A Transfer Learning Model Proposal for Country Border Security Using Aerial Thermal Images

Santiago Felipe Luna Romero*. Charles Roberto Stempniak.**
Mauren Abreu de Souza*. Gilberto Reynoso-Meza******

**Graduate Program on Health Technology, Pontifícia Universidade Católica do Paraná, 80215-901, Curitiba Brazil (Tel: 41999851040; e-mail: sanfeluro@gmail.com).*
*** CEO at AUTOMA VISION Tecnologias de Inteligência Artificial Ltda., Paraná, Curitiba Brazil (e-mail: charles@automa.com.br)}*
**** Graduate Program on Health Technology, Pontifícia Universidade Católica do Paraná, 80215-901, Curitiba Brazil (e-mail: mauren.souza@pucpr.br)*
***** Industrial and Systems Engineering Graduate Program, Pontifícia Universidade Católica do Paraná, 80215-901,Curitiba Brazil (e-mail: : g.reynosomeza@pucpr.br)*

Abstract: Smugglers use a variety of tactics to bring goods into a country without paying taxes. These bandits usually operate at night, when the weather and lack of light make it easier for them to move goods. Smugglers' entry points are typically rural areas with a vast distance between them and the city. Thermal cameras and drones are the most effective tools for detecting smugglers because they can identify items that are unaffected by weather, lighting, or body posture, and they can cover enormous amounts of terrain quickly. This research proposes a new thermal model that uses transfer learning from a small hand-labeled database and another model created from a public thermal database to recognize 49 objects with 524,700 thermal images, as well as define and propose a final model of pedestrian recognition in thermal images captured with a drone that has a 94% accuracy.

Resumo: Os contrabandistas utilizam uma variedade de táticas para trazer mercadorias para um país sem pagar impostos. Esses bandidos costumam operar à noite, quando o clima e a falta de luz facilitam a movimentação de mercadorias. Os pontos de entrada dos contrabandistas são tipicamente áreas rurais com uma grande distância entre eles e a cidade. Câmeras térmicas e drones são as ferramentas mais eficazes para detectar contrabandistas porque podem identificar itens que não são afetados pelo clima, iluminação ou postura corporal e podem cobrir enormes quantidades de terreno rapidamente. Esta pesquisa propõe um novo modelo térmico que utiliza aprendizado de transferência de um pequeno banco de dados rotulado à mão e outro modelo criado a partir de um banco de dados térmico público para reconhecer 49 objetos com 524.700 imagens térmicas, bem como definir e propor um modelo final de reconhecimento de pedestres em imagens capturadas com um drone com 94% de precisão.

Keywords: **Pedestrian detection, Thermal imaging, Unnamed aerial vehicle (UAV), Transfer Learning, Yolo, Computer Vision.**

Palavras-chaves: Detecção de pedestres, Imagem térmica, Veículo Aéreo, Transferência de Aprendizagem, Yolo, Visão Computacional.

1. INTRODUCTION

Concerns about smuggling across a country's borders have increased as a result of the rise of smugglers using different methods for bringing merchandise into a country without paying taxes. These contrabandists commonly operate at night, where weather conditions and the absence of light favor them to transport merchandise. The country authorities carry out efforts using all possible technology improvements to prevent negative events and protect the interests of a country and its property (Kristo et al., 2020). The entry routes for the smugglers are generally long extension rural regions very far from the urban areas, which makes it difficult for the authorities to detect them and take control over them. In addition, in these places it is generally not possible to use terrestrial vehicles.

So up to this point there are two problems to detect these smugglers, on the one hand the detection at night with its different climatic conditions; and on the other hand the coverage of large areas efficiently. The first problem warrants the use of thermal cameras because thermal or thermal infrared images (TIR) can make objects stand out due to their temperature difference from the environment, making them immune to weather, lighting conditions or body pose (Gomez et al., 2018) (Bebis et al., 2006). The second problem points to the use of an unmanned aerial vehicles (UAVs, i.e., drones equipped with cameras) that allows covering large areas of land quickly (Sambolek & Ivašić-Kos, 2020).

