IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 February 2023, accepted 5 March 2023, date of publication 9 March 2023, date of current version 16 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3255110

SURVEY

On the Role of Thermal Imaging in Automotive Applications: A Critical Review

MUHAMMAD ALI FAROOQ[1], WASEEM SHARIFF[1,2], DAVID O'CALLAGHAN[2], ARCANGELO MERLA[3], AND PETER CORCORAN[1], (Fellow, IEEE)

[1]School of Engineering, University of Galway, Galway, H91 TK33 Ireland
[2]Xperi Inc., Galway, H91V0TX Ireland
[3]Department of Engineering and Geology, G. d'Annunzio University of Chieti–Pescara, 66100 Chieti, Italy

Corresponding author: Muhammad Ali Farooq (m.farooq3@nuigalway.ie)

This work was supported in part by the Electronic Components and Systems for European Leadership (ECSEL) Joint Undertaking (JU) through the Heliaus European Union Project (website: www.heliaus.eu) under Grant Agreement No 826131; and in part by the European Union's Horizon 2020 Research and Innovation Program and the national funding agencies of France, Germany, Ireland, and Italy.

ABSTRACT For decades, the number of automobiles in urban areas around the world has been increasing. It causes serious challenges such as traffic congestion, accidents, and pollution, which have a social, economic, and environmental impact on widespread urban cities. To overcome these challenges, we need to explore smart AI-based perception systems for vehicular applications. Such types of systems can provide improved situational awareness to the driver and generate early alarm about upcoming obstacles and road incidents. In this study, we have presented the effective use of uncooled thermal IR sensors for designing smart thermal perception systems as an alternative to CMOS visible imaging by presenting state-of-the-art studies for in-cabin and out-cabin vehicular applications with potential long-term benefits for the automotive industry. The key rationale for selecting thermal IR sensors over conventional image sensors is that visible cameras are highly dependent on lighting conditions and performance is degraded significantly in low-lighting scenarios and harsh weather conditions. Contrary to this, thermal sensors remain largely unaffected by external lighting conditions or most environmental conditions, making them a perfect optical sensor choice for all-weather and harsh environmental conditions. This study presents a review of the current state of the art for automotive thermal imaging with a focus on the contributions and advances achieved by the EU-funded project 'HELIAUS' in the domain of AI-based thermal imaging pipelines for safer and reliable road journeys.

INDEX TERMS Thermal-infrared, AI, in-cabin, out-cabin monitoring, advanced driver-assistance system (ADAS), deep learning, optimization.

I. INTRODUCTION

Advanced Driver Assistance Systems (ADAS) are a collection of digital technologies that help drivers with safer driving and enhanced security features for reliable road journeys. ADAS improves automotive and road safety by providing a safe human-machine interface. It uses automated technologies such as vehicle sensors and cameras to identify surrounding impediments and driver faults and react correspondingly. Since most traffic accidents are caused by human mistakes [1], ADAS is designed to optimize, adapt, and improve automobile safety thus providing a reliable road journey experience. By reducing human error, ADAS has been shown to minimize fatal accidents [2].

As mentioned before safety features are intended to prevent mishaps and collisions by integrating technologies with the existing vehicular system that notifies the driver of potential hazards, executes protections, and, if required, takes control over the vehicle. Such features include sensor fusion for real-time data logging and object/obstacle detection and tracking system deployment using advanced machine learning algorithms are two key essential technologies directly associated

The associate editor coordinating the review of this manuscript and approving it for publication was Hong-Mei Zhang.

with smart car systems. This will eventually enable drivers to monitor the physical factors, recognize external objects, and forecast occurrences that they should be aware of, giving them a better perspective of the roadside environment and its surroundings. Furthermore, lidar and radar are categorized as typical sensors commonly used in the formation of smart perception systems for automotive sensor suites. Both visible imaging solutions and a variety of hardware sensors are frequently used in conjunction with providing enhanced monitoring systems. However, visible imaging has some restrictions, such that in poor lighting situations, in the nighttime, with sun glare, and glim from the headlight beam, the RGB camera does not provide satisfactory results. Moreover, in computer vision applications, common automotive sensors (radar and lidar) exploit some flaws [3].

Thermal imaging sensor technology overcomes most of these flaws. This study will examine and analyze vision-based smart perception systems, with a core focus on the use of uncooled thermal cameras for advanced driver assistant systems. Thermal imagers are the type of infrared camera, that collects and generates images using infrared thermal radiation generated from the surface temperature of an object. Infrared thermography can be used as an effective method to overcome the limitations of visible or RGB imaging. The real-time operating capabilities of thermal cameras are not affected by low lighting scenarios sun glare or vehicle headlight beam reflection. Further, it has immunity to visual limitations and is considered a reliable solution in harsh weather conditions such as snowy and foggy weather. Uncooled thermal imaging sensors have emerged as low-cost yet effective optical sensors due to recent advancements in microbolometer technology. These optical sensors in the automotive suite can supplement or even replace current technology, with the added benefit of sensing the thermal emissivity of objects and operating independently of illumination conditions, giving it a more consistent option for improved environmental perception systems.

