New Thermal Automotive Dataset for Object Detection

Tomasz Balon*, Mateusz Knapik†‡ and Bogusław Cyganek†

*Department of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering,
Email: *tbalon@student.agh.edu.pl*
†Department of Computer Science, Electronics and Telecommunication
Email: *mknapik@agh.edu.pl, cyganek@agh.edu.pl*
AGH University of Science and Technology,
Al. Mickiewicza 30, 30–059 Kraków, Poland
‡MyLED Inc.
Email: *m.knapik@myled.pl*
Ul. W. Łokietka 14/2, 30–016 Kraków, Poland

Abstract—Although there are many efficient deep learning methods, object detection and classification in visible spectrum have many limitations especially in case of poor light conditions. To fill this gap, we created a novel thermal video database containing few thousands of frames with annotated objects acquired in far infrared thermal spectrum. Thanks to this we were able to show its usability in the traffic object recognition based on the YOLOv5 network, properly trained to gain maximal performance on thermal images, which contain many small objects and are characteristic of different properties than the visible spectrum counterparts. The proposed thermal database, as well as the fully trained model are main contributions of this paper. These are made available free for other researchers. Additionally, based on the highly efficient car detector we show its application in the car speed measurement based exclusively on thermal images. The proposed system can be also used in the Advanced Driver-Assistance Systems (ADAS), and help autonomous driving.

I. Introduction

ARTIFICIAL intelligence (AI) and machine learning (ML) are two of the fastest developing technologies nowadays. New and novel architectures are developed to be faster, more accurate and reliable. Image classification and object detection is very active field of research and many innovative techniques were proposed recently. Range of possible applications is very wide and autonomous driving is one of them. It gained much of an interest from scientists and companies recently. Vision systems based on a visible lights have limitations when used in a moving vehicle caused by wide range of lighting conditions that can occur. Low light during the night time as well as very high amounts of light during the day pose a challenge to hardware and software modules. On the other hand, thermal imaging in recent years gains popularity, both in industrial solutions, as well as in research projects.

However, the development of the image analysis methods might be rapid for images acquired using conventional RGB sensors, other imaging technologies operating in spectrum beyond visible light still fall short mostly due to the lack of publicly available sufficiently large training datasets.

To help alleviate this problem, in this paper we present:

- A new novel traffic dataset acquired using thermal imaging camera.
- Pretrained object detection model based on YOLOv5 architecture.
- An exemplary application based on detections: speed measurement in thermal spectrum.
- Examples of potential further applications.

Our dataset contains videos with close to 30,000 hand-annotated objects, many of small size, which makes them difficult to detect. Our second contribution is pretrained object detection model based on YOLOv5 architecture. Primary use case intended for this model is detecting four classes of objects in thermal images, as well as car speed measurement, which is the third contribution provided in this paper.

The paper is organized as follows. Section II describes the related works. In Section III process of acquiring the data and model training is explained. Section III-B presents the structure and provides more insight into dataset. Section IV shows example of how acquired data might be used in calculating vehicle's speed. In Section V more of future development possibilities are discussed. Finally, Section VI concludes the paper.

II. Related works

Object detection combined with thermal imaging gained a lot attention in recent years. Thermal imaging is based on observing infrared waves emitted by warm objects [1]. It allows user to see infrared spectrum which is invisible with naked eye. Hence it's willingly used not only during daylight, but especially during nighttime or difficult weather conditions [2] [3]. In this section an overview of the influencing works related to the processing of thermal images, analysis and detection in infrared spectrum is presented and discussed.

A. Object detection in thermal images

Knapik et al. [4] presented eye detection in thermal images scheme using the virtual high dynamic range technique, to enhance performance of the dense grid of scale-invariant

feature descriptors, combined with the bag-of-visual-words approach.

Redmon et al. proposed a series of improved versions of the YOLO architecture, i.e. YOLOv2 [5] and YOLOv3 [6]. Deep convolutional backend network, along with techniques like residual skip connections, residual blocks and upsampling, it is still one of the fastest object detection techniques, while achieving very respectable accuracy. Recently, a thorough refreshment of the YOLO architecture, named YOLOv5, was presented by Jocher et al. [7].

