# Robust Temperature Measurement in Dynamic Environments through Detection and Tracking

by R. Usamentiaga[*], P. Venegas[**], S. Sfarra[***], H. Zhang[****]

[*] *Department of Computer Engineering, University of Oviedo, Campus de Viesques, 33204 Gijón, Spain, rusamentiaga@uniovi.es*
[**] *Aeronautical Technology Centre (CTA), Parque Tecnológico de Álava, Juan de la Cierva 1, 01510, Miñano, Spain, pablo.venegas@cta.aero*
[***] *Department of Industrial and Information Engineering and Economics (DIIIE), University of L'Aquila, L'Aquila, I-67100, Italy, stefano.sfarra@univaq.it*
[****] *Department of Electrical and Computer Engineering, Computer Vision and Systems Laboratory (CVSL), Laval University, Quebec G1V 0A6, Canada, hai.zhang.1@ulaval.ca*

**Abstract**

This study presents a novel approach for robust temperature measurement in dynamic environments. By integrating deep learning-based object detection and tracking techniques, the proposed methodology eliminates reliance on unreliable GPS coordinates. An active learning methodology is employed to iteratively enhance detection and tracking performance over time. Testing across diverse scenarios demonstrates the effectiveness of the approach in achieving accurate temperature measurements, even in challenging conditions. Results underscore the potential for real-world applications in dynamic settings. This innovative method offers a promising solution for precise temperature monitoring in industrial and infrastructure inspection scenarios, where traditional approaches may fall short.

## 1. Introduction

Temperature measurement is a fundamental aspect of scientific inquiry, industrial processes, healthcare, environmental monitoring, and countless other domains. The ability to quantify temperature accurately and reliably is crucial for understanding and controlling a wide range of phenomena, from the microscopic scale of molecular interactions to the macroscopic world of climate systems. In the landscape of temperature sensing technologies, a myriad of solutions has been developed to meet diverse application needs. Among these, infrared thermography-based temperature sensors stand out for their distinctive advantages over other sensor types. The non-contact nature of infrared thermography allows for remote temperature measurement, eliminating the need for physical contact with the object or surface. Additionally, infrared sensors offer rapid and real-time monitoring capabilities, covering a wide temperature range with high spatial resolution and providing a visual representation of temperature distributions [1].

Leveraging the principles of non-contact temperature measurement and thermal imaging, infrared thermography plays a pivotal role in quality control and defect detection, offering a dynamic and non-intrusive approach to analyze materials and processes in various industries [2, 3]. This technology proves invaluable in identifying subtle temperature variations that may signal defects or irregularities within manufactured components. The ability to capture detailed thermal images and analyze temperature distributions facilitates the early detection of anomalies, contributing to proactive quality control measures. This has resulted in a wide variety of applications that involve either exploiting naturally occurring temperature differences or introducing an external stimulus to identify defects in materials such as composites [4], metals [5], and ceramics [6]. This capability has proven indispensable in predictive maintenance and condition monitoring for identifying potential equipment failures by monitoring temperature variations in machinery components [7]. Moreover, the advent of technological advancements has opened new frontiers in temperature measurement, allowing for real-time monitoring, remote sensing, and the measurement of temperature in dynamic and challenging environments.

In cases where the infrared camera maintains a stable position in relation to the monitored object, temperature measurement only requires the adjustment of the field of view and focal length. Subsequently, the determination of intensity within a designated region of the image is calculated from the sequence of images, providing the necessary temperature time history. This process incorporates suitable compensation for factors like emissivity, reflected temperature, and other relevant parameters. This information is crucial for understanding and monitoring the thermal behaviour of the object or system under observation over a specific period. This historical record of temperature variations provides valuable insights into the dynamic aspects of the object's thermal performance, enabling the identification of patterns, trends, and anomalies.

In dynamic environments where either the target object or the camera is in motion, temperature measurement becomes inherently more challenging. The introduction of variability through movement poses distinct challenges to achieving precision and reliability in temperature assessment. Changes in the relative positions of the camera and the target object during the measurement process can distort the thermal data, leading to inaccuracies and complicating the analysis. This challenge has become particularly prominent with the surge in the use of unmanned aerial vehicles for temperature monitoring.

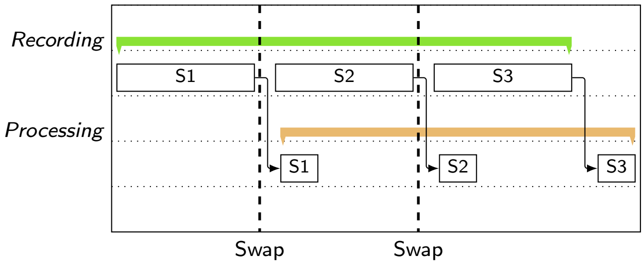



**Fig. 1.** *Recording and processing running in parallel for different sections of a large structure*

In scenarios where aerial surveillance is employed, the inherent mobility of both the drone-mounted camera and the monitored objects introduces complexities that demand specialized approaches [8]. Factors such as altitude fluctuations, varying distances, and rapid changes in perspective require the application of advanced image processing and motion compensation algorithms. Therefore, dynamic scenarios demand sophisticated techniques to account for motion-related challenges. This includes image registration methods to align images captured at different time points, motion correction algorithms to compensate for object or camera movement, and advanced data processing approaches to extract meaningful temperature information despite the dynamic nature of the environment.