In this research, we have focused on the introduction of thermal imaging which can be beneficial for the design and deployment of thermal perception systems for advanced vehicular systems. Figure 1 shows a comprehensive block diagram representation of various in-cabin and forward-vision applications and vehicular applications using thermal imaging.

The rest of the paper is organized as follows. Section II describes the difference between CMOS and thermal imaging sensors whereas Section III describes the thermal data acquisition pre-processing pipelines and lists the publicly available large-scale thermal datasets. Sections IV and V present published research studies regarding the latest advances in thermal imaging for in-cabin driver and occupant monitoring and out-cabin road monitoring systems. Section VI will elaborate on and detail the significant contributions we made while taking part in the Horizon 2020 HELIAUS [107] project, which is funded by the EU under grant agreement no 826131. The project mainly aims to develop and deploy smart thermal perceptual systems for in-cabin driver monitoring systems and vision-based advanced driver assistance systems thus effectively addressing the inside and outside challenges. The project focuses on creating low-cost and innovative technology thus validating the performance of developed prototypes in perceptual systems for automotive applications. It will measure the added value of thermal sensing and promote the advantages that such systems can offer for autonomous driving. Lastly, section VII presents the overall conclusions drawn based on this study and future possibilities in this domain for the research community.

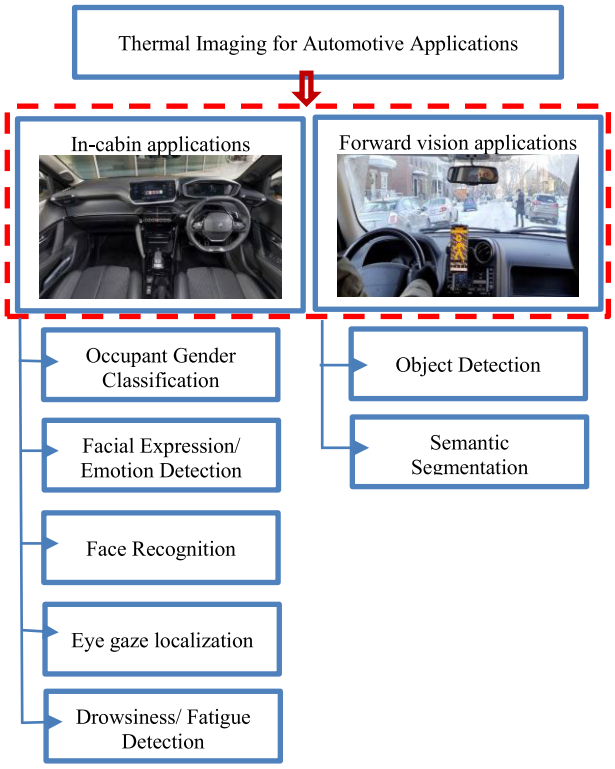


FIGURE 1. Block diagram representation of In-cabin and out-cabin vehicular applications.

II. DIFFERENCE BETWEEN CMOS AND THERMAL IMAGING SENSORS

A Complementary Metal Oxide Semiconductor (CMOS) camera sensor is a type of imager that collects visible light ranging from 400∼700nm band [4] (which is the same spectrum that the human eye perceives). The CMOS sensor works on the theory of the photoelectric effect to convert the photons into electrons using the Analog to Digital (A/D) conversion methodology. In the next stage, it organizes that information to render image frames and sequences of frames. Image sensors assembled into today's digital/ RGB cameras, mobile phone cameras, and CCTVs mostly use either the CCD (charge-coupled device) or CMOS technology. Visible cameras are designed to create images, capturing light in red, green, and blue wavelengths (RGB) for accurate colour representation. As compared to the human eye which requires

visible light, RGB cameras also require light in the visible spectrum to generate images with lower noise levels. Due to this reason visible cameras are considered unfavourable for producing adequate outputs in low-lighting or zero-lighting conditions. Their performance is also significantly degraded by rough atmospheric conditions such as fog, haze, smoke, heat waves, and smog. This limits their usage and applications to daytime and clear weather conditions mostly for real-time applications. Moreover, the CMOS image sensor has the disadvantage of having numerous active devices in the readout path that might cause time-varying noise. Furthermore, fabrication errors can cause charge-to-voltage amplifier imbalance between distinct pixels. Fixed-pattern noise is the result of this, in which distinct pixels output different values despite being exposed to homogeneous illumination [5].

Thermal infrared cameras, in contrast, do not require any additional external lighting conditions to operate and can produce high-image-quality data even in low-light circumstances. As a result, thermal cameras can be used inconspicuously while still being quite effective. This makes thermal cameras the optimal option for applications that are needed both during the day and at night or in low-light conditions. Furthermore, thermal cameras function well in a variety of environmental conditions, such as fog, haze, smoke, or sandstorms, which can hinder visible cameras' performance and render them useless in adverse environmental situations. In this paper, the development of intelligent systems that should stay functional and effective regardless of lighting conditions is the prime motivation for using thermal imaging technology for vehicular applications.