In [8] Bhattarai and MartíNez-Ramón presented intelligent system for real-time object detection and recognition for firefighters during an emergency response. They trained deep Convolutional Neural Network (CNN) to improve situational awareness by identifying objects of interest from thermal imagery in real-time.

In an article from 2020, Gong et al. [9] employed thermal camera for vehicle detection task. In order to achieve faster detection time, the modified YOLOv3-tiny architecture, by recalculating anchor box priors as well as deepening the network structure.

Thermal images are also used to enhance other modalities. Zhou et al. in [10] presented feature fusion network for salient object detection (SOD) task, merging foreground and background information from RGB camera and thermal sensor. Proposed architecture outperforms 12 state-of-the-art methods under different evaluation indicators.

Some researchers propose custom network architectures, designed specifically for infrared images, like Dai et al. in [11]. They proposed TIRNet architecture, which consist of lightweight feature extractor as well as residual branch for regression and classification.

B. Thermal imaging and datasets

One of the biggest problems of thermal imaging is low resolution. To mitigate this problem, Rivadeneira et al. presented novel super-resolution architecture for thermal images based on CycleGAN network [12]. Authors created they own dataset for network training.

Yeduri et al. presented novel low resolution thermal images dataset in [13]. Containing 3200 images of sign language digits captured with very low-res thermal sensor can be used to build human input devices for people with disabilities.

Kristo et al. [2] compared night vision to thermal imaging in their paper and emphasized benefits of using infrared thermal imaging approach over standard RGB. Their research was focused on difficult weather conditions, as described, their dataset was captured during winter in different weather conditions, such as rain, fog or clear weather, during the night. In their paper, YOLOv3 model was trained on custom dataset to detect objects (people) even from far away (up to 215m).

System proposed by Knapik and Cyganek in [14] proved that thermal imaging can be successfully applied to driver's fatigue detection task based on yawn detection. Face alignment is done by detection of eye corners. Then, yawns are detected based on the proposed yawning thermal model.

Thermal imaging was proposed by Farooq et al. [15] to support Advanced Driver-Assistance Systems (ADAS). In their research, thermal imaging was used to capture different objects that are likely to be met on road, such as person, dog, bicycle, car, bike, etc. They also proposed YOLOv5 model trained on their custom dataset. In their work, also a comparison between several available YOLOv5 models was made.

III. EXPERIMENTAL PART

A. Data acquisition

The data provided with this article was collected using the thermal imaging camera FLIR® A35. Acquisition took place in the afternoon, between 3.30 PM and 4 PM with cloudy weather and temperature around -3°C. Videos contain real-life traffic with cars, trucks, buses and people. Camera was placed at elevated footbridge above the street, our setup is shown in Figure 1/ Figure 2 shows RGB image of field of view the setup had. It also depicts exact weather conditions and approximate time of the day.

Figure 1: Acquisition setup
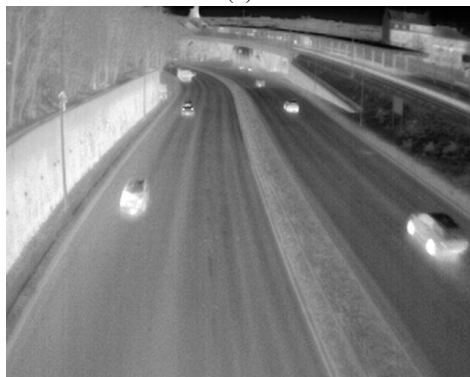

Figure 2: Camera's field of view

Images were manually labeled to provide ground-truth data which is used for training and evaluation. Sample images from the dataset are presented in Figure 3.

(a)

(b)

(c)

(d)

Figure 3: Sample images from the dataset

B. Dataset description

Dataset contains over 6000 annotated images with more than 30000 object instances. Within dataset, 4 classes are annotated, as shown in Table I. Frames were annotated using DarkLabel [16] software. To maintain consistency with COCO dataset, we used the same class IDs. Figure 4 presents number of the instances per class.

