2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) | 979-8-3503-7091-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/SITIS61268.2023.00056

Object Detection in Very Low-Resolution Thermal Images through a Guided-Based Super-Resolution Approach

Rafael E. Rivadeneira[1]
[1]*ESPOL Polytechnic University*
FIEC - CIDIS
Guayaquil, Ecuador
rrivaden@espol.edu.ec

Henry O. Velesaca[1,2]
[2]*Software Engineering Department*
University of Granada
18014, Granada, Spain
hvelesaca@correo.ugr.es

Angel Sappa[1,3]
[3]*Computer Vision Center*
08193-Bellaterra
Barcelona, Spain
sappa@ieee.org

***Abstract*—This work proposes a novel approach that integrates super-resolution techniques with off-the-shelf object detection methods to tackle the problem of handling very low-resolution thermal images. The suggested approach begins by enhancing the low-resolution (LR) thermal images through a guided super-resolution strategy, leveraging a high-resolution (HR) visible spectrum image. Subsequently, object detection is performed on the high-resolution thermal image. The experimental results demonstrate tremendous improvements in comparison with both scenarios: when object detection is performed on the LR thermal image alone, as well as when object detection is conducted on the up-sampled LR thermal image. Moreover, the proposed approach proves highly valuable in camouflaged scenarios where objects might remain undetected in visible spectrum images.**

***Index Terms*—Thermal imaging, super-resolution, deep learning, object recognition, computer vision, low-resolution images, yolo v8, camouflage detection.**

I. Introduction

In several applications, including surveillance, search and rescue efforts, and industrial inspections, thermal imaging has developed into a useful technology. Thermal cameras help us to see beyond the range of visible light by catching the infrared radiation that things release. The inherent low-resolution of the images that are recorded, however, is a key downside of thermal imaging and can make the object detection task more difficult. This restriction is a significant obstacle in a variety of real-world settings where accurate object identification is essential.

Researchers have looked into using super-resolution techniques to increase the resolution of thermal imaging as a way to get around this problem (e.g., [1], [2], [3]). A higher level of visual fidelity is provided by super-resolution, which seeks to recover high-frequency details and fine characteristics from low-resolution images. Parallel to this, deep learning-based object identification techniques, such as the well-known You Only Look Once (YOLO) architecture, have proven very effective at identifying and localizing items in photos.

To solve the problem of low-resolution thermal image-based object recognition, a novel method is offered in this study that blends super-resolution methods with deep learning-based object recognition. The goal is to improve object identification by enhancing the resolution of thermal images by utilizing the strength of super-resolution algorithms. To achieve super-resolution, a deep neural network architecture is used, which makes use of both the statistical characteristics of thermal imagery and contextual information. The high-resolution thermal images are then sent into a deep learning model that has already been trained to detect objects, doing away with the requirement for specialist thermal object detection architectures.

This study's main objective is to assess how well the suggested method works for boosting thermal image resolution and increasing object detection precision. Tests are done on a variety of images to assess how well the proposed approach outperforms more established low-resolution thermal image-based object recognition methods. The enormous improvements made possible by combining super-resolution and deep learning-based object recognition are shown by the experimental findings. The integrated strategy offers a full answer for precisely identifying objects in thermal photography in addition to improving the resolution of thermal images.

The remainder of the paper is divided into the following sections: A review of pertinent literature in the fields of super-resolution, thermal image processing, and object recognition is given in Section II. The proposed methodology is explained in Section III, along with the steps involved in super-resolution and integration with deep learning-based object recognition. The experimental setup is described in Section IV, together with information on the dataset, the selected evaluation measures, and the overall experimental arrangement. Finally, conclusions are given in Section V.

II. Related Work

This section presents a summary of the most recent and relevant contributions to the topics tackled in the current work. Section II-A summarizes the state-of-art on image super-resolution, going from SR on visible spectrum images to SR on thermal images. Then, Section II-B presents object detection approaches and finally, Section II-C describes recent thermal image datasets freely available in the literature.