Due to video surveillance systems and the capability of area controlling in border countries are allowing more sophisticated coverage. A significant factor in this development is the improvements in computer vision technologies. Security concerns the domain of personal security, national security, border protection due to global terrorism threats and illegal migrations, as well as the security of important government and private infrastructure (Ivašić-Kos et al., 2019).

This work proposes the creation of a pedestrian recognition model in aerial thermal images taken by an unmanned aerial vehicle (Drone). The model uses transfer learning from a small hand-labeled database and another model created from a public database to recognize 49 objects with 524,700 properly labeled thermal images.

This work is divided into 5 parts, the introduction is offered in the first part, the second part presents a background and related works, the third part discusses the recognition model proposed, the fourth part gives the achieved findings, and the fifth section presents this research's conclusions.

2. BACKGROUND AND RELATED WORKS

The goal of object detection is to classify certain objects in images and provide their exact position. Many successful machine learning algorithms have been developed for the detection of full human figures in RGB images (Ivašić-Kos et al., 2019).

The IR radiation is electromagnetic radiation emitted in proportion to the heat generated/ reflected by an object and, therefore thermal imaging is referred to as thermal imaging. The wavelengths of IR are longer than those of visible light, ranging from 400 nm to 1400 nm, so IR is not visible to humans (Krišto & Ivasic-Kos, 2018).

Thermal sensors form imagery of the environment or object solely by the detected amount of thermal energy emission of recorded object. They are, unlike the visible sensors, invariant to illuminating conditions and robust to a wide range of light variations and weather conditions. Thermal cameras can be used in security applications in weather conditions where regular RGB cameras produce poor results (Wu et al., 2014).

Thermal cameras are important for surveillance and security because they can be used in weather conditions that ordinary RGB cameras cannot, such as in the night and darkness. Thermal cameras are now more than necessary in video surveillance systems that take care of the safety of people and objects in not urban areas (Kristo et al., 2020).

High-precision object identification is required for effective surveillance and increased security. Because of their application in climatic circumstances, thermal cameras are crucial for security and surveillance (Goel et al., 2021). Several profitable machine learning techniques for tracking and identifying things like pedestrians have been developed (Dalal & Triggs, n.d.; Eveland et al., 2003; Viola & Jones, n.d.).

Traditional object detection systems typically rely on manual annotation, in order to obtain pedestrian features, which have a low accuracy problem. Deep learning has steadily stepped into the challenge of object detection as science and technology have progressed, and has produced good results (Kieu et al., 2021)

Pedestrian targets are non-rigid objects that are easily influenced by a variety of factors such as posture, angle of view, target occlusion, clothing type and thickness. The primary methods of pedestrian detection at night include currently visible image, Lidar, infrared image. The development of infrared pedestrian detection technology has a lot of market potential as well (Kieu et al., 2021).

Traditional approaches, feature-based extraction methods, multi-information fusion methods, and developing in depth learning methods are among the most regularly used pedestrian detection methods. Traditional detection approaches include drawbacks such as limited detection accuracy, generalization ability, and robustness. Feature-based approaches are more accurate, but they involve manual feature extraction and take a long time to be processed (Kieu et al., 2021).

The most common application of thermal imaging is to detect the presence of individuals under varied lighting conditions (Ghose et al., 2019). M. Ivasic-Kos et al. reported that a thermal imaging dataset was used to train YOLO detectors for human detection and examining different weather conditions (Heo et al., 2018; Ivašić-Kos et al., 2019). The weakness in these studies is that they only analyze security camera images that are static; they make no attempt to assess aerial photographs that are moving, such those taken by drones.

Shahid et al., 2019 describes a system for recognizing items in thermal pictures under varied illumination situations that is both accurate and efficient. Background modeling is implemented using modified running Gaussian averaging for real-time execution, and CNN-based human classification is employed just for foreground items.

