Thermal Image Super-Resolution Using Real-ESRGAN for Human Detection

Vinícius H. G. Correa[1][a], Peter Funk[2][b], Nils Sundelius[2], Rickard Sohlberg[2],
Mastura Ab Wahid[3][c] and Alexandre C. B. Ramos[4][d]

[1] *Institute of Mechanical Engineering, Federal University of Itajubá, Itajubá, Brazil*
[2] *School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden*
[3] *School of Mechanical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*
[4] *Institute of Mathematics and Computing, Federal University of Itajubá, Itajubá, Brazil*
{correa, ramos}@unifei.edu.br, {peter.funk, nils.sundelius, rickard.sohlberg}@mdu.se, mastura@mail.fkm.utm.my

Keywords: Digital Image Processing, Generative Adversarial Networks, Target Detection, Search and Rescue.

Abstract: Unmanned Aerial Vehicles (UAVs) are increasingly crucial in Search and Rescue (SAR) operations due to their ability to enhance efficiency and reduce costs. Search and Rescue is a vital activity as it directly impacts the preservation of life and safety in critical situations, such as locating and rescuing individuals in perilous or remote environments. However, the effectiveness of these operations heavily depends on the quality of sensor data for accurate target detection. This study investigates the application of the Real Enhanced Super-Resolution Generative Adversarial Networks (Real-ESRGAN) algorithm to enhance the resolution and detail of infrared images captured by UAV sensors. By improving image quality through super-resolution, we then assess the performance of the YOLOv8 target detection algorithm on these enhanced images. Preliminary results indicate that Real-ESRGAN significantly improves the quality of low-resolution infrared data, even when using pre-trained models not specifically tailored to our dataset, this highlights a considerable potential of applying the algorithm in the preprocessing stages of images generated by UAVs for search and rescue operations.

1 INTRODUCTION

The use of UAVs (Unmanned Aerial Vehicles) like drones has been extensively studied and applied in SAR (Search and Rescue) operations, as presented on (Dousai and Lončarić, 2022), (Svedin et al., 2021), (Lygouras et al., 2019), (Kulkarni et al., 2020). This technology helps save resources and makes it easier to locate targets in hard-to-reach areas, natural disasters, or accidents. Furthermore, the capability to detect victims covered in mud or other debris during searches makes multispectral sensors important tools in this field of operation (Pensieri et al., 2020) (Schoonmaker et al., 2010), and given that the human and animal body emits electromagnetic waves in the infrared spectrum, the use of cameras in thermal spectrum can greatly aid identification in dark environments.

1.1 Related Work

Some studies are being conducted on target detection in the infrared spectrum using YOLO. (Shen et al., 2023) introduces an enhanced method called DBD-YOLOv8 designed specifically to address challenges associated with low signal-to-noise ratio and lack of texture detail in infrared images. This method incorporates innovative modules like BiRA and Dyheads to improve object detection accuracy across different scales, and handle occluded and small objects. Moreover, the proposed model significantly enhances multi-scale feature representation, filters out irrelevant regions, and improves feature fusion, resulting in a notable increase in average detection accuracy. These improvements enable DBD-YOLOv8 to meet real-time detection requirements, despite a slight increase in model complexity and minor inference time

[a] https://orcid.org/0000-0002-7578-559X
[b] https://orcid.org/0000-0002-5562-1424
[c] https://orcid.org/0000-0003-0515-0932
[d] https://orcid.org/0000-0001-8844-5116

Correa, V. H. G., Funk, P., Sundelius, N., Sohlberg, R., Wahid, M. A. and Ramos, A. C. B.
Thermal Image Super-Resolution Using Real-ESRGAN for Human Detection.
DOI: 10.5220/0013078800003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 247-254
ISBN: 978-989-758-728-3; ISSN: 2184-4321

overhead. The authors utilize various datasets tailored for target detection during training.