979-8-3503-7091-1/23/$31.00 ©2023 IEEE
DOI 10.1109/SITIS61268.2023.00056

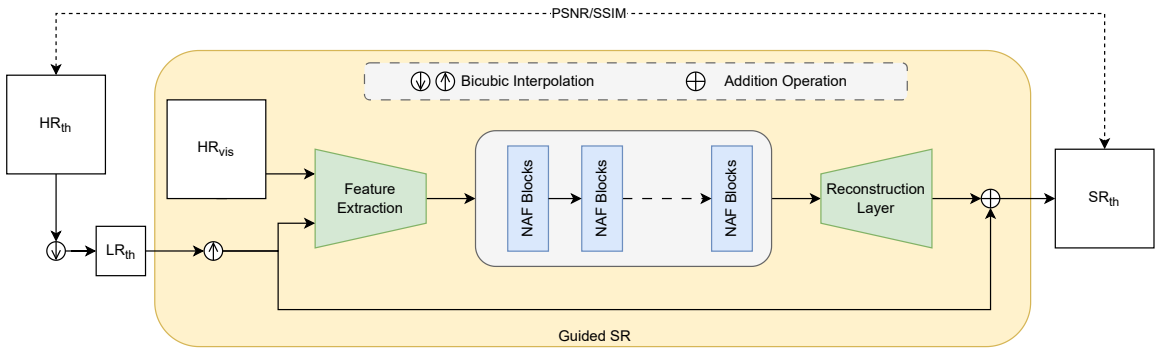

Fig. 1. Guided Thermal Image Super-resolution approach presented in [4].

A. *Image Super-Resolution*

The literature on single image super-resolution (SISR) has witnessed significant advancements over the years, with a recent focus on leveraging deep learning techniques to achieve improved results compared to conventional methods. Convolutional neural networks (CNNs) have particularly demonstrated remarkable capabilities in enhancing the quality of super-resolution (SR) outcomes. Dong et al. [5] introduce the SRCNN, which pioneered the concept of end-to-end mapping between interpolated LR images and their corresponding HR counterparts, achieving state-of-the-art performance. Building upon this work, [1] proposes FSRCNN, which extracts feature maps from LR images and performs the up-sampling in the final layer, leading to further performance improvements.

Inspired by the success of SRCNN, subsequent studies started to explore deeper networks with residual learning, [6] and [7] employ deep CNN architectures with residual connections to enhance SR accuracy. To expedite the training process, [8] introduces EDSR, a network that eliminates the batch-normalization layer and leverages residual learning [9]. It is worth noting that most of these CNN-based approaches aim to minimize the mean-square error (MSE) between the SR and ground truth (GT) images, which can sometimes result in the suppression of high-frequency details. This supervised training process typically requires pairs of pixel-wise registered SR and GT images to compute the MSE.

All the aforementioned approaches have been initially designed for visible spectrum images. Over the last decade, novel strategies have emerged to address the thermal image super-resolution challenge. In [2], the authors introduce TISR-DCNN, a supervised CNN architecture featuring both residual and dense connections. This architecture effectively captures high-level features and facilitates the reconstruction of thermal images. Several unsupervised architectures have been proposed (e.g., [10], [11], [12]) in recent years to tackle the thermal image SR problem. These unsupervised super-resolution approaches leverage unpaired images to overcome the constraint of pixel-wise registration without relying on specific degradation assumptions. They achieve this through the utilization of CycleGAN [13] architectures, which proficiently map information from one domain (LR thermal image) to another domain (HR thermal image). To promote the advancement of novel techniques within this domain, a thermal image super-resolution challenge has been annually organized in the Perception Beyond the Visible Spectrum workshop, held in conjunction with the CVPR conference (see [10], [14], [15]).

Although interesting results have been made in achieving $\times 2$ and $\times 4$ super-resolutions, the task of restoring information with a large up-scaling requires more attention. In recent years, guided super-resolution (GSR) approaches have been proposed as a solution for such a challenging problem. Guided super-resolution approaches use information from one domain, usually a HR cheap visible spectrum camera, as a guidance (e.g., [16], [17], [18]) to drive the SR process of LR images. This strategy has been used in different SR domains (e.g., depth-map SR, infrared SR, thermal SR, and some others). In the current work, the guided thermal image SR problem is considered. In this context, the authors of [19] propose to overcome the limitation of thermal imaging using a multiconditioned guidance network (MGNet). The MGNet uses HR visible spectrum images to enhance the super-resolution of thermal UAV images. HR visible spectrum images contribute rich information from the given scene (e.g., texture and edge features, semantic details) that is useful for the SR process. To leverage this information, the authors introduce the multicue guidance module (MGM) to effectively integrate this image information from visible images to guide the process of thermal UAV image super-resolution. In [20] the authors propose to use a generative model, referred to as Dual-IRT-GAN, to simultaneously tackle the super-resolution and defect detection problems in thermal images. The visibility of flawed areas in the resulting high-resolution images is enhanced by using defect-aware attention maps derived from segmented defect images. Finally, in [4] several GSR models for thermal images have been proposed in the context of a challenge organized in the 2023 Perception Beyond the Visible Spectrum workshop; from this challenge and inspired by ChaSNet [21] and Swin Transformer [22], the ANT INS