This work proposes a robust approach to temperature measurement in dynamic environments combining detection and tracking techniques. Detection leverages deep learning techniques, capitalizing on the sophisticated capabilities of neural networks to detect objects in dynamic environments. Deep learning models, particularly convolutional neural networks (CNNs), excel in extracting intricate patterns and features from complex datasets [9]. In the context of infrared thermography, these models can effectively identify and isolate thermal signatures amidst varying environmental conditions and dynamic scenarios. The versatility of deep learning enables it to learn and adjust to a broad spectrum of temperature patterns, rendering it resilient in identifying objects of interest from infrared images acquired in complex environments. The objects of interest need not be confined solely to the monitored objects; rather, they encompass regions within the image that facilitate the interpretation of its content. These delineated regions play a pivotal role in pinpointing the specific location where temperature measurement is necessitated. Different state-of-the-art models for object detection are considered in this work. These real-time object detection systems are single-stage, end-to-end object detectors that use a convolutional neural network (CNN) to predict bounding boxes and class probabilities directly from full images in one forward pass.

The method proposed for robust temperature measurement in this study undergoes evaluation using a laboratory prototype, where various configurations are tested, incorporating synthetic false positives and missed detections. This prototype serves as a testbed to fine-tune and assess the effectiveness of the proposed approach. Subsequently, the system is tested in an industrial facility with a large pipeline system, necessitating temperature measurement across all pipeline sections. In this scenario, an autonomous aerial vehicle operates without human interaction, conducting fully autonomous flights. This case study also introduces a mission planning proposal, outlining a detailed plan for the execution of the drone flight. Overall, the present study addresses a spectrum of challenges inherent in automated inspection using infrared thermography, encompassing issues related to navigation inaccuracies, weather-induced variability, processing speed, image quality, data management, and object recognition and tracking. Through the seamless integration of detection and tracking techniques, coupled with the advancement of robust measurement methods, the proposed approach markedly expands the capabilities for collecting temperature data. This broadening of scope extends to scenarios that were once deemed inaccessible or impractical, highlighting the comprehensive and transformative nature of the proposed methodology.

## 2. Proposed approach

In the considered approach, the first step entails conducting the recording, followed by data processing. Simultaneously, while one recording is being processed, a new recording can commence. This enables both steps to be seamlessly overlapped in time, enhancing efficiency. For example, considering the inspection of a large structure, such as a large pipeline, three different recordings can be carried out. Fig. 1 shows the recording and processing tasks for the inspection of a long pipeline divided into three sections (S1, S2 and S3). In this approach, multiple tasks (recording and processing) applied to different sections are carried out at the same time. Upon the depletion of the drone's battery, both the battery and storage card are replaced, and the data processing from the previous recording commences. This procedure is repeated until the inspection is completed.

Pipeline inspection is a good example of structure that requires to be inspected for defects or leaks, as they are used to transport hazardous materials, and a leak or failure in the pipeline could have serious consequences for the surrounding environment and community [10]. In addition, regular inspection can help to identify potential issues with the pipeline before they become major problems, such as corrosion, insulation issues, cracks, or other types of damage [11]. Therefore, it can reduce the risk of accidents and disruptions to service. Moreover, regular inspection can help to extend the lifespan of the

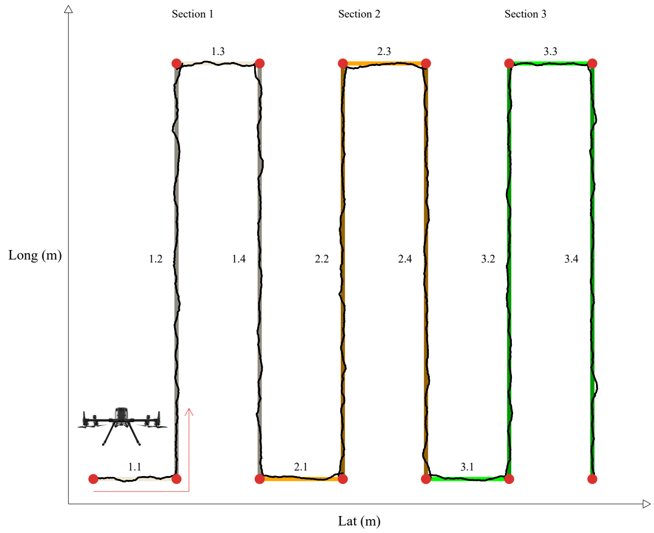



**Fig. 2.** *Sections and segments in the pipeline inspection. The colored thick lines represent the flight plan connecting the waypoints. The thin, dark line depicted in the figure represents the GPS coordinates recorded by the drone during the flight, which include measurement errors and inaccuracies.*

pipeline by identifying and addressing any issues that may cause premature failure. Pipeline inspection can be a reactive process, where getting ahead of an issue can enable a proactive approach to the problem.

In the case of pipeline inspection, the first step is mission planning, i.e., the creation of a detailed plan for the execution of the drone flight. Mission planning for pipeline inspection using drones involves a number of specific considerations to ensure the safe and efficient inspection of the pipeline. The main objective of the mission planning is to create a flight plan that allows the drone to safely and effectively inspect the pipeline while minimizing risks to the drone and the pipeline itself. Some of the key considerations for mission planning for pipeline inspection include: (i) The flight path to provide the best possible view of the pipeline while avoiding obstacles and other hazards, (ii) the altitude that allows the drone to safely inspect the pipeline while also minimizing the risk of collision with other aircraft or obstacles, (iii) the flight time considering the duration of the battery of the drone.

Pipelines are static infrastructures, meaning that they are built and remain in a fixed location. Once a pipeline is constructed, it is typically not moved or repositioned. Therefore, a particular flight plan will remain valid over a long time, and re-planning will not be needed unless there are significant changes to the environment or the pipeline infrastructure. The same would occur for the inspection of most structures.

