Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 225 (2023) 2234–2243

Procedia Computer Science

www.elsevier.com/locate/procedia

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

Object Detection and Localisation in Thermal Images by means of UAV/ Drone

Fabio Martinelli[a], Francesco Mercaldo[b,a,*], Antonella Santone[b]

[a]*Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy*
[b]*University of Molise, Campobasso, Italy*

Abstract

Object detection is one of the crucial tasks that has made deep learning of fundamental importance in last years, also thanks to the use of drones and unmanned aerial vehicles able to obtain images and videos in real-time from any location. In the absence of daylight or artificial light for object detection, it is necessary to resort to thermal images, obtained by converting infrared radiation (i.e., heat) into visible images that depict the spatial distribution of temperature differences in a scene viewed by a thermal camera. However, object detection in this type of image and video stream is still challenging due to the complicated scene information and coarse resolution compared to a visible image or video. In this paper, we propose a method aimed to detect objects in thermal images, in particular, the proposed method is aimed to identify persons and dogs acquired from a thermal camera installed, for instance, on drones or unmanned aerial vehicles. We employ an object detection model, i.e., the "You only look once" one, for the automatic localization of objects in thermal images. In the evaluation of a dataset composed of 203 images with 257 annotations, the proposed method obtains a precision of 0.897, a recall equal to 0.904, and a mean Average Precision value (with an Intersection over Union greater than 0.5) equal to 0.924, showing the effectiveness of the proposed method for the identification and location of persons and dogs from images directly acquired with thermal cameras.

Keywords: object detection, deep learning, thermal images, drone, UAV, YOLO

1. Introduction and Related Work

Object detection is a common task in computer vision: its main purpose is to classify and locate a specific object in an image. Recently, object detection has been performed on spaceborne, aerial, and ground remote sensing images [33, 17, 6]. These models are widely adopted to offer numerous optically labelled datasets but relatively few thermal infrared datasets for multiple objects at the ground level. Drones and unmanned aerial vehicles (UAV) can acquire

* Corresponding author: Francesco Mercaldo
E-mail address: francesco.mercaldo@unimol.it, francesco.mercaldo@iit.cnr.it

1877-0509
10.1016/j.procs.2023.10.214

high tempera-spatial resolution images and videos, which makes up for the shortcomings of remote sensing satellites that are unavailability of high-resolution thermal images due to the limitations of satellite sensors. Compared with optical sensors, thermal infrared sensors (TIR) can capture images in both day and night scenes. However, current object detection has not been extensively conducted on UAV TIR images and videos. Some efforts have been made on ground thermal infrared pedestrian detection [8, 18, 34] by using computer vision and deep learning. Ships [19], vehicles [21, 20, 2, 22], and thermal bridges in buildings [9], have also been explored. For instance, Khalifa et al. [16] surveyed different application and surveillance systems with embedded devices to detect human presence, while authors in [12] proposed a method to detect vehicles by using TIR images.

Overall, object detection systems with drones and UAVs are promising and growing technologies with many application scenarios not only in computer vision and deep learning but also in other areas, such as surveillance cases, human detection in search and rescue [15, 28, 31]. Although UAV TIR remote sensing has been applied in the above fields, object detection from UAV TIR images and videos still encounters numerous difficulties because of the complex image background, low resolution, long imaging distance, flight angles, and TIR detection for multiple scenes and objects [13].

Researchers have explored artificial intelligence for object detection through computer vision, with particular regard by resorting to machine learning [1, 22] and deep learning [2, 21, 3, 26, 5] generally considering images and videos with normal daylight conditions.

Starting from these considerations, in this paper, we propose a deep learning model aimed to automatically detect and localise objects in thermal images, with particular regard to humans and dogs (but the proposed method can be applied also to other kinds of objects).

We focus on thermal images considering their wide array of applications, for instance, monitoring machine performance, and seeing in low light conditions. As a matter of fact, infrared imaging is useful in security, wildlife detection, and hunting/outdoor recreation.

The paper proceeds as follows: in the next section the proposed method is presented, the experimental analysis is presented in Section 3, and, in the last section, the conclusion and future research lines are drawn.

2. Object Detection from Thermal Images

In this section, the proposed method aimed to detect and localize objects in thermal images, with particular regard to humans and dogs, is presented.

In detail, we propose a method aimed to automatically detect objects directly from thermal images, taken for instance from UAV and drones, by adopting deep learning techniques.

Figure 1 shows the proposed approach: for each detected object the proposed method shows also the prediction percentage of the detection (a measure of the model confidence in how confident it is of a given prediction).

In order to build a deep learning model that is effective in object detection from thermal images, we need a dataset composed of thermal images, acquired for instance by drone/UAV (i.e., images obtained from different distances). To build a model aimed able to identify not only the presence of objects in the image but also to detect where the object is in the image, we need a set of images with the detail of the position of the object in the image: for this reason, the images captured by thermal imaging camera are manually labeled and annotated with the aim to mark the area(s) of the images under analysis where the objects are present (i.e., *bounding box* in Figure 1). The classes (i.e., *classes* in Figure 1) for the detection of the bounding box are people and dogs (we contextualize the proposed model to the detection of these two classes, but can be applied also to other detection contexts).

