Seeing in the Dark: A Different Approach to Night Vision Face Detection with Thermal IR Images

Kinshuk Gaurav Singh[1,*,†], Charulkumar Chodvadiya[2,†], Chintan Bhatt[3,*,†], Pooja Shah[4,†] and Alessandro Bruno[5,*,†]

Abstract

Identifying faces in low-light situations can be a difficult feat because of the reduced visibility and substandard image quality. Conventional methods of face detection rely on visible light, which is insufficient in environments with low-light conditions. Our paper introduces a fresh approach to identifying faces in night vision. We make use of thermal infrared (IR) images to detect faces. Thermal IR images capture the thermal signatures of objects, which remain unaffected by low-light conditions, providing valuable information for accurate face detection. Our method utilizes a deep learning model trained on thermal IR images to detect faces in conditions with low lighting. To evaluate our approach, we tested it on a dataset of thermal IR images captured in different lighting scenarios and compared its performance with traditional face detection methods. The results of our experiments indicate that our proposed approach surpasses traditional face detection methods in low-light conditions, achieving high accuracy in detecting faces. Through a qualitative analysis of thermal IR images, we delved into the key factors that lead to our approach's success. Our findings show that the thermal signature of the face offers valuable insights that aid in precise face detection even in low-light environments. Furthermore, we assessed the effectiveness of our approach in detecting faces under different pose and expression variations, and the results indicate that our method is highly efficient in detecting faces in various pose and expression conditions. Our solution presents a novel and efficient method for detecting faces in low-light conditions using thermal infrared images, which has the potential to be utilized in various applications including surveillance, security and law enforcement.

Keywords

Face detection, Thermal images, U-net, dlib, yolo, CNN

1. Introduction

Face detection is a critical task in computer vision with a wide range of applications such as surveillance, security, and law enforcement. However, detecting faces in low-light conditions is a valuable asset and is a challenging problem due to limited visibility and poor image quality. Traditional methods of face detection rely on visible light, which is inadequate in low-light environments. Thus, there is a need for approaches to detect faces in low-light conditions. The ability to see in the dark is a valuable asset in many situations, including security, law enforcement, and search and rescue. Traditional night vision methods are limited by the amount

*Corresponding author.
kgaurav11286@gmail.com (K. G. Singh); charulp2499@gmail.com (C. Chodvadiya); chintan.bhatt@sot.pdpu.ac.in (C. Bhatt); pooja.shah@sot.pdpu.ac.in (P. Shah); alessandrobruno10@gmail.com (A. Bruno)
0000-0002-6511-9219 (K. G. Singh); 0000-0002-6263-4514 (C. Chodvadiya)

of light available, which can make it difficult to see objects in complete darkness, or in low-light conditions.

Night vision face detection is an area of research that has gained significant attention in recent years. Various approaches have been proposed for this problem, such as using visible light cameras with high-intensity infrared illuminators or active 3D imaging systems. While these approaches have shown promising results, they have some limitations, such as high power consumption, limited range, and high cost.

In recent years, thermal infrared (IR) imaging has emerged as a promising technology for night vision face detection. Thermal IR images are created by detecting the heat emitted by objects. This makes them ideal for night vision, as they can be used to see objects in complete darkness, which are not affected by low-light conditions, and can provide valuable information for face detection. Thermal IR cameras are also less affected by environmental factors such as fog, smoke, and dust, making them suitable for outdoor applications. However, using thermal IR images for face detection requires the development of new methods and models that can effectively utilize this modality.

In this paper, we propose a novel approach to night vision face detection using thermal IR images. Our approach involves the use of a deep learning model trained on thermal IR images to detect faces in low-light conditions. We evaluated our method using a dataset of thermal IR images captured in various lighting conditions and compared the performance of our approach with traditional face detection methods. Our approach has a number of advantages over traditional night vision methods. First, it is not affected by light conditions. This means that it can be used to see faces in complete darkness, or in low-light conditions. Second, it can be used to see through camouflage. This means that it can be used to identify people who are trying to hide their faces.

To investigate the factors that contribute to the success of our approach, we conducted a qualitative analysis of the thermal IR images. Our analysis reveals that the thermal signature of the face provides valuable information that can be used for accurate face detection in low-light conditions. We also evaluated the robustness of our approach to different pose and expression variations and found that our method is effective in detecting faces under various pose and expression conditions. Night vision face detection with thermal IR images has a number of advantages over traditional night vision methods. For example, it is not affected by light conditions, and it can be used to see through camouflage. Night vision face detection with thermal IR images has a number of potential applications in security and privacy. For example, it can be used to identify people in dark environments, and it can be used to detect people who are trying to hide their faces.