The ITD-YOLOv8, as discussed in (Zhao et al., 2024), is tailored for detecting infrared targets in complex scenarios and across different scales, all while reducing computational complexity. It incorporates improvements in YOLOv8's feature extraction backbone, including modules like GhostHGNetV2, AKConv, VoVGSCSP, and the CoordAtt attention mechanism. These enhancements are aimed at enhancing multi-scale feature extraction capabilities and effectively detecting hidden infrared targets in challenging environments. Moreover, ITD-YOLOv8 introduces the XIoU loss function to improve the accuracy of target localization, thereby reducing the rates of missed and false detections. Experimental results highlighted in the study show that ITD-YOLOv8 outperforms YOLOv8n significantly by notably decreasing the number of missed and false detections. Additionally, the model achieves a reduction in both model parameters and floating-point operations. The average precision (mAP) achieved is 93.5%, confirming the model's efficacy in detecting infrared targets in UAV applications.

(Luo and Tian, 2024b) presents YOLOv8-EGP, which proposes improvements and optimizations based on the original YOLOv8 to overcome challenges such as low detection accuracy, low robustness, and missed detections in infrared images. The enhancements include replacing the C2f module with a more flexible and adaptive convolution module called SCConv to improve feature diversity in the output. It also incorporates the dyhead detection header combined with multi-attention to enhance the expression capability of the detection head for infrared targets, along with adding a small target detection layer (min) to reduce missed detections of small targets and improve overall detection precision. These optimizations and improvements have led to a significant increase in detection accuracy, precision, and recall compared to the original YOLOv8, demonstrating the effectiveness of the enhanced model in detecting targets in the infrared spectrum.

(Luo and Tian, 2024a) incorporates the CPCA attention mechanism to enhance the model's focus on specific areas of infrared images. The authors replace the original downsampling layer with the CGBD module to preserve edge information and effectively handle local and contextual features. Additionally, they adopt the Weighted Intersection over Union (WIoU) loss function to more accurately assess target box coverage, thereby improving evaluation precision. These improvements result in a 1.4% increase in average precision (mAP) compared to the YOLOv8s model, demonstrating significant enhancements in precision and recall for detecting targets in the infrared spectrum. Moreover, the enhanced model is better suited for practical applications such as assisted driving and road monitoring platforms.

(Wang et al., 2023) introduces the Small Target Detection (STC) structure in the network, which serves as a bridge between shallow and deep features to enhance semantic information gathering for small targets and improve detection accuracy. Additionally, the algorithm incorporates the Global Attention Mechanism (GAM) to capture multi-dimensional feature information, thereby enhancing detection performance by integrating features from different dimensions. These improvements in small target detection in the infrared spectrum are crucial for addressing challenges faced by UAVs, such as detecting small targets and mitigating the blur caused by high flight speeds. Therefore, the algorithm proposed in this article has the potential to significantly enhance object detection in the infrared spectrum by UAVs, contributing to improved performance in detecting small targets in real-world scenarios of industrial inspection by UAVs.

Tables 1 and 2 present the YOLO metrics results for related studies, this comparison should be considered only for contextual understanding, as the datasets in the studies presented in the tables differ. The ITD-YOLOv8 model on (Zhao et al., 2024) was trained using 300 epochs on the HIT-UAV dataset, which consists of 2008 training images, 571 testing images, and 287 validation images. The dataset includes three classes: people, bicycles, and vehicles; The YOLOv8-EGP in (Luo and Tian, 2024b) was trained over 300 epochs using a dataset of 10,467 infrared images. The dataset was divided into training, testing, and validation sets, with 7,326 images for training, 2,094 images for testing, and 1,047 images for validation. The model proposed on (Luo and Tian, 2024a) was also trained over 300 epochs using the same dataset and configuration as (Luo and Tian, 2024b) and the classes of objects detected in the training dataset included person, bike, car, bus, light, and sign. These classes were carefully selected to improve the model's generalization ability for real-world scenarios in infrared object detection tasks.

Table 1: Comparison of Precision(P) and Recall(R) metrics from other studies. Here Eps. means the number of epochs applied to YOLO training.

Study	P(%).	R(%).	Eps.
(Zhao et al., 2024)	90.3	**88.6**	300
(Luo and Tian, 2024b)	85.6	74.0	300
(Luo and Tian, 2024a)	84.2	70.2	300

Table 2: Comparison of mAP50 and F1 Score metrics from other studies.