Moreover, to build a model efficient but also able to correctly predict unseen images, the images have been taken from different angles, in different conditions. All of the images should have different sizes, but once obtained the images, we need to perform a preprocessing step aimed to resize the images to the same dimension.

The next step is devoted to increment the number of images (i.e., the *image augmentation* in Figure 1): to perform this task we consider a set of techniques that expand the available dataset without actually collecting new elements: data augmentation applies controlled random changes to existing images, creating modified copies. It is used for the automatic learning of artificial neural networks, which ”learn” more and more precisely as the available training dataset increases.

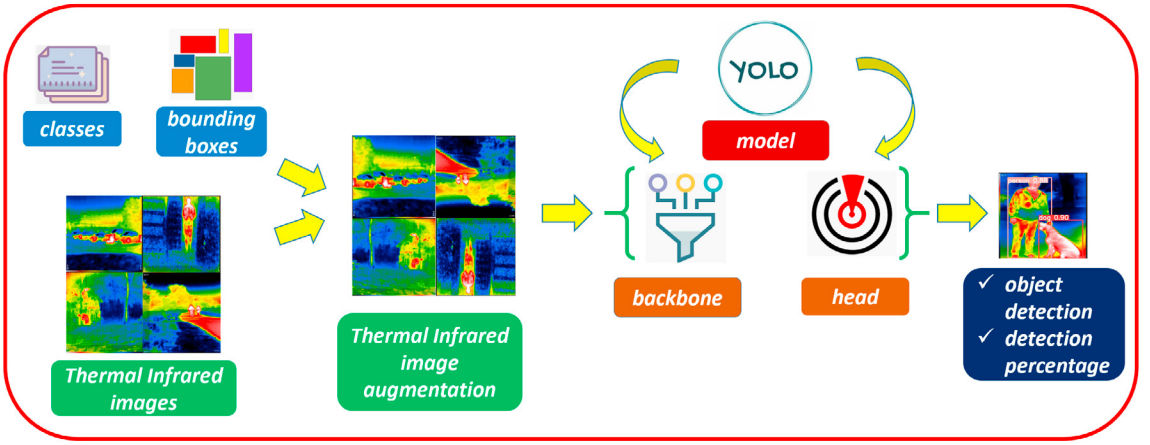

Fig. 1. The proposed method.

In particular, we apply data augmentation to generate thermal images with controlled random changes such as rotations, flips, cuts, and trims [32]. The idea behind the data augmentation application in this case is to make the model suitable for carrying out an effective reconnaissance regardless of the position in the image where the humans and dogs are present. Moreover, augmented data are used to solve the problem of overfitting, the overfitting of the statistical model to the observed data sample, which occurs when the model has too many parameters compared to the number of observations performed. Structured to recognize recurring patterns starting from the proposed data, the artificial neural network learns from the images that sees, but it cannot find a generalized rule, so it easily mistakes patterns not yet viewed: to avoid this aspect we resort to data augmentation.

Once obtained the (augmented) images acquired by the thermal imaging camera, with the related details about the (person and dog) classes and the related bounding boxes, we need a deep learning model (i.e. *Object Detection model* in Figure 1).

In this paper, we resort to the ”You only look once”(i.e. YOLO) model. YOLO [27] was proposed by J. Redmon et al. in 2016: it represents the first one-stage deep learning detector. YOLO is an object detection model, it is aimed to classify the images and to detect their correct positioning within them [24].

The main difference from other models is that YOLO considers a pipeline to carry out the whole process in a completely independent way [23], where each image under analysis is divided into a matrix of SxS cells, where one cell is responsible for the object if it falls in the center of the cell itself [25, 14].

Compared to pre-existing object detection models, YOLO is significantly faster, as demonstrated in [30, 29].

This is mainly possible due to the fact that YOLO does not divide the recognition into several phases, but predicts bounding boxes, probability, and classes of objects present in the input image in a single phase.

The reason why we resort to this model is that compared to other object detection deep learning models, considering that we are aware that YOLO makes more localization errors [11], at the same time is less likely to recognize false positives in the background of the image, as well as being significantly faster [10, 14]: this is the reason why YOLO is considered one of the best convolutional neural network models for object detection. There are several versions of the YOLO model, in this paper we implement the model with the PyTorch[1] framework the YOLOv5s[2] version of the YOLO model.

The YOLO network consists of a backbone i.e., a convolutional neural network that aggregates and forms image features at different granularities, and a Head, aimed to consume features from the neck and take a box and class prediction steps. Between the backbone and the head, there is the neck i.e., a series of layers to mix and combine image features to pass them forward to prediction.