team devised TCP-SRNet, a SR Network blending Channel Split Convolutions for extracting local features and Swin Transformers for capturing spatial relationships. This network enhances super-resolution by combining both. It is trained with inputs semi-matched, first with $\times 2$ upscaling using L1 loss, and then with additional LSGAN and SSIM losses. The final output is an average of these two models' outputs. On the other hand, the GUIDED-SR team introduces a two-stream network for enhancing LR thermal images using HR RGB images as guides. Shallow features from LR and RGB images are concatenated and fused using cascaded NAF Blocks [23]. HR image reconstruction yields a super-resolution outcome. The training involves L1 and MSE loss in two steps. Finally, the TU-PAC team tackles guided super-resolution challenges using an Attention-based Pixel Adaptive Convolution (APAC) layer [24], enhancing misaligned thermal images with guide images. Their network involves Encoder, Guide, and Decoder branches, refining guides with attention mechanisms and upscaling thermal features. The model that reaches the best result from the challenge has been selected and is going to be used in the current work.

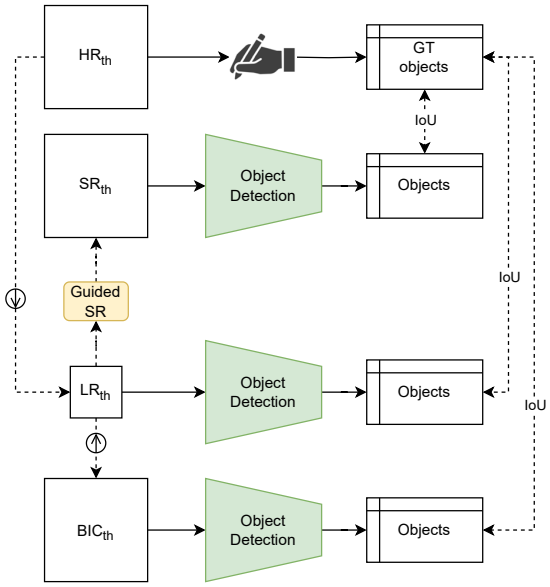

Fig. 2. Object detection evaluation.

B. Object Recognition

Object detection is a fundamental task in computer vision that involves localizing and identifying objects within an image. Over the years, numerous methods have been proposed to address this challenge, with significant advancements achieved through the use of deep learning techniques. One influential approach in object detection is the You Only Look Once (YOLO) architecture, which has gained widespread popularity due to its real-time processing capabilities and high accuracy. The YOLO framework, introduced by [25], revolutionized object detection by formulating it as a regression problem. Unlike traditional methods that employ region proposal algorithms, YOLO directly predicts bounding boxes and class probabilities in a single pass through the network. This design allows YOLO to achieve remarkable speed while maintaining competitive accuracy. Several iterations and improvements have been made to the original YOLO architecture, including from version 2 to the latest YOLO v8 [26], which is part of the present work and which has further enhanced the detection performance.

YOLO uses a deep convolutional neural network (CNN) to extract meaningful features from the input image. It divides the image into a grid and applies a set of predefined anchor boxes to predict bounding boxes and associated class probabilities for each grid cell. This efficient design enables YOLO to detect multiple objects simultaneously across different spatial locations. Additionally, YOLO incorporates techniques like Feature Pyramid Networks (FPN) [27] and multi-scale training to improve detection accuracy for objects of various sizes.

The YOLO framework has demonstrated outstanding performance across a wide range of object detection applications, including pedestrian detection (e.g., [28], [29]), vehicle detection (e.g., [30], [31]), video surveillance (e.g., [32], [33]) and general object detection (e.g., [34], [35]). Its real-time processing capability has made it especially valuable for real-time video analysis and applications that require fast and accurate object detection. In the current work, YOLO is used as the object detection component to perform accurate object recognition on thermal images. While YOLO has been extensively used in the context of visible light images, its application to thermal imagery is a novel direction that holds great potential. By combining YOLO with the proposed super-resolution technique, the research aims to enhance the resolution of thermal images and improve the accuracy of object identification, opening up new possibilities for object detection in thermal imaging domains.