Study	mAP50(%)	F1(%)
(Zhao et al., 2024)	**93.5**	**89.4**
(Luo and Tian, 2024b)	82.9	79.3
(Luo and Tian, 2024a)	78.2	76.5

1.2 Super-Resolution

Enhancing images is a critical preprocessing step for raw data because factors like lighting conditions, sensor movement during capture, and camera quality can introduce artifacts and degradation. Improving these images is essential to mitigate these issues for their subsequent use, such as target detection in search and rescue operations conducted by unmanned aerial vehicles.

Classical algorithms for image enhancement, such as Nearest Neighbour; Bilinear, Bicubic (Rahim et al., 2015) and Lanczos (Fadnavis, 2014) increase an image's scale by interpolating pixel values through different mathematical methods. Nearest Neighbour duplicates the nearest pixel values, leading to a blocky look, while Bilinear averages the four closest pixels for smoother results. Bicubic interpolation, which considers a 4x4 grid of pixels, offers even smoother and sharper images, and Lanczos, using a sinc function, delivers high-quality results with minimal blurring and aliasing. Importantly, these methods do not involve artificial intelligence; they rely purely on mathematical computations rather than learning-based approaches. Each method has its strengths and trade-offs, impacting the final image quality.

Regarding the use of Convolutional Neural Networks (CNNs) for super-resolution, especially Generative Adversarial Networks (GANs) applied to UAV images, (Correa et al., 2024) provides a comprehensive systematic literature review of the application of GANs to drone images, analyzing trends and methodologies to improve Search and Rescue operations. The study highlights the effectiveness of GANs in enhancing image quality through super-resolution, which is crucial for improving target detection in SAR missions. Additionally, it discusses the integration of GANs with traditional object detection algorithms, such as YOLO and Faster R-CNN, to enhance the identification of targets in images captured by UAVs. The authors also propose areas for further investigation, including the use of pre-trained models and real-time applications of GANs in SAR operations, emphasizing the need for tailored datasets and methodologies. Overall, the findings suggest that GANs can significantly improve UAV sensor capabilities, enabling more effective and timely identification of targets in various operational conditions.

1.3 Real-ESRGAN

A widely employed algorithm for image enhancement that utilizes super-resolution techniques with GAN approach is Real-ESRGAN (Wang et al., 2021b), an upgraded work of the ESRGAN algorithm (Wang et al., 2018). Its generator utilizes deep neural networks with Residual-in-Residual Dense Blocks (RRDB) to capture fine details and produce high-quality images from low-resolution inputs. The training process involves comparing high-resolution images with their degraded counterparts to teach the model how to reconstruct high-quality details. Real-ESRGAN's training is more complex than its predecessor, ESRGAN, due to a broader range of image degradations and the use of a U-Net architecture combined with Spectral Normalization (SN) for improved stability and performance. Additionally, a pixel-unshuffle technique is employed to reduce computational load by decreasing the spatial size of inputs and increasing the channel size, which optimizes GPU memory usage. Real-ESRGAN's training on synthetic images allows it to handle diverse degradation scenarios, enhancing its performance on real-world images.

State-of-the-art (SOA) metrics like Precision, Recall, and mAP are critical for evaluating object detection performance, especially in tasks such as human detection. These metrics, commonly used in models like YOLO, provide insights into the accuracy and reliability of the detection system. Precision and Recall focus on identifying true positives and minimizing errors, while mAP measures overall performance across multiple detection thresholds. In applications like search and rescue, where accurate human detection is vital, SOA metrics are essential for tracking and improving detection effectiveness.

In contrast, traditional image quality metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are often used to evaluate super-resolution algorithms. While these metrics assess pixel-level similarity between original and enhanced images, they may not fully reflect improvements in visual quality, which are more relevant for tasks like object detection. GAN-based super-resolution methods often result in lower PSNR and SSIM values (Xue et al., 2020); (Zhang et al., 2021); (Wang et al., 2021a); (Lucas et al., 2019). However, this does not indicate a flaw in the algorithm but highlights a trade-off: GANs are designed to optimize perceptual quality, improving visual details that are more important for object detection, even at the cost of lower pixel accuracy (Wang et al., 2020).