### 2.1 Infrared video processing

In the video, each frame or image is associated with geographic coordinates by using the GPS device of the drone that records the location during the flight. These GPS coordinates are attached to each frame of the video. However, they are unreliable due to a number of factors, such as low-quality GPS receivers or poor signal reception. Therefore, when the drone moves at high speeds, an image may not be accurately located using the attached GPS coordinates, resulting in a discrepancy between the location where the image was acquired and the GPS coordinates attached to it. This represents a serious issue as the information extracted from a particular image cannot be directly associated with a specific pipe in the pipeline.

Fig. 2 shows an example of pipeline inspection. Due to the pipeline's length, the inspection is divided into three sections, requiring three flights. The resulting video for each section is processed to detect regions of interest and then it is divided into segments, each identified by the section number and its position within that section.

#### *2.1.1 Detection of objects of interest in the infrared images*

Object detection in images is a challenging task due to numerous factors, such as objects being partially or fully occluded, the similarity between objects of interest and the background, variations in lighting and camera angles, and complexity of the objects themselves. Additionally, in infrared images, the task becomes even more difficult due to additional issues such as low contrast and resolution, variations in the spectral characteristics of objects, and complex backgrounds.

The data processing stage encompasses two primary tasks: firstly, the detection of objects of interest within the infrared images, and secondly, the computation of temperatures corresponding to the detected infrared radiation measurements. In the context of pipeline inspection, a key discriminative object of interest is the pipeline joints. These junctions play a crucial role in the integrity and functionality of the pipeline system, making their detection and assessment a priority during inspection processes. The temperature can be measured from these objects or from the pipes connecting consecutive joints. The pipeline joints have distinct shapes and features that make them easier to detect and recognize compared to the pipes themselves, particularly in the infrared spectrum. A pipeline joint is a section of a pipeline that is designed to connect two separate pieces of pipe. Thus by detecting pipeline joints, the position of an individual pipe can also be determined and located, making it an effective approach for object detection.

There are many different types of pipeline joints, and the specific type used will depend on the material of the pipe, the operating conditions of the pipeline, and other factors. Some common types of pipeline joints include flanged joints, which consist of two flanges, one on each end of the pipe, that are connected by bolts; welded joints, which are created by welding the two pieces of pipe together at the joint; and threaded joints, which consist of threads that are cut into the ends of the pipe, which allow the pipes to be screwed together at the joint. Overall, the type of pipeline joint used depends on the specific requirements of the pipeline and the materials being transported through it. In all cases, pipeline joints possess distinct shapes and features that make them more easily recognizable and simpler to detect in comparison to individual pipes.

Different state-of-the-art models for object detection are considered in this work, including YOLOv5 and YOLOv8. These real-time object detection systems are single-stage, end-to-end object detectors that use a convolutional neural network (CNN) to predict bounding boxes and class probabilities directly from full images in one forward pass. YOLOv5 is the continuation of previous YOLO models [12], in this case, implemented using PyTorch. YOLOv5 is a mature object detector with a proven track record of success in a wide range of applications. It has been used in a variety of fields including self-driving cars, surveillance systems, robotics, and more. YOLOv5 improves upon previous versions with a more efficient architecture, improved training techniques, and a better balance between speed and accuracy. YOLOv5 also includes an ecosystem of tools to help make object detection tasks easier and more efficient, including pre-trained models that can be fine-tuned to work on custom datasets. Due to its high performance and ease of use, YOLOv5 is one of the most popular choices for object detection tasks. YOLOv8 is a recent version in the YOLO series for object detection models that introduces new features and improvements to further boost performance and flexibility. It includes a new backbone network, a new anchor-free detection head, and a new loss function. These features enable YOLOv8 to handle a wide range of object detection tasks with high accuracy and fast inference time.

#### *2.1.2 Tracking and identification*

This study relies on object detection for image interpretation; however, deriving temperature measurements solely from the information contained in a single image may introduce inaccuracies and errors. Despite the advancements in state-of-the-art object detection models, they remain susceptible to issues such as false positives and missed detections. In the context of infrared images, the challenges are compounded by low resolution and noise, with outdoor inspections further impacted by potential reflections. Leveraging the capability of infrared cameras to capture images at high frame rates, typically exceeding 30 images per second, this study introduces a tracking procedure to markedly improve the resilience of temperature measurements. Departing from the conventional approach of providing measurements for individual images, the proposed methodology establishes measurement tracks. This allows the tracking procedure to effectively bridge detection gaps and disregard false positives. Consequently, measurements are derived from a compiled list of single measurements calculated while the monitored region remains within the camera's field of view. To ensure robustness and mitigate the impact of outliers, the proposed methodology incorporates the use of robust statistics, such as the median, in the data analysis process. This strategic inclusion enhances the reliability of the obtained measurements. The presented methodology introduces a tracking procedure aligned with the tracking-by-detection paradigm, recognized as the predominant approach in multi-object tracking. In this proposed approach, the object detector is employed on each video frame, followed by the tracking method associating these detections with tracks. This association is facilitated through a motion model that incorporates predictions regarding the future positions of the objects, coupled with semantic information about the environment, including previously known geometric details about the area where the inspection is being performed. Motion prediction is computed utilizing Kalman filtering within the image space, while frame-by-frame data association relies on the Hungarian method, taking into account bounding box overlap. The tracking procedure is crafted to operate effectively in environments characterized by repetitive patterns, such as photovoltaic plants featuring a vast array of symmetrically arranged panels or pipelines exhibiting recurring sections during inspections. These scenarios represent some of the prevalent applications of infrared thermography utilizing unmanned aerial vehicles.

