Supporting the safety of people and car intelligence with the use of automatic pedestrian detection in thermal image sequences

Aleksandra Górska, Patrycja Guzal, Iga Namiotko, Aleksandra Wędołowska, Martyna Włoszczyńska, Jacek Rumiński

***Abstract*—Over one million people die in car accidents worldwide each year. A solution that will be able to reduce situations in which pedestrian safety is at risk has been sought for a long time. One of the techniques for detecting pedestrians on the road is the use of artificial intelligence in connection with thermal imaging. The purpose of this work was to design a system to assist safety of people and car intelligence with the use of automatic detection pedestrians in thermal image sequences. The system consists of hardware part (ready-made camera modules and RPi computer modules), the software part (implemented AI algorithms) and the database part. The dataset consists of 9178 thermal images presenting pedestrians on the streets of Gdansk. It was shown that the use of transfer learning based on the features learned from the night images results in mAP greater than 85.00% for all six of the investigated algorithms. The best model turned out to be Faster R-CNN ResNet50 FPN which mAP was 94.00%. The designed system finds practical application in increasing road safety through the use of autonomous cars and city monitoring.**

***Index Terms*—deep learning, artificial intelligence, thermal imaging, pedestrian detection, artificial neural network**

I. Introduction

Pedestrian detection plays a major role in ensuring human safety in road conditions. Modern solutions based on artificial intelligence could be a future prospect for every car or its end user. According to WHO approximately 1.3 billion people die every year in road traffic accidents [1]. Frequently accidents happen as a result of driver error. Many of those casualties could be prevented with artificial intelligence-assisted systems, responsible for pedestrian detection. Such systems could assist the driver and contribute to preventing serious accidents. The aim of the research project is to develop, implement and evaluate a system responsible for the contribution to people safety and car intelligence with the use of automatic pedestrian detection in thermal image sequences. The system will consist of a hardware part build with a thermal imaging camera and a computer module. The other part would be the software segment composed of trained AI algorithms. The superiority of thermal imaging in pedestrian detection over RGB images could be proven in irregular conditions, such as in bad weather or nighttime pedestrian detection. Solutions based only on visible spectrum imagery are much less successful in hard conditions [2]. Therefore the system's importance is immense as it has the potential to assure safer conditions on the roads and as a result a prospective of saving multiple lives.

II. Related work

The problem of pedestrian detection on images is well-known, widely investigated, and approached in countless ways. Traditionally the only way to spot a person on the road was to physically notice them with a driver's eyes. Till now this is the most commonly used way, however, the technological progress resulting in installing a pedestrian detection set based on AI solutions within vehicles is promising. Preceding practical applications numerous studies have been conducted in the field. One of the solutions was proposed in a paper [3]. The authors suggested a novel outlook on a traditional problem. Proposed system used high-level features from multiple tasks and multiple data sources, such as backpacks or scene attributes. Finally, the features were combined with a deep learning algorithm. The proposed model, able to learn multiple tasks and data coming from various sources was task-assistant CNN (TACNN). The results of this technique outperformed state-of the-art deep learning models and received a significantly lowerfalse-negative rate than traditional solutions.

Another paper proposes a strategy to use a pre-trained CNN framework on RGB images and to introduce a novel domain [4]. The authors suggest that pedestrian detection based on thermal imagery suffers from poor accuracy, that is why this perspective is introduced. This approach uses two steps strategy. Firstly, the IR images are appropriately prepossessed to receive transformed data as similar to the RGB domain images as possible. This itself results in a satisfactory outcome. However, researchers have also addressed the remaining domain gap by fine-tuning the model with the IR dataset. Presented architecture notably exceeds other commonly used solutions.