There has been a lot of research on night vision face detection in recent years. One of the most common approaches is to use a thermal IR camera to capture images of faces. The images are then processed using a computer vision algorithm to detect faces. One of the challenges of night vision face detection is that thermal IR images can be noisy. This can make it difficult for computer vision algorithms to detect faces. Another challenge is that faces can be obscured by objects, such as hats or sunglasses.

2. Literature Review

The method uses the reliable Hough transform approach to retrieve objects while filtering out background activity noise. By using LC-Harris to extract 2D features, the depth of each detected item is calculated. For efficient avoidance, the asynchronous adaptive collision avoidance (AACA) method is used [1]. In [2], The hotspot and background removal techniques were used on a dataset of thermal images. When compared to the hot-spot approach, the accuracy of the background subtraction algorithm was higher at 79%. The background removal approach proved more resilient to variations in lighting and brightness conditions than the other technique, which both employed dynamic thresholds.

While the tracking phase incorporates mean shift tracking and Kalman filter prediction, the detection phase uses a support vector machine (SVM) that uses size-normalized pedestrian candidates. To improve the detection phase, the road-detection module verifies pedestrian observations. using an onboard forward-looking infrared (FLIR) camera, an autonomous vehicle detection system for pedestrians in poor light. To distinguish between infrared pedestrians, low-level Haar-like characteristics are used, while the AdaBoost learning algorithm selects the most pertinent information. A keypoint-based region of interest (ROI) selection approach in IR imageries is suggested in order to efficiently scan sub-windows [4, 5]. By combining Haar features and HOG features in Cascade to classify and validate pedestrians, the system is made robust, reducing the false alarm rate during pedestrian detection and eliminating non-pedestrians in cluttered background situations [10]. A face detector based on HOG-SVM, landmark identification methods based on feature-based active appearance models, deep alignment networks, a deep shape regression network, and an emotion recognition system are all included. Unconstrained input can be transformed into frontal images using facial fractalization algorithms [16].

CNN designs employ hierarchical, discriminative processing for facial recognition and mark regression. The hierarchical filtering approach uses both global and local filtering and addresses the hard "face drifting" and "landmark shaking" challenges. The Landmark quality evaluation system and the Kalman filter both ensure the stability and durability of regional face components [7, 9]. Facial emotion recognition method divides the face into parts and employs active regions (ARs) like eyes and lips. Employing Convolutional Neural Networks (CNNs) and ten-fold cross-validation improves recognition accuracy, further aided by parallelism techniques cutting processing time in half. Decision-level fusion results in a 96.87% recognition accuracy, proving the scheme's robustness and usefulness in adverse conditions. The optimized approach increases average recognition rates by 1%, addressing natural disasters, harsh weather, and low-light scenarios. The technique effectively tracks ARs and enhances pose prediction, leading to improved authentication accuracy [18].

The Viola-Jones approach has the shortest detection time and an acceptable accuracy of 90.5±4.34% The pattern-matching method has an accuracy of 91.6±0.03%, while the active-contour approach has an accuracy of 89.8±0.06%. A Viola-Jones and pattern-matching algorithm combination might improve system accuracy [13, 14]. The stereo-based technology and long-wave infrared-based technology systems [12] can robustly identify the location of the head with high reliability, which can then be utilized to select how the airbag should be deployed. Both algorithms identify and track the head with great precision, with success rates of 96.4%

and 90.1% for varied occupant categories.

The camera processes images using OpenCV software and a Raspberry Pi (RPI), a tiny computer about the size of a credit card. It uses control algorithms to handle alerts and sends recorded photographs through Wi-Fi to the user's email. For both human and smoke detection, the system has an accuracy of 83.56% and 83.33 % respectively. Comparing OpenCV and dlib, the OpenCV library is more effective, performs better for face identification and detection, and has greater productivity [4, 6]. The night-vision system employs a dedicated ODROID XU4 microprocessor running the Ubuntu MATE operating system to process thermal images.The deep learning strategy outperformed the Haar+Adaboost algorithm in terms of false detections [8].