When determining if a tentative pipeline joint can be confirmed, various features of the pipeline are taken into account:

1. Object detectors provide a confidence level, or score, for each detection, indicating the level of confidence that the object has been correctly identified. This score is used in this work to filter out low-confidence detections and to assign significance to the detections in subsequent processing. The first two detections require high confidence to be considered valid pipeline joints. Initially, when there is no prior information about the pipeline, the tracking method requires a high level of confidence in the detection to start tracking the object. This is because the initial detections serve as the starting point for the tracking, and if the detection is incorrect or uncertain, the tracking may quickly become unreliable. As more information is gathered over time, the required level of confidence for detections is lowered.

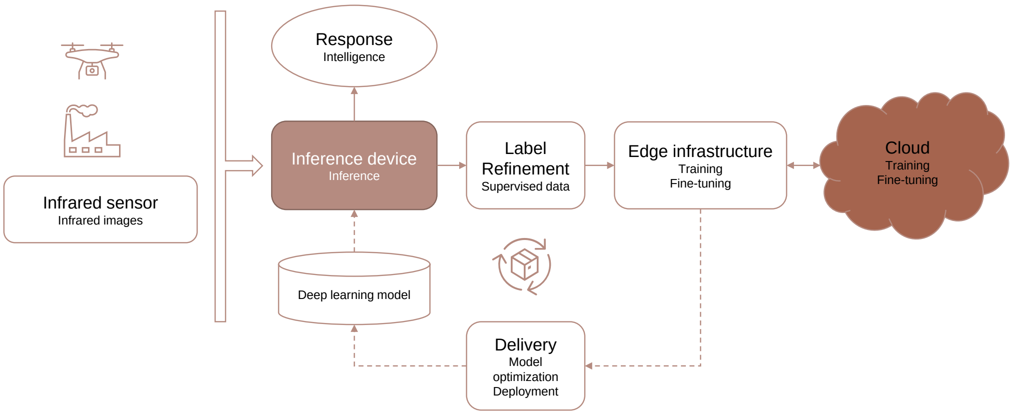



**Fig. 3.** *Active learning architecture.*

2. The tracking procedure is applied to linear segments of the pipeline. Therefore, all pipeline joints are aligned on the same line. As soon as two pipeline joints are confirmed, the line connecting these two objects in the image is also tracked. This enables the identification of false positive detections by ignoring any other detections that do not lie on this line.

3. Since the drone flight is planned along the path where the pipeline is located, new pipeline joints must appear in the image in the direction the drone is moving towards. Additionally, the region in the image where new valid pipeline joints appear must be located downstream of the last confirmed detection considering the direction advance. Since the drone flight is not perfectly aligned with the pipeline, and sudden motion corrections may occur during the flight navigation, an uncertainty factor is taken into account when calculating the valid region for new pipeline joint detections. Any detections outside this region are considered false positives and ignored, as they are unlikely to be actual pipeline joints.

4. In a pipeline, the pipes have a fixed length. Thus, the pipeline joints, necessary to connect the pipes, can be found periodically along the length of the pipeline. Once two pipeline joints are confirmed, the distance in image coordinates between flight joints can be calculated. Based on this distance and considering the flight advance, the position of the next pipeline joint in the image can be estimated. When a tentative joint is located in the estimated position of the next joint, the confidence threshold and the number of required consecutive detections are decreased. The tracking method is able to quickly confirm a tentative pipeline joint based on the prior gathered information. This helps ensure that the tracking is accurate and reliable, even when the detection is uncertain.

#### 2.1.3 Active learning

When utilizing deep learning, having a high-quality dataset of labeled samples is essential. The dataset is utilized by the model to identify patterns and connections in the data and make precise predictions on unseen data. The creation of the dataset represents a necessary but often costly process in the development of deep learning models. The dataset must be reliable and representative of the population it is intended to model, as a lack of diversity can lead to poor performance on new data. The dataset must also be accurate and complete, as poorly labeled data can lead to inaccurate and unreliable models. The process of obtaining and annotating the dataset can be costly and time-consuming. However, having an accurate and varied dataset is of utmost importance as it has a direct impact on the accuracy of the trained model. In practical systems that solve real problems, having a high-quality dataset and a procedure to annotate and incorporate new data quickly is crucial for maintaining and improving the performance of the system.

The proposed approach combines concepts from active learning, incremental learning and continual learning [13]. Incremental learning (also called online learning) is a technique used to add new data to a pre-trained model, and updating the model's parameters based on the new data. Continual learning (also known as lifelong learning) is a technique used to train a model on a stream of tasks, where the model can learn and adapt to new tasks without forgetting previous tasks.

Fig. 3 illustrates the architecture of this approach. The proposed methodology streamlines the annotation process through assisted labeling. Initially, a model is trained on a small labeled dataset, which it uses to generate initial labels for the remaining unlabeled data. Human annotators then review and correct these generated labels. This iterative process continues with new data, gradually improving the detection model's accuracy by incorporating human feedback. This approach accelerates annotation, enhances label accuracy, and reduces costs. Additionally, an interactive image annotation tool simplifies the process further.