The research by Goel et al., 2021 offered an approach for object recognition in challenging settings as rain, fog, hazy weather, and darkness. Using threshold segmentation, a framework is meant to segment the objects of the photographs. Finally, the Faster-RCNN is used to recognize pedestrians in thermal images under different lighting conditions.

Using a drone to observe through a high field of view is cost-effective and requires no highly trained personnel. Using a small drone is useful because it allows you to observe in a variety of different fields at the same time, without any special equipment or specialist knowledge (Kumar et al., 2001; Lee et al., 2016).

Thermal cameras can be used in non-visible environments or at night because they inspect objects with different temperatures (Yeom, 2016). Long wavelength infrared (LWIR) has low attenuation in air and it is suitable for detecting ground objects near 300 K that emit most radiation (Moon et al., 2016). The image quality can be varied by the climate and surrounding objects (Park & Yeom, 2021).

People in thermal images captured by a multirotor are detected and tracked in various environments (Park & Yeom, 2021). Humans and animals were detected in difficult weather conditions using YOLO, but no human posture was detected. It is noted that the existing methods of estimating posture usually use visible light images with clear shapes and texture (Park & Yeom, 2021).

When flying at low altitudes, drone records more details on objects of interest. Changes in lighting (day, night and weather) drastically affect the visibility and display of an object. In a single video, a drone can record an object from the front, side, or a bird's eye. The challenge in detecting and monitoring is also posed by the rapid movements of the camera, the occlusions, and the relative movement between the camera and the object (Sambolek & Ivašić-Kos, 2020).

Objects captured by drones are often too small to be seen by humans, so the post-processing need to be automated. The most popular deep learning-based detectors are based on convolutional neural networks (CNN), but they are not equally successful for drone-recorded images (Sambolek & Ivašić-Kos, 2020).

Images taken by drones are being used to improve the results of object detection on drone imagery. Until recently, there were no publicly available datasets recorded by drones for researchers to use in their works (Sambolek & Ivašić-Kos, 2020).

The majority of openly available thermal datasets are for tracking or categorization purposes. They are short sequences with few variations in the scene, such as weather, light, and heat radiation from people. The generalization of detectors suffers as a result of this flaw. The best method would be to look at a huge amount of data and use the transfer learning tool to learn from previous public data (Huda et al., 2020).

A machine learning strategy for transferring information from one domain to another is known as transfer learning. Therefore, in the next field, the data features and model parameters learned previously can be optimized. Transfer learning helps to speed up model convergence, fine-tune data characteristic classification, and enhance classification accuracy (Hu et al., 2020).

There are numerous advantages of transfer learning, including a low demand for data. On similar and different data sets, good generalization and outcomes can also be produced. That is why accelerating the model's convergence speed is beneficial. The training model is consistent and reliable, as well as simple to debug and capable of improving network performance (Hu et al., 2020).

3. PROPOSAL

3.1 Thermal pedestrian object detection model

The proposed model's development process is represented in Fig. 1. It is divided into three phases: the first involves creating a general thermal image identification model using a public database, the second involves creating a small database for this specific problem, and the third involves using transfer learning to define a final model of pedestrian recognition based on drone thermal images.

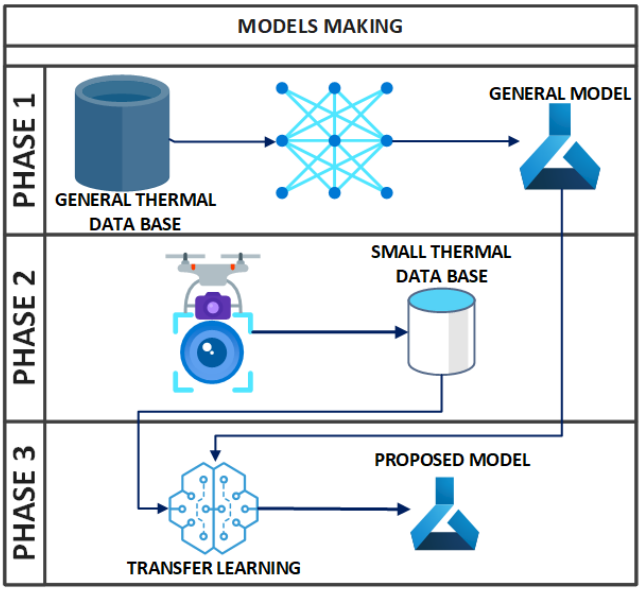

Fig. 1: Making Process for thermal pedestrian object detection model.