In this study, we sought to assess the effectiveness of Real-ESRGAN in enhancing infrared images, as this spectrum often exhibits higher noise levels and is crucial for detecting human body heat signatures. We present an overview of the YOLOv8 model applied to our dataset, comparing the results between standard and super-resolved image samples. We then analyze and compare the super-resolved images produced by Real-ESRGAN with those generated using traditional non-AI super-resolution techniques, including Bilinear, Bicubic, Lanczos, and Nearest Neighbor Interpolation methods.

2 METHODOLOGY

To evaluate the performance of Real-ESRGAN using a pretrained model for enhancing infrared spectrum images, we propose the following methodological steps:

1. Extraction of frames from our infrared video sequences to compile a dataset for analysis.
2. Application of Real-ESRGAN Algorithm on the Dataset Samples previously obtained.
3. Assessing the effectiveness of image enhancement by comparing person detection results using YOLOv8 on both the original and enhanced images.
4. Comparison of the performance of Real-ESRGAN with classical super-resolution algorithms by evaluating Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics.

The initial phase of this research involved the creation of a dataset, which was constructed from a series of frames, each measuring 640x512 pixels. These frames were extracted from three videos captured by an infrared camera and represent the thermal signatures of individuals in an outdoor setting. Figures 1 and 2 provides examples of the frames used in the dataset construction.

In total, 10,043 frames ranging from 7 through 14 μm in wavelength were extracted from the 3 videos, forming our initial dataset. Subsequently, we created collections to facilitate our tests. Collection I was composed of 500 randomly selected frames from the first and second videos. Collection II consists of 100 randomly selected frames from all three videos. Collection III consists of 100 randomly selected frames, which were subsequently enhanced using the Real-ESRGAN algorithm with 4x upscaling through the RealESRGAN_x4plus model.

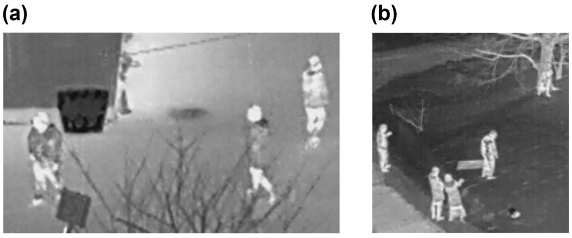

Figure 1: Examples of frames captured from the videos recorded by an infrared camera. Plot (a) is from the first video, plot (b) from the second.

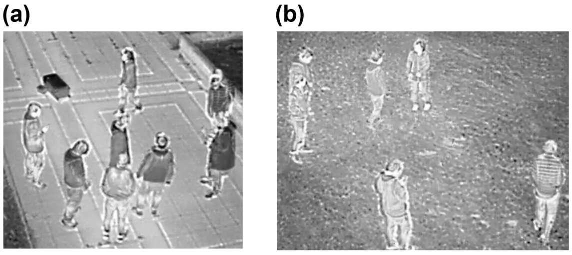

Figure 2: Examples of frames captured from the videos. Both plots are from the third video.

Collection I was used as training data for detecting people with YOLOv8. Except for collection I, collections II and III, which were used for validating, consist of frames extracted from all three videos.

After completing the super-resolution and collection creation phases, we proceeded with target detection using the YOLOv8 tool. During the annotation process, we labeled all the targets as ”human” in the dataset. We then selected an image from our dataset and applied classical interpolation methods, including Nearest Neighbor, Bilinear, Bicubic, and Lanczos, to enhance the image. Each interpolation method was evaluated for its impact on image quality. Next, the same image was enhanced using the Real-ESRGAN for comparison with the classical algoritms.

Following the enhancement process, we compared the PSNR and SSIM metrics of the images enhanced by the classical interpolation methods with that enhanced by the Real-ESRGAN algorithm.

3 RESULTS

After applying the Real-ESRGAN algorithm to our dataset, we observed significant qualitative improvements in the images. The enhanced images exhibited sharper details, improved clarity, and more defined features. Figure 3 illustrates a qualitative result of the algorithm on one of the images.