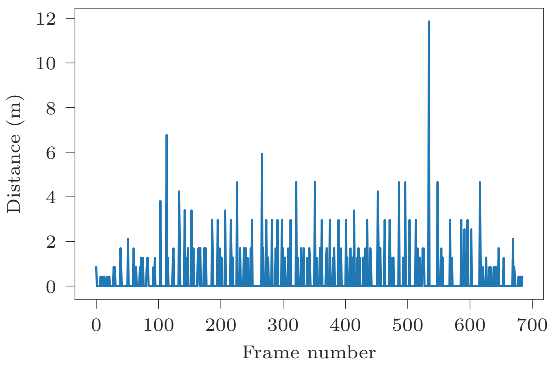



**Fig. 4.** *Distance between consecutive frames in the first video segment as determined by the GPS coordinates attached to each frame.*

The proposed approach in this work aims to actively select the most informative examples to be labeled by using the result of the tracking procedure, rather than relying on the uncertainty of the object detector alone. This approach leverages the information provided by the tracking procedure to identify the instances where the model is making errors, and then queries the user for additional information or clarification about these instances. Additionally, this approach is used in conjunction with uncertainty sampling to further improve the performance of the object detector.

The video segment is analyzed to identify individual pipes, and the data collected for this section of the pipeline includes a series of pipes located at specific intervals along the pipeline. As such, the number of pipes identified in the video must match the number of pipes expected for the length of the inspected pipeline in this segment. In case of discrepancy, it indicates errors in either the detection or the tracking methods. To improve the accuracy of the detection and tracking methods, random frames from this video segment are extracted, and the current object detector is used to detect and annotate pipeline joints. Additionally, this process is also applied to frames with low-confidence detections. The human annotator is then required to review the generated labels and correct any errors, resulting in an extended and more diverse dataset. The object detector is retrained using the extended dataset, which includes the newly labeled frames. This process involves fine-tuning the model's performance, resulting in a more accurate and robust object detector.

In the considered approach, the training procedure divides the available dataset into three subsets: the training subset used for training with 60% of the samples, the validation subset with 20% of the samples to monitor the training and prevent overfitting, and the test subset with 20% to the evaluate the performance of the resulting model.

The proposed active learning approach is an iterative process, where the object detector is retrained multiple times, each time using a larger and more diverse dataset. This allows the model to continue to improve its accuracy and performance over time, as it is exposed to more varied data and different scenarios. During the retraining process, the updated model is evaluated against a test set which is kept separate from the training set, to measure its performance, the ability to remember old information and to identify any areas where the model may still need improvement.

### 2.2 Results

Tests for the proposed approach used the drone model DJI Matrice 300 RTK, paired with a Zenmuse XT2 camera. This camera boasts a dual-sensor design comprising both thermal and visual capabilities. The FLIR infrared camera, a component of the Zenmuse XT2, offers a resolution of 640 x 512 pixels and operates at a frame rate of 30 Hz. Notably, the manufacturer reports a sensitivity of less than 50 mK for this camera. It utilizes an uncooled microbolometer detector and operates within the long-wave infrared spectrum, specifically in the range of 7.5-13.5 micrometers.

#### *2.2.1 Unreliability of the GPS*

Figure 4 demonstrates the inaccuracies of the geographic coordinates recorded during the flight. The figure shows the distance between consecutive frames in the first video segment, as determined by the GPS coordinates attached to each frame. The variability in distance is substantial, erratically ranging from 0 to 12 meters. Additionally, in more than 77% of the frames, the distance is 0, even when the drone is moving at high speeds. This suggests that the camera and GPS receiver may not be operating at the same frequency. This supports the approach taken in this work of not relying on GPS for pipe identification. The GPS coordinates measured do not provide a reliable and accurate reflection of the drone's position and movement, especially when the drone is flying at high speeds. This makes GPS unreliable for identifying pipes.

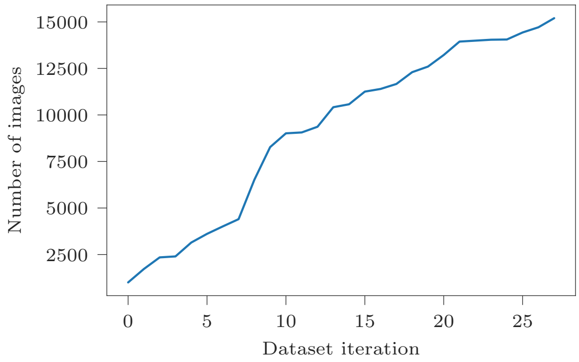



**Fig. 5.** *Size of the dataset during the implementation of the active learning approach for one year.*

**Table 1.** *Object detection performance for different model architectures.*

| Model | Par. (M) | $P$ | $R$ | $F_1$ | $AP_{50}$ | $AP_{50:95}$ |
|---|---|---|---|---|---|---|
| YOLOv5n | 1.77 | 0.993 | 0.991 | 0.992 | 0.995 | 0.869 |
| YOLOv5s | 7.23 | 0.992 | 0.992 | 0.992 | 0.995 | 0.880 |
| YOLOv8n | 3.01 | 0.993 | 0.989 | 0.991 | 0.995 | 0.881 |

#### *2.2.2 Evolution of the dataset using active learning*

The active learning approach proposed in this study was executed over the course of one year, during which the dataset steadily grew in size. Figure 5 illustrates the progression of both the dataset size and the iterations within the active learning framework, providing insights into the evolution of the dataset over time.

The process of bootstrapping a dataset was also considered since, initially, there was no model to automatically annotate the samples. An image processing technique that combined Gabor filters and thresholding was used to solve this problem. This method was used to automatically annotate the images in the first version of the dataset, which consisted of 1000 images. The detection method had an accuracy of around 90%, so the necessary review and correction of annotation errors represented a low overhead for the human annotator.

#### *2.2.3 Object detection*

Table 1 presents the performance metrics of object detection across various model architectures, evaluated on the final version of the dataset. This table offers a comprehensive overview of the detection capabilities of different models.