[1] https://pytorch.org/
[2] https://github.com/ultralytics/yolov5

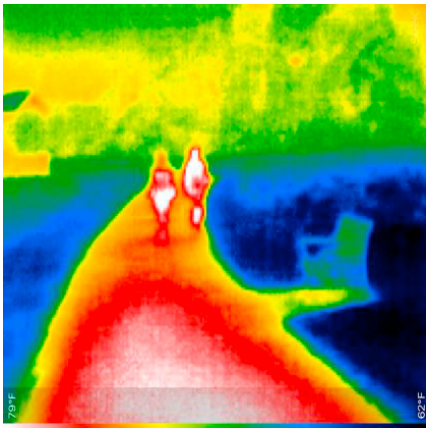

Fig. 2. An example of a thermal image with two persons belonging to the analysed dataset.

3. Experimental Analysis

In this section, we present the results of the experimental analysis aimed to show the effectiveness of the proposed YOLO model for object detection using thermal images.

We gathered images from the Thermal Dogs and People dataset, a dataset aimed to build models to detect people and dogs from thermal images. The dataset is freely available for research purposes[3].

The Thermal Dogs and People dataset is composed by 203 images with 257 annotations (1.3 per image (average), across two classes: person (140 annotations) and dog (117 annotations) We consider the adoption of a public dataset for results replicability. The thermal infrared images were captured at various distances from people and dogs in a park and near home. Some images are deliberately unannotated as they do not contain a person or dog. Images were captured in both portrait and landscape.

The thermal images were captured using the Seek Compact XR Extra Range Thermal Imaging Camera for iPhone[4]. The selected color palette is Spectra.

The thermal images are stored in the jpg file format with a 416 x 416 resolution. We split the images in the following way: 142 images for training, 41 for validation, and the remaining 20 for testing.

Figure 2 shows an example of a thermal image considered in the analysed dataset: it is possible to note the presence of two persons in the central part of the image.

The dataset we obtained is annotated i.e., each image has the detail about the bounding box around each person and dog.

We performed the image augmentation by exploiting the Roboflow web application[5] by randomly rotating the pictures 90°clockwise, 90°counterclockwise, and upside down.

After the application of the data augmentation, we obtained the final dataset. All the images exhibit a dimension equal to 416x416. Relating to the model parameters, a batch size equal to 16 is considered and we set the epoch number equal to 500 and an initial learning rate equal to 0.01.

For the model training, we exploited a machine equipped with an NVIDIA Tesla T4 GPU card with 16 GB of memory.

The model training employed 0.324 hours (this value is related to the wall clock time). This time is referred to 500 epochs.

Figures 3 show the experimental results obtained by the proposed method in several plots. In the first line of plots in Figure 3 there are: the train/box_loss (i.e., the box_loss trend during the training: a loss that measures how ”tight” the

[3] https://public.roboflow.com/object-detection/thermal-dogs-and-people
[4] https://www.thermal.com/compact-series.html
[5] https://roboflow.com/

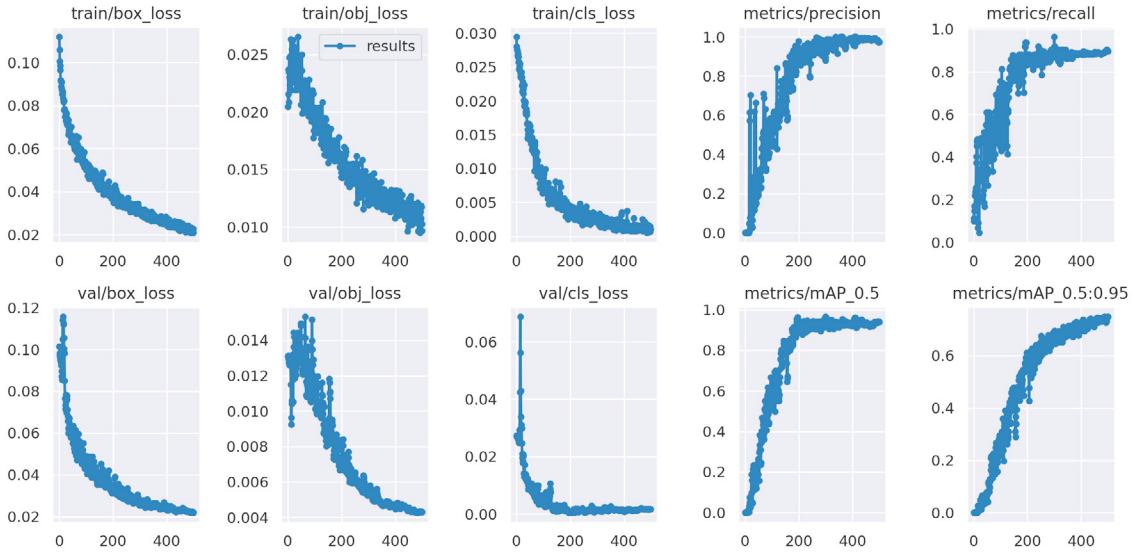

Fig. 3. Experimental analysis results.

predicted bounding boxes are to the ground truth object), the train/obj_loss (i.e., the obj_loss trend during the training: the objectness determines whether an object exists at an anchor), the train/cls_loss trend (i.e., the cls_loss trend during the training: the cls_loss is a loss that measures the correctness of the classification of each predicted bounding box; each box may contain an object class or a "background". This loss is usually called cross-entropy loss), the precision trend, and the recall trend. In the second line of plots in Figure 3 there are: the val/box_loss (i.e., the trend of the box_loss in validation), the val/obj_loss (i.e., the trend of the obj_loss in validation), the mean Average Precision when Intersection over Union is equal to 0.5 (mAP_0.5) and when the Intersection over Union is between 0.5 and 0.95 (mAP_0.5:0.95).