III. PRINCIPLE OF THERMAL IMAGING TECHNOLOGY AND EFFECTIVE USAGE IN COMMERCIAL VEHICLES

This section will highlight the working principle of LWIR thermal cameras, thermal camera configurations, and types of commercially available thermal cameras. Further, this section presents publicly available large-scale thermal and synthetic thermal datasets which can be effectively used in vehicular applications for training and validation purposes and challenges associated with thermal data. The last part of this section highlights the effective usage of thermal sensing in the automotive sensor suite for night vision systems adapted by commercial car manufacturers.

A. WORKING METHODOLOGY OF THERMAL SENSORS

The thermal camera has internal measuring devices that capture infrared radiation, called microbolometers. Thermal infrared radiation is a type of electromagnetic radiation which is comprised of minute particles referred to as photons. Such types of radiation are emitted by all the objects at a surface temperature above absolute zero. From there, the microbolometer records the temperature and then assigns that pixel to an appropriate colour, thus mapping the complete heat map representing the temperature intensities which can be viewed on the camera screen or camera GUI (graphical user interface).

B. THERMAL SENSOR CONFIGURATION AND IMAGE CORRECTION PIPELINE

Thermal cameras based on microbolometer technology generate thermal images by applying a colour palette to the different intensities of Infrared Radiation. However, the way data is gathered and processed is directly influenced by internal sensor configurations. Most thermal cameras come with onboard initial image pre-processing pipelines before recording the data. Some of the commonly used pre-processing steps are as follows.

- Sensor Calibration (shutter/shutterless)
- Re-scaling and denoising techniques
- Automatic gain correction (AGC)
- Bad Pixel Replacement
- Temporal Denoising

Figure 2 shows the three-stage image correction/ processing pipeline once the camera is calibrated using the shutterless algorithm.

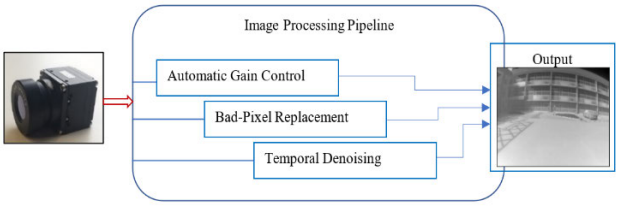


FIGURE 2. Three stages of image correction/ processing pipeline to produce high-quality thermal data.

C. COOLED AND UNCOOLED THERMAL CAMERA

Commercially available thermal cameras can be divided into two distinct categories. This includes cooled and uncooled thermal cameras respectively.

1) COOLED THERMAL CAMERA

Thermal imaging sensor that is coupled with a cryocooler is a core feature of cooled thermal cameras. It is the kind of system that lowers the sensor's temperature to cryogenic levels. To eventually lower the thermally actuated noise to a level below that of the sign from the scene being captured, it is required to lower the sensor temperature. Helium gas gradually pushes past gas seals in cryocoolers, and the moving parts that are designed to extremely tight mechanical tolerances eventually wear out. For applications that demand precise and consistent data, cooled thermal imaging technology is widely regarded as the most sensitive sort of thermal imaging technique. It can even detect very minute temperature changes between objects. They can generate images in the wavelength spectrum of the mid-wave infrared (MWIR) typically with wavelengths of 3-5μm (3000nm to 5000nm) and long-wave infrared (LWIR) range where the thermal complexity is high because of blackbody material. However, cooled MWIR thermal cameras are more costly as compared to the standard uncooled LWIR cameras.

2) UNCOOLED THERMAL CAMERAS

These cameras are developed utilizing technology that eliminates the need for cryogenic cooling of the imaging sensor. A standard uncooled thermal detector relies on the microbolometer sensor array, a silicon component with a large surface area, low heat limit, high thermal segregation, and a moderate vanadium oxide resistor with a big temperature coefficient. The bolometer temperature fluctuates as a result of variations in scene temperature, which is the main working principle of uncooled thermal cameras. The electrical impulses that result from these temperature variations are then processed to create an image. Ferroelectric technology is counted as a different type of microbolometer for the development of uncooled thermal cameras. In this case, slight variations in the material's temperature led to significant variations in electrical polarization. Metal barium strontium titanate is used to make ferroelectric microbolometers (BST). Uncooled sensors are made to operate in the longwave infrared range of 7 to 14 microns, where the majority of the infrared energy is emitted by terrestrial temperature targets. Uncooled cameras are more affordable as compared to cooled thermal cameras.

D. LARGE-SCALE PUBLICALLY AVAILABLE THERMAL DATASETS

Dataset size plays a critical role in the training of deep neural networks (DNN). The bigger the amount of training data the better a DNN can generalize and regularize which can be further used for cross-validation on unseen test data. When coming to the supervised learning methodology for training DNN, datasets must be organized according to their respective class labels. In this section, we will highlight large-scale publicly available thermal datasets along with their respective attributes which can be used for training of DNN or specifically, convolutional neural networks (CNN). However, it is important to mention that as compared to visible imaging datasets we cannot find many large-scale 2D thermal datasets specifically for automotive applications on the open internet. The first part will show the facial thermal datasets whereas the second portion will underline the object detection datasets in the thermal spectrum. These datasets have been acquired using different types of thermal sensors in different environmental conditions as discussed in Table 1 & Table 2 respectively. These datasets are used for training a wide range of pre-trained CNN for various in-cabin occupant monitoring as well as out-cabin vehicular applications. These applications include face detection, thermal gender classification, and object detection in the thermal spectrum on GPU and edge-GPU devices for the automotive sensor suite. This subsection will list the number of face thermal datasets in tabular form (Table 1) which are published and available on the open internet for non-commercial purposes.