Table I: Class IDs

Number	Name
0	person
2	car
5	bus
7	truck

Images in the dataset are in .jpg and .bmp format. The resolution of single image is 320x256 pixels with 8-bit grayscale values. Annotation files are stored in text files with .txt extension in YOLO format [7].
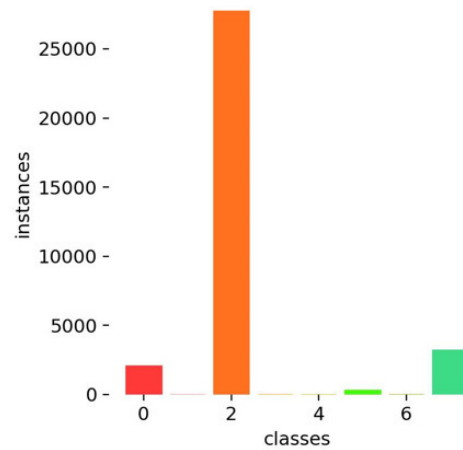

Figure 4: Number of instances in each class.
Classes: 0 - person, 2 - car, 5 - bus, 7 - truck

Dataset is publicly available for all researchers to download from our website: https://home.agh.edu.pl/~cyganek/AutomotiveThermo.zip.

C. Data structure

Dateset contains images and labels as well as trained YOLOv5 object detector. Images are divided into train, validate and test subsets, each stored in a separate folders.

D. Object detection model training

To evaluate the dataset, we decided to train and test object detection model based on YOLO architecture. We chose open-source implementation provided by Ultralytics company [7]. This architecture was chosen due to its availability and ease of use and high quality of code. Due to dataset size and computation speed we decided to use YOLOv5m variant out of other available models (Figure 5). Training was executed in several runs, each with slightly different variables, such as

e.g. epochs number. Finally, after comparing results of all runs, model was trained for 50 epochs, to observe if this would lead to model overfitting. Training results are presented in Figure 6. These results contain graphs of loss subfunctions: box loss, objects loss and classification loss as well as precision, recall and mean average precision (mAP). It is clearly visible, that the more epochs pass by, the more accurate the model is. Figures 7a - 7d show precision, recall and overall score of the model's accuracy. After being trained, model was later tested on new, unseen but labeled images. Figure 8 shows predictions of labeled objects on test data. A closer look of a predictions made by trained model and the confidence levels of detected objects are shown in Figure 9.
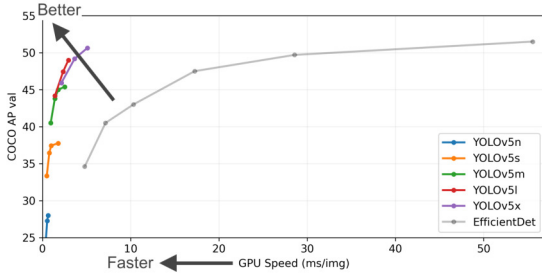

Figure 5: Family of YOLOv5 models
Source: [7]

IV. VEHICLE SPEED MEASUREMENT

Presented dataset might be used for vehicle speed measurement. This can be achieved by calculating distance travelled by a vehicle within some portion of time. It's relatively easy to get the timestamps, as images come with exact date with minutes and seconds in their name. When it comes to getting distance out of collected data, let's take into consideration 2 photos (Figure 10). Timestamp provided with left image is 15:50:44, and timestamp provided with right image is 15:50:45, which means, that exactly one second passed between taking those two images. Let's also consider car marked in yellow circle.

Now, to measure the distance the vehicle has travelled within this time, we need to have some reference. This can be done in several ways, but for our sample application we decided to use the line marks between right and middle lane. Although they are not clearly visible, they still can act as a reference in this experiment. Based on knowledge where the recordings took place and the standards according to which the stripes are painted [17], we can conclude that stripe itself is 2m long, and the gap between 2 stripes is 4m long.