C. Datasets

The field of super-resolution thermal imaging has seen the emergence of various datasets, although with different sizes and purposes. While some datasets are suited for object detection, tracking, biometrics, or medical applications, only a few are specifically intended for super-resolution tasks. Among the available thermal datasets, the data set presented by Davis et al. [36] consists of 284 thermal images captured on a college campus using a Raytheon 300D camera. Olmeda et al. [37] propose a dataset of 15224 thermal images acquired of urban scenes with a vehicle-mounted Indigo Omega imager. Hwang et al. [38] use a FLIR-A35 camera to capture more than 41,500 thermal images. Wu et al. [39] presents a high-resolution dataset with seven scenes captured with a FLIR SC8000 camera, which provides a resolution of 1024×1024 pixels.

The combination of visible and infrared images has gained significant attention in the field of computer vision, especially for object detection tasks. The M^{3}FD [40] dataset has emerged

as a valuable resource for researchers in this area. The M^{3}FD dataset includes synchronized systems consisting of binocular infrared sensors and an optical camera. It offers a diverse range of scenarios captured at locations such as the Dalian University of Technology campus and the State Tourism Holiday Resort at Golden Stone Beach in Dalian, China. With over 8,400 images available for fusion, detection, and fused-based detection, as well as an additional 600 independent scene images for fusion, the dataset provides ample data for comprehensive evaluation.

Furthermore, the M^{3}FD dataset incorporates manual labeling for 34407 instances across six target categories, including People, Cars, Buses, Motorcycles, Lamps, and Trucks. This labeling facilitates object detection research, although the dataset creators acknowledge the possibility of some labeling errors or omissions due to limited human resources.

III. Proposed Approach

This section presents the approach proposed in the current work to address object detection in very LR thermal images, using a guided super-resolution model. The proposed approach combines a guided super-resolution technique with the YOLO network for object recognition. By integrating these two components, the proposed approach aims to enhance the resolution of thermal images while simultaneously improving the capabilities of object detection.

The first task is to define and configure the dataset to use in the next sections. For the training and testing phase, the **M^{3}FD dataset** [40] is used with the following distribution training=3360 images, validation=740 images, and testing=100 images. In addition, for the test images, downsampling to the size of 40x30 pixels is performed (i.e., $\times 16$ factor). Some examples of LR thermal images used as input to detect objects are shown in Fig. 3 (*left col.*). These images are the starting point for the super-resolution techniques used in the current work.

Regarding the guided super-resolution approach, in the current work, a model for super-resolving a LR thermal image with the aid of an HR RGB image is considered. Given that the imaging pipelines used to acquire RGB and thermal images differ, the proposed method employs a two-stream network to enhance the LR thermal image. The architecture proposed by the winning team in **Track 2** of the **Thermal Image Super-resolution Challenge** - PBVS2023 [4], within the context of CVPRW, serves as inspiration. The network begins by inputting the RGB and bicubic up-sampled LR images into a layer responsible for extracting shallow features from each image type. Subsequently, these extracted features are concatenated and transmitted through feature fusion layers, which are specifically designed to combine multi-modal data. The feature fusion layers employ cascaded NAF Blocks [23]. Finally, the super-resolution output is reconstructed using a HR image reconstruction layer. Figure 1 illustrates the network architecture.

To enhance the resolution of the LR thermal image, the proposed model leverages the complementary data provided by the RGB and thermal images. By integrating the shallow feature extraction, feature fusion, and HR image reconstruction procedures, the model effectively improves the visual quality and details of the super-resolved thermal image. The next section provides a detailed analysis and evaluation of this approach's experimental results.

Finally, the object detection task in super-resolver thermal images is performed by using the YOLO v8 [26] framework, the model *YOLOv8x*, and the pre-trained weights *yolov8x.pt* are used, in addition, a fine-tunning is applied using thermal images of the M^{3}FD dataset described in Section II-C. The code and the pre-trained weights are obtained from the YOLO official page[1]. YOLOv8 framework is employed due its proven real-time processing capabilities and high accuracy in object detection tasks. YOLOv8 consistently outperformed them in terms of speed and precision, especially when dealing with the intricacies of thermal images. Its architecture, which formulates object detection as a regression problem, proved to be particularly effective for the dataset. It should be mentioned that the proposed strategy is also valid if any other object detection model is used.

IV. Experimental Results

This section presents the results of the proposed approach, objects detected in SR thermal images, and compares them with results obtained when HR and LR thermal images are considered. The effectiveness of each component in the proposed work is illustrated, providing insights into the contribution of each module to the overall performance. Figure 2 depicts the process of comparing the number of detected objects, the Intersection over Union (IoU) metric between the predicted bounding box and ground truth bounding box when the overlapping area is $\geq 50\%$ (IoU_{50} metric) is considered.