Multiprocess method based on networking middleware that enables real-time face tracking and emotion identification in thermal infrared photos. Approch can detect facial landmarks and recognize facial emotions, even when the faces move around. Furthermore, the approach can recognize face and head postures, which increases its capacity to handle arbitrary head poses [15]. The inherent spatial and spectral aspects of several histogram-based feature descriptors and a set of classifiers to classify thermal emotion. The existence of mixed emotions and inter-person variability are just two of the most likely causes of low accuracy and precision [19].

Dynamic group difference coding is a screening method based on examining influencing factors (D.G.D.C.). D.G.D.C. computes temperature differences between the target person and the recently passed crowd (dynamic group) by describing facial temperature with a face temperature encoder (FTE) and constructing an embedding feature difference matrix. MLPs are used to capture intrinsic information by describing the difference matrix in vertical and horizontal directions [17].

3. Methodology

In this section, we explain the approach to achieving our objectives by applying advanced computer vision techniques to thermal-visual image data while constantly fine-tuning our techniques through rigorous testing and experimentation. We will briefly overview the datasets we use and where they come from to provide some context. Then, we will discuss the primary methods we use, such as Dlib, Haar-cascade, and Yolo v8, and explain why they are important for our analysis process.

3.1. Data

The dataset consists of a total of 2556 pairs of thermal-visual images. Each image pair features participants with meticulously labelled facial areas, including manually defined face bounding boxes and precisely positioned 54 facial landmarks. The dataset was constructed from our large-scale (SpeakingFaces dataset) [20].

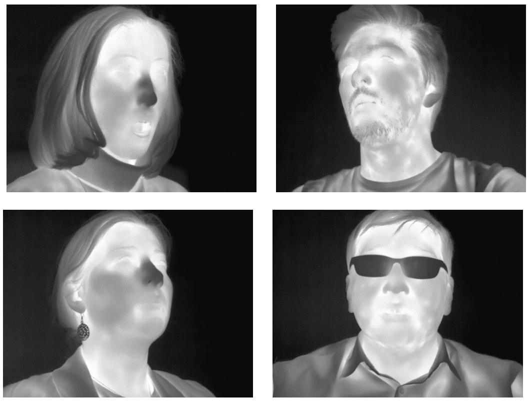


Figure 1: Data Preview : The Thermal-visual Image Data

3.2. Dlib Approach

A complete and potent open-source software ml toolkit, the dlib library provides a wide range of features and capabilities for computer vision applications. The face identification capabilities of the dlib package are effective and precise. Modern algorithms for identifying faces in pictures or video streams are used. The Histogram of Oriented Gradients (HOG) approach and the Convolutional Neural Network (CNN)-based method are the two main face identification techniques offered by dlib.

3.2.1. Histogram of Oriented Gradients (HOG)

In order to capture the distinguishing characteristics of faces, the HOG-based face identification method implemented in dlib makes use of the idea of local picture gradient orientations. It entails calculating gradient orientations in tiny picture areas, creating histograms of these orientations, and representing the local image structure using these histograms. To categorise areas as either face or non-facial, these histograms are then given into a classifier, such as a support vector machine (SVM). The HOG-based technique in dlib is renowned for its success in identifying frontal and near-frontal faces and functions well in a variety of illumination situations[23].

The gradient is obtained by combining magnitude and angle from the image. First Gx and Gy is calculated using the formulae below for each pixel value .

$$G_x(r,c) = I(r,c+1) - I(r,c-1) \tag{1}$$

$$G_y(r,c) = I(r-1,c) - I(r+1,c) \tag{2}$$

where r, c refer to rows and columns respectively. After calculating Gx and, the magnitude and angle of each pixel is calculated using the formulae mentioned below.

$$\text{Magnitude}(\mu) = \sqrt{G_x^2 + G_y^2} \tag{3}$$

$$\text{Angle}(\theta) = \left|\tan^{-1}\left(G_y/G_x\right)\right| \tag{4}$$

3.2.2. Convolutional Neural Networks (CNN)

The CNN-based face identification technique in dlib makes use of deep neural networks to extract distinguishing characteristics from the raw image data directly. A convolutional neural network is trained using methods like convolutional layers, pooling layers, and fully connected layers using a sizable dataset of labeled face photos. The next step is to use the trained network to identify whether a section of an input picture contains a face or not. When detecting faces with complicated postures, occlusions, and scale fluctuations, the CNN-based method in dlib performs very well. It can record both subtle features and broad facial structures, producing extremely accurate face detection outcomes.