This section will list the number of out-cabin object detection thermal datasets in tabular form (Table 2) which are published and available on the open internet for non-commercial purposes.

TABLE 1. Facial thermal datasets attributes.

DATASET NAME	DATASET DETAILS	THERMAL CAMERA USED	DATASET SIZE
Tufts thermal dataset [6-7]	• This dataset is comprised of 113 unique subjects and includes images from six different image modalities that incorporate visible, near-infrared, thermal, computerized sketch, a recorded video, and 3D images of both male and female classes. • The data was collected in an indoor environment under controlled lighting circumstances with diffused lighting. • The images were acquired using a FLIR Vue Pro camera by mounting them at a fixed distance and height.	FLIR Vue Pro Camera	≈ 256 MB
Laval Face Motion and Time-Lapse Video Dataset [8]	• This dataset consists of a total of 238 subjects, which makes it the biggest thermal database. • It is divided into two main categories which include genuine thermal subjects having a total number of 134 and impostor thermal subjects having a total number of 104. • The overall dataset is acquired with a variety of facial poses, expressions, ethnicity, ageing, time-lapse, and opaqueness to eyeglasses thus making it a unique thermal facial dataset.	MWIR Phoenix Indigo IR camera manufactured by FLIR	66 GB
University of Notre Dame (UND) thermal dataset [9]	• This dataset contains LWIR and visible spectrum facial images with an image resolution of 320x240 pixels resolution. • The dataset is acquired using the facial information of 241 people under two illumination conditions. • The database was collected in multiple sessions, and it contains a total of 2492 images.	Merlin-Uncooled thermal camera from Indigo Systems	3 GB

TABLE 1. *(Continued.)* **Facial thermal datasets attributes.**

CARL thermal dataset [10-11]	• This dataset is acquired in two different imaging spectrums which include visible and thermal. • The overall thermal dataset is recorded in 160x120 resolution size. • The complete dataset is consisting of 41 different subjects among which people 32 are male subjects and 9 are female subjects. This dataset consists of 7380 thermal frames.	Thermographic camera TESTO 880-3 which is equipped with an uncooled detector	NA

E. SYNTHETIC THERMAL DATASETS

Convolutional Neural Networks (CNN), which come under the bigger umbrella of Deep Neural Networks, have significantly improved discriminative tasks and are bridging the automation gap. But even so, many computer-vision applications require significant amounts of training data to obtain optimum training results and reliable validation outcomes. Sophisticated pre-trained architectures demand large amounts of training data, such as annotated data, in order to train object detection models. However, this method is expensive, prone to mistakes, difficult, and time-consuming, especially in highly complex, dynamic production environments. This barrier can be addressed by generating and including synthetic data to speed up the training phase of DL from suitable training datasets as the seed data.

Instead of being generated by actual events, synthetic data is a form of information that is derived from a set of real data. It is usually produced with the aid of algorithms and is applied to a variety of tasks in order to supplement and enhance the number of variations in the training data and, as a result, to facilitate the best possible training of deep learning architectures. Moreover, manufacturers can use synthetic data for software testing and quality assurance. Synthetic data can help professionals and researchers to build data repositories that are required to train the networks from scratch and even fine-tune machine learning models, a technique referred to as transfer learning. As discussed in section 3D we cannot find enough large-scale training datasets in thermal imaging modality therefore synthetic data plays a vital role at this point for optimal generalization of deep learning architectures. In this work, we have highlighted various methods for generating synthetic thermal data using the existing thermal datasets listed in table 1 and table 2. These methods include data augmentation or data transformation, generating fake thermal data using Generative Adversarial Networks (GANs), image-to-image translation method, and 2D to 3D

TABLE 2. **Object detection thermal datasets attributes.**

Dataset Name	Dataset Details	Thermal Camera Used	Dataset Size
OSU Thermal [12]	• This dataset is acquired in the day and nighttime environmental conditions. • The overall dataset is consisting of different objects which include persons, cars, and poles • The dataset is recorded with an image resolution of 360x240 with a total of 284 thermal frames.	Raytheon 300D thermal sensor core 75 mm lens	≈ 18 MB
CVC [13]	• This database is acquired during the day and nighttime environmental conditions. • The overall dataset is consisting of different objects which include persons, cars, poles, bicycles, bikes, and buses. • The dataset is recorded with an image resolution of 640x480 with a total of 11 thousand thermal frames.	LWIR FLIR Tau2	5 GB
LITIV [14]	• The overall dataset is consisting of a single class object i.e., person/ pedestrian. • The dataset is recorded with an image resolution of 320x240 with a total of 6 thousand thermal frames.	FLIR A40 LWIR camera	746 MB
TIV [15]	• The overall dataset is consisting of objects from different classes which include persons, cars, bicycles • Images are captured with a resolution of 1024x1024 with a total of 63 thousand thermal frames. • The TIV dataset consists of seven different scenes, among which two of them were recorded in indoor scenes.	FLIR SC8000 cameras	5.43 GB
FLIR [16]	• FLIR dataset is acquired in the day (60%) and nighttime (40%) environmental conditions. • The dataset includes ground truth annotations. • It contains objects from six different classes which include persons, Cars, Poles, Bicycles, buses, and Dogs.	LWIR FLIR IR Tau2	16 GB
	• The complete dataset consisted of 14 thousand thermal frames with a recorded frame size of 640x512.		
KAIST [17]	• Likewise, the FLIR dataset and KAIST dataset is also acquired in the day and nighttime environmental conditions and provides ground truth data annotations. • It contains objects from five different classes which include persons, cars, poles, bicycles, and bus • The dataset is comprised of 95 thousand thermal frames with a frame size of 320×256.	FLIR-A35 thermal camera	≈ 40 GB