Despite notable improvements in image clarity, minimal contrast differences were observed in some

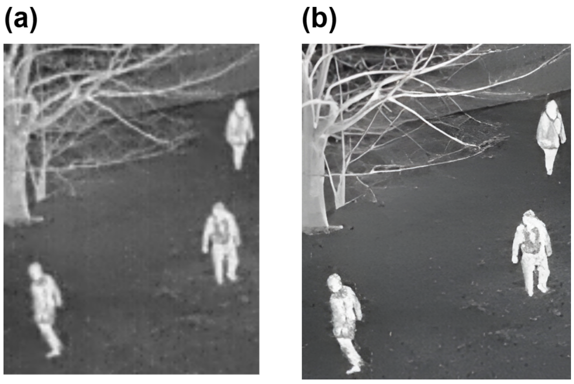

Figure 3: Example of satisfactory contrast enhancement in image resolution by Real-ESRGAN. (a) Image before enhancement. (b) Image after upgrading (Correa, 2024).

noisy frames even with the increased scaling, as illustrated in Figure 4 showing a struggle to improve contrast in regions dominated by noise, as its focus is on texture refinement rather than contrast enhancement. Additionally, artifacts were present in specific areas of the images, as depicted in Figure 5. These artifacts, also observed in (Wang et al., 2021b), were attributed to aliasing by the authors.

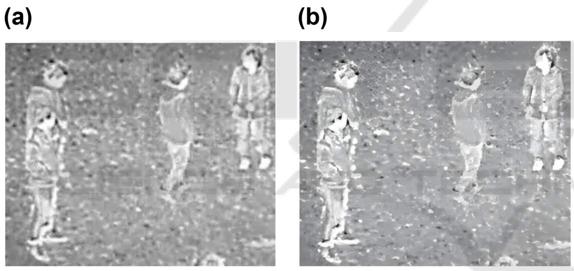

Figure 4: Example of minimal contrast improvement in an image after super-resolution enhancement. (a) Image before enhancement. (b) Same image after enhancement (Correa, 2024).

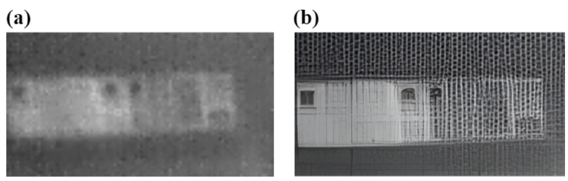

Figure 5: Artifacts observed in a region of an image after applying the enhancement algorithm. (a) Original image depicting a window. (b) Observed artifacts (Correa, 2024).

Tables 3 and 4 shows the performance of SOA metrics from our tests. Test 1 entailed training with the 500 frames from collection I and validating with the 100 frames from collection II. Test 2 involved training with the same 500 frames from collection I and validating with the 100 frames from collection III. We conducted training for 300 epochs in both tests 1 and 2.

Table 3: Comparison of Precision(P) and Recall(R) metrics from tests 1 and 2. (Correa, 2024).

Study	**P(%).**	**R(%).**	**Eps.**
Test 1	55.6	57.7	300
Test 2	**92.9**	**71.6**	300

Table 4: Comparison of mAP50 and F1 Score metrics from tests 1 and 2 (Correa, 2024).

Study	**mAP50(%)**	**F1(%)**
Test 1	48.7	56.6
Test 2	**83.4**	**80.9**

Figures 6 and 7 display the evolution of the performance metrics across epochs for Tests 1 and 2. In these figures, we observe a significant improvement in Test 2. Specifically, the images show higher values for Precision, Recall, and mAP50 in Test 2. This indicates that the model performed better in Test 2, where the images were enhanced using the Real-ESRGAN algorithm. The increased precision, recall, mAP50 and mAP50-95 suggest that the super-resolution technique contributed positively to the detection performance, helping the model achieve more accurate and reliable results in identifying human targets.