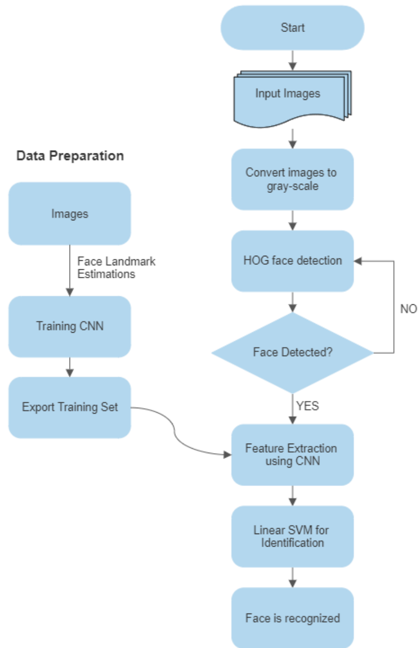


Figure 2: Flow Chart: Dlib Approach for Night Vision Face Detection with Thermal IR Images

3.3. Haar-cascade Approach

By examining the characteristics of an item, the Haar cascade approach uses machine learning to find it. It employs a collection of basic rectangular characteristics known as Haar-like features that are computed at various sizes and locations throughout a picture. The intensity changes between neighboring rectangular picture sections are the source of these characteristics.

The Haar-like features capture Local image fluctuations, which are then fed into a classifier built on the AdaBoost algorithm—as input. AdaBoost creates a strong classifier by combining a number of weak classifiers, successfully teaching the strong classifier how to distinguish between different types of objects. The Haar cascade approach arranges the features in a cascade structure, with a subset of all the characteristics comprising each step. Because the cascade structure may swiftly reject portions of the picture that are unlikely to contain the item being recognized, the processing is efficient[22].

At each cascade step, the classifier assesses the Haar-like characteristics and decides based on a threshold. The area moves on to the next step if the features meet the threshold for possible detection. If not, the algorithm rejects it and continues to the following area of the image.

At each stage, the Haar-like features are calculated for the image's region of interest (ROI). Let's denote the ROI as $I(x, y, w, h)$, where (x, y) is the top-left corner coordinates of the ROI, and (w, h) is its width and height.

The difference between the sums of pixel intensities in two rectangular sections inside the ROI is used to calculate the Haar-like features. A single Haar-like feature can be described by the following formula:

$$f = \sum \text{(pixels in the white rectangle)} - \sum \text{(pixels in the black rectangle)} \tag{5}$$

where $\sum$ denotes the sum of pixel intensities.

A collection of weak classifiers, often based on straightforward decision functions, make up the classifier at each level. The weak classifier may be modelled as follows:

$$h(f) = \begin{cases} \alpha & \text{if } f \geq \theta \\ \beta & \text{if } f < \theta \end{cases} \tag{6}$$

where $h(f)$ is the output of the weak classifier, α and β are the weights assigned to positive and negative detections, respectively, f is the computed Haar-like feature, and θ is the threshold.

The outputs of the weak classifiers are combined at each stage of the cascade structure. Applying a combination rule, such as a weighted sum or a voting mechanism, to the outputs of all steps will yield the final choice.

3.4. YoloV8 Architecture

An important development in real-time object identification, the YOLOv8 architecture provides outstanding performance. The backbone, head, and neck, which together comprise its design and are essential to its outstanding skills, are its three main parts. The backbone uses Powerful convolutional layers to extract high-level information from the input picture, allowing the model to collect detailed characteristics and semantic data required for accurate object detection. YOLOv8 creates a strong foundation for following processing stages by using a backbone network that is resilient[21].

The head module is essential for the extracted characteristics to be optimised and for reliable predictions to be produced. It uses cutting-edge methods to predict bounding boxes, class probabilities, and objectness ratings. The detecting head successfully accumulates object properties at multiple sizes using a variety of convolutional layers, including 1x1 convolutions, allowing for accurate localization and classification. The neck module in YOLOv8 also improves the model's capacity to recognise objects of various sizes and aspect ratios. This intermediate component extends the backbone network's feature collection by including extra context and semantic data. To further improve the model's capacity to handle objects of various sizes and looks, feature pyramid networks (FPN) and spatial pyramid pooling (SPP) are often used for the neck module.