In Figure 11 red lines show there the stripes are, and blue lines represent car's front in regard to the stripes. Now, distance can finally be measured. In Figure 11, considered vehicle travelled 3 gaps and 2.5 stripes, which is equal to

$$3 \cdot 4m + 2.5 \cdot 2m = 17m \qquad (1)$$

All necessary data to calculate the velocity is now available,

(a) Training box loss

(b) Training object loss

(c) Training class loss

(d) Validation box loss

(e) Validation object loss

(f) Validation class loss

(g) Metrics mAP_0.5

(h) Metrics mAP_0.5:0.95

Figure 6: YOLO model train results

thus

$$\frac{17m}{1s} \cdot \frac{3600\frac{s}{h}}{1000\frac{m}{km}} = 61.2\frac{km}{h} \qquad (2)$$

Although no radar data was acquired to back up this calculation, the result is believable and within allowed speed limit on this road, which leads to conclusion, that velocity can be measured by calculating distance travelled within certain timestamped frames.
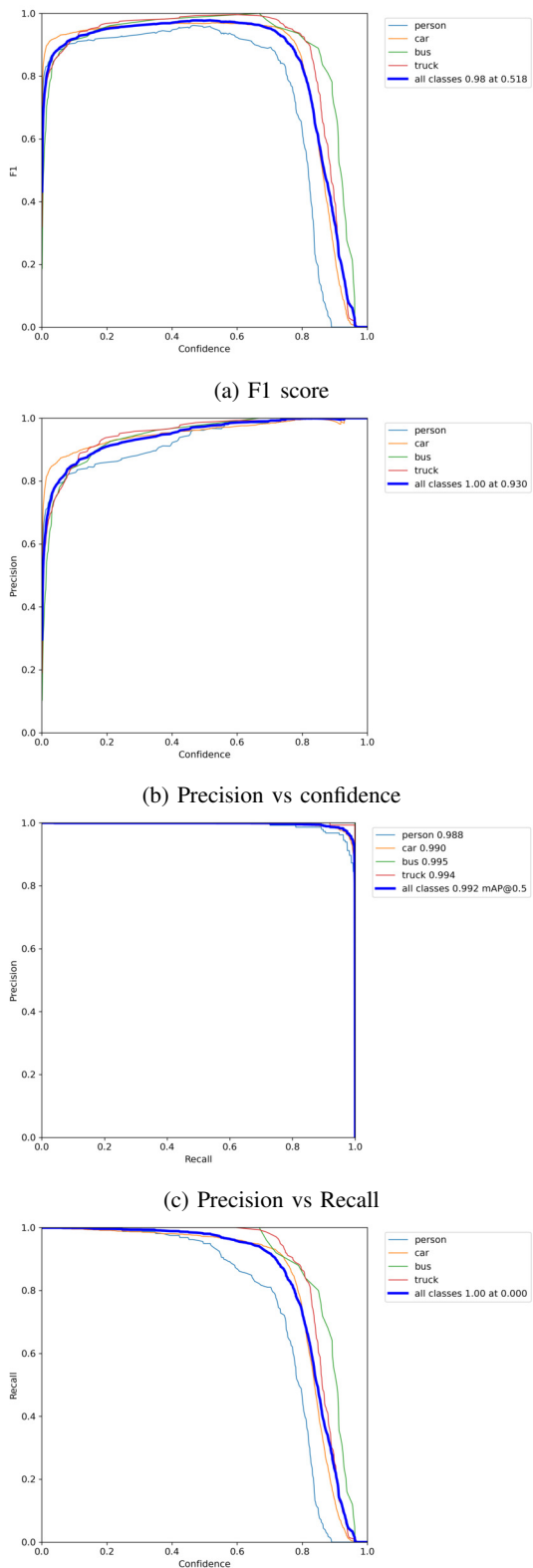
(a) F1 score