3.1.1 General Data base: LSOTB-TIR: A Large-Scale High-Diversity Thermal Infrared Object Tracking Benchmark

The database for the general model was built as part of a project to create a large-scale, extremely diversified general TIR object tracking benchmark that covers real-world occurrences and challenges (Liu et al., 2020). This database has five different types of moving objects of interest (humans, animals, vehicles, planes, and ships) in four different contexts, yielding 1,400 sequences with over 600,000 frames, 49 item subclasses with a total of 524,700 properly labeled thermal images.

3.1.2 Artificial Neural Networks

Machine learning (ML) and deep learning (DL) are methods for generalizing behaviors (models) based on data (data) (Contreras Ortiz & Mora Alvarez, 2018). In particular, supervised machine learning employs a network of artificial neurons to model the connection between a collection of input and output signals using a model that mimics how the brain responds to external stimuli. The literature shows that the most models for **object detection (section 3.1.2)** used widespread at present is convolutional neural networks (CNN).

CNN's overall architecture is made up of a convolutional layer, a grouping layer, and a Full fully connected layer for modeling complex data (J. Li et al., 2019). In general, CNN has multiple convolutional and clustered layer hierarchies, in which multiple convolution runs are performed to extract important characteristics from the input data. In the convolutional layer, neurons from different layers of the network are locally connected through a weight-sharing technique. Furthermore, the convolutional layer comprises forms the core of CNN, which performs convolution and activation operations on the input data to create a feature map, and finally, a Full fully connected layer is used to connect adjacent layers and integrate features to provide a linear output (Lu et al., 2019).

In a convolutional operation, weight sharing is used to link neurons in various layers locally. Kernels are a group of parameters found in convolutional layers. Every kernel has receptive fields that correspond to the input's entire depth. A kernel completes a convolution operation in the feedforward process by calculating a dot product between the kernel and the input to provide a 2D representation of the kernel called a feature map (Lu et al., 2019). The feature map is formed by the representation after an activation function operation, which is computed as showed in Equation 1:

$$y = f(x_i \otimes w_i + b_i) \quad (1)$$

Where:

y: output convolutional operation

x = input

w = kernel

b = bias

$\otimes$ indicates the convolutional operation

3.1.3 Object detection

The major duty in the system is object detection, which determines the bounding boxes and categorization probability for each object. On the MS COCO dataset, **YOLOv4** outperformed all other methods in terms of detection speed (FPS) and detection accuracy (mAP) as the state-of-the-art object detector (Yu et al., 2020). As illustrated in Fig. 2, the YOLOv4 model structure is made up of CSPDarknet-53, Spatial Pyramid Pooling in Deep Convolutional Networks (SPPnet), Path Aggregation Network (PANet), and three YOLO heads. CSPDarknet-53, as the backbone of YOLOv4, is in charge of extracting deep features from the input image via 5 Resblock bodies (C1-C5). Each convolution layer is coupled to a batch normalization (BN) layer and a Mish activation layer in the network, which has 53 convolution layers with sizes ranging from 11 to 33. In addition, in YOLOv4, all activation functions have been replaced with leaky-ReLU, which requires less processing. SPPnet significantly extended the model's receptive field by using different max-pooling layers with sizes of 5, 9, and 13, and PANet extracted features repeatedly using top-down and bottom-up approaches. To detect objects of various sizes, three YOLO heads with sizes of 1919,3838, and 7676 are employed to fuse and interact with feature maps of various scales as illustrated in Fig. 2 (Y. Li et al., 2020).