The number of parameters and the GFLOPs (giga floating-point operations per second) are measures that indicate the complexity of the model. As the GFLOPs increase, a more powerful GPU is required to operate the model for the same duration. This is because higher GFLOPs indicate that the model is more computationally intensive, and therefore requires more processing power to run at the same speed. A more complex model is generally associated with improved accuracy. This is because a more complex model has a greater number of parameters, thereby enabling it to extract more information from the data, leading to more accurate predictions. However, it is essential to bear in mind that as the complexity of the model increases, so does the risk of overfitting. Overfitting occurs when the model becomes too closely attuned to the training data, resulting in deficient performance on new, unseen data. Additionally, it should also be noted that there is a trade-off between model complexity and computational cost, as more complex models may necessitate greater computational resources and larger amounts of data to generalize effectively.

A comparison of the inference times using this configuration can be seen in Fig. 6. The inference time for YOLOv5n takes around 0.5 ms per image, which is half the inference time required for YOLOv8n. These results are closely related to the number of parameters of the model, but also the GPU used for inference (RTX 3090).

The findings presented in Table 1 suggest that there is no significant improvement in performance when employing a more complex model for the given dataset. This observation underscores the importance of selecting a model architecture that strikes a balance between complexity and performance for optimal results.

The results show that YOLOv5n not only has the lowest computational cost, but also performs similarly in terms of accuracy to the other models evaluated. Based on these results, the less demanding model, YOLOv5n, is used in the inspection system.

Training hyperparameters control various aspects of training, such as the learning rate, batch size, and many oth-

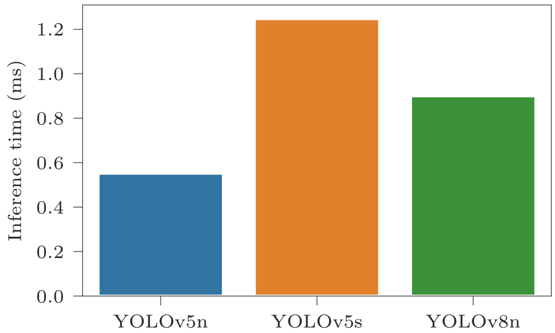



**Fig. 6.** *Inference time for various model architectures using a batch size of 16.*

ers. These hyperparameters can greatly affect the performance of the model. However, finding the optimal combination of hyperparameters is a major challenge as the number of possible combinations can be very large. Moreover, evaluating each combination can be extremely time-consuming, making it hard to find the best set of hyperparameters in a reasonable amount of time. Traditional methods, such as grid search, have limited capabilities due to the high dimensional search space and the unknown correlations among the dimensions. In this work an alternative option was explored: automated tuning using genetic algorithms.

The automated tuning approach looks for the best configuration based on an initial population of candidate solutions, each representing a combination of 30 training hyperparameters. Evolution is performed based on a fitness function defined as the weighted combination $AP_{50}$ (10%) and $AP_{50:95}$ (90%). To evaluate the fitness function, each individual is trained for a small number of epochs, as the procedure is expensive and time-consuming. Then, based on the principles of crossover, mutation, and selection, the population evolves over multiple generations.

The tuning approach was applied to YOLOv5n. The results provided an optimal configuration of training hyperparameters with a 0.0102 learning rate, 0.98 momentum and an alternative configuration for the augmentation. Results indicate that the performance of the model improved by only 1.38% in the metric $AP_{50:95}$. This suggests that while the automated tuning approach was able to find a better configuration of hyperparameters, the performance improvement was limited. However, the slight increase in performance makes the performance of YOLOv5n when considering $AP_{50:95}$ comparable to that of the default configuration of the more complex model, YOLOv8n, and even surpasses it when considering $F_1$.

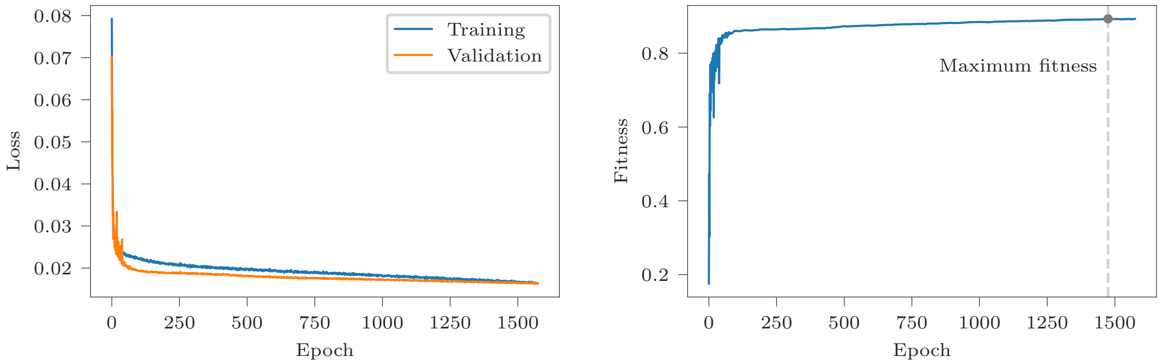
Fig. 7a shows the minimization of the loss function when training YOLOv5n using the optimal hyperparameter configuration. Data augmentation is employed to improve the generalization of the model, thereby reducing overfitting. However, as the training progresses, the validation loss stops improving even though the training loss continues to decrease. At this point, the early stopping approach terminates the training procedure. The selected model is not the one obtained at the last epoch, but rather the one that maximizes the fitness function, which takes into account the performance of the model on the validation subset. The evolution of the fitness and the epoch at which the maximum value is reached can be seen in Fig. 7b. The training process in total takes approximately 10 hours.

