YOLOv3-tiny object detection algorithm. In 2021 IEEE International Conference on Robotics and Biomimetics (ROBIO), Snowmass, Sanya China, 27-31 December 2021, 1636-1641.
62. Ahmed, M; Khan, N; Ovi, P; Roy, N; Purushotham, S; Gangopadhyay, A; You, S. Gadan: Generative adversarial domain adaptation network for debris detection using drone. In 2022 18th International Conference on Distributed Computing in Sensor Systems (DCOSS), Los Angeles, USA, 30 May - 1 June 2022, 277-282.
63. Ma, D; Fang, H; Wang, N; Zhang, C; Dong, J; Hu, H. LighterGAN: Automatic detection and counting system for pavement cracks based on PCGAN and YOLO-MF. *IEEE Trans. Intell. Transp. Syst.***23(11)**, *2022*, 22166-22178.
64. Wang, W; Huang, W; Zhao, H; Zhang, M; Qiao, J; Zhang, Y. Generative adversarial domain adaptation network for debris detection using drone. Data Enhancement Method Based on Generative Adversarial Network for Small Transmission Line Detection. In International Conference on Neural Computing for Advanced Applications, Jinan, China, 8-10 July 2022, 413-426.
65. Park, M; Bak, J; Park, S. LighterGAN: Advanced wildfire detection using generative adversarial network-based augmented datasets and weakly supervised object localization. *Int. J. Appl. Earth Obs. Geoinf.***114**, *2022*, 103052.
66. Avola, D; Cannistraci, I; Cascio, M; Cinque, L; Diko, A; Fagioli, A; Pannone, D. A novel GAN-based anomaly detection and localization method for aerial video surveillance at low altitude. *Remote Sens.***14(16)**, *2022*, 4110.
67. Ren, K; Gao, Y; Wan, M; Gu, G; Chen, Q. Infrared small target detection via region super resolution generative adversarial network. *Appl. Intell.***52(10)**, *2022*, 11725-11737.
68. Hu, J; Zhang, Y; Zhao, D; Yang, G; Chen, F; Zhou, C; Chen, W. A robust deep learning approach for the quantitative characterization and clustering of peach tree crowns based on UAV images. *IEEE Trans. Geosci. Remote Sens.***60**, *2022*, 4408613.
69. Balachandran, V; Sarath, S. A novel approach to detect unmanned aerial vehicle using Pix2Pix generative adversarial network. In 2022 second international conference on artificial intelligence and smart energy (ICAIS), Coimbatore, India, 23-25 February 2022, 1368-1373.
70. Xu, Y; Luan, F; Liu, X; Li, X. Edge4fr: A novel device-edge collaborative framework for facial recognition in smart uav delivery systems. In 2022 IEEE 8th International Conference on Cloud Computing and Intelligent Systems (CCIS), Chengdu, Chinaa, 26-28 November 2022, 95-101.
71. Shimada, T; Nishikawa, H; Kong, X; Tomiyama, H. Depth Estimation from Monocular Infrared Images for Autonomous Flight of Drones. In 2022 International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Korea, Republic of, 06-09 February 2022, 1-6.
72. Marathe, A; Jain, P; Walambe, R; Kotecha, K. Restorex-ai: A contrastive approach towards guiding image restoration via explainable ai systems. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 019-20 June 2022, 3030-3039.
73. Du, Y; Yi, Y; Guo, H; Tian, X. Restorex-ai: A contrastive approach towards guiding image restoration via explainable ai systems. Vehicle detection in UAV traffic videos using GAN online augmentation: a transfer learning approach. In Third International Conference on Computer Vision and Data Mining (ICCVDM), Hulunbuir China, 19-21 August 2022, 1251132.
74. Zhu, B; Lv, Q; Tan, Z. Adaptive Multi-Scale Fusion Blind Deblurred Generative Adversarial Network Method for Sharpening Image Data. *Drones*.**7(2)**, *2022*, 96.
75. Sigillo, L; Grassucci, E; Comminiello, D. StawGAN: Structural-aware generative adversarial networks for infrared image translation. In 2023 IEEE International Symposium on Circuits and Systems (ISCAS), Monterey, CA, USA, 21-25 May 2023, 1-5.
76. Li, R; Peng, Y; Yang, Q. Fusion enhancement: UAV target detection based on multi-modal GAN. In 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 15-17 September 2023, 1953-1957.
77. Wu, H. Research on Motion Trend Enhanced 2D Detection on Drones. In 2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI), Taiyuan, China, 26-28 May 2023, 221-226.

APPENDIX B – 1st round of the manuscript review

For review article

Response to Reviewer 1 Comments

1. Summary

Dear reviewer, thank you very much indeed for taking the time to review this manuscript. We appreciate the comments you gave us and hope that the changes made to the document respond to your feedback as best as possible. Please find the detailed responses below and the corresponding revisions/corrections highlighted/in red in the re-submitted files.