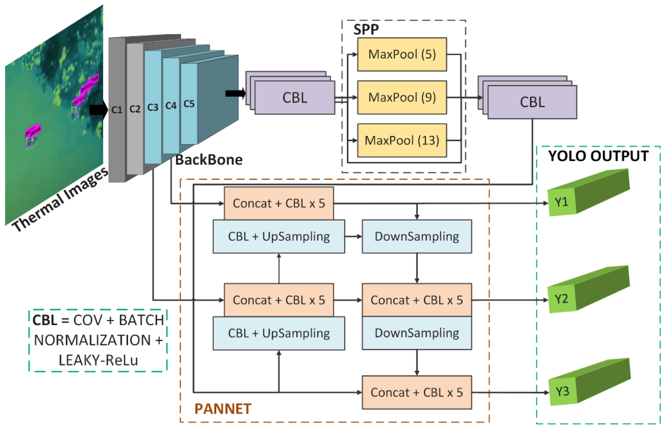

Fig. 2: General structure of YOLOv4 (Adapted from Li et al., 2020).

3.2 Evaluation Metrics

The evaluation technique for a trained Yolo model mostly uses standard object detection metrics like intersection over union (IoU) and mean average precision (mAP).

IoU According to **Equation 2**, the most widely employed indicator for object prediction is the intersection of the expected result and the ground truth over the union. If IoU > 0.5, for instance, the prediction of a bounded box is successful if its IoU is greater than 50%.

$$IoU\ (A,B) = (A \cap B)|\ (A \cup B) \quad (2)$$

mAP: IoU is used as the criterion in Eq. (2) to compute each category's Accuracy. The expected proposal is consistent with the ground truth, as indicated by the fact that TP(c) is a true positive in each class. Then, using Eq 4, mAP, the average of all computed accuracy across all classes is obtained.

$$Accuracy(e) = \frac{\mathrm{Tp}(e)}{Tp(e)+FP(e)} \quad (3)$$

Where **e** represents each class.

$$mAP = \frac{1}{|Num\ Classes|} \sum_{e \in Classes}^{n} \frac{\mathrm{Tp}(e)}{Tp(e)+FP(e)} \quad (4)$$

3.3 Making a database for the specific problem

The images for this study were captured on the border between Brazil and Paraguay in the cities of Foz do Iguazu and Ciudad del Este, by using an aerial vehicle equipped with thermal vision.

To accomplish high-quality image labeling, we start by recording the position of an object with a bounding box. True occlusion occurs when an object is occluded by more than 50% or is out of the vision by more than 50%. This characteristic is useful for training deep models because it may be used to filter out visible noise. We double-check the annotations frame by frame. We ended up with 2920 accurately labeled images. Fig. 3 illustrates some annotated regions from the images data set.

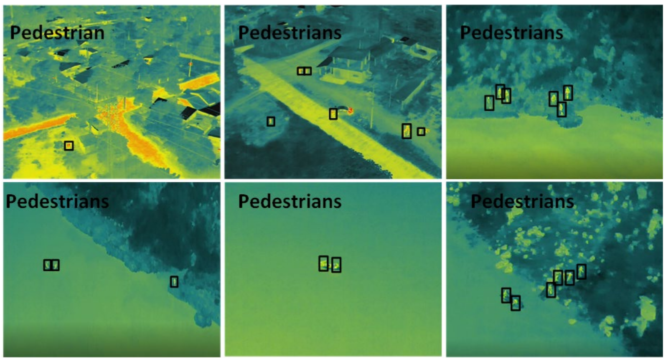

Fig. 3: Making Process for thermal pedestrian object detection model.

3.4 Transfer learning fine tuning to achieve a desired model

One of the most essential and difficult aspects of computer vision is object detection. The majority of existing datasets focus on visible images, whereas thermal infrared images can help with detection even in low-light conditions. To improve the accuracy of infrared pedestrian detection, we apply transfer learning. We believe that the transfer learning method can be used for other infrared sensing applications not only with static images, which is the case in many works in the literature, but also with moving thermal images (taken by a drone) and thus driving. to new advances. (Hu et al., 2020).