All the metrics exhibit the expected trends: as a matter of fact precision, recall, mAP_0.5, and mAP_0.5:0.95 should exhibit a trend that grows with the increase of the epochs, symptomatic that the model is correctly learning to detect objects from thermal images. The other metrics exhibit an opposite trend i.e., show a decreasing trend when the epoch number is increasing and this is the second confirmation that the model is correctly learning from thermal images. In fact, the loss metrics indicate, generically, when a model is wrong in recognizing a specific object, therefore the loss values are usually very high in the very first epochs but with the increase of the epochs they are expected to decrease as the model learns to detect objects of interest.

In the following, more details are provided about the precision, recall, mAP_0.5, and mAP_0.5:0.95 metrics.

Precision gives the proportion of positive predictions that are actually correct. It takes into account false positives, which are cases that were incorrectly flagged for inclusion. Precision can be computed as:

$Precision = \frac{TP}{TP+FP}$

From the precision trend in Figure 3 we can note that this trend is growing in the various epochs until it reaches stability (approximately from the 250th epoch): this trend is symptomatic that the network, during the various eras, is correctly learning to distinguish between the presence and the absence of persons and dogs in thermal images.

The second metric we compute to evaluate the effectiveness of the proposed method is recall: the idea behind this metric is to compute the proportion of actual positives that were predicted correctly. It takes into account false negatives, which are cases that should have been flagged for inclusion but were not. Recall can be computed as:

$Recall = \frac{TP}{TP+FN}$

From the recall trend, shown in Figure 3 we can note a behaviour similar to the one we highlight for the recall: as a matter of fact precision and recall as the number of epochs increases should show an increasing trend. In fact, this trend is shown in the Figure 3 plots for both metrics and, considering that precision and recall vary from 0 to 1, promising performances are achieved. Similar to precision, recall also has a growing trend with increasing epochs.

Precision and recall are metrics typically exploited to evaluate the effectiveness of a model in classification tasks: clearly, to evaluate whether the model is able to localise the object of interest from thermal images in the right part of the image under analysis are metrics are needed, one of these is the AP (Average precision, a classic metric devoted to measure the accuracy of object detectors (as, for instance, the YOLO one we exploited). Average precision computes the average precision value for recall value over 0 to 1.

We are interested in the computation of mean Average Precision (mAP), which requires Intersection over Union (IOU), Precision, Recall, Precision-Recall Curve, and AP. Object detection models predict the bounding box and category of objects in an image. IOU is used to determine if the bounding box was correctly predicted.

The IOU indicates how much bounding boxes overlap. This ratio of overlap between the regions of two bounding boxes becomes 1.0 in the case of an exact match and 0.0 if there is no overlap. The IOU formula is shown below:

$IOU = \frac{ao}{au}$

where *ai* is the area of overlap and *au* represents the area of union. In the evaluation of object detection models, it is necessary to define how much overlap of bounding boxes with respect to the ground truth data should be considered as successful recognition. For this purpose, IOUs are used, and mAP_0.5 is the accuracy when IOU=50, i.e., if there is more than 50% overlap, the detection is considered successful. The larger the IOU, the more accurate the bounding box needs to be detected and the more difficult it becomes. For example, the value of mAP_0.75 is lower than the value of mAP_0.5.

The mAP is an average of the AP values, which is a further average of the APs for all classes.

In Figures 3 are shown, respectively in the metrics/mAP_0.5 and the metrics/mAP_0.5:0.95 plots the mAP value for IOU=50 and IOU ranging from 50 and 95 (i.e., this value represents different IoU thresholds from 0.5 to 0.95, with a step size equal to 0.05) on average mAP.).

We observe that the trends for the metrics/mAP_0.5 and the metrics/mAP_0.5:0.95 plots in Figure 3 both show an increasing trend, therefore also with regard to the task of locating the humans and dogs in thermal images, the model is able to learn the point of the image in which to look in order to correctly highlight the object to detect.

Figure 1 shows the values obtained for Precision, Recall, mAP_0.5, and mAP_0.5:0.95 metrics (detailed for the single classes i.e., person and dog, and for both the classes).

Table 1. Classification results.

Class	Image	Labels	Precision	Recall	mAP_0.5	mAP_0.5:0.95
all	41	49	0.973	0.896	0.94	0.749
person	41	22	1	0.94	0.963	0.834
dog	41	27	0.946	0.852	0.916	0.663

From Table 1 we can note that the Precision and the Recall are respectively equal to 0.973 and to 0.896 (for both the classes, shown in the *all* label in the Class row).

Furthermore, to better evaluate the proposed method, in Figure 4 we report the precision and recall values on the Precision-Recall graph.