(b) Precision vs confidence

(c) Precision vs Recall

(d) Recall vs confidence

Figure 7: Training results.
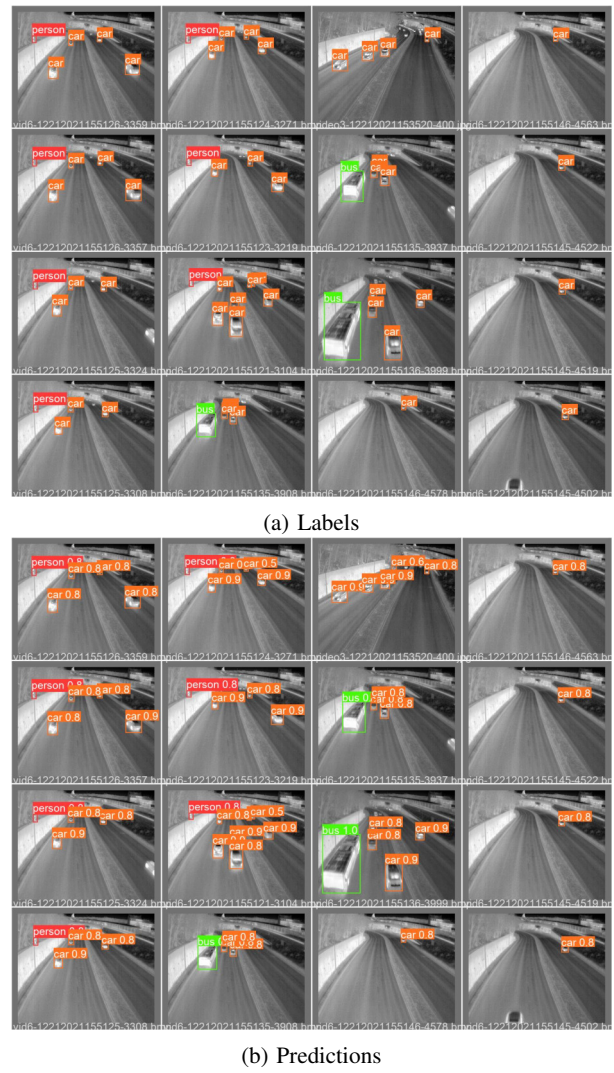

(a) Labels

(b) Predictions
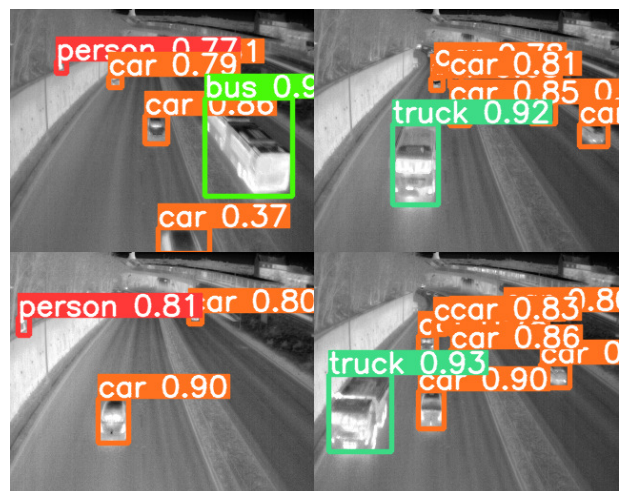
Figure 8: Detection results on test data
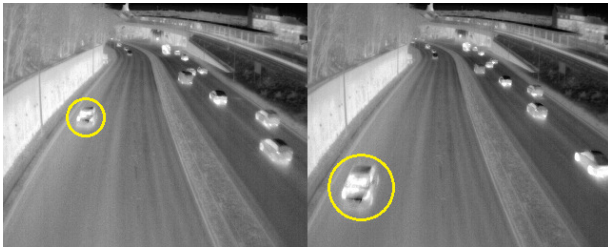

Figure 9: Sample detections

Figure 10: Two pictures, right taken exactly
1 second after left

Figure 11: Highlighted stripes (red) and car's front bumper
position in regard to the stripes (blue)

V. Future possibilities

Dataset presented in this paper can be used to develop computer vision systems for numerous applications, like traffic monitoring, traffic management for smart cities as well as surveillance and advanced driver assistance systems. Thanks to usage of long-wave infrared imaging, such systems will be immune to even the harshest lighting conditions, providing the same level of accuracy in day and night.