As presented in Table I, the total number of objects in the GT images is 993. In contrast, the results in super-resolved thermal images achieve a recognition of 405 objects, which represents a significant improvement when compared to the lack of recognition in the LR images. It should be mentioned that these 405 detected objects have, on average, an $IoU_{50} = 0.8176$, which is an excellent value in object detection tasks—0 means completely fails and 1 a perfect detection. Another value used to evaluate the quality of detection is the Precision of the detection task in this case an average value of $Acc. = 0.6819$ is obtained. While in the Mean Average Precision value, an average of $mAP = 0.355$ is obtained. The values displayed on the three evaluations correspond to the average computed on the testing set (100 images) taking into account all the categories and showing the relevance of the proposed solution.

The proposed super-resolution approach demonstrates notable enhancements in object recognition performance when compared to the LR images.

Qualitative results are depicted in Fig. 3, which shows a comparison between the annotated objects in the ground truth, and those detected in the bicubic interpolation and

[1]https://docs.ultralytics.com/models/yolov8/

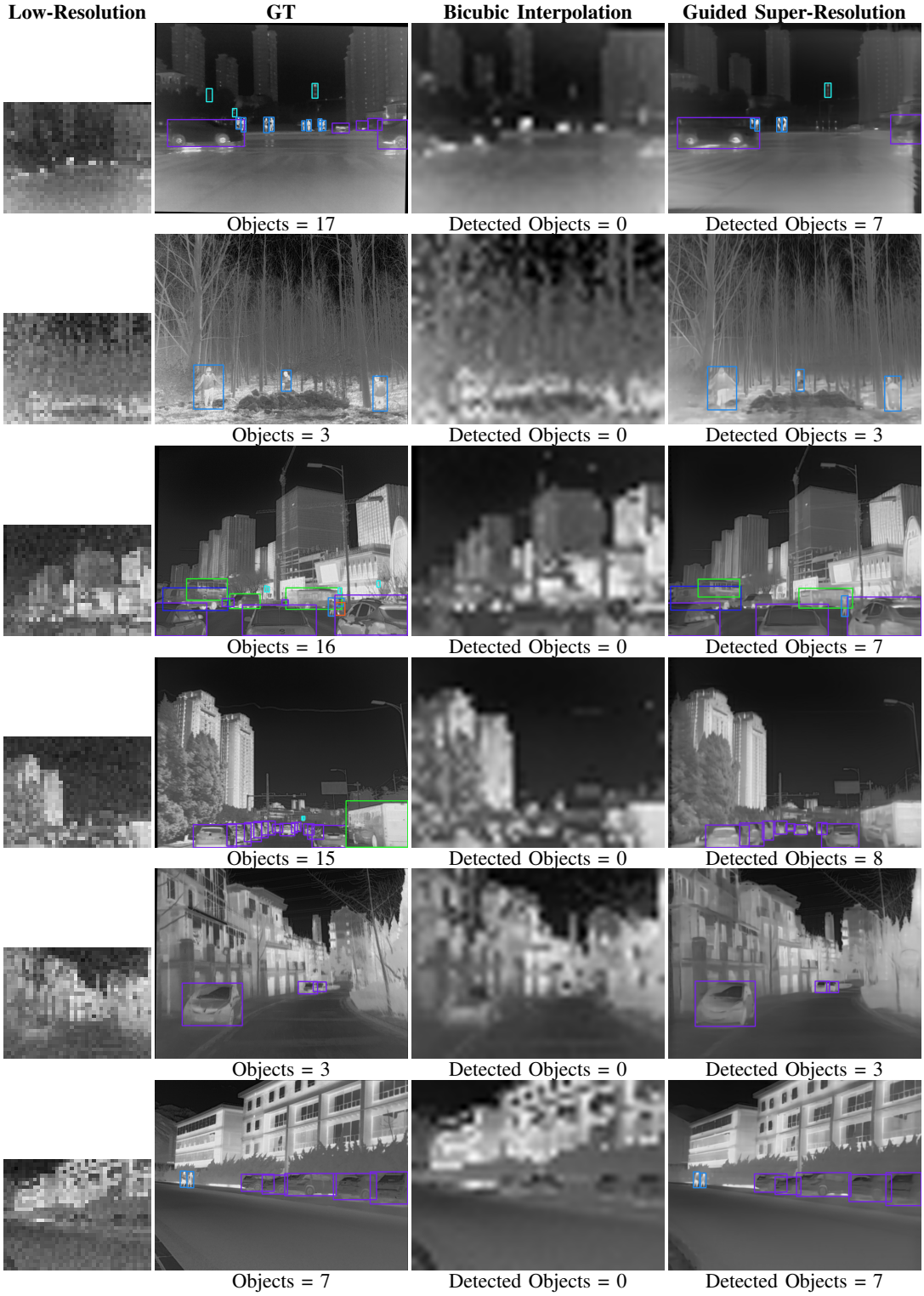

Fig. 3. Detected objects in LR thermal images (testing set) when a $\times 16$ SR factor is considered.