Machine learning handles the challenge of a machine learning from data and applying that knowledge to new problems. Transfer learning is a method of learning that makes use of similarities in data, tasks, or models to adapt models learnt in one domain to another. There are few labeled infrared picture datasets for the job of infrared image detection. As a result, we employed a large number of annotated visible images to train a model that was already trained. Then transfer the training data to our task (Hu et al., 2020).

The proposed transfer learning is as follows, all synaptic weights from the network trained with the public database are imported, and the synaptic weights of the Resblock bodies (C1-C5) in Fig. 2 are frozen. After then, fine tuning is given to the other blocks of the neural network by training with its own database, and the results are then evaluated in an evaluation data set.

4. RESULTS & DISCUSSION

The 524,700 thermal images in the database (Liu et al., 2020) were divided as follows: 80% (419,760 images) for training with K-fold cross-validation (K=5) and 20% (104940 images) for test. This model can distinguish between 49 different objects. The result of evaluation step is showed in Table 1. According to the metrics acquired, the problem has been solved by attaining an accuracy of 99.65%, IoU of 84.81%, precision of 97%, and an F1 score of 98% in this model.

Table 1. A general model for recognize 49 objects in thermal images.

Total images			104,940		
Precision	*0.97*	**Recall**	0.99	**F1-score**	0.98
TP	*129,025*	**FP**	4352	**FN**	1075
average IoU			84.81 %		
Mean Average Precision (mAP)			0.996524, or 99.65 %		

In this part, some qualitative results with successful samples are shown in Fig. 4, and then the overall system performance is evaluated in terms of accuracy and other object detection metrics (see further details in Table 2).

The test images obtained by the thermal camera of the aerial vehicle (DRONE) in various authentic scenes under varied weather conditions are used in this inquiry. The suggested system's test results show that it can accurately detect the majority of pedestrians in each image, including some pedestrians who are quite far from the camera.

The final proposed model was developed using a total of 2920 images, in the same way, this data base was divided as follows: 80% (2920 images) for training with K-fold cross-validation (K=5) and 20% for test (584 images). The result of evaluation step is showed in Table 2. According to the metrics acquired, the model is relatively efficient in identifying pedestrians in drone thermal images, with an accuracy of 94.33%, precision of 82%, and an F1 score of 87% in this fine tunned model.

Table 2. Transfer learning fine tuning in order to create a new model.

Total images			584		
Precision	*0.82*	**Recall**	0.92	**F1-score**	0.87
TP	*2,728*	**FP**	600	**FN**	238
average IoU			62.82 %		
Mean Average Precision (mAP)			0.943346, or 94.33 %		

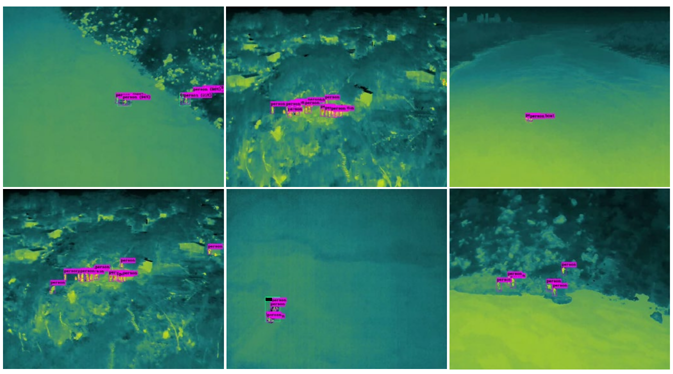

Fig. 4: Successful test cases of the proposed model were visualized in multiple scenarios and under various environmental conditions.

Although some images were taken in various weather circumstances, if the weather becomes severe, the testing process becomes more difficult due to low image quality and bad lighting conditions. Because some of the pedestrians in the second shot are little or hidden, identifying each part of their body for future pose estimation and intention recognition work is tough.