Nighttime pedestrian detection is a particularly challenging task to achieve while proceeding detection only on visible spectrum images. The authors of the paper [5] have especially targeted nighttime vision human detection inadequacies. Described model is based on the Faster R-CNN and VGG-16 structures. Within the project's structure, the RDB phase was introduced predicating the masking stage responsible for reducing the decline in detection rate caused by occlusion. After that step, the segmentation part was introduced. The experiment results showed extremely stable performance, superior to the previous solutions, under harsh conditions, such

as night time and occluded images.

The most expected solutions thanks to their prosperity are those exploiting thermal images in combination with various data registered in the visible spectrum. Multiple pieces [6], [7] based on the idea was written to support the predicated results. Diverse multi-spectral systems were implemented, the majority based on convolutional networks used for detection. KAIST multispectral pedestrian benchmark dataset was used for those experiments [8] and experiment findings could be easily compared with state-of-the-art results achieved on this dataset.

However there are some solutions which are still based on a singular input. In this case the input were IR images [9]. The authors proposed to use Thermal Position Intensity Histogram of Oriented Gradients and the additive kernel SVM in order to receive decent pedestrian detection results. The experiments were focused on detecting humans during nighttime which is the biggest fragility of RGB spectrum images.

Another up-to-date experiment was discussed in paper [10] where authors have introduced the saliency maps extracted from thermal images for the model to achieve a better understanding of image contents through pixel-level context. This has the potential to improve the performance of human detection on thermal imagery, especially during daytime. Next, the PiCANet and R3 architectures. The obtained results show that the miss rates from models with vanilla thermal images compared to thermal images and saliency maps are lower. Concluding this novel application is very promising as it already introduces improvements to traditional methods.

Most recently the recommended approach consists of using the generative algorithm to augment data. The common issue of algorithms computed to detect humans on thermal images is the dataset size. To overcome this Least-Squares Generative Adversarial Network model could be used to synthesize thermal images on RGB images inputs. In the paper [11]after augmenting the dataset to the appropriate size, the researchers have applied YOLOv3 [12] neural network previously trained to detect pedestrians on thermal images. The experiments results achieved in this project show that artificially generated thermal images are a good source of pedestrian detection as real thermal images. Also, the results obtained by the model are within the state-of-the-art results range. This shows that lacking data is an issue that could easily be overcome.

III. Materials and Methods

A. Datasets

1) Data acquisition: To conduct an analysis similar to the one presented in the previous part of this article, it was decided to collect dataset consisting of thermal photos of pedestrians in the street independently. In order to do this, it was necessary to select electronic equipment, design an appropriate measurement module and write software that would capture images from the camera. In order to construct the measuring module the electronic equipment was chosen. It was decided to use the Google Coral Development Board, the PureThermal 2 Smart I/O Module and the FLIR Lepton 3.5 Micro Thermal Camera for this task. The next step was to protect the electronic equipment against the wind, moisture, and other unfavorable weather conditions. For this reason, the cover for the module was deigned. The case was then printed using 3D technology. Figure 1 shows final version of the cover.

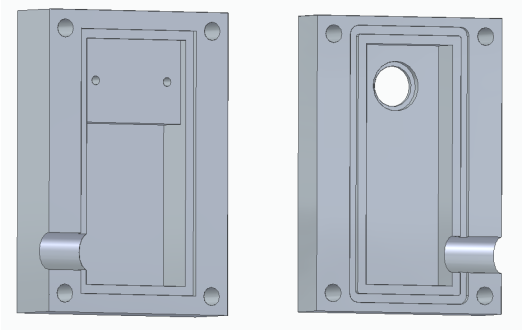

Fig. 1. Cover designed for the camera module.

The cover design also included the selection of appropriate elements ensuring complete tightness of the material. At the same time, it should be reusable, which means that electronic modules could be easily removed and put back in, while still being leakproof. For this reason, it was decided to choose specialized sealing elements and substances, such as a liquid gasket or rubber sealing modules.