face transformation using end-to-end deep learning networks. The generated synthetic data using these methods can be effectively used for the training purposes of pretrained CNN architectures for thermal classification, segmentation, and detection tasks. Table 3 lists the number of publicly available synthetic datasets along with their respective attributes.

TABLE 3. Synthetic thermal datasets attributes.

Dataset Name	Dataset Details	Thermal Camera Used	Dataset Size
Synthetic Depth & Thermal (SDT) Dataset [18]	• This Synthetic Depth & Thermal (SDT) dataset consists of 40k synthetic and 8k real depth and thermal stereo images, showing human behaviour in indoor environmental conditions. • Included samples show uniquely posed lying, sitting, and standing persons within four different room types (living room, bedroom, bathroom, and kitchen), recorded from an elevated position. • Both parts of the SDT dataset contain balanced sets of these four classes and room types. • The synthetic portion of the dataset is proposed to be used as training (and validation) data for single/multi-modal pose classification or person detection models.	FLIR Lepton 3.5 thermal camera	19.2 GB

F. CHALLENGES OF USING THERMAL DATA FOR MACHINE LEARNING ALGORITHMS

While discussing the key advantages of thermal sensing technology for advanced vehicular systems, there are also some challenges associated with thermal data, especially when using it for training and validation of deep learning models. Some of these challenges are as follows.

- Limited number of large-scale publicly available thermal datasets as compared to visible imaging datasets.
- Some of the constraints in publicly available thermal datasets also lead towards the challenge of training dense models on thermal imaging. Such that some of the publicly available thermal data consist of video sequences however with little variability in the scene, i.e., weather conditions, light conditions, and person heat radiation. This drawback reduces the generalization of object detection algorithms.
- Moreover, optimal training and fine-tuning of deep learning (DL) networks on thermal data is a challenging task especially when the networks are pre-trained using RGB data.

G. THERMAL IMAGING FOR NIGHT-VISION SYSTEMS IN COMMERCIAL VEHICLES

Accidental statistics have proved that nighttime driving carries a significant risk. Approximately 50% of fatal car accidents in Germany occurred at night, despite the fact that driving is typically done 75% during the daytime. The same issue occurs in United States as well where 55% of all fatal accidents happen at night, with driving accounting for 28% of all incidents [126]. Keeping this in view, large-scale commercial car manufacturers are more focused towards deploying efficient night vision systems based on thermal imaging technology. The most prominent auto brand in the world Bayerische Motoren Werke (BMW) which is a symbol for high-quality, secure, environmentally friendly and technologically advanced vehicles has already deployed an effective night vision system in its vehicles to avoid fatal collisions. Their night vision systems integrate high-quality FLIR thermal imaging modules capable of detecting and classifying living objects, which include pedestrians and animals in low lighting conditions and harsh weather conditions [127]. Similarly, another Swedish auto-technology group Veoneer had won a production contract with leading car automakers to manufacture thermal cameras for the autonomous vehicle to stimulate additional safety features [128].

IV. IN-CABIN THERMAL MONITORING SYSTEMS

Cabin monitoring goes beyond traditional driver monitoring to include not just the driver, but also passengers and the entire cabin environment. It can safely identify the presence of individuals and objects, as well as assess seat occupancy and seatbelt wearing. For autonomous driving, the car must be aware of not just its passengers' presence, but also their position and state. The car, for example, needs to know if the driver is paying attention and has both hands on the wheel. Customizing driving experience based on the driver's identity, drowsiness/fatigue monitoring, eye localization, or health and safety criteria are further examples of Comfort-related use cases. In this section, we will briefly discuss the in-cabin applications developed using thermal imaging to date. The applications below include both published articles for intelligent vehicles as well as additional relevant publications that can be utilized in intelligent vehicles.

A. GENDER CLASSIFICATION

In addition to smart in-cabin driver monitoring systems, human-computer interaction systems [121], video communications systems [122], psychological analysis [123], [124], and human age and gender classification [125] based on advanced computational algorithms have found various applications for in-cabin driver comfort and safety. Researchers have already developed a gender categorization method based

on visible spectrum images of the human face. However, a variety of factors influence the effectiveness of these systems, including lighting, shadow, occlusions, and time of day. Chen and Ross [19] presented the use of local binary pattern histograms (LBPH) to deduce face gender categorization in thermal and NIR (near-infrared) images, as well as the significance of machine learning algorithms like SVM, Adaboost, and LDA for significantly better gender recognition.