Decreasing the confidence threshold would increase the recall of the model, as it would result in more positive predictions being made. However, this would also increase the number of false positives, leading to a drop in precision. The balance between precision and recall can be adjusted by varying the confidence threshold used for making predictions.

This research also evaluated the integration of deep learning with infrared imaging. While many object detection models are designed for color images, infrared images typically encode information differently, utilizing a single 16-bit number for radiation measurement. To adapt, infrared images can be transformed into color images via temperature-based color mapping or into grayscale images with a single channel. This study thoroughly evaluated and compared both approaches in terms of detection accuracy. Interestingly, the results revealed no significant differences between the two approaches, indicating that changing the image format did not substantially impact detection performance.

#### *2.2.4 Tracking*

Figure 8 illustrates a tracking and identification process. Initially, pipeline joints are detected with confidence, marked as tentative (red boxes) with assigned tentative identifications. As the system confirms these joints over consecutive frames (green boxes), identification progresses considering the direction of advance. Semantic information is extracted from the joints, aiding in estimating the position of the next joint and mitigating false positive detections. Subsequently, pipes are segmented for further analysis, with measurements such as median temperature displayed. The process remains robust despite small rotations and changes in trajectory. In Fig.8f, the tracking method continues updating positions even with missing detections, while Fig.8g demonstrates handling false positive detections using semantic information. Moreover, multiple measurements taken for each pipe along the pipeline path are combined to enhance accuracy and reduce outliers, providing a more reliable




**(a)** **(b)**

**Fig. 7.** *Training evolution. (a) Loss during training. (b) Fitness function during training*

estimate of measured values.

#### *2.2.5 Test in the industrial facility*

Tests were conducted at an industrial facility, with Fig. 9 providing an aerial view showcasing the pipeline locations within the facility.

Figure 10 displays a temperature map of a section of the inspected pipeline. The figure illustrates the average temperature within each pipe. A comparison of temperatures across pipes highlights an anomaly in pipe 4.6 (segment 4, number 6), which exhibits a temperature deviation from the others. Subsequent investigation confirmed the presence of a defect in this pipe, specifically related to insulation issues.

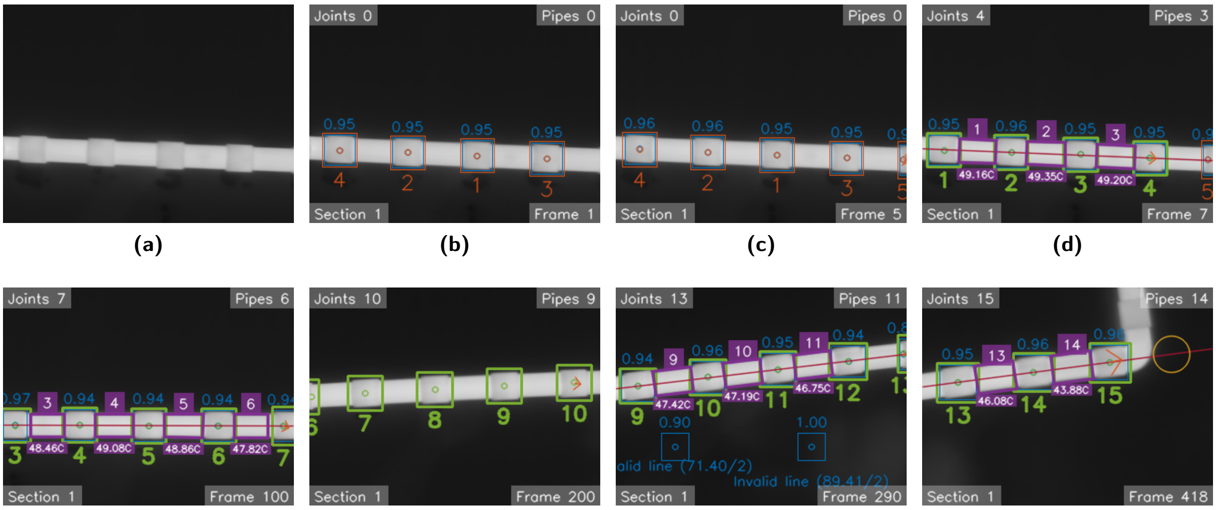
## 3. Conclusions

This work proposes a method for robust temperature measurement in dynamic environments through detection and tracking. The approach incorporates deep learning for detecting objects of interest in images and tracking their movement, eliminating the reliance on unreliable GPS coordinates. By combining object detection and tracking techniques, the method ensures accurate temperature measurements in challenging, dynamic settings. In addition, an active learning methodology is applied to iteratively improve the detection and tracking performance over time. By leveraging active learning, the system efficiently learns from human feedback, enhancing its ability to accurately detect and track objects in dynamic environments.

Testing demonstrates the effectiveness of the proposed approach in achieving accurate temperature measurement in dynamic environments. The system's performance is evaluated through rigorous testing across various scenarios. Results indicate that the combined use of deep learning-based object detection and tracking techniques yields reliable temperature measurements, even in challenging situations. The demonstrated effectiveness underscores the potential of the proposed methodology for real-world applications in dynamic settings.

## Acknowledgments

This research was partially funded by the Spanish National Program for Mobility of Professors and Researchers, reference PRX22/00165.