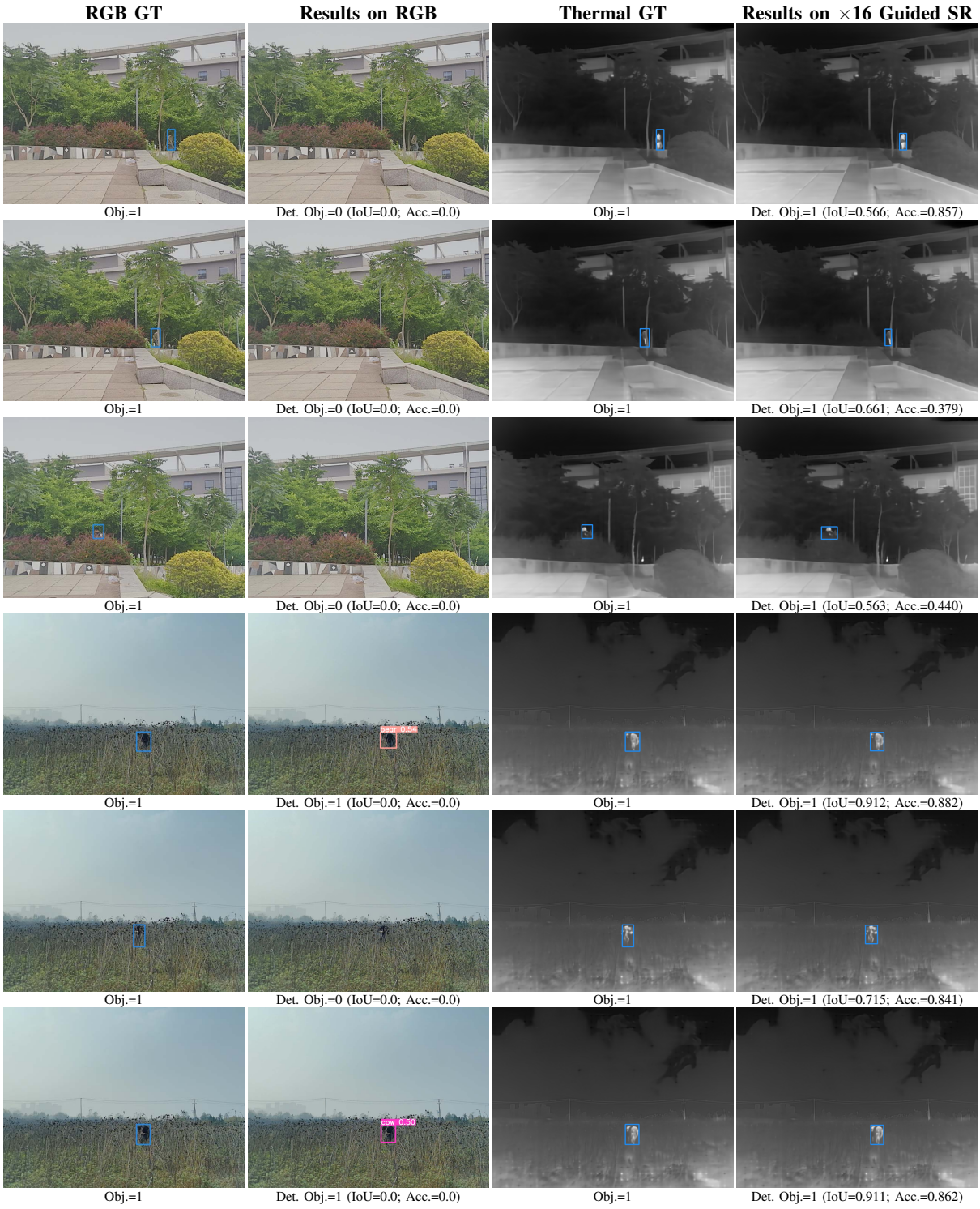

Fig. 4. Comparative results between visible (RGB) and thermal images in an object detection tasks in camouflaged environments—note that in the right column (thermal image from the guided super-resolution approach) all pedestrians are correctly detected. These example images are part of the testing set.

