Infrared video-based object recognition for agricultural research field uses

Damir A. Guzaerov[1], *Igor D. Sergunin*[1], and *Vlada V.* Kugurakova[1*]

[1]Kazan Federal University, Institute of Information Technology and Intelligent Systems, 420008, 35 Kremlevskaya st., Kazan, Russia.

Abstract. The paper presents a study of an object recognition system based on infrared (IR) video stream using the YOLOv8 model. The aim of the work was to develop and evaluate the performance of a computer vision algorithm for analysing infrared images under various conditions including low light. While the use of RGB images can provide rich information in normal conditions, it becomes inefficient to use a visible spectrum camera in total darkness or in the presence of clutter. The study was based on an extensive dataset containing about 176,000 annotated frames from an infrared camera. The YOLOv8 model was adapted and trained to recognise 15 classes of objects, including people, animals and different types of vehicles. The results showed high performance of the model in recognising the most represented classes in the dataset, such as cars, people with F1-score up to 0.8. However, the problem of unbalanced classes was identified, leading to lower recognition accuracy for rarely seen objects. The study demonstrates the potential of using infrared cameras in computer vision tasks and opens new directions for further developments in the field of object recognition in limited visibility conditions. The results of the work may find application in a wide range of disciplines, including medicine, agriculture and environmental monitoring.

1 Introduction

Computer vision is rapidly becoming popular in various fields, which stimulates the search for new ways to apply it and improve the technology to increase its efficiency and functionality, such as in low-light conditions [1]. Object recognition is a technology that allows systems to identify and classify objects in images or video. The development and optimisation of object recognition algorithms, as well as their use in the infrared range, create new perspectives for the implementation of security, navigation and object monitoring systems in a variety of environments.

Infrared video streaming-based object recognition is designed to work with a variety of video data sources. Infrared spectrum technology is used in a variety of applications such as night vision, searching for inflammatory processes in medicine, identifying fruit among foliage or detecting temperature changes in animals in pastures.

* Corresponding author: vlada.kugurakova@gmail.com

Despite all the advantages of this approach, the use of IR cameras in computer vision still has a rather high chance of error. However, there are currently insufficient datasets in the public domain to train artificial neural networks to operate on IR images [2, 3].

2 Related Works

Object recognition using infrared cameras has attracted considerable research attention in recent years due to a number of advantages this technology offers. Unlike visible-spectrum cameras, IR cameras can operate effectively in low-light conditions and in a variety of weather conditions.

A comprehensive review of the current state of the art of research in object detection using near-infrared and thermal imagery is presented in [4]. The authors analyse various methods and algorithms used in this field, including the application of IR spectrum. From the results, it is found that the lack of publicly available annotated IR image datasets makes it difficult to develop and compare algorithms. In addition, a promising direction for improving performance by fusing different wave spectra (IR, RGB, LiDAR) is discussed [5].

In the context of autonomous driving [6], the use of thermal imagery for object detection in low-light conditions has been investigated. The authors note that the lack of marked-up data limits the use of artificial intelligence in the area of object recognition from thermal images.

The paper [7] investigates the application of non-destructive IR spectrum and deep learning techniques for object detection, in particular for robotic inspection and harvesting of chilli peppers. The application of IR radiation shows advantages over the use of RGB spectrum alone. Due to the fruit's property of effective heat retention and the sensitivity of the IR camera to temperature differences between foliage and fruit, it is possible to effectively distinguish green fruit from foliage, especially in low-light conditions or in the presence of a large amount of visual noise.

In [8], a comprehensive review of thermal camera applications in various applications including surveillance, medical and industrial is presented, with a high appreciation of the potential of IR technology. Particular attention is paid to the effectiveness of thermal cameras in low visibility environments, making them indispensable for security and surveillance applications. Despite the effectiveness of IR in identifying people and animals, the technology is limited by low image resolution, the high cost of IR cameras, and problems in distinguishing objects with close temperatures.

In [9], a new benchmark dataset is created and the developed methodology for evaluating object tracking algorithms on thermal images is presented, which provides researchers with valuable resources for comparing and improving object recognition methods in the infrared spectrum. The study shows that ranking tracking algorithms on visible spectrum images and thermal images yields different results, with some methods that are effective for the visible range proving less successful in the infrared. The results of this work indicate the need for the development of specialised approaches for tracking in thermal images and further research in this area.