Another important aspect was to protect the camera lens against harsh weather conditions and general scratching. For this purpose, the research was conducted in order to select the appropriate glass, which will be able to protect and transmit infrared rays at the same time [13]. The best type meeting the assumed requirements turned out to be germanium glass. Because of the needed size this material was difficult to access. The case was designed for lens of a very small size (a piece with a diameter and thickness of 10 mm and 1 mm respectively).

The measuring module required placing it in the appropriate place of the vehicle. The best solution was to place the camera on the roof rack. The camera cover has been attached to the rack using an universal holder for GoPro cameras. As for the computer module, there was no need to place it outside the car, so it was placed inside (on the dashboard), and both elements of the measurement set were connected with a waterproof USB cable with a length of 3 m. For data acquisition, software capturing images from a camera located on the car's roof was also necessary. For this purpose, the generally available software of FLIR camera manufacturers (UV Capturesb12) has been modified so that it saves individual frames in the device memory in the appropriate format (RAW). The system is shown on figure 2.

2) Data Annotation: From the collected images, only those in which pedestrians were present were selected. A publicly available tool, MakeSenseAI [15], was used to tag pedestrians in the images. This tool has many features, but among other things, it has the option to mark the data with bounding boxes and export the annotations to different formats. The

Fig. 2. Data acquisition system.

formats that were used for subsequent algorithm training were PASCAL VOC XML, MS COCO JSON and YOLO. It should be noted that each of the images of the prepared database was manually tagged.

The collected and tagged data consisted of a total of 9178 thermal images. This data was divided into a training set and a test set - the training set contains 7801 files with 15448 tagged pedestrians, while the test set contains 1377 images with 2731 tagged pedestrians. Created dataset was named AI Thermal Pedestrians (AITP). Figure 3 shows example images (with their annotations) from the mentioned dataset.

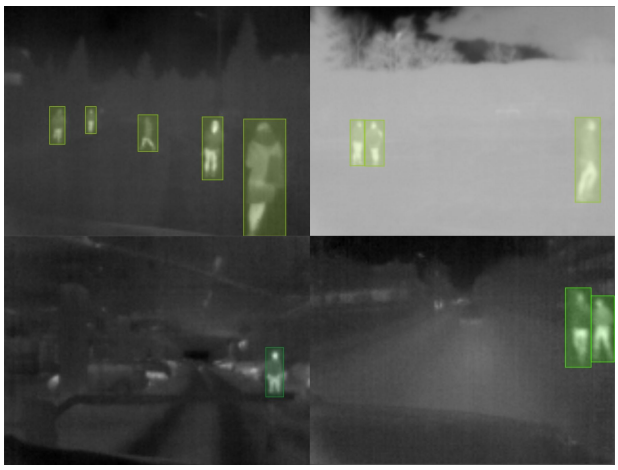

Fig. 3. Images from AITP dataset.

3) Additional datasets: One of the most important stages of the entire project was the selection of an artificial intelligence algorithm to detect pedestrians in thermal images. The process of selecting the best algorithm was not only highly labor-intensive but also needed proper resources. Initially, the collected dataset did not contain a sufficient number of elements necessary to compare the metrics achieved by the algorithms, so it was necessary to find thermal image databases containing a sufficient number of elements that would allow for proper training of neural networks.

The chosen dataset was the KAIST Multispectral Pedestrian Detection Dataset [8]. It covers a greater range of drivable regions, from urban to residential, for autonomous systems. It provides different perspectives of the world captured by day and night. The KAIST Multispectral Pedestrian Dataset consists of 95k color-thermal pairs (640x480, 20Hz) taken from a vehicle. All of the pairs are manually annotated (person, people, cyclist) for a total of 103,128 dense annotations and 1,182 unique pedestrians. The annotation includes temporal correspondence between bounding boxes like Caltech Pedestrian Dataset [16]. Annotations are in the Caltech format [17].

4) Data preprocessing: In order to improve the quality of the images, tests were conducted related to the implementation of the super resolution method in the databases held [18]. Unfortunately, no improvement of the results on the set was noticed, so finally the method was abandoned.