**(a)** **(b)** **(c)** **(d)**

**(e)** **(f)** **(g)** **(h)**

**Fig. 8.** *Evolution of the tracking and identification approach. (a) First infrared image acquired. (b) Four pipeline joints are detected and marked as tentative joints in the first frame. (c) As the drone moves the systems keep track of the detections, matching detections and tentative joints. (d) After some consecutive matches, the pipeline joints are confirmed and the pipe identification number is assigned considering the movement direction. (e)-(h) The process continues as the drones move along the pipeline path dealing with missed detections and false positive detections.*

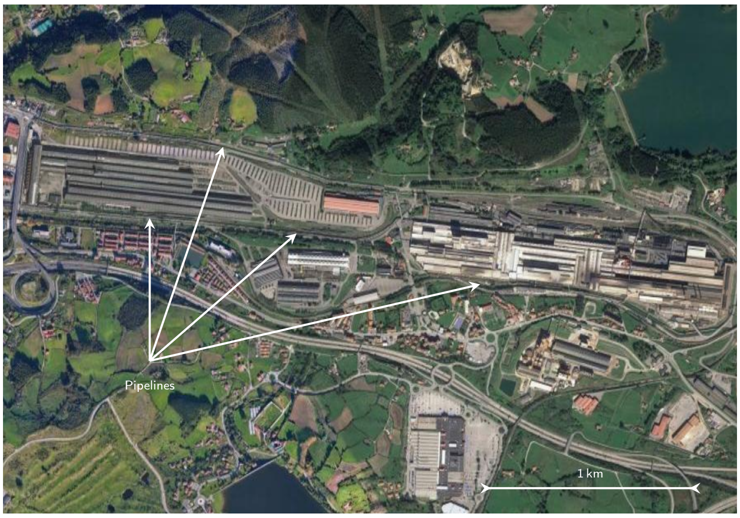



**Fig. 9.** *Aerial view showcasing the pipeline locations within the facility where tests are performed.*

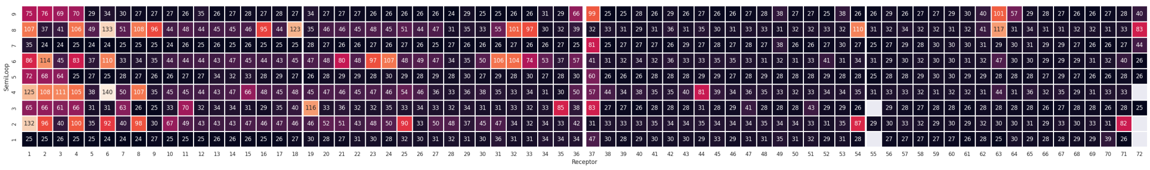


**Fig. 10.** *Temperature measurement results for different sections of the pipeline system. All values are in Celsius.*

**References**


[1] Rubén Usamentiaga, Pablo Venegas, Jon Guerediaga, Laura Vega, Julio Molleda, and Francisco G Bulnes. Infrared thermography for temperature measurement and non-destructive testing. *Sensors*, 14(7):12305–12348, 2014.

[2] Christoph Tuschl, Beata Oswald-Tranta, Timotheos Agathocleous, and Sven Eck. Scanning inductive pulse phase thermography with changing scanning speed for non-destructive testing. *Quantitative InfraRed Thermography Journal*, pages 1–16, 2023.

[3] Karol M Golasiński, Michał Maj, Leszek Urbański, Maria Staszczak, Arkadiusz Gradys, and Elżbieta A Pieczyska. Experimental study of thermomechanical behaviour of gum metal during cyclic tensile loadings: the quantitative contribution of irt and dic. *Quantitative InfraRed Thermography Journal*, pages 1–18, 2023.

[4] Carosena Meola, Simone Boccardi, and Giovanni Maria Carlomagno. *Infrared thermography in the evaluation of aerospace composite materials: infrared thermography to composites*. Woodhead Publishing, 2016.

[5] Rubén Usamentiaga, Daniel F Garcia, and Jesús M Pérez. High-speed temperature monitoring for steel strips using infrared line scanners. *IEEE Transactions on Industry Applications*, 56(3):3261–3271, 2020.

[6] J Manara, M Arduini-Schuster, and M Keller. Infrared-optical characteristics of ceramics at elevated temperatures. *Infrared Physics & Technology*, 54(5):395–402, 2011.

[7] Ruben Usamentiaga, Miguel Angel Fernandez, Alberto Fernandez Villan, and Juan Luis Carus. Temperature monitoring for electrical substations using infrared thermography: architecture for industrial internet of things. *IEEE transactions on industrial informatics*, 14(12):5667–5677, 2018.

[8] Alberto Fernández, Rubén Usamentiaga, Pedro de Arquer, Miguel Ángel Fernández, D Fernández, Juan Luis Carús, and Manés Fernández. Robust detection, classification and localization of defects in large photovoltaic plants based on unmanned aerial vehicles and infrared thermography. *Applied Sciences*, 10(17):5948, 2020.

[9] Ella Mahoro and Moulay A Akhloufi. Breast cancer classification on thermograms using deep cnn and transformers. *Quantitative InfraRed Thermography Journal*, 21(1):30–49, 2024.

[10] Hassan Iqbal, Solomon Tesfamariam, Husnain Haider, and Rehan Sadiq. Inspection and maintenance of oil & gas pipelines: a review of policies. *Structure and Infrastructure Engineering*, 13(6):794–815, 2017.

[11] Mina Fahimipirehgalin, Emanuel Trunzer, Matthias Odenweller, and Birgit Vogel-Heuser. Automatic visual leakage detection and localization from pipelines in chemical process plants using machine vision techniques. *Engineering*, 7(6):758–776, 2021.

[12] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[13] Gido M Van de Ven, Tinne Tuytelaars, and Andreas S Tolias. Three types of incremental learning. *Nature Machine Intelligence*, 4(12):1185–1197, 2022.