In [20], the authors proposed a Bayesian network with a feature selection method for the explicit and implicit fusion of visible and thermal images to further classify gender. Finally, they tested the presented approaches on the Equinox face database, and the Natural Visible and Infrared facial Expression spontaneous database. The outcomes of the studies indicated that combining feature-level and decision-level fusion improves gender recognition performance when compared to using only one modality. Nguyen and Park [21], proposed a similar idea with HOG and MLBP (multi-level local binary pattern) methods to classify body-based gender using images both from the visible and thermal camera. Further, the same author upgraded their system to use a convolutional neural network to classify males and females from surveillance systems in [22]. Instead of just utilizing video or images, Abouelenien et al. [23] offered a multimodal dataset containing audio-visual, thermal, and physiological readings of males and females to classify gender. They also demonstrated how non-contact physiological measures, such as thermography readings, may improve existing systems that rely on audio or visual input. According to one of the studies [24], the thermal condition and gender of a person can be determined by monitoring physiological indicators from non-intrusive body areas with wearable sensor technologies like humidity sensation, airflow sensation, thermal preferences, and thermal comfort. Further to focus more on just deep learning, [25] proposed a system where they merged multiple CNN models to perform more robustly towards occlusion and low-resolution degradation as well as demonstrate competitive performance.

Using the GENDER-FERET face dataset, Dwivedi and Singh [26] offered a detailed evaluation of deep learning approaches for robust gender identification. They also demonstrated that Convolutional Neural Networks (CNNs) are increasingly being used for feature extraction and classification in various vision applications and that they are suitable because of their high performance. In [27] authors presented a new method for classifying gender that relied on the temperature distribution of the person's ear. It has been discovered that the colder area on the ear is greater in percentage for males than for women, further to train on using simplest neural networks [27]. Koukiou and Anastassopoulos [28] suggested that the selected features of the thermal image can be based on the mean value of the pixels of specified areas on the face, which is a relatively simple way for gender discrimination utilizing thermal infrared images of the person's face. They also demonstrated that discrimination can be accomplished either using simple visualization in the feature space or a reasonably simple neural network [28]. To recognize people's gender in outdoor places where it is difficult or impossible to guard all roads, especially in dim illumination conditions or in the dark, [29] suggested a model that was developed and evaluated utilizing a controlled UAV flight that captured images of humans.

B. FACIAL EXPRESSION/EMOTION DETECTION

Human curiosity leads to a thorough examination of computational models for modelling psychological states and estimating emotion. Human emotion is a pure qualitative entity to be investigated, as the term implies. Basu et al. [30], proposed suggested a non-invasive method for classifying human emotion using thermal images. Hu's moment about distinct patches has been combined with a statistical feature called a histogram and utilized as resilient features in the multi-class support vector machine classification method. In [31], the authors developed a non-invasive technology that relied on thermal value and not its intensity, further image processing techniques that make it possible to identify the difference between the subject and the environment, and a cropped region of interest to better recognize emotions in the thermal spectrum. As an outcome of the research, a smart-thermal system for diagnosing emotions was designed and evaluated on twenty-five people (625 thermograms). This test achieved an overall success rate of 89.9%. Goulart et al. [32] utilizing emissivity variation designed an experiment to analyse emotions in children's thermal images. The research results demonstrate the effectiveness of a design of experiments, including a link between valence and nose thermal decrement; disgust and happiness as effective triggers of facial emissivity variations; and significant emissivity variations in the nose, cheeks, and regions around the eye associated with various emotions. Furthermore, face thermal asymmetry was discovered, with a particular thermal tendency in the cheeks, and classification accuracy was more than 85% on average.

Authors in [33] adapted the Yolo algorithm and proposed heat-map-based face recognition and emotion recognition from thermal images. Further, the algorithm performance was compared with ResNet and DenseNet in terms of precision and intersection over union (IOU). Another promising solution was developed, which combined long-wave infrared imaging (LWIR) with a parallel deep emotion net to improve robustness and accuracy [34]. Authors in [35] improved a Yolo algorithm to assess emotions hidden in the face, such as stress and anxiety and then estimated thermal images using thermal values of pixels rather than intensity values of pixels. The authors found that the modified YOLO-v3 algorithm is an effective method for predicting human emotions. They also claimed that the issue was due to a lack of thermal datasets. In the future, an attempt could be made to create a good dataset that could provide a more accurate result in everyday life, assisting in the prediction of various people human's psychology [35]. Authors in [36]

proposed TFSRNet, a super-resolution network to enhance low-quality thermal images for thermal facial emotion recognition. They used the Convolutional Block Attention Module (CBAM) in both super-resolution architectures to highlight the most significant aspects of each facial emotion while suppressing unimportant elements. Low-resolution thermal facial expression images are enhanced using the suggested super-resolution frameworks, which are obtained using three distinct degradation models: bi-cubic down-sampling (BI), blurring backed by bi-cubic down-sampling (BD), and bi-cubic down-sampling followed by adding random noise (DN). Residual networks are easy to tune and can benefit from increasing depth to improve accuracy. Authors in [37] used the Natural Visible and Infrared Facial Expression (NVIE) dataset with a pre-trained customized ResNet152 to train thermal facial images to predict distinct emotions. Further, Al Qudah et al. [38] surveyed and reviewed a comprehensive analysis of thermal-based imaging and particularly focused on emotions in the thermal spectrum. This study could also assist newcomers to the field of thermal imaging and emotion recognition by allowing them to investigate the various methodologies utilized by researchers to build an affective state system based on thermal imaging. Understanding the current state of humans may help not just with human-to-human communication, but also with human-computer connection (HCI) [38].