Also, tests were carried out to see if data augmentation would affect the results obtained. The publicly available Albumentations library was used for that purpose [19]. It turned out that the results after augmentation are better, so this technique was used on the KAIST database.

B. Models

1) Adaptation of Models: The first tested algorithm is YOLOv3 (You Only Look Once) to perform object detection in real time. Model was initially pretrained on COCO dataset [20] and obtained weights were used to perform transfer-learning. YOLOv3 applies a single neutral network to the image, which divides the image and predicts bounding boxes and probabilities for each region [21].

Another model that was taken into consideration was Task Conditioned network for Pedestrian Detection task in Thermal Domain [22]. This model introduces task conditioned implementation of YOLOv3 algorithm. During training the learning rate was adjusted appropriately to the current loss gained by model. Model solves two related tasks which are classification and detection. It was pretrained on the KAIST thermal dataset.

The next algorithm was SSD MobileNet V2 FPNLite 640x640. It comes from TensorFlow 2 Detection Model Zoo [23] and it was pretrained on COCO dataset [20]. It is a single-stage detector. In a given algorithm, MobileNet and Feature Pyramid Network (FPN) have the role of feature extractor [24].

The faster R-CNN ResNet101 V1 640x640 is another algorithm which has been adapted and which is included in TensorFlow 2 Zoo detection model [23]. Like the previous algorithm, it was trained on the COCO set [20]. The method includes the faster R-CNN network and ResNet101 as a feature extractor. It is a two-stage detector [24].

Another approach at using Faster R-CNN architecture was with assist of Faster R-CNN ResNet50 FPN, the model pre-

viously trained on COCO Dataset comes from the Facebook AI Detectron2 library [25]. Mentioned algorithm consists of a ResNet-50 Feature Pyramid Network (FPN) backbone.

The last adapted algorithm, EfficientDet D1 640x640, was also a part of TensorFlow 2 Detection Model Zoo [23]. As the rest of the models from this collection it was pretrained on COCO dataset [20]. The model is a weighted bi-directional Feature Pyramid Network (BiFPN) that uses EfficientNet as it's backbone [26].

2) Models Testing Plan: A step-by-step approach to the detection problem was used. In order to test how the algorithms behave in relation to thermal and visible data, attempts were conducted to pre-train the algorithms on existing databases with pedestrian images (Caltech [16], FLIR [27], and KAIST [8]). The following training configurations were tested: visible data, thermal data, and visible data mixed with thermal data.

Unfortunately, numerous databases had flaws that affected model training. These included problems with inaccurate annotations or poor quality images. Ultimately, it was decided to discard the Caltech and FLIR databases and filter the KAIST thermal database in terms of mentioned issues. This resulted in obtaining subset of the KAIST dataset, which was used in later processes.

Finally, three training configurations were selected. The first was to train the algorithms on the filtered KAIST database, the second was to train only on the AITP database, and the third configuration was to train on a mixed dataset consisting of both subset of KAIST and AITP.

Training for each configuration was conducted separately. All things considered, 18 trained network models were obtained. Evaluations were performed on every one of them on a common test set derived from the AITP set. For each of the 6 algorithms trained on the 3 datasets, the model that performed best on the AITP test set was selected.

Table I presents parameters used in training of each model.

TABLE I
TRAINING PARAMETERS OF SELECTED MODELS.

Model name	Base model	Batch size	Optimizer	Initial learning rate
YOLOv3	YOLOv3	32	SGD	0.001
Task Conditioned YOLOv3	YOLOv3	32	SGD	0.001
SSD MobileNet V2 FPNLite 640x640	SSD	8	SGD	0.07999
Faster R-CNN ResNet101 V1 640x640	Faster R-CNN	8	SGD	0.04
Faster R-CNN ResNet50 FPN	Faster R-CNN	128	SGD	0.0025
EfficientDet D1 640x640	EfficientNet	16	SGD	0.07999

3) Metrics: In order to evaluate our addapted models it was decided to use four metrics that are commonly used in detection tasks, those being: mean Average Precision (mAP), True Positives (TP), False positives (FP) and False Negatives (FN).