In [10], a novel method for adapting neural networks for human detection in IR images is proposed. This approach solves the critical problem of limited training data in the IR domain by adapting models initially trained on visible spectrum images. Thus, this approach allows detectors pre-trained on RGB data to be efficiently used for IR object detection tasks without the need for large IR datasets for training. It is mentioned, however, that this approach may result in the loss of specific loss of IR information, and the need for continuous pre-training increases the computational requirements and limits the scalability of the system.

Research in the field of object recognition using infrared cameras shows considerable potential. Promising directions are the development of specialised algorithms for IR images,

creation of methods for adapting models from RGB to IR domain and combining data from different spectra. At the same time, application of RGB spectrum is impossible in some cases, such as detection of unripe fruits among foliage or low light conditions. However, the development of this area is hampered by the lack of publicly available annotated IR image datasets. Key challenges include low resolution of IR images, high camera costs, and difficulty in distinguishing objects with close temperatures. Despite the challenges, IR technology remains promising for a variety of applications including agriculture, autonomous driving, security systems, and medicine.

3 Model Development

In the developed system of object recognition based on infrared video stream it is supposed to be possible to work with different sources of video data. The input is the path to the video file or the address of the camera, which allows it to be used flexibly both for analysing pre-recorded video clips and for processing video streams in real time. Consequently, the following requirements were set when developing the system:

- Uploading recorded video or real-time recognition from the camera.
- Processing each frame of the video stream using a trained neural network.
- Adding a Bounding Box on each recognised object.
- Saving the system results to a new video file.

Deep learning, especially using convolutional neural networks (CNNs), is an advanced approach in object recognition, distinguished by its robustness to changes in the environment, its ability to generalise and its adaptability. Despite its advantages, this method has several drawbacks, including the need for large amounts of high-quality training data, high computational resource requirements, and difficulty in interpreting model performance.

The YOLO (You Only Look Once) model was selected for the task of object detection on infrared video stream. As of 2023, YOLOv8 was the most advanced model within the object detection task [11]. YOLO is able to process images and recognise objects in real time due to its architecture that performs predictions in a single pass over the image. The model performs well in terms of accuracy, recognising objects with a high degree of reliability, although it has problems in recognising small objects [12]. Another advantage of YOLO is its versatility in being able to process different types of data including infrared images.

Since the efficiency and accuracy of the neural network directly depends on the size and quality of the training dataset, it is necessary to collect the most extensive collection of images from the infrared camera. Self-collection of the dataset using IR camera is a time-consuming and labour-intensive task, given that annotation of each frame must be done manually. Another option is to generate synthetic datasets based on simulation of the environment in the form of computer graphics [13], which can help automate the image acquisition and annotation process. At the same time, the development of a variety of 3D graphics, animations and virtual environment algorithm also requires a large amount of time, and the experience of 3d artists directly affects the quality of the resulting synthetic datasets.

Thus, if existing datasets are available, their use is more appropriate, as it allows to test the approach in a short time. A rather extensive labelled dataset of IR camera recordings is the ADAS dataset [4], which contains annotations for 15 types of objects, including people, different types of vehicles and traffic control devices. The problem with this dataset is the lack of training data for some object classes, which is solved by adding other image sets to the training sample [14]. The final dataset was about 176 thousand annotated frames from the infrared camera.

After preparing the dataset, we should proceed to the selection of optimal hyperparameters responsible for learning control, which affects the speed and quality of learning:

1. Number of epochs (epoch) – the number of times the entire dataset is used during training. Often, there is not enough training data to train the model in one epoch, so the number of epochs can reach hundreds or thousands. This number varies in each case and should be based on the preliminary results of training.
2. Image set size (batch) – the number of images the model processes at a one time before changing the weights. A large value means more memory consumption, and infrequent change of weights can lead to incorrect response to the data. At the same time, too small a value entails a slower learning speed, and the influence of each image on the final result is too unpredictable.
3. Horizontal image reflection probability (fliplr) is a parameter used to increase the diversity of training data and increase stability when processing images for different angles.
4. Image splice probability (copy_paste) - a parameter defining the probability of adding objects from one image to the original dataset in order to improve noise tolerance and add variability to the data.
5. Image mixing probability (mixup) - the probability of applying colour mixing between two images, creating a new, blurred frame that will allow the model to navigate even in the face of data uncertainty.
6. Image size (imgsz) is the size of the image the model will work with. This parameter directly affects the speed of processing one image due to the image detail.
7. Processing device (device) and number of parallel threads (workers) are parameters for selecting a technical device for image processing (CPU or GPU) and its workload.

To evaluate the accuracy of the model, the F1-score metric was used (Figure 1-2), which is a harmonic average between Precision and Recall, which allows both metrics to be combined into one metric.