The trend of this plot is expected to be monotonically decreasing: in fact, there is always a trade-off between precision and recall. Increasing one will decrease the other. Sometimes the precision-recall graph is not always monotonically decreasing due to certain exceptions and/or lack of data but from the plot in Figure 4 we can see that this plot exhibits a decreasing trend for the involved labels. The precision-recall plot shows also the Area Under the Curve (AUC) values related to the involved classes (i.e., dog and person) and the identification of all classes with mmAP_0.5. As previously stated the precision-recall trend is expected to be monotonically decreasing: this behaviour is shown from the precision-recall plot related to all classes with mAP_0.5 (with an AUC equal to 0.940). This value is the media of the AUC value of all considered classes: from the precision-recall graph in Figure 4 we can note that relating to the dog class an AUC equal to 0.963 is obtained, while relating to the person class the value reached is 0.916. Considering that these metrics range from 0 to 1, these values can be considered symptomatic that the proposed model is able to detect persons and dogs in thermal images.

In order to visually validate the proposed method and to confirm its effectiveness in a real-world environment in Figures 5 and 6 we respectively show examples of thermal images with the annotation manually performed (shown

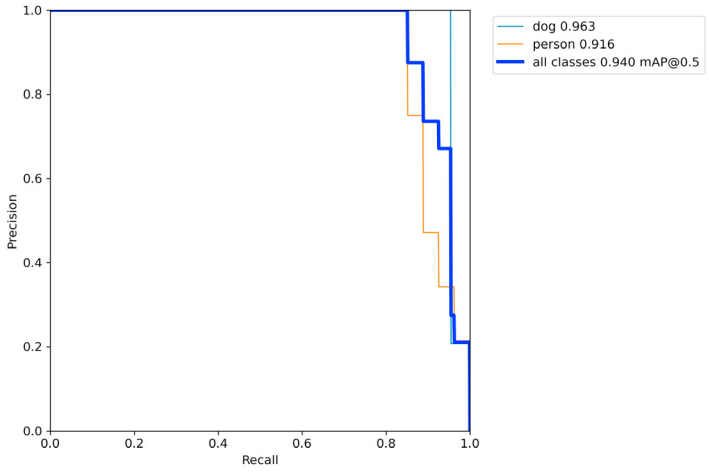

Fig. 4. The Precision-Recall graph.

in Figure 5) and the same images with the detection and the bounding box (shown in Figure 6): in this way we can directly compare the manual annotation with the ones added by the proposed model.

As shown in Figure 5 we consider thermal images obtained from different angulations and with subjects at different distances (in this way we maximize the possibility to consider a model generalizable as possible). We can note that the proposed model is able to detect more objects (i.e., persons and dogs) in the same images and that the color background is not affecting object detection. In Figure 5 for each image there is the detail of the bounding box for the classes involved in the experiment (in red) with the relative name (i.e., dog and person).

In Figure 6 we show the prediction and the bounding boxes added by the proposed model in testing (clearly in testing the images are submitted to the model without the bounding box).

From Figure 6 it emerges that the proposed exhibits a good ability in the right localisation of persons and dogs area: as a matter of fact, for most of the thermal images the bounding boxes can be considered equal to the ones shown in Figure 5.

For example, in the first image on the left (in the first row of images in Figure 6) people were correctly detected with a detection percentage equal to 0.8 (considering that this value varies from 0 to 1, we can state that the model is confident in this prediction). In the penultimate image from the right (in the first row of images shown in Figure 6), we can see that the person, despite being in the distance, is detected with a detection percentage of 0.9, in a similar way to what happens in the first image of the second row of images in Figure 6 (starting from left), where we have a detection percentage of 0.9. In Figure 6 there are also several cases of subjects (persons but also dogs) that occupy a large part of the image, which are taken with high probabilities (for instance the dogs detected in the third rows of images in Figure 6), to demonstrate that the proposed model is capable of recognizing persons and dogs at different distances and in different camera framing (as typically happens when we consider images obtained from drones and UAV that there is no fixed and a priori framing considering that they are vehicles in constant motion).

4. Conclusion and Future Work

Object detection is one of the most important tasks, used for security but also for military purposes. The quality of the images is of fundamental importance for optimal detection, in fact, most of the methods that propose the detection of objects are tested on images with optimal light conditions. Considering the importance of detecting objects in dark conditions, in this paper we proposed a model able to identify objects on thermal images. Thermal images acquired at different distances and in different camera framing are considered, thus making the proposed model able to work from images acquired from UAVs and drones. We resort to the YOLO model for object detection from thermal images: in particular, we focus on the detection of human subjects and dogs (but the proposed method can be extended to detect also other classes of objects), by obtaining a precision and a recall respectively equal to 0.973 and 0.896 for the person