The performance of the entire object recognition model is evaluated in terms of accuracy, precision and F1 score. Table 2 shows that all recognition tasks can achieve an accuracy of over 94.33 percent. The YOLOv4 model was used for the thermal pedestrian drone image detection challenge, and the model's operation may achieve good performance, according to a complete evaluation, and it is suggested for real-time applications.

5. CONCLUSIONS

A thermal vision-based pedestrian identification and recognition framework in drone images was proposed in this study. The task of creating a general thermal object recognition model based on a large database, manual labeling of a small database associated with the specific problem, and finally the application of transfer learning to generate a Pedestrian detection model in thermal images captured with a drone are all included in the proposed framework. The model is built on the use of YOLOv4, which can process data quickly and accurately. In the theme of break smuggling of merchandise at country borders, the pedestrians observed in the images are particularly crucial.

Thermal cameras have the advantage of being able to work in both day and night lighting settings. Especially in different situations, for example: (1) in the presence of non-human moving objects in the scene, such as vehicles and animals, (2) background clutter, such as trees and lamp posts, (3) lack of rigidity of the human body, (4) camera noise, (5) occlusions, (6) human appearance and disappearance, (7) lighting, and (8) scale changes all complicate the task of developing a reliable and robust thermal imaging camera-based pedestrian detection algorithm to operate in real-world open environments.

In the future, in order to improve system performance, a pruning of the YOLOv4 model could be used to improve system performance by reducing processing time and making it suitable for usage in embedded systems that can be implemented directly in the aerial vehicle.

ACKNOWLEDGMENTS

We acknowledge Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) CNPq/MAI-DAI CP nº 12/2020. Additionally, we acknowledge CNPq, by projects 310079/2019-5-PQ2, 4408164/2021-2-Universal and Fundação Araucária (FAPPR) - project 51432/2018-PPP.

REFERENCES

Bebis, G., Gyaourova, A., Singh, S., & Pavlidis, I. (2006). Face recognition by fusing thermal infrared and visible imagery. *Image and Vision Computing*, *24*(7), 727–742.

Contreras Ortiz, P. S., & Mora Alvarez, M. X. (2018). *Aplicación de técnicas de clustering y detección de anomal{\'\i}as para la clasificación de consumos de energ{\'\i}a eléctrica* [{B.S.} thesis].

Dalal, N., & Triggs, B. (n.d.). Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 886–893. https://doi.org/10.1109/CVPR.2005.177

Eveland, C. K., Socolinsky, D. A., & Wolff, L. B. (2003). Tracking human faces in infrared video. *Image and Vision Computing*, *21*(7), 579–590. https://doi.org/10.1016/S0262-8856(03)00056-8

Ghose, D., Desai, S. M., Bhattacharya, S., Chakraborty, D., Fiterau, M., & Rahman, T. (2019). Pedestrian detection in thermal images using saliency maps. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 0.

Goel, R., Sharma, A., & Kapoor, R. (2021). Deep Learning Based Thermal Object Recognition under Different Illumination Conditions. *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 1227–1233. https://doi.org/10.1109/ICESC51422.2021.9532715

Gomez, A., Conti, F., & Benini, L. (2018). Thermal image-based CNN's for ultra-low power people recognition. *Proceedings of the 15th ACM International Conference on Computing Frontiers*, 326–331. https://doi.org/10.1145/3203217.3204465

Heo, D., Lee, E., & Ko, B. C. (2018). Pedestrian detection at night using deep neural networks and saliency maps. *Electronic Imaging*, *2018*(17), 60401–60403.

Hu, J., Zhao, Y., & Zhang, X. (2020). Application of Transfer Learning in Infrared Pedestrian Detection. *2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC)*, 1–4. https://doi.org/10.1109/ICIVC50857.2020.9177438

Huda, N. U., Hansen, B. D., Gade, R., & Moeslund, T. B. (2020). The Effect of a Diverse Dataset for Transfer Learning in Thermal Person Detection. *Sensors*, *20*(7), 1982. https://doi.org/10.3390/s20071982

Ivašić-Kos, M., Krišto, M., & Pobar, M. (2019). Human detection in thermal imaging using YOLO. *Proceedings of the 2019 5th International Conference on Computer and Technology Applications*, 20–24.