This metric allows the model to be estimated even when the training sample is unbalanced across classes. In cases where one class is highly dominant (e.g., 95% negative and 5% positive examples), the model may show high accuracy without predicting positive cases at all. F1-score helps to avoid the situation where the model seems to be effective only because of high accuracy, while failing to consider the significance of positive examples.

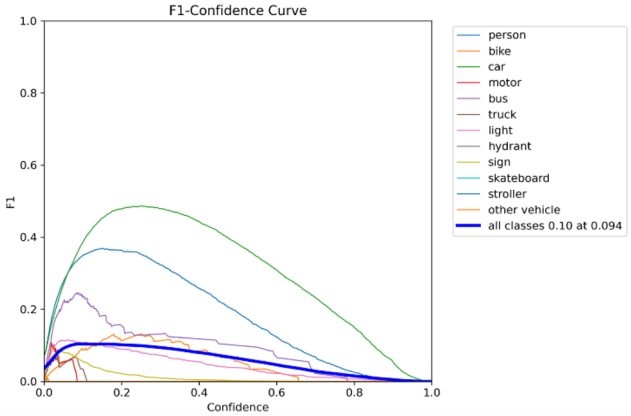

Fig. 1. F1-score metric in the initial stages of training.

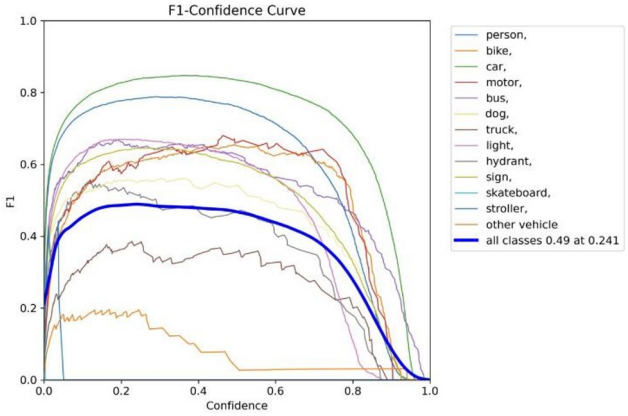

Fig. 2. F1-score metric after finalising the model.

According to the obtained results, the model is most effective in searching for objects of the classes “car”, “person” reaching 0.8 in the peak. This indicates its ability to accurately recognise key objects under sufficient training data. Meanwhile, other classes have lower scores, which is expected for classes with less data in the dataset (more than 10 times). Thus, the model performs well for a limited number of classes, but suffers greatly when dealing with rare objects, which significantly reduces its overall performance.

The resulting system allows the user to run the recognition system (Fig. 3) with different arguments. In the real-time recognition mode, the system processes the input video file, marks recognised objects on it, and saves it in a separate directory. It is also possible to recognise objects in the background.

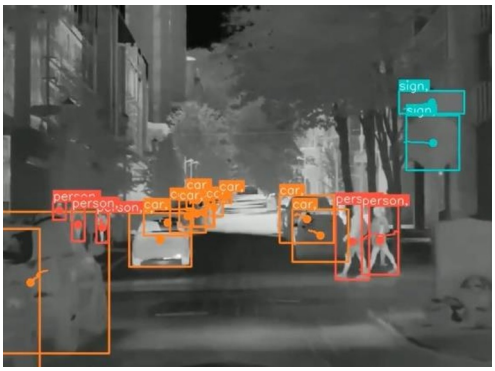

Fig. 3. A frame from the processed video with recognised objects.

4 Conclusions

In the course of the research, an object recognition system based on infrared video stream using the YOLOv8 model was developed and tested. The analysis of the results showed that the developed model demonstrates high efficiency in recognising the most represented object classes in the dataset, reaching F1-score values up to 0,8. This indicates the significant potential of using infrared cameras in computer vision systems.

However, a significant reduction in model performance was found when dealing with infrequently occurring object classes, indicating the problem of unbalanced training dataset. In order to solve this problem and improve the overall efficiency of the model, several areas of improvement are proposed. First, expanding the dataset for rare classes (around 10 thousand images and more) can significantly improve the overall model performance. Second, applying data augmentation techniques such as mixup and copy-paste, especially for classes with few examples, can improve the generalisation ability of the model. Third, individual adjustment of the confidence threshold for each class can optimise the balance between recognition accuracy and completeness.