IV. RESULTS

All six algorithms were tested on test set of collected AITP database. As can be seen in table II the results for each algorithm vary depending on training database. Models trained only on KAIST dataset performed poorly on AITP test images. For each algorithm there is a significant improvement in all metrics when they are trained with AITP training set, the most noticeable one being for SSD MobileNet V2, where the mAP metric improved by over 65 %.

TABLE II
RESULTS OBTAINED BY THE ALGORITHMS.

Algorithm	Training set	mAP50	TP	FP	FN
YOLOv3	KAIST	48.7	1492	268	1239
	AITP	85.33	2366	182	365
	KAIST + AITP	78.63	2229	231	502
Task Conditioned YOLOv3	KAIST	64.12	2049	804	682
	AITP	86.99	2448	288	283
	KAIST + AITP	86.77	2461	398	270
SSD MobileNet V2 FPNLite 640x640	KAIST	22.15	702	202	1669
	AITP	87.96	2538	190	193
	KAIST + AITP	62.50	1769	205	962
Faster R-CNN ResNet101 V1 640x640	KAIST	53.99	1672	362	1059
	AITP	85.78	2407	174	324
	KAIST + AITP	86.57	2469	330	262
Faster R-CNN ResNet50 FPN	KAIST	64.92	1828	163	903
	AITP	92.38	2590	499	141
	KAIST + AITP	**94.00**	**2625**	453	**106**
EfficientDet D1 640x640	KAIST	75,26	2142	312	589
	AITP	91,17	2513	**128**	218
	KAIST + AITP	89.60	2472	164	259

Most algorithms achieve slightly better results when trained only on AITP database, however, the highest mAP value was achieved while training on both KAIST and AITP images. Faster R-CNN ResNet50 FPN was trained on both datasets and reached mAP 94% with the highest number of truly positive recognitions (2625) and lowest count of missing detections (106). Figure 4 shows examples of model predictions. All algorithms, when trained using AITP images, achieved mAP over 85%, with EfficientDet D1 algorithm having as little as 128 False Positive detections.

V. DISCUSSION

Achieved results in pedestrian recognition on thermal imagery are distinct that the consideration of using them in the working products could be initiated. Faster R-CNN ResNet50 FPN with mAP 94% shows a promise of achieving even better precision, if a way to reduce false positive detections was found.

When analyzing the results which included the preprocessing such as super resolution or data augumentations, no improvements were found in models score. Another important aspect that could be noticed is the result obtained on easily

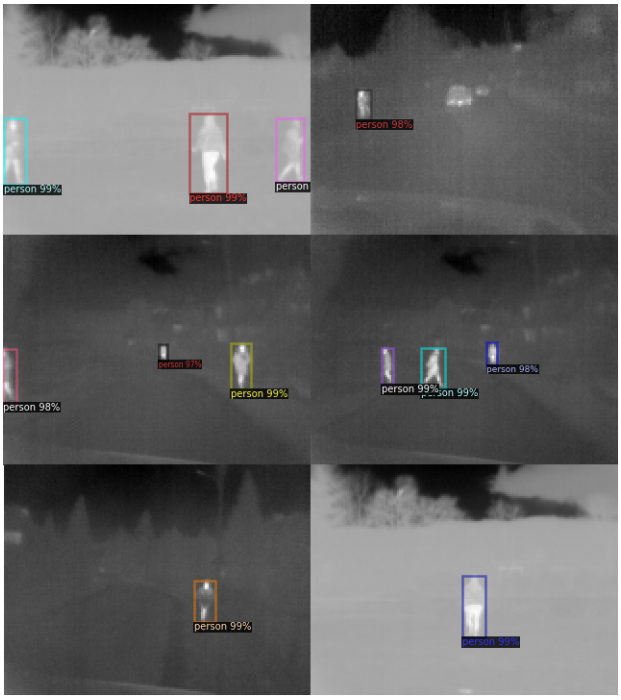

Fig. 4. Predictions of the best model, Faster R-CNN ResNet50 FPN.

accessible datasets such as KAIST. Those were significantly poorer than results that were received on AITP dataset. During the previous works in this project other datasets were also initially evaluated. None of the experiments were satisfactory. This concluded the need of creating a new database, which is well prepared, with more accurately annotated pedestrians and less unclear scenes.