3.5. Hybrid Approach

Our research proposes a hybrid approach that combines the strengths of Dlib and YOLOv8 for object detection, with the aim of optimizing performance in diverse lighting conditions. Initially trained a Dlib-based detector using a carefully annotated dataset collected specifically for low-light, night conditions. This model underwent thorough training and validation to improve its performance in nocturnal scenarios. The Dlib model was then integrated into the YOLOv8 framework, and additional training was carried out using datasets. In the second phase, we fine-tuned the model using a separate dataset captured in daylight conditions, allowing YOLOv8 to adapt and specialize for optimal performance during the day while leveraging the knowledge acquired during the initial Dlib training. This process leverages the strengths of each model to create a hybrid architecture that exhibits superior object detection proficiency in both day and night settings. We took care to maintain consistency in annotation format and hyperparameter tuning for enhanced accuracy. Comprehensive evaluations on independent test sets, including both day and night scenarios, validate the effectiveness of this hybrid model in achieving superior robust detection across diverse lighting environments.

4. Result

Our results comprise using multiple models to perform better at night and day. Face detection can extend beyond day only or night only; the best way is to integrate multi-model architecture trained on various data domains. This multi-headed model can perform well in many conditions.

The U-net results can be observed in fig.3 for shape prediction of the face, which can be helpful in recognizing persons. The red dots show the predicted points, and the green points are the original points.

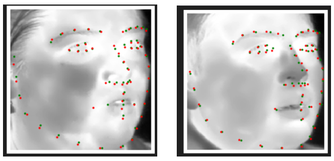


Figure 3: Resulting Image from U-net Shape Predictor

The dlib face detector is trained and tested on [20] with a shape predictor for night images and shows the accurate test results.

Figure 4: Result: dlib Face Detection

The dlib face detector, on top of its results, can be observed in fig.5 for detecting the persons or people's faces at night. It is trained with [20]. The model predicts faces for the night and thermal data on random images for testing model performance.

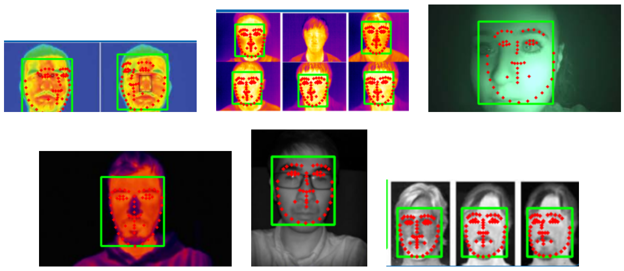


Figure 5: Different Space Test Images of dlib Face Detection

For daylight conditions, this model fails to predict all faces. Thus for the day condition, YoloV8 is used, fig 6. Which performs good for daylight conditions.

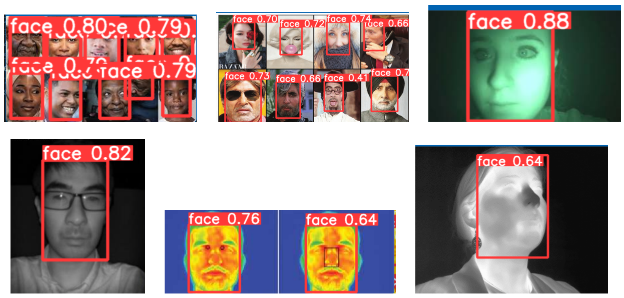


Figure 6: Result: Yolov8 Architecture

Training on a common dataset for day and light can lead to limitations like sometimes detecting noise instead of purely detecting faces, and some background clutter or occlusions can lead to failures in predicting. Combining both models can give face detection far better results than training on common ground, fig. 7 shows promising results.Our hybrid model combines day and night features through multiple models, resulting in a more robust model.

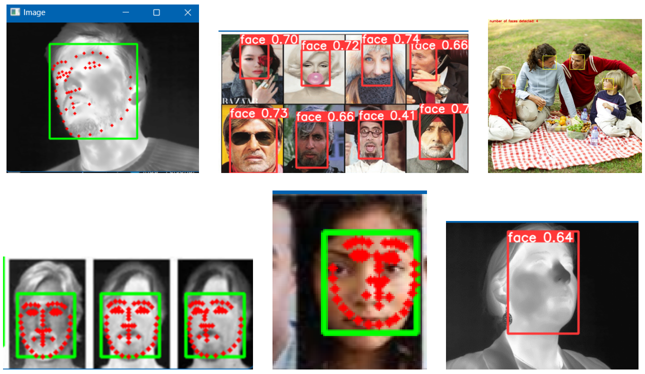


Figure 7: Result Images: Hybrid Model

4.1. Model Performance

4.1.1. Dlib Approach

Testing revealed encouraging results for the face identification system built using the Dlib library, which combines HoG and CNN algorithms. The test error was slightly greater at 3.29 than the train error, which was measured at 2.72. This shows that the model has learned to generalize to new data quite effectively, even at night.