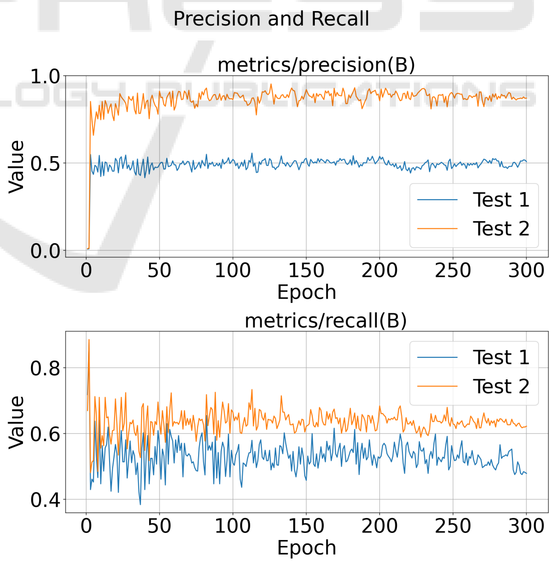

Figure 6: Precision and recall evolution for Tests 1 and 2 (Correa, 2024).

Based on the literature review by (Correa et al., 2024), we found that PSNR and SSIM are commonly used metrics in many studies that utilize GANs for super-resolution. These metrics compare values between an original image and its modified version, validating the degree of similarity between them. Thus,

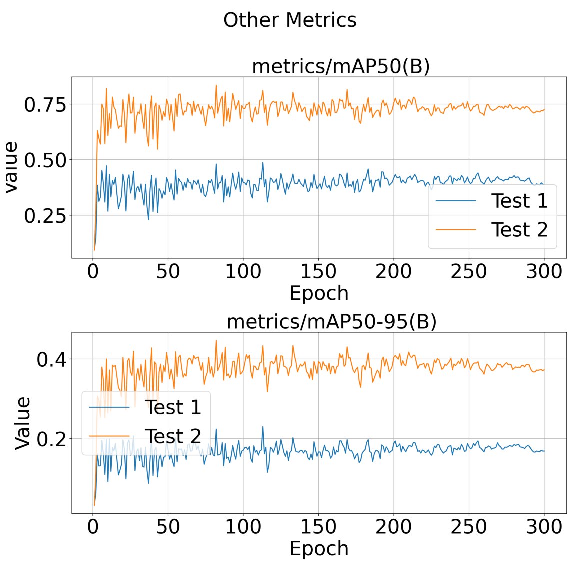

Figure 7: mAP50 and mAP50-95 evolution for Tests 1 and 2 (Correa, 2024).

we compared PSNR and SSIM across various interpolation methods, including Nearest Neighbor, Bicubic, Bilinear, and Lanczos. These classical interpolation methods were implemented using the OpenCV library and Python. The PSNR and SSIM metrics require that the compared images be of the same size. Since the original frames in the dataset have dimensions of 640x512 pixels, all interpolation methods were applied with a scaling factor of 4x to ensure that the metrics could be compared accurately against each other. Tables 5, 6, 7, 8 and 9 present the results of the comparison of these metrics.

Table 5: Comparison of metrics against Bicubic Interpolation (Correa, 2024).

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 6: Comparison of metrics against Bilinear Interpolation (Correa, 2024).

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 7: Comparison of metrics against Lanczos Interpolation (Correa, 2024).

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 8: Comparison of metrics against Nearest Neighbour Interpolation (Correa, 2024).

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

4 DISCUSSION

Infrared images often contain more noise, and the perceptual improvements offered by Real-ESRGAN can be particularly beneficial in enhancing features such as human shapes or body heat, which are critical for human detection. Despite struggling with some noisy images and the presence of artifacts in certain frames, as shown in our results, the use of Real-ESRGAN significantly improved sharpness and body delineation, demonstrating its effectiveness in enhancing key features for detection tasks.

As shown in Figures 6 and 7, the metric curves for Test 2 demonstrated better performance. This can be attributed to the differing image qualities in Collection II and Collection III. In Test 1, where the model was validated on low-resolution images, the model encountered difficulties in extracting detailed features due to the lower image quality. As a result, performance during training was lower, as reflected in the corresponding curves.

In contrast, Test 2, which used the super-resolution images from Collection III for validation, showed a significantly improved performance. The enhanced images, with more detail and clarity, allowed the detection model to more effectively extract relevant features, leading to faster convergence and better performance, as indicated by the higher precision, recall, and mAP50 scores from Tables 3 and 4 and curves from Figures 6 and 7. These findings emphasize the impact of image resolution on object detection performance.