Even if the project would continue to progress mostly in thermal imagery detection it could be significantly improved with a use of an extended dataset. This part of the project is collected database with a substantial amount of data, but as any AI project, pedestrian detection will perform better with bigger database. Collecting images with cameras of various resolutions would allow creating database consisting of images ranging in quality, which would make AITP database more diverse and that would translate to better pedestrian detection results.

Another aspect worth pointing in this project is that compared to the standard optical images, obtained thermal images offer a better human recognition on night-time images and videos. Even though some gathered images contain visual artifacts the results obtained by algorithms are outstanding. In further project development those could be fully reduced and scores additionally improved.

VI. Conclusions

In this work the problem of pedestrian detection on thermal images was addressed. In the process of developing this project, the fully functioning thermal imagery database was formed. In order to obtain the data, the image acquisition system was assembled. System consisting of Lepton Flir 3,5 camera, self designed and 3D printed case and Google Coral AI Board was developed in order to collect thermal images. As a result created database contains 9178 thermal images, with 18,179 pedestrians annotated. This dataset was then used to train six individual deep learning algorithms. Those have resulted in mAP over 85% on gathered images. Comparing the results of other openly distributed datasets and AITP on six used datasets the conclusions has been made that AITP receives significantly better score using the same techniques. Also during the project it was shown that special data preprocessing such as data augumentation or super resolution is not positively affecting pedestrian detection on thermal imagery.

In the future, more algorithms can be trained and tested for their performance in pedestrian recognition to find one that can achieve even better mAP result. Other interesting approach to improve pedestrian detection algorithm results would be combining recognition of people in thermal imaging and visible spectrum. It would require advancing the image acquisition system to one that can capture both types of images, however it would make it possible to create more innovative and universal database.

Use of pedestrian detection is not limited to autonomous cars and driver support. It could be adapted to city monitoring as well as home security and analyzing pedestrian traffic [28]. With further development of the system such implementations could be resolved.

References

[1] WHO | Death on the roads. Available online: https://extranet.who.int/roadsafety/death-on-the-roads/ (accessed on 12.01.2022)

[2] F. Altay and S. Velipasalar, "Pedestrian Detection from Thermal Images Incorporating Saliency Features," 2020 54th Asilomar Conference on Signals, Systems, and Computers, 2020, pp. 1548-1552, doi: 10.1109/IEEECONF51394.2020.9443411.

[3] Y. Tian, P. Luo, X. Wang and X. Tang, "Pedestrian detection aided by deep learning semantic tasks," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5079–5087, doi: 10.1109/CVPR.2015.7299143.

[4] J. Beyerer, M. Ruf, en C. Herrmann, "CNN-based thermal infrared person detection by domain adaptation", 05 2018, bl 8.

[5] J. Baek, S. Hong, J. Kim, en E. Kim, "Efficient Pedestrian Detection at Nighttime Using a Thermal Camera", Sensors, vol 17, bl 1850, 08 2017.

[6] C. Li, D. Song, R. Tong, en M. Tang, "Illumination-aware Faster R-CNN for Robust Multispectral Pedestrian Detection", Pattern Recognition, vol 85, 03 2018.

[7] Y. Hou, Y. Song, X. Hao, Y. Shen and M. Qian, "Multispectral pedestrian detection based on deep convolutional neural networks," 2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2017, pp. 1-4, doi: 10.1109/ICSPCC.2017.8242507.