4.1.2. Haar cascade Approach

As we see in 1, it is clear that the Haar cascade algorithm exhibits a 48.15% accuracy in identifying faces in test photos that contain both day and nighttime faces, correctly identifying 13 out of 27 faces. However, the algorithm's accuracy increases to 50%, correctly identifying 9 out of 18 faces, when examining only nighttime photos. These probabilities show the algorithm's varying performance with respect to the time of day and indicate the need for more advancements to increase face detection rates in both situations.

Table 1

Table variables	image result	
	all test image	***only night test images***
No. of correctly detected faces	13/27	9/18
No. of incorrectly detected faces	14/27	9/18

Result of Haar cascade Approach

4.1.3. Final Result Comparison

The table 2 contains all the details for the day and night general case, comparing YoloV8, dlib, and hybrid models.

Table 2

Table 2(i): 1 - Confidence/Probability Result for *dlib*

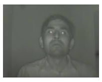
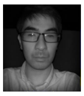
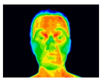
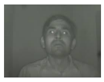
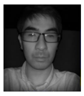
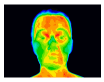
sr.no	*test image*	*dlib*
1.		0.258
2.		0.209
3.		0.815
4.		0.103
5.		0.135
6.		0.1120
7.		0.113
8.		0.1185
9.		0.185
10.		0.115

Table 2(ii): Confidence/Probability Result for *YoloV8* and *hybrid*

sr.no	*test image*	*YoloV8*	*hybrid*
1.		0.76	0.9
2.		0.64	0.9
3.		0.94	0.27
4.		0.32-0.82	0.69
5.		0.32-0.82	0.85
6.		0.32-0.82	0.7
7.		0.32-0.82	0.8
8.		0.32-0.82	0.85
9.		0.32-0.82	0.77
10.		0.32-0.82	0.68

Result Comparison: Table 2(i) and Table 2(ii) serve as subparts of this overarching table. table 2(i) presents the results of the dlib approach, while table 2(ii) provides a results comparison between the hybrid approach and YoloV8 Architecture

5. Conclusions

This research significantly contributes to the advancement of night vision face detection in thermal infrared (IR) imaging. A comprehensive examination of various techniques, including Haar Cascade, HOG+SVM, YOLOv8, and hybrid provides valuable insights for professionals in computer vision and surveillance technologies. Whereas the Yolo trained on the night data performs quite well than others, combining the weights of day and night light as a hybrid model can make a generalized and more accurate ground truth model. The study underscores the importance of method selection based on parameters like speed, accuracy, and computational demands while emphasizing the pivotal role of high-quality, diverse training datasets. The resulting improvements in accurate and efficient face detection systems for low-light environments have the potential to meaningfully enhance security and safety measures, aligning with the overarching objectives of this research endeavour.

References

[1] Yasin, J. N., Mohamed, S. A., Haghbayan, M. H., Heikkonen, J., Tenhunen, H., Yasin, M. M., Plosila, J. (2020, October). Night vision obstacle detection and avoidance based on Bio-Inspired Vision Sensors. In 2020 IEEE SENSORS (pp. 1-4). IEEE.

[2] Sharma, S. K., Agrawal, R., Srivastava, S., Singh, D. K. (2017, March). Review of human detection techniques in night vision. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 2216-2220). IEEE.

[3] Abaya, W. F., Basa, J., Sy, M., Abad, A. C., Dadios, E. P. (2014, November). Low cost smart security camera with night vision capability using Raspberry Pi and OpenCV. In 2014 International conference on humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM) (pp. 1-6). IEEE.

[4] Xu, F., Liu, X., Fujimura, K. (2005). Pedestrian detection and tracking with night vision. IEEE Transactions on Intelligent Transportation Systems, 6(1), 63-71.