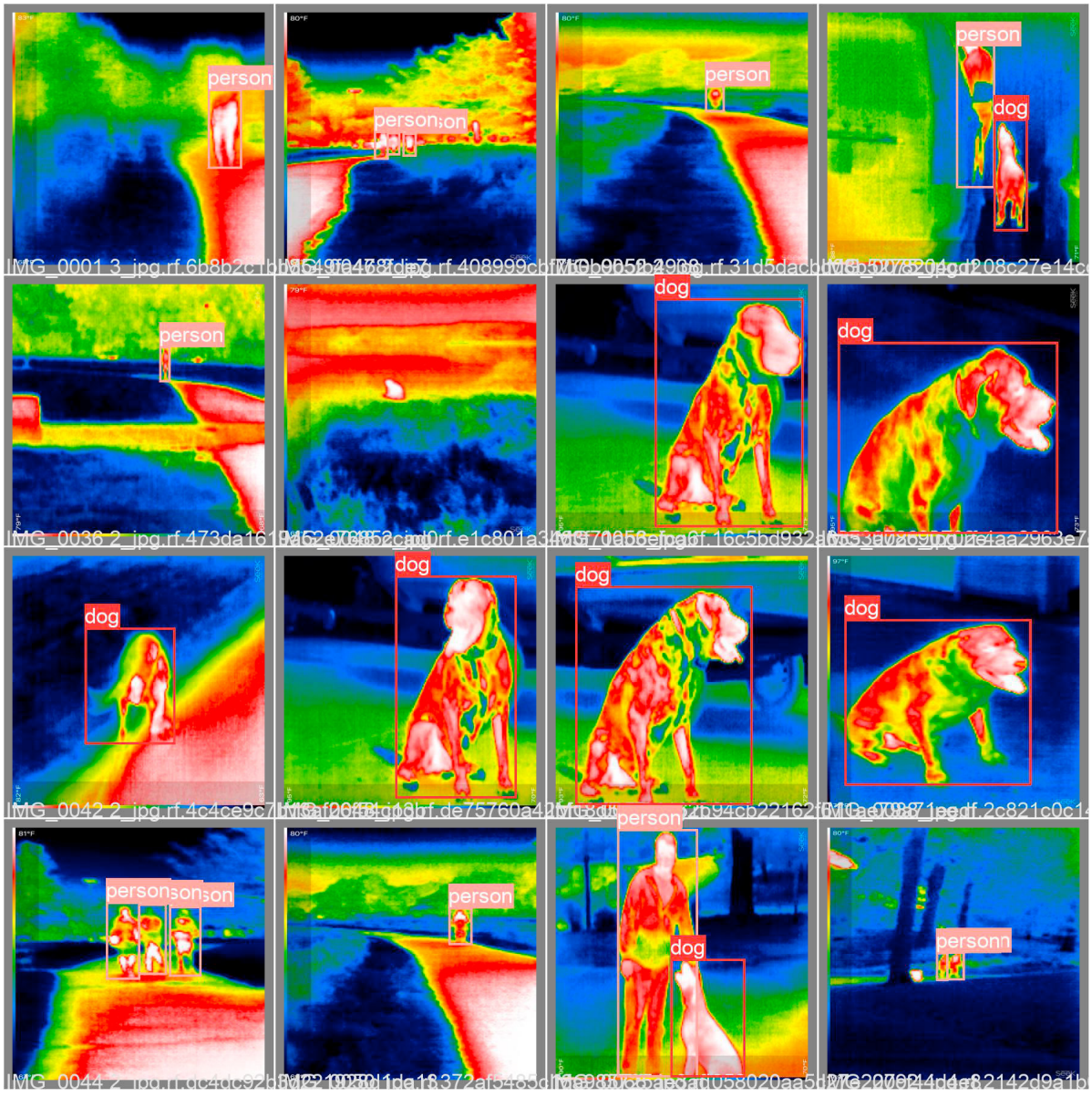

Fig. 5. Example of images with the related bounding box around the persons and dogs, manually added for model building.

and dog detection while, with regard to the localisation, a mAP_0.5 value equal to 0.94 is obtained, thus showing the effectiveness of the proposed model in person and dog detection and localisation from thermal images.

As future research work, we plan to consider other versions of the YOLO model (for instance, the seventh and eighth ones) with the aim to try to obtain better performances but also other object detection models for instance, Single Shot MultiBox and Faster R-CNN. Moreover, we plan to apply model checking, considering that in other research fields (from software security[7] to medical image analysis[4]), demonstrated its effectiveness.