TABLE I
COMPARATIVE RESULT OF DETECTED OBJECTS ON THERMAL IMAGES. THE VALUES $Acc.$, IoU_{50} AND mAP ARE THE AVERAGE VALUES COMPUTED ON THE TESTING SET OF **M^{3}FD DATASET** ON EACH CLASS OF THE SR RESULTS.

Class	GT	Bic LR Results	SR Results			
		# of detect. obj.	# of detect. obj.	$Acc.$	IoU_{50}	mAP
People	111	0	33	0.692	0.753	0.351
Car	662	0	335	0.685	0.820	0.659
Bus	33	0	12	0.804	0.879	0.396
Motorcycle	18	0	1	0.267	0.794	0.315
Lamp	115	0	12	0.587	0.823	0.117
Truck	58	0	12	0.573	0.852	0.293
Total	997	0	405	0.6819	0.8176	0.355

SR-resolved image respectively. The thermal LR images are shown and the improvement in quality performance can be observed with the use of the guided super-resolution technique and the benefit when performing the object detection task in different scenes compared with the classical bicubic interpolation technique. Figure 4 provides illustrations of visible spectrum images and the corresponding SR thermal images. These visualizations further validate the effectiveness of the proposed super-resolution approach in achieving better object recognition results in conditions where state-of-the-art YOLO v8 is not capable to detect or wrongly detect objects. One noteworthy application is the detection of camouflaged objects, as illustrated in the six images in Fig. 4, where individuals hidden in vegetation are either not correctly detected or remain undetected.

V. CONCLUSIONS

To overcome the challenge of LR thermal image-based object recognition, this research presents a novel strategy that combines a SR method with deep learning-based object recognition. The experimental results demonstrate the effectiveness and superiority of the proposed strategy, showcasing significant improvements in object identification precision and quality images. By leveraging the power of super-resolution algorithms and deep learning models, the resolution of thermal images is enhanced, enabling the recovery of fine features and accurate object identification. The outcomes of this study contribute to the field of computer vision by expanding the potential applications of thermal imaging technology and providing new opportunities for enhanced object detection in thermal images. These findings hold practical implications for various industries, as well as for sectors such as security and public safety. Future research can delve deeper into advancing super-resolution techniques and exploring novel deep-learning architectures for more sophisticated thermal image analysis.

Finally, it can be mentioned that the proposed methodology is not without challenges. One of the most important constraints of the present work is the need to have the perfectly registered pair of thermal-visible images available to perform the super-resolution technique. Also, the advantages of the method include its ability to handle very low-resolution thermal images and its integration with the YOLO network for enhanced object detection.

ACKNOWLEDGEMENTS

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-22-1-0261; and partially supported by the Grant PID2021-128945NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe"; the "CERCA Programme / Generalitat de Catalunya"; and the ESPOL project CIDIS-12-2022.

REFERENCES

[1] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*, pp. 391–407, Springer, 2016.

[2] R. E. Rivadeneira, P. L. Suárez, A. D. Sappa, and B. X. Vintimilla, "Thermal image superresolution through deep convolutional neural network," in *International Conference on Image Analysis and Recognition*, pp. 417–426, Springer, 2019.

[3] A. Mehri, P. B. Ardakani, and A. D. Sappa, "Mprnet: Multi-path residual network for lightweight image super resolution," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2704–2713, 2021.

[4] R. E. Rivadeneira, A. D. Sappa, B. X. Vintimilla, D. Bin, L. Ruodi, L. Shengye, Z. Zhong, X. Liu, J. Jiang, and C. Wang, "Thermal image super-resolution challenge results-pbvs 2023," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, June 2023.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[6] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016.

[7] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3929–3938, 2017.

[8] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136–144, 2017.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[10] R. Rivadeneira, A. Sappa, B. Vintimilla, L. Guo, J. Hou, A. Mehri, P. Ardakani, H. Patel, V. Chudasama, K. Prajapati, *et al.*, "Thermal image superresolution challenge," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 432–439, 2020.

[11] R. E. Rivadeneira, A. D. Sappa, and B. X. Vintimilla, "Thermal image super-resolution: A novel architecture and dataset.," in *VISIGRAPP (4: VISAPP)*, pp. 111–119, 2020.

[12] R. E. Rivadeneira, A. D. Sappa, and B. X. Vintimilla, "Thermal image super-resolution: A novel unsupervised approach," in *International Joint Conference on Computer Vision, Imaging and Computer Graphics*, pp. 495–506, Springer, 2020.

[13] H. Chang, J. Lu, F. Yu, and A. Finkelstein, "Pairedcyclegan: Asymmetric style transfer for applying and removing makeup," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 40–48, 2018.