[5] Sun, H., Wang, C., Wang, B. (2011, J'anuary). Night vision pedestrian detection using a forward-looking infrared camera. In 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping (pp. 1-4). IEEE.

[6] N. Boyko, O. Basystiuk and N. Shakhovska, "Performance Evaluation and Comparison of Software for Face Recognition, Based on Dlib and Opencv Library," 2018 IEEE Second International Conference on Data Stream Mining Processing (DSMP), Lviv, Ukraine, 2018, pp. 478-482, doi: 10.1109/DSMP.2018.8478556.

[7] Yi Jin, Xingyan Guo, Yidong Li, Junliang Xing, Hui Tian, "Towards stabilizing facial landmark detection and tracking via hierarchical filtering: A new method," Journal of the Franklin Institute, Volume 357, Issue 5, 2020, Pages 3019-3037, ISSN 0016-0032, https://doi.org/10.1016/j.jfranklin.2019.12.043.

[8] Nowosielski, A., Małecki, K., Forczmański, P., Smoliński, A., Krzywicki, K. (2020). Embedded night-vision system for pedestrian detection. IEEE Sensors Journal, 20(16), 9293-9304.

[9] Y. Wu, T. Hassner, K. Kim, G. Medioni and P. Natarajan, "Facial Landmark Detection with Tweaked Convolutional Neural Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 12, pp. 3067-3074, 1 Dec. 2018, doi: 10.1109/TPAMI.2017.2787130.

[10] Govardhan, P., Pati, U. C. (2014, May). NIR image based pedestrian detection in night vision with cascade classification and validation. In 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies (pp. 1435-1438). IEEE.

[11] Kopaczka, M., Nestler, J., Merhof, D. (2017). Face detection in thermal infrared images: A comparison of algorithm-and machine-learning-based approaches. In Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18 (pp. 518-529). Springer International Publishing.

[12] Krotosky, S. J., Cheng, S. Y., Trivedi, M. M. (2004, October). Face detection and head tracking using stereo and thermal infrared cameras for" smart" airbags: a comparative analysis. In

Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749) (pp. 17-22). IEEE.
[13] Ribeiro, R. F., Fernandes, J. M., Neves, A. J. (2017). Face detection on infrared thermal image. Signal, 45.
[14] Kwaśniewska, A., Rumiński, J. (2016, July). Face detection in image sequences using a portable thermal camera. In Proceedings of the 13th Quantitative Infrared Thermography Conference (pp. 4-8).
[15] Kopaczka, M., Schock, J., Nestler, J., Kielholz, K., Merhof, D. (2018, October). A combined modular system for face detection, head pose estimation, face tracking and emotion recognition in thermal infrared images. In 2018 IEEE International Conference on Imaging Systems and Techniques (IST) (pp. 1-6). IEEE.
[16] Kopaczka, M., Breuer, L., Schock, J., Merhof, D. (2019). A modular system for detection, tracking and analysis of human faces in thermal infrared recordings. Sensors, 19(19), 4135.
[17] Yan, M., Qian, J., Wang, R., Gao, S., Yang, J. (2023). Dynamic Group Difference Coding based on Thermal Infrared Face Image for Fever Screening. IEEE Transactions on Instrumentation and Measurement.
[18] Assiri, B., Hossain, M. A. (2023). Face emotion recognition based on infrared thermal imagery by applying machine learning and parallelism. Mathematical Biosciences and Engineering, 20(1), 913-929.
[19] Rooj, S., Routray, A., Mandal, M. K. (2023). Feature based analysis of thermal images for emotion recognition. Engineering Applications of Artificial Intelligence, 120, 105809.
[20] Kuzdeuov, A., Koishigarina, D., Aubakirova, D., Abushakimova, S., Varol, H. A. (2022). SF-TL54: A Thermal Facial Landmark Dataset with Visual Pairs. In 2022 IEEE/SICE International Symposium on System Integration, SII 2022 (pp. 748-753). (2022 IEEE/SICE International Symposium on System Integration, SII 2022). Institute of Electrical and Electronics Engineers Inc.
[21] Glenn,J.Ultralytics, YOLOv8.https://github.com/ultralytics/ultralytics(2023)
[22] L. Cuimei, Q. Zhiliang, J. Nan and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," 2017 13th IEEE International Conference on Electronic Measurement Instruments (ICEMI), Yangzhou, China, 2017, pp. 483-487, doi: 10.1109/ICEMI.2017.8265863.
[23] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.