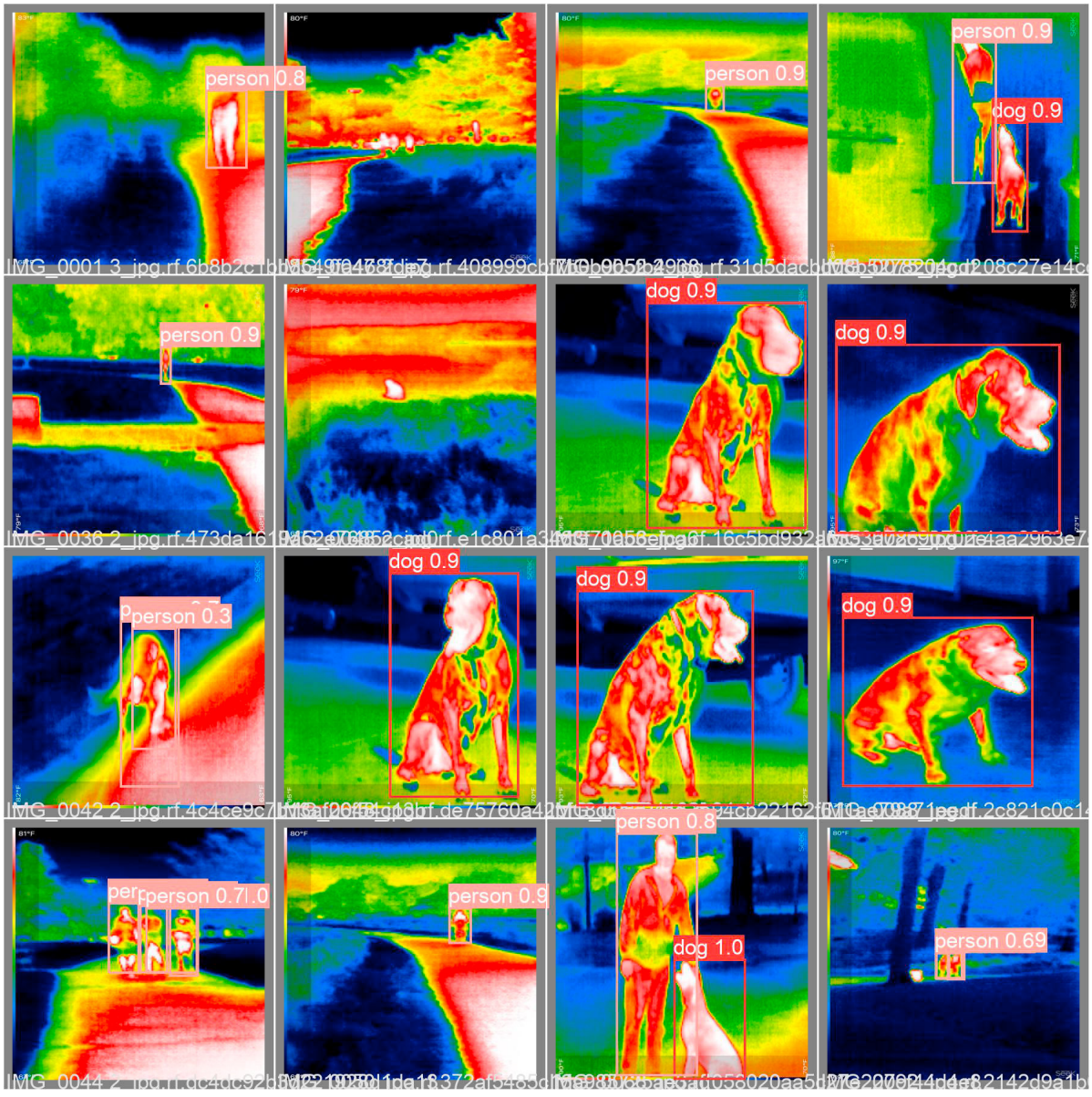

Fig. 6. Example of predictions automatically performed by the proposed method.

Acknowledgment

This work has been partially supported by EU DUCA, EU CyberSecPro, EU E-CORRIDOR, PTR 22-24 P2.01 (Cybersecurity) and SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the EU - NextGenerationEU projects.

References

[1] Ai, D., Jiang, G., Lam, S.K., He, P., Li, C., 2023. Computer vision framework for crack detection of civil infrastructure—a review. Engineering Applications of Artificial Intelligence 117, 105478.