[14] R. E. Rivadeneira, A. D. Sappa, B. X. Vintimilla, S. Nathan, P. Kansal, A. Mehri, P. B. Ardakani, A. Dalal, A. Akula, D. Sharma, *et al.*, "Thermal image super-resolution challenge," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 4359–4367, 2021.

[15] R. E. Rivadeneira, A. D. Sappa, B. X. Vintimilla, J. Kim, D. Kim, Z. Li, Y. Jian, B. Yan, L. Cao, F. Qi, *et al.*, "Thermal image super-resolution challenge results," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 418–426, 2022.

[16] H. Gupta and K. Mitra, "Toward unaligned guided thermal super-resolution," *IEEE Transactions on Image Processing*, vol. 31, pp. 433–445, 2021.

[17] R. d. Lutio, S. D'aronco, J. D. Wegner, and K. Schindler, "Guided super-resolution as pixel-to-pixel transformation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 8829–8837, 2019.

[18] H. Gupta and K. Mitra, "Pyramidal edge-maps and attention based guided thermal super-resolution," in *Proceedings of the European Conference on Computer Vision Workshops*, pp. 698–715, Springer, 2020.

[19] Z. Zhao, Y. Zhang, C. Li, Y. Xiao, and J. Tang, "Thermal uav image super-resolution guided by multiple visible cues," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[20] L. Cheng and M. Kersemans, "Dual-IRT-GAN: A defect-aware deep adversarial network to perform super-resolution tasks in infrared thermographic inspection," *Composites Part B: Engineering*, vol. 247, 2022.

[21] K. Prajapati, V. Chudasama, H. Patel, A. Sarvaiya, K. P. Upla, K. Raja, R. Ramachandra, and C. Busch, "Channel split convolutional neural network (chasnet) for thermal image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4368–4377, 2021.

[22] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1833–1844, 2021.

[23] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII*, pp. 17–33, Springer, 2022.

[24] H. Su, V. Jampani, D. Sun, O. Gallo, E. Learned-Miller, and J. Kautz, "Pixel-adaptive convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11166–11175, 2019.

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

[26] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023.

[27] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

[28] W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian detection based on yolo network model," in *2018 IEEE international conference on mechatronics and automation (ICMA)*, pp. 1547–1551, IEEE, 2018.

[29] W.-Y. Hsu and W.-Y. Lin, "Ratio-and-scale-aware yolo for pedestrian detection," *IEEE transactions on image processing*, vol. 30, pp. 934–947, 2020.

[30] J. Lu, C. Ma, L. Li, X. Xing, Y. Zhang, Z. Wang, and J. Xu, "A vehicle detection method for aerial image based on yolo," *Journal of Computer and Communications*, vol. 6, no. 11, pp. 98–107, 2018.

[31] Y. Zhang, Z. Guo, J. Wu, Y. Tian, H. Tang, and X. Guo, "Real-time vehicle detection based on improved yolo v5," *Sustainability*, vol. 14, no. 19, p. 12274, 2022.

[32] H. O. Velesaca, S. Araujo, P. L. Suárez, A. Sánchez, and A. D. Sappa, "Off-the-shelf based system for urban environment video analytics," in *Int. Conf. on Systems, Signals and Image Processing*, pp. 459–464, IEEE, 2020.

[33] J. L. Charco, A. D. Sappa, B. Vintimilla, and H. O. Velesaca, "Camera pose estimation in multi-view environments: From virtual scenarios to the real world," *Image and Vision Computing*, vol. 110, p. 104182, 2021.

[34] H. O. Velesaca, J. Vulgarin, and B. X. Vintimilla, "Deep learning-based human height estimation from a stereo vision system," in *2023 IEEE 13th International Conference on Pattern Recognition Systems (ICPRS)*, pp. 1–7, 2023.

[35] J. L. Charco, A. D. Sappa, B. X. Vintimilla, and H. O. Velesaca, "Transfer learning from synthetic data in the camera pose estimation problem.," in *VISIGRAPP*, pp. 498–505, 2020.

[36] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1*, vol. 1, pp. 364–369, IEEE, 2005.

[37] D. Olmeda, C. Premebida, U. Nunes, J. M. Armingol, and A. de la Escalera, "Pedestrian detection in far infrared images," *Integrated Computer-Aided Engineering*, vol. 20, no. 4, pp. 347–360, 2013.

[38] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1037–1045, 2015.

[39] Z. Wu, N. Fuller, D. Theriault, and M. Betke, "A thermal infrared video benchmark for visual analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 201–208, 2014.

[40] J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, and Z. Luo, "Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5802–5811, 2022.