VI. Conclusion

Main contribution of this paper is novel thermal imaging dataset with automotive scenes. It contains several thousands images with hand annotated objects. Second contribution of this paper is trained object detector based on YOLOv5 architecture that shows high accuracy in small object detection alongside with the high speed of operation. Both, the automotive thermal database, as well as the pretrained automotive thermal object detection model, are available free for further research on our website. Moreover, we present sample application of our model for car speed measurement based on thermal images. Clear advantages of such approach are also presented. Finally, in the future we plan to further extend our database, as well as develop more resilient trackers that can reliably operate in dense road conditions and with thermal images.

References

[1] J. M. Lloyd, *Thermal imaging systems*. Springer Science & Business Media, 2013.

[2] M. Krišto, M. Ivasic-Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using yolo," *IEEE access*, vol. 8, pp. 125 459–125 476, 2020.

[3] J. Gąsienica-Józkowy, M. Knapik, and B. Cyganek, "An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance," *Integrated Computer-Aided Engineering*, vol. 28, pp. 221–235, 2021, 3.

[4] M. Knapik and B. Cyganek, "Fast eyes detection in thermal images," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3601–3621, 2021.

[5] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, 2017.

[6] ——, "Yolov3: An incremental improvement," *ArXiv*, vol. abs/1804.02767, 2018.

[7] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, TaoXie, J. Fang, imyhxy, K. Michael, Lorna, A. V, D. Montes, J. Nadar, Laughing, tkianai, yxNONG, P. Skalski, Z. Wang, A. Hogan, C. Fati, L. Mammana, AlexWang1900, D. Patel, D. Yiwei, F. You, J. Hajek, L. Diaconu, and M. T. Minh, "ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference," Feb. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.6222936

[8] M. Bhattarai and M. Martinez-Ramon, "A deep learning framework for detection of targets in thermal images to improve firefighting," *IEEE Access*, vol. 8, pp. 88 308–88 321, 2020.

[9] J. Gong, J. Zhao, F. Li, and H. Zhang, "Vehicle detection in thermal images with an improved yolov3-tiny," in *2020 IEEE international conference on power, intelligent computing and systems (ICPICS)*. IEEE, 2020, pp. 253–256.

[10] W. Zhou, Q. Guo, J. Lei, L. Yu, and J.-N. Hwang, "Ecffnet: Effective and consistent feature fusion network for rgb-t salient object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[11] X. Dai, X. Yuan, and X. Wei, "Tirnet: Object detection in thermal infrared images for autonomous driving," *Applied Intelligence*, vol. 51, no. 3, pp. 1244–1261, 2021.

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

[13] S. R. Yeduri, D. S. Breland, S. B. Skriubakken, O. J. Pandey, and L. R. Cenkeramaddi, "Low resolution thermal imaging dataset of sign language digits," *Data in Brief*, vol. 41, p. 107977, 2022.

[14] M. Knapik and B. Cyganek, "Driver's fatigue recognition based on yawn detection in thermal images," *Neurocomputing*, vol. 338, pp. 274–292, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231219302280

[15] M. A. Farooq, P. Corcoran, C. Rotariu, and W. Shariff, "Object detection in thermal spectrum for advanced driver-assistance systems (adas)," *IEEE Access*, vol. 9, pp. 156 465–156 481, 2021.

[16] "Darklabel annotation software," https://github.com/darkpgmr/DarkLabel, accessed: 2022-06-07.

[17] "Dz. u. 26.11.2019, position 2311, szczegółowe warunki techniczne dla znaków drogowych poziomych i warunki ich umieszczania na drogach, section 2.2.1.2," https://sip.lex.pl/akty-prawne/dzu-dziennik-ustaw/szczegolowe-warunki-techniczne-dla-znakow-i-sygnalow-drogowych-oraz-17066287, accessed: 2022-06-07.