2. Questions for General Evaluation	**Reviewer's Evaluation**	**Response and Revisions**
Is the work a significant contribution to the field?	★★★☆☆	Thank you very much for the evaluation.
Is the work well organized and comprehensively described?	★★★☆☆	
Is the work scientifically sound and not misleading?	★★★☆☆	
Are there appropriate and adequate references to related and previous work?	★★★☆☆	
Is the English used correct and readable?	★★★★☆	

3. Point-by-point response to Comments and Suggestions for Authors

Comments 1: Optimize the title of Figure 1. The content of the title can be explained in the body of the paper.

Response 1 Thank you for pointing this out, we have updated Fig 1's title to " Research Overview for this Systematic Review".

Comments 2: Clarify the meaning of the short-term SAR.

Response 2: Certainly. In line 89, we enclosed the acronym in parentheses and expanded it to "search and rescue." Additionally, in the introductory paragraph, we refined the definition of the term. Lines [15 to 22] were revised to provide a clearer explanation of "search and rescue.

Comments 3: EEE Xplore in Figure 4 should be IEEE Xplore.

Response 3: Thank you for pointing this out. We have revised and corrected Figure 4 which had this misspelled word.

Comments 4: In the result part, a large number of literature are listed and introduced, but the summary and analysis of the literature are lacking. It is suggested to list the advantages and disadvantages of different method systems and their applicability in a table.
Response 4: Thank you for this contribution, the methodologies, main advantage, main disadvantage, and possible application of the study to aid search and rescue were summarized in Tables 6 and 7 in the results section. We also add a cluster column on those tables to group the studies into similar investigations. We explain about this new table on lines [599 to 608]. We also discuss about the importance of this clusterization on the discussion section(lines 661 to 678).

Comments 5: The shortcomings of the existing research are summarized, and the research prospects are added.
Response 5: Thank you very much for pointing this out.

Comments 6: Many paragraphs in the paper lack full stops.
Response 6: Thanks we took a second reading on the paragraphs and did some modifications to this matter.

4. Response to Comments on the Quality of English Language
Point 1: Minor editing of English language required.
Response 1: Thank you for this feedback, The language was revised. We hope it has enhanced the clarity and readability.

For review article

Response to Reviewer 2 Comments

1. Summary

Dear reviewer, thank you very much indeed for taking the time to review this manuscript. We appreciate the comments you gave us and hope that the changes made to the document respond to your feedback as best as possible. Please find the detailed responses below and the corresponding revisions/corrections highlighted/in red in the re-submitted files.

2. Questions for General Evaluation	**Reviewer's Evaluation**	**Response and Revisions**
Is the work a significant contribution to the field?		Thank you very much indeed, we appreciate your generous evaluation.
Is the work well organized and comprehensively described?		
Is the work scientifically sound and not misleading?		
Are there appropriate and adequate references to related and previous work?		
Is the English used correct and readable?		

3. Point-by-point response to Comments and Suggestions for Authors

Comments 1: SAR in the paper is a shorthand for search and rescue work, and a clear explanation should be provided in the abstract or elsewhere. Because in the field of radar, SAR stands for synthetic aperture. This can easily lead to misunderstandings among readers. Need to make changes to the SAR of the title of the paper.

Response 1: Thank you for pointing this out, we were not familiar to this other meaning(synthetic aperture) for the SAR acronym. So, to aid this issue, we put the search and rescue before the (SAR) acronym in the abstract to address the term correctly. We did the same in line 89. In the introduction, we give more info about its meaning on lines [15 to 22].

Comments 2: What is the reason for the separate paragraph of the four literature work introductions in lines 43-44 and 104-105 of the article? If the work of

the four papers is important, it is recommended to elaborate on it in
detail.

Response 2: Thanks for the observation, We're sorry for the mistake and poor explanation. Lines 43 and 44 was intended to address 2 works from the literature that deals with contrast issues and super-resolution on the images without the usage of GANs in order to give background for the paper. We redid the lines [47 to 57] with more discussion from those cited works and merged them with the upmost closest paragraph.

Lines 104 and 105 presented a study that we decided to remove according to EX2, given that it's not a primary study, but a secondary work. It addresses a review done on different techniques (including GANs) used to identify vehicles on images generated by UAVs. We added this explanation in more detail in lines [124 to 127] and merged those new lines into the closest paragraph above.

Comments 3: I think the work shown in Figures 8 and 9 is only a technical breakthrough in the relevant literature and is not very related to the theme and background of the article's application. Can we replace the technical work that is more relevant to the application background?