Overall, the results demonstrate the importance of high-quality images, especially in the case of infrared

Table 9: Comparison of metrics against Real-ESRGAN super-resolution (Correa, 2024).

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

imagery, where enhanced details are crucial for accurate detection. This supports the effectiveness of Real-ESRGAN as a preprocessing step for object detection tasks, particularly when dealing with images of varying resolutions.

While classical filters are effective for enhancing image quality, they are limited to manipulating existing pixels through mathematical transformations. In contrast, GANs reconstruct the image from the latent space, making more substantial pixel-level modifications to enhance details. However, these algorithms are constrained by the dataset used for training. In our study, applying the Real-ESRGAN algorithm with pre-trained models for infrared image enhancement proved effective for some images, despite the pre-trained model not being specifically related to our dataset. This demonstrates a useful capability, particularly in situations where obtaining data for training generative models is challenging.

YOLOv8 showed consistent performance and metrics throughout the training process. The SOA metrics show that, despite using a smaller dataset than other studies, the super-resolution method achieved high precision, recall, and mAP50 scores, underscoring the effectiveness of Real-ESRGAN in image preprocessing. Our key contribution lies in proposing image enhancement as a preprocessing step prior to object detection, especially for YOLO, which requires intermediate steps such as image annotation and class creation. We recommend enhancing the images first, followed by annotation and YOLO training for human detection.

By comparing Real-ESRGAN with classical interpolation algorithms, it can be observed in Tables 5, 6, 7, 8, and 9 that Real-ESRGAN resulted in lower PSNR and SSIM values. The lower values for these metrics have also been noted in other studies that utilize GANs for super-resolution. While this might initially suggest that the algorithm is less effective, it is important to note that classical interpolation methods typically make only minor adjustments to the image, preserving much of the original structure. In contrast, Real-ESRGAN enhances the image by reconstructing it from learned feature representations in the latent space, which can result in significant modifications at the pixel level. These modifications, while improving perceptual quality, may introduce dissimilarities from the original image, leading to lower PSNR and SSIM values.

Compared to traditional interpolation methods, Real-ESRGAN is significantly more computationally intensive. Enhancing 1,000 frames requires approximately 90 minutes on standard hardware, such as an RX570 graphics card. This extended processing time presents a substantial limitation for real-time or automated applications, particularly in critical fields like search and rescue operations. Therefore, further research is necessary to optimize super-resolution algorithms for real-time deployment.

5 CONCLUSIONS

This study enabled us to assess the use of a super-resolution algorithm on a low-resolution dataset, with a particular focus on improving thermal images for human detection. By comparing the results before and after data enhancement using YOLOv8 as the benchmark, we observed a significant improvement in both image quality and SOA metrics. The application of the Real-ESRGAN algorithm demonstrated considerable potential for enhancing thermal images, which is especially beneficial for human detection in UAV-based applications.

Our study also revealed that the image enhancement process via Real-ESRGAN, which entails pixel reconstruction and modification, can result in decreased PSNR and SSIM values. However, human detection performance improved compared to using the original images. This highlights the trade-off: while the similarity metrics decreased, detection performance increased, suggesting that the enhancement process benefits detection tasks, even if traditional quality metrics are lower.

The images in this study were captured at low altitudes using an infrared camera designed for short-range detection, while drones typically operate at much higher elevations. We also observed challenges in enhancing some noisy images and the presence of artifacts in certain regions, highlighting potential flaws and areas for improvement in the algorithm. Additionally, the dataset used was relatively small, comprising 500 training images and 100 validation images per test. As a result, future research could focus on improving the quality of high-altitude images and expanding the dataset size to enable more robust training, leading to more accurate results. Furthermore, exploring other GAN-based algorithms for super-resolution is recommended to address the noise and artifact issues, as different models may offer varying performance trade-offs. Additionally, training a GAN from scratch, rather than relying on pre-trained models, could offer alternative approaches that may better suit the specific dataset and improve performance.

ACKNOWLEDGEMENTS

The authors thank the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), project APQ00234-18, for the financial support that made this research possible.