[2] Bernardi, M.L., Cimitile, M., Martinelli, F., Mercaldo, F., 2018. Driver and path detection through time-series classification. Journal of Advanced Transportation 2018.
[3] Brunese, L., Martinelli, F., Mercaldo, F., Santone, A., 2020a. Deep learning for heart disease detection through cardiac sounds. Procedia Computer Science 176, 2202–2211.
[4] Brunese, L., Mercaldo, F., Reginelli, A., Santone, A., 2019. Prostate gleason score detection and cancer treatment through real-time formal verification. IEEE Access 7, 186236–186246.
[5] Brunese, L., Mercaldo, F., Reginelli, A., Santone, A., 2020b. Radiomics for gleason score detection through deep learning. Sensors 20, 5411.
[6] Campanile, L., Iacono, M., Marulli, F., Mastroianni, M., 2020. Privacy regulations challenges on data-centric and iot systems: A case study for smart vehicles., in: IoTBDS, pp. 507–518.
[7] Canfora, G., Martinelli, F., Mercaldo, F., Nardone, V., Santone, A., Visaggio, C.A., 2018. Leila: formal tool for identifying mobile malicious behaviour. IEEE Transactions on Software Engineering 45, 1230–1252.
[8] Chen, Y., Zhang, C., Qiao, T., Xiong, J., Liu, B., 2021. Ship detection in optical sensing images based on yolov5, in: Twelfth International Conference on Graphics and Image Processing (ICGIP 2020), SPIE. pp. 102–106.
[9] Garrido, I., Lagüela, S., Arias, P., Balado, J., 2018. Thermal-based analysis for the automatic detection and characterization of thermal bridges in buildings. Energy and Buildings 158, 1358–1367.
[10] Horak, K., Sablatnig, R., 2019. Deep learning concepts and datasets for image recognition: overview 2019, in: Eleventh international conference on digital image processing (ICDIP 2019), SPIE. pp. 484–491.
[11] Hurtik, P., Molek, V., Hula, J., Vajgl, M., Vlasanek, P., Nejezchleba, T., 2022. Poly-yolo: higher speed, more precise detection and instance segmentation for yolov3. Neural Computing and Applications 34, 8275–8290.
[12] Iwasaki, Y., Misumi, M., Nakamiya, T., 2013. Robust vehicle detection under various environmental conditions using an infrared thermal camera and its application to road traffic flow monitoring. Sensors 13, 7756–7773.
[13] Jiang, C., Ren, H., Ye, X., Zhu, J., Zeng, H., Nan, Y., Sun, M., Ren, X., Huo, H., 2022a. Object detection from uav thermal infrared images and videos using yolo models. International Journal of Applied Earth Observation and Geoinformation 112, 102912.
[14] Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B., 2022b. A review of yolo algorithm developments. Procedia Computer Science 199, 1066–1073.
[15] Kanistras, K., Martins, G., Rutherford, M.J., Valavanis, K.P., 2013. A survey of unmanned aerial vehicles (uavs) for traffic monitoring, in: 2013 International Conference on Unmanned Aircraft Systems (ICUAS), IEEE. pp. 221–234.
[16] Khalifa, A.F., Badr, E., Elmahdy, H.N., 2019. A survey on human detection surveillance systems for raspberry pi. Image and Vision Computing 85, 1–13.
[17] Khanal, S., Fulton, J., Shearer, S., 2017. An overview of current and potential applications of thermal remote sensing in precision agriculture. Computers and Electronics in Agriculture 139, 22–32.
[18] Krišto, M., Ivasic-Kos, M., Pobar, M., 2020. Thermal object detection in difficult weather conditions using yolo. IEEE access 8, 125459–125476.
[19] Leira, F.S., Johansen, T.A., Fossen, T.I., 2015. Automatic detection, classification and tracking of objects in the ocean surface from uavs using a thermal camera, in: 2015 IEEE aerospace conference, IEEE. pp. 1–10.
[20] Martinelli, F., Marulli, F., Mercaldo, F., Santone, A., 2021a. Neural networks for driver behavior analysis. Electronics 10, 342.
[21] Martinelli, F., Mercaldo, F., Nardone, V., Orlando, A., Santone, A., 2018. Who's driving my car? a machine learning based approach to driver identification., in: ICISSP, pp. 367–372.
[22] Martinelli, F., Mercaldo, F., Nardone, V., Santone, A., 2021b. Driver identification through formal methods. IEEE Transactions on Intelligent Transportation Systems 23, 5625–5637.
[23] Martinelli, F., Mercaldo, F., Santone, A., 2022. Smart grid monitoring through deep learning for image-based automatic dial meter reading, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE. pp. 4534–4542.
[24] Martinelli, F., Mercaldo, F., Santone, A., 2023. Water meter reading for smart grid monitoring. Sensors 23, 75.
[25] Mercaldo, F., Martinelli, F., Santone, A., Cesarelli, M., 2022. Blood cells counting and localisation through deep learning object detection, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE. pp. 4400–4409.
[26] Mercaldo, F., Santone, A., 2021. Transfer learning for mobile real-time face mask detection and localization. Journal of the American Medical Informatics Association 28, 1548–1554.
[27] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788.
[28] Rudol, P., Doherty, P., 2008. Human body detection and geolocalization for uav search and rescue missions using color and thermal imagery, in: 2008 IEEE aerospace conference, Ieee. pp. 1–8.
[29] Sah, S., Shringi, A., Ptucha, R., Burry, A.M., Loce, R.P., 2017. Video redaction: a survey and comparison of enabling technologies. Journal of Electronic Imaging 26, 051406.
[30] Sanchez, S., Romero, H., Morales, A., 2020. A review: Comparison of performance metrics of pretrained models for object detection using the tensorflow framework, in: IOP Conference Series: Materials Science and Engineering, IOP Publishing. p. 012024.
[31] Shao, Z., Cheng, G., Ma, J., Wang, Z., Wang, J., Li, D., 2021. Real-time and accurate uav pedestrian detection for social distancing monitoring in covid-19 pandemic. IEEE transactions on multimedia 24, 2069–2083.
[32] Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. Journal of big data 6, 1–48.
[33] Xiang, T.Z., Xia, G.S., Zhang, L., 2019. Mini-unmanned aerial vehicle-based remote sensing: Techniques, applications, and prospects. IEEE geoscience and remote sensing magazine 7, 29–63.
[34] Xu, Z., Zhuang, J., Liu, Q., Zhou, J., Peng, S., 2019. Benchmarking a large-scale fir dataset for on-road pedestrian detection. Infrared Physics & Technology 96, 199–208.