Response 3: Thank you very much. figures 8 and 9 indeed provide additional information that is non-related to the proposed systematic literature review, we agree that they should be removed and used in future work.

Therefore, we removed them and also removed the lines [588 to 596] that address them and we also modified the text to remove its references.

For review article

Response to Reviewer 3 Comments

1. Summary

Dear reviewer, thank you very much indeed for taking the time to review this manuscript. We appreciate the comments you gave us and hope that the changes made on the document respond to your feedback as best as possible. Please find the detailed responses below and the corresponding revisions/corrections highlighted/in red in the re-submitted files.

2. Questions for General Evaluation	**Reviewer's Evaluation**	**Response and Revisions**
Is the work a significant contribution to the field?	★★★☆☆	Thank you for the evaluation.
Is the work well organized and comprehensively described?	★★☆☆☆	
Is the work scientifically sound and not misleading?	★★★☆☆	
Are there appropriate and adequate references to related and previous work?	★★★☆☆	
Is the English used correct and readable?	★★★☆☆	

3. Point-by-point response to Comments and Suggestions for Authors

Comments 1: The topic is innovative and of recent interest, however the introduction is too brief. The authors must, above all, briefly describe what the performance limits of current object detection approaches in images generated by UAVs are, highlighting how the use of GAN algorithms allows these limits to be overcome.

Response 1 Thank you for pointing this out, we appreciate this comment. The use of GAN algorithms has the property of creating new images from a latent space, based on a specific trained dataset. This alternative differs from algorithms not oriented to generation kind, since they use filters to improve the image, not necessarily creating new data in the process. With this ability of GANs to generate new data, these algorithms have great potential to replace some of these filters, generating new and improved images from low-resolution ones. Finally, with these improved images, one can use algorithms like YOLO to detect objects, edges, people, and animals, which can be very useful in search and rescue operations. So, this systematic literature review was meant to study about the solutions that apply GANs on UAV images. " We have modified the introduction, specifically lines[23-30] and [58-71] to briefly describe those gaps and the benefits of using GANs to address it".

Comments 2: In Tab. 3 it is necessary to add a column that specifies the target of the search, for example, detection of people or animals or other objects, detection of objects in the infrared bands, etc.

Response 2: Thank you very much for pointing it out. We created this target column in Table 3.

Comments 3: In section 3 the description of the literature searches should be reorganized into a more structured way. I suggest discussing individual research by grouping them by topic (for example, type of GAN algorithms used, detection targets, etc.) highlighting the significant results of the research

Response 3: Thank you, we appreciate your suggestion. We divided section 3 into Human and animals detection; Object detection; Infrared spectrum; YOLO versions and Pre-trained models subsections, later on Tables 6 and 7, we divided the selected studies into clusters with similar objectives, those tables also indicate the main advantage and disadvantage of each study, including a column with a possible application of the study for search and rescue assistance. Later on discussion(lines 661 to 678), we explained the importance of the clusterization to guide future works on upon the studies analyzed on this SLR.

Comments 4: In addition to the type of metrics used in the individual searches (TAbs 4 and 5), a further table is needed that shows the value of the measurement obtained, so that the performance results of the different methods can be compared.

Response 4: Thank you for this suggestion, we use the checkmark only to qualitatively inform the reader that the article uses this specific metric in its work, since this systematic review returned several kinds of applicability, each one using the metrics differently, for example, one study uses accuracy to comparing results of a technique in different datasets, another article uses the same metric to compare the same dataset, but in different algorithms, other authors use accuracy in YOLO. We decide only to point out that the metric listed was used on the given study. We added lines [652 to 660] to explain it to the reader.

We also moved the PSNR and SSIM metrics columns from Table 5 to Table 4, as they are widely used in many articles. We also moved AR and ROC columns from Table 4 to Table 5, since there were only one study on each one.

Comments 5: A more in-depth discussion needs to be had on the performance benefits of the algorithms deemed to be better. For example, it is necessary to have a discussion on the reasons for evaluating RSRGAN as the best algorithm for detecting individual or animals.

Response 5: Sorry, lines[627 to 631] were misguided, Our intention was to highlight that the paper referencing RSRGAN serves as a valuable reference within the review. This is because it not only employs RSRGAN for image super-resolution but also introduces its own detection approach, thus establishing itself as a benchmark for guiding future research endeavors. We have clarified this further in lines [644 to 649] for better comprehension.

4. Response to Comments on the Quality of English Language

Point 1: Some grammar typos are present in the document. They must be removed.

Response 1: Thank you for this feedback, the language was revised. We hope it has enhanced the clarity and readability.