Kieu, M., Berlincioni, L., Galteri, L., Bertini, M., Bagdanov, A. D., & del Bimbo, A. (2021). Robust pedestrian detection in thermal imagery using synthesized images. *2020 25th International Conference on Pattern Recognition (ICPR)*, 8804–8811.

Krišto, M., & Ivasic-Kos, M. (2018). An overview of thermal face recognition methods. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1098–1103.

Kristo, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal Object Detection in Difficult Weather Conditions Using

YOLO. *IEEE Access*, *8*, 125459–125476. https://doi.org/10.1109/ACCESS.2020.3007481

Kumar, R., Sawhney, H., Samarasekera, S., Hsu, S., Tao, H., Guo, Y., Hanna, K., Pope, A., Wildes, R., Hirvonen, D., & others. (2001). Aerial video surveillance and exploitation. *Proceedings of the IEEE*, *89*(10), 1518–1539.

Lee, E. J., Shin, S. Y., Ko, B. C., & Chang, C. (2016). Early sinkhole detection using a drone-based thermal camera and image processing. *Infrared Physics & Technology*, *78*, 223–232.

Li, J., Li, X., & He, D. (2019). A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction. *IEEE Access*, *7*, 75464–75475.

Li, Y., Wang, H., Dang, L. M., Nguyen, T. N., Han, D., Lee, A., Jang, I., & Moon, H. (2020). A Deep Learning-Based Hybrid Framework for Object Detection and Recognition in Autonomous Driving. *IEEE Access*, *8*, 194228–194239. https://doi.org/10.1109/ACCESS.2020.3033289

Liu, Q., Li, X., He, Z., Li, C., Li, J., Zhou, Z., Yuan, D., Li, J., Yang, K., Fan, N., & Zheng, F. (2020). LSOTB-TIR: A Large-Scale High-Diversity Thermal Infrared Object Tracking Benchmark. *Proceedings of the 28th ACM International Conference on Multimedia*, 3847–3856. https://doi.org/10.1145/3394171.3413922

Lu, X., Lin, P., Cheng, S., Lin, Y., Chen, Z., Wu, L., & Zheng, Q. (2019). Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph. *Energy Conversion and Management*, *196*, 950–965.

Moon, H.-S., Kim, C. Y., & Lee, W. S. (2016). Drone-based Location Detection Technique for Buried Persons in Collapsed Area. *The Magazine of the Korean Society of Civil Engineers*, *64*(4), 32–35.

Park, D.-M., & Yeom, S. (2021). Human Detection and Posture Estimation with IR Thermal Images by a Drone. *2021 4th Artificial Intelligence and Cloud Computing Conference*, 115–121.

Sambolek, S., & Ivašić-Kos, M. (2020). Detecting objects in drone imagery: a brief overview of recent progress. *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1052–1057.

Shahid, N., Yu, G.-H., Trinh, T. D., Sin, D.-S., & Kim, J.-Y. (2019). Real-time implementation of human detection in thermal imagery based on CNN. 한국정보기술학회논문지, *17*(1), 107–121.

Viola, P., & Jones, M. (n.d.). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, I-511-I–518. https://doi.org/10.1109/CVPR.2001.990517

Wu, Z., Fuller, N., Theriault, D., & Betke, M. (2014). A thermal infrared video benchmark for visual analysis. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 201–208.

Yeom, S. (2016). Multi-level segmentation of infrared images with region of interest extraction. *International Journal of Fuzzy Logic and Intelligent Systems*, *16*(4), 246–253.

Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., & Darrell, T. (2020). Bdd100k: A diverse driving dataset for heterogeneous multitask learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2636–2645.