Future work will focus on solving the class imbalance problem, optimising data augmentation techniques and developing more flexible recognition algorithms and integration within augmented reality applications [15], although current augmented reality technologies are not adapted for IR camera applications. Another application that seems logical is the use of this technology in fully autonomous or remotely controlled robotic devices in order to gather more information about the environment [16]. Such improvements will allow to extend the application of the system in various fields, including the agro-industrial sector. In particular, the developed technology can find application in physiology for non-invasive monitoring of animal body temperature, in botany and agriculture for assessing plant health and detecting diseases, and in zoology for observing animal behaviour at night.

Acknowledgments

This paper has been supported by the Kazan Federal University Strategic Academic Leadership Program ("PRIORITY-2030").

References

1. MordorIntelligence, IR Camera Market Size & Share Analysis - Growth Trends & Forecasts (2024 – 2029). URL: https://www.mordorintelligence.com/industry-reports/ir-camera-market.
2. L. Budach, et al, *The Effects of Data Quality on Machine Learning Performance*. [Online]. Available at: https://arxiv.org/abs/2207.14529 (2022)
3. O. Zendel, M. Murschitz, M. Humenberger, and W. Herzner, *How Good Is My Test Data? Introducing Safety Analysis for Computer Vision*, International Journal of Computer Vision, **125**. doi: 10.1007/s11263-017-1020-z (2017)
4. N. Bustos, M. Mashhadi, S. K. Lai-Yuen, S. Sarkar, and T. K. Das, *A systematic literature review on object detection using near infrared and thermal images*, Neurocomputing, **560**. doi: 10.1016/j.neucom.2023.126804 (2023)
5. A. Banuls, A. Mandow, R. Vazquez-Martin, J. Morales, and A. Garcia-Cerezo, *Object Detection from Thermal Infrared and Visible Light Cameras in Search and Rescue Scenes*, In *2020* IEEE International Symposium on Safety, Security, and Rescue Robotics, SSRR 2020. doi:10.1109/SSRR50563.2020.9292593 (2020)
6. F. Munir, et al, *Exploring thermal images for object detection in underexposure regions for autonomous driving,* Applied Soft Computing, **121**. doi: 10.1016/j.asoc.2022.108793 (2022)
7. S. C. Hespeler, H. Nemati, and E. Dehghan-Niri, *Non-destructive thermal imaging for object detection via advanced deep learning for robotic inspection and harvesting of*

chili peppers, Artificial Intelligence in Agriculture, **5**. doi: 10.1016/j.aiia.2021.05.003 (2021)

8. R. Gade, and T. B. Moeslund, *Thermal cameras and applications: a survey, Machine Vision and Applications*, **25**, pp.245-262. doi: 10.1007/s00138-013-0570-5 (2014)
9. A. Berg, J. Ahlberg, and M. Felsberg, *A Thermal Object Tracking Benchmark*, In 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1-6. doi: 10.1109/AVSS.2015.7301772 (2015)
10. C. Herrmann, M. Ruf, and J. Beyerer, *CNN-based thermal infrared person detection by domain adaptation*, Proc. SPIE 10643, Automatic Target Recognition XXVIII, no. 1064308. doi: 10.1117/12.2304400 (2018)
11. J. Terven, D. M. Cordova-Esparza, and J. A. Romero-Gonzalez, *A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS*, Machine Learning and Knowledge Extraction, **5**. doi: 10.3390/make5040083 (2023)
12. T. Diwan, G. Anirudh, and J. Tembhurne, *Object detection using YOLO: challenges, architectural successors, datasets and applications*, Multimedia Tools and Applications **82**. doi: 0.1016/j.aiia.2021.05.003 (2023)
13. V. V. Kugurakova, et al, *Generation of Three-Dimensional Synthetic Datasets*, Russian Digital Libraries Journal, **24**, no. 4. doi: 10.26907/1562-5419-2021-24-4-622-652 (2021)
14. SensationTI, *FLIR_IR_Expansion*. [Online]. Available: https://github.com/sensationTI/FLIR_IR_Expansion/commits/master/ (2020)
15. D. Gabdullina, E. Zykov, and V. Kugurakova, *Using OpenCV for real-time image recognition through augmented reality devices*, In Proceedings of International Conference on Artificial Life and Robotics. doi: 10.5954/icarob.2023.gs2-4 (2023)
16. V. V. Kugurakova, et al, *VR-Telecontrol of Multi-Arm Devices: Problems, Hypotheses, Problem Statement*, Russian Digital Libraries Journal, 25, no.5. doi: 10.26907/1562-5419-2022-25-5-441-471 (2022).