Annex

ANNEX A – Certificate of acceptance of SLR publication

drones
an Open Access Journal by MDPI

CITESCORE 5.6

IMPACT FACTOR 4.4

CERTIFICATE OF PUBLICATION

The certificate of publication for the article titled:

Applications of GANs to Aid Target Detection in SAR Operations: A Systematic Literature Review

Authored by:

Vinícius Correa; Peter Funk; Nils Sundelius; Rickard Sohlberg; Alexandre Ramos

Published in:

Drones **2024**, Volume 8, Issue 9, 448

MDPI Academic Open Access Publishing since 1996

Basel, September 2024

Prof. Dr. Diego González-Aguilera
Editor-in-Chief

Bibliography

1 SANTOS, H.; OLIVEIRA, S.; PORTELLA, S. Between the craft and the mud: An analysis of the work and health relationship of firefighters in the breach of the córrego do feijão dam—brumadinho/mg. *Journal of Psychological Research*, v. 4, n. 4, 2022. 15

2 YU, S. et al. Study of the sheyuegou dam breach–experience with the post-failure investigation and back analysis. *Engineering Failure Analysis*, Elsevier, v. 125, p. 105441, 2021. 15

3 PRIMO, P. P. B. et al. Mining dam failures in brazil: comparing legal post-disaster decisions. *International journal of environmental research and public health*, MDPI, v. 18, n. 21, p. 11346, 2021. 15

4 FABRÍCIO, S. A.; FERREIRA, D. D. M.; BORBA, J. A. A panorama of mariana and brumadinho disasters: What do we know so far? *REAd. Revista Eletrônica de Administração (Porto Alegre)*, SciELO Brasil, v. 27, p. 128–152, 2021. 15

5 TAIB, Z. M.; JAHARUDDIN, N. S.; MANSOR, Z. A review of flood disaster and disaster management in malaysia. *International Journal of Accounting & Business Management*, v. 4, n. 3, 2016. 15

6 TINGSANCHALI, T. Urban flood disaster management. *Procedia engineering*, Elsevier, v. 32, p. 25–37, 2012. 15

7 JONKMAN, S. N.; KELMAN, I. An analysis of the causes and circumstances of flood disaster deaths. *Disasters*, Wiley Online Library, v. 29, n. 1, p. 75–97, 2005. 15

8 RAHIM, A. N. A. et al. An analysis of interpolation methods for super resolution images. In: IEEE. *2015 IEEE Student Conference on Research and Development (SCOReD)*. [S.l.], 2015. p. 72–77. 15

9 BOK, V.; LANGR, J. *GANs in action: deep learning with generative adversarial networks*. [S.l.]: Simon and Schuster, 2019. 17, 26

10 GOODFELLOW, I. et al. Generative adversarial nets. *Advances in neural information processing systems*, v. 27, 2014. 17

11 GOODFELLOW, I. et al. Generative adversarial networks. *Communications of the ACM*, ACM New York, NY, USA, v. 63, n. 11, p. 139–144, 2020. 17

12 WANG, X. et al. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: *Proceedings of the IEEE/CVF international conference on computer vision*. [S.l.: s.n.], 2021. p. 1905–1914. 18, 20, 22

13 WANG, X. et al. Esrgan: Enhanced super-resolution generative adversarial networks. In: *Proceedings of the European conference on computer vision (ECCV) workshops*. [S.l.: s.n.], 2018. p. 0–0. 19

14 HUANG, Y. et al. Unfolding the alternating optimization for blind super resolution. *Advances in Neural Information Processing Systems*, v. 33, p. 5632–5643, 2020. 19

15 WEI, P. et al. Component divide-and-conquer for real-world image super-resolution. In: SPRINGER. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*. [S.l.], 2020. p. 101–117. 19

16 JI, X. et al. Real-world super-resolution via kernel estimation and noise injection. In: *proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. [S.l.: s.n.], 2020. p. 466–467. 19

17 ZHANG, K. et al. Designing a practical degradation model for deep blind image super-resolution. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2021. p. 4791–4800. 19

18 SHI, W. et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 1874–1883. 19

19 ZHAO, X. et al. Itd-yolov8: An infrared target detection model based on yolov8 for unmanned aerial vehicles. *Drones*, MDPI, v. 8, n. 4, p. 161, 2024. 24, 40

20 LUO, Z.; TIAN, Y. Infrared road object detection based on improved yolov8. *IAENG International Journal of Computer Science*, v. 51, n. 3, 2024. 24, 40

21 LUO, Z.; TIAN, Y. Improved infrared road object detection algorithm based on attention mechanism in yolov8. *IAENG International Journal of Computer Science*, v. 51, n. 6, 2024. 24, 40

22 CORREA, V. et al. Applications of gans to aid target detection in sar operations: A systematic literature review. *Drones*, MDPI, v. 8, n. 9, p. 448, 2024. 32