[8] S. Hwang, J. Park, N. Kim, Y. Choi, en I. S. Kweon, "Multispectral Pedestrian Detection: Benchmark Dataset and Baselines", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[9] V. John, S. Mita, Z. Liu and B. Qi, "Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks," 2015 14th IAPR International Conference on Machine Vision Applications (MVA), 2015, pp. 246-249, doi: 10.1109/MVA.2015.7153177.

[10] D. Ghose, S. Desai, S. Bhattacharya, D. Chakraborty, M. Fiterau, en T. Rahman, "Pedestrian Detection in Thermal Images Using Saliency Maps", 06 2019.

[11] M. Kieu, L. Berlincioni, L. Galteri, M. Bertini, A. D. Bagdanov and A. del Bimbo, "Robust pedestrian detection in thermal imagery using synthesized images," 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 8804-8811, doi: 10.1109/ICPR48806.2021.9412764.
[12] J. Redmon en A. Farhadi, “YOLOv3: An Incremental Improvement”, ArXiv, vol abs/1804.02767, 2018.
[13] C. Salzberg and J. Villa, "Infrared Refractive Indexes of Silicon Germanium and Modified Selenium Glass*," J. Opt. Soc. Am. 47, 244-246 (1957).
[14] PureThermal UVC Capture. Available online: https://github.com/groupgets/purethermal1-uvc-capture (accessed on 9 December 2021)
[15] Make sense AI - Image Tagging Online Software. Available online: https://www.makesense.ai/ (accessed on 12.01.2022)
[16] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian detection: A benchmark," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 304-311, doi: 10.1109/CVPR.2009.5206631.
[17] S. Hwang, J. Park, N. Kim, Y. Choi, I. Kweon, "Multispectral Pedestrian Detection: Benchmark Dataset and Baseline", Integrated Comput.-Aided Eng.. 20. 10.1109/CVPR.2015.7298706.
[18] A. Kwasniewska, J. Ruminski, M. Szankin, M. Kaczmarek, Super-resolved thermal imagery for high-accuracy facial areas detection and analysis, Engineering Applications of Artificial Intelligence, Volume 87, 2020, 103263, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2019.103263.
[19] Albumentations: fast and flexible image augmentations. Available online: https://albumentations.ai (accessed on 10.01.2022)
[20] COCO-Common Objects in Context. Available online: https://cocodataset.org/#home (accessed on 12.02.2021)
[21] YOLO: Real-Time Object Detection. Available online: https://pjreddie.com/darknet/yolo/ (accessed on 02.12.2021)
[22] K. My, A. Bagdanov, M. Bertini, en A. Bimbo, “Task-Conditioned Domain Adaptation for Pedestrian Detection in Thermal Imagery”, 11 2020, bll 546–562.
[23] Tensorflow zoo. Available online: https://github.com/tensorflow/models/blob/master/research/object{_}detection/g3doc/tf2{_}detection{_}zoo.md (accessed on 21.12.2021)
[24] V. Soloviev, F. Farahnakian, L. Zelioli, B. Iancu, J. Lilius and J. Heikkonen, "Comparing CNN-Based Object Detectors on Two Novel Maritime Datasets," 2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2020, pp. 1-6, doi: 10.1109/ICMEW46912.2020.9106019.
[25] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, en R. Girshick, “Detectron2”, 2019. Available online: https://github.com/facebookresearch/detectron2 (accessed on 22.12.2021)
[26] M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10778-10787, doi: 10.1109/CVPR42600.2020.01079.
[27] FLIR Thermal Dataset for Algorithm Training. Available online: https://www.flir.com/oem/adas/adas-dataset-form/ (accessed on 12.01.2022)
[28] D. Kim and D. Kwon, "Pedestrian detection and tracking in thermal images using shape features," 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 2015, pp. 22-25, doi: 10.1109/URAI.2015.7358920.