Moreover, researchers are also working on building algorithms to detect various illnesses from thermal images based on emotions. Authors in [39] proposed a method to detect attention-deficit hyperactivity disorder (ADHD) syndrome (a sign of behavioural or emotional abnormalities) by employing data fusion analysis for face expression in thermal imaging and deep reinforcement learning to treat behavioural issues. Another example is the evaluation of the problems of patients who are unable to communicate their emotions, such as those with autism. Ganesh et al. [40] proposed a ResNet50 network, a deep-learning technique to detect autism disorders based on thermal imaging. Authors in [40] proposed a stress recognition system using biological signals and thermal images. When a person is anxious, the deep neural network gets facial landmarks as input to take use of the fact that eye, mouth, and head movements are different than usual.

C. FACE RECOGNITION IN THE THERMAL IMAGE

Thermal infrared (IR) images emphasize temperature variations in facial muscles and blood vessels. Temperature variations can be considered texture elements in face images for thermal face recognition [41]. Bhattacharjee et al. [41] proposed a comparative survey of thermal face recognition based on local binary pattern (LBP) and Haar wavelet transform. For each induvial/person the temperature of face muscle and blood flow varies significantly. Authors in [42] published an overview of thermal facial characteristics and approaches that have been successful in face identification, recognition, and verification. The usage of convolutional neural networks and the merging of visual and thermal images were then highlighted as advances in the development of monitoring and surveillance systems. Wu et al. [43], proposed a novel convolutional neural network (CNN)architecture for thermal face recognition. When compared to standard recognition methods like local binary pattern (LBP), histogram of oriented gradients (HOG), and moments invariant, their recommended CNN architecture achieved a higher recognition rate. Further authors in [44], due to advances in CNN, proposed an optimized technique with a short processing time to recognize and detect faces from low-resolution thermal images. The advantages of the suggested network were experimentally tested using thermal video sequences obtained in various settings to overcome potential limits of remote diagnostics, such as the mobility of the person doing the diagnosis and the movements of the person being inspected. The research indicated that the state-of-the-art at that time in image classification and facial detection in thermography had been significantly outperformed. Authors in [45], proposed a thermal to the visible generative adversarial network (TV-GAN). The network was able to transform thermal face images into their corresponding visible light domain images and then perform recognition.

Recognition Systems have gained a lot of interest in the previous few years from academics, entertainment, biomedical, and business groups, among other places. Biometric authentication technologies have risen to prominence as a potential alternative to traditional identification methods. Thermal imaging for facial recognition is used in some systems. The heat transfer action created by the flow of warm arterial blood in arteries is known as convection and for each induvial/person the convection (temperature) of face muscle and blood flow varies significantly [46]. Tamboli and Desai [46] proposed a framework to read vein structure from the thermal face to further extract unique features, as it differs from person to person. Authors in [47], proposed a biometric identification technique, a fusion of visible and thermal images for face recognition. Blood perfusion measurements are defined by localized blood circulation in human tissue, and so are not fully dependent on ambient temperature. A person's blood vessel distribution pattern is unique to them, therefore a collection of extracted notable features from blood perfusion data of a human face should be distinct to that face as well [46], [48]. Following this, authors in [48] presented a neural network based on minutiae (trivial detail of blood vessels) to distinguish faces in thermal images with 91.47% accuracy. Authors in [49] proposed a pose-invariant physiological model for face recognition in the thermal spectrum. It uses image morphology to locate the superficial blood vascular tissue structure. As mentioned before, the contour shapes generated by the recovered vascular tissue are unique for each individual. Data acquired from different poses and the skeletonized vascular tissue branching points are then developed for face recognition in the thermal [49]. Weidlich [50] further reviewed research regarding temperature fluctuations, mathematical formulae, wave kinds, and approaches in thermal infrared face

identification. The authors also proposed that the blood vessel structure and facial vascular networks be exploited for unique biometric characteristics, resulting in a thermal map of the face image. Thermal feature extraction from face images could be achieved by executing morphological procedures such as opening and top-hat segmentation to produce heat signs.