ChatGPT was carefully used to refine and clarify the English expressions in the text, ensuring the ideas were communicated accurately. Additionally, ChatPDF was thoughtfully employed to identify key sections of the references, which helped the authors address the findings effectively and make well-informed comparisons of the results.

REFERENCES

Correa, V., Funk, P., Sundelius, N., Sohlberg, R., and Ramos, A. (2024). Applications of gans to aid target detection in sar operations: A systematic literature review. *Drones*, 8(9):448.

Correa, V. H. G. (2024). Application of real-esrgan in improving ir sensor images for use in sar operations. Master's thesis, Federal University of Itajubá, Brazil.

Dousai, N. M. K. and Lončarić, S. (2022). Detecting humans in search and rescue operations based on ensemble learning. *IEEE access*, 10:26481–26492.

Fadnavis, S. (2014). Image interpolation techniques in digital image processing: an overview. *International Journal of Engineering Research and Applications*, 4(10):70–73.

Kulkarni, S., Chaphekar, V., Chowdhury, M. M. U., Erden, F., and Guvenc, I. (2020). Uav aided search and rescue operation using reinforcement learning. In *2020 SoutheastCon*, volume 2, pages 1–8. IEEE.

Lucas, A., Lopez-Tapia, S., Molina, R., and Katsaggelos, A. K. (2019). Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Transactions on Image Processing*, 28(7):3312–3327.

Luo, Z. and Tian, Y. (2024a). Improved infrared road object detection algorithm based on attention mechanism in yolov8. *IAENG International Journal of Computer Science*, 51(6).

Luo, Z. and Tian, Y. (2024b). Infrared road object detection based on improved yolov8. *IAENG International Journal of Computer Science*, 51(3).

Lygouras, E., Santavas, N., Taitzoglou, A., Tarchanidis, K., Mitropoulos, A., and Gasteratos, A. (2019). Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations. *Sensors*, 19(16):3542.

Pensieri, M. G., Garau, M., and Barone, P. M. (2020). Drones as an integral part of remote sensing technologies to help missing people. *Drones*, 4(2):15.

Rahim, A. N. A., Yaakob, S. N., Ngadiran, R., and Nasruddin, M. W. (2015). An analysis of interpolation methods for super resolution images. In *2015 IEEE Student Conference on Research and Development (SCOReD)*, pages 72–77. IEEE.

Schoonmaker, J., Reed, S., Podobna, Y., Vazquez, J., and Boucher, C. (2010). A multispectral automatic target recognition application for maritime surveillance, search, and rescue. In *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX*, volume 7666, pages 374–384. SPIE.

Shen, L., Lang, B., and Song, Z. (2023). Infrared object detection method based on dbd-yolov8. *IEEE Access*.

Svedin, J., Bernland, A., Gustafsson, A., Claar, E., and Luong, J. (2021). Small uav-based sar system using low-cost radar, position, and attitude sensors with onboard imaging capability. *International Journal of Microwave and Wireless Technologies*, 13(6):602–613.

Wang, F., Wang, H., Qin, Z., and Tang, J. (2023). Uav target detection algorithm based on improved yolov8. *IEEE Access*.

Wang, J., Gao, K., Zhang, Z., Ni, C., Hu, Z., Chen, D., and Wu, Q. (2021a). Multisensor remote sensing imagery super-resolution with conditional gan. *Journal of Remote Sensing*.

Wang, X., Xie, L., Dong, C., and Shan, Y. (2021b). Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0.

Wang, Z., Jiang, K., Yi, P., Han, Z., and He, Z. (2020). Ultra-dense gan for satellite imagery super-resolution. *Neurocomputing*, 398:328–337.

Xue, X., Zhang, X., Li, H., and Wang, W. (2020). Research on gan-based image super-resolution method. In *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pages 602–605. IEEE.

Zhang, X., Feng, C., Wang, A., Yang, L., and Hao, Y. (2021). Ct super-resolution using multiple dense residual block based gan. *Signal, Image and Video Processing*, 15:725–733.

Zhao, X., Zhang, W., Zhang, H., Zheng, C., Ma, J., and Zhang, Z. (2024). Itd-yolov8: An infrared target detection model based on yolov8 for unmanned aerial vehicles. *Drones*, 8(4):161.