Authors in [51] compared the performance of a convolutional neural network with the conventional random forest algorithm. The evaluation was carried out in a variety of settings, namely normal, with noise while wearing both a facemask and glasses. Furthermore, the research results indicated that the model based on convolutional neural networks performed better in various challenges. Poster et al. [52] proposed visible-to-thermal facial landmark detection in thermal images based on transfer learning. Authors in [53] proposed a transformation model based on multi-scale image synthesis for thermal face recognition. This transformation model is based on a generative model (GAN), with multi-scale categorization and multiple loss functions such as features anchoring, identification conservation, and face landmark-guided texture generation as basic concepts. The results of the analysis indicate that the suggested strategy surpasses the current state of the art. The results of the analysis indicate that the suggested strategy surpasses the current state of the art. Hermosilla et al. [54] proposed StyleGAN-based thermal face generation and further to validate the implementation of the synthetic thermal database, researchers trained six pre-trained deep learning models for face recognition, achieving 99.98% accuracy.

D. EYE-GAZE LOCALIZATION AND ESTIMATION

Almost every driver monitoring system includes eye gaze estimation as a crucial component. The goal of gaze estimation is to determine the point of gaze, or ''where is the person or driver looking.'' This can assist in determining if the driver is paying attention to the road or is distracted. Eye-tracking and localization are difficult in the thermal spectrum. Jin et al. [55] proposed a quick method to locate an eye in infrared images. To begin, they utilized a homomorphic filter to boost image contrast, then used an iterative threshold selection algorithm and integral projection function to segment the human face. They got the eye location approximately using their understanding of the eye and facial geometry. Furthermore, they accurately captured the human eye area using the RAMF (Ranked-order Adaptive Median Filtering) approach. Authors in [56], proposed a novel algorithm for recognition and localization of the face and eyes in thermal images to monitor the temperature of the human body by measuring the eye corner (inner canthus). In the localization phase, the algorithm employs a mixture of layout, knowledge-based, and morphological approaches, especially the modified Randomized Hough Transform (RHT), as well as increasing segmentation to improve the system's accuracy. Further authors in [57], improved the existing technology where the thermal camera detects eye corners and measures the temperatures. The authors developed a device that would automatically measure people's body temperatures as they passed by. People will not need to stop and gaze into a sensor one by one, as they do with existing systems. Multiple people can be scanned at the same time. Authors in [58], to test the effectiveness of thermal eye-tracking, the authors invited ten participants and used passive thermal imaging at 60 Hz to observe their corneal motions. The cornea was then segmented from other regions of the eye in thermal images using a combination of shape models of eyes and an intensity threshold. For 5-point calibration/validation 5 times, they employed an animation sequence as a calibration target. Their results were evaluated to data obtained simultaneously using an SR EyeLink eye tracker at 500 Hz, indicating that eye-tracking using thermal images is possible.

Marzec et al. [59] proposed a fast algorithm for eye localization from thermal images. The algorithm begins with a block for creating characteristics that describe eye regions. The second stage is a neural network-based decision block that allows for the accurate categorization of pre-designated locations. A sophisticated combination of these blocks in a single system allows for accurate analysis of images with a wide geometric variety (size, location, aspect) and brightness distribution, with more than 91%correct localizations and analysis times in a few seconds. Authors in [60], proposed a two-eye detection approach and evaluation in thermal images. A comparison of performance was done on three distinct features: Haar, Histogram of Oriented Gradients (HoG), and Local Binary Patterns (LBP). The HoG function provided the best detection accuracy. Based on a thermal image, the authors in [61] proposed an effective approach for detecting eyes. A unique virtual high dynamic range approach is used for image pre-processing, which considerably improves thermal image contrast and enables the more reliable generation of sparse image descriptors. Their technique was also compared to the YOLO-v3 deep learning model, further, the proposed model achieved robust accuracy and rapid responsiveness in real-world situations without the computational complexity of deep neural networks or the need for a large dataset. Authors in [62] tested the results of two sparse image descriptors for eye recognition in the long-range infrared spectrum. Sparse descriptors of the training images were generated and utilized to build feature vocabulary throughout the training phase. Final detections were made with a bag-of-words technique and a geometrical constraints heuristic. Ferrari et al. [63] proposed an automated method for locating the inner eye canthus (inner eye-corner) in thermal images. They start by detecting five facial key points that correspond to the centre of the eyes, the tip of the nose, and the ears. Then, using a 3D Morphable Face Model, they calculate a sparse 2D-3D point correspondence. Using the YOLO v2 object detector, the authors in [64], presented an automated eye localization approach using IR thermal images. For test images, eye localization in IR thermal images using

YOLO v2 achieved an mAP of 97% and a mean intersection over union (IoU) of 90%.

Most of the previously suggested eye localization algorithms relied solely on frontal positioning. Authors in [65] proposed a novel algorithm for eye localization and face detection in cattle based on multi-view. The authors used HOG filters to understand the features and support vector machines for classification. The paper's results show that the suggested technique had a high level of accuracy, with an average sensitivity of 0.9780, precision of 0.7212, F measure of 0.8024, and misclassification of 0.0455.

E. DROWSINESS/FATIGUE DETECTION IN THERMAL IMAGING

Drowsy driving is a leading cause of deadly car accidents across the world, and it may be avoided with early detection. There are a few key aspects that give thermal imaging an advantage in detecting drowsiness. For starters, unlike visible cameras, thermal sensors are not sensitive to light and do not rely on lighting. Second, drowsiness in the driver causes a decrease in blood flow and a change in face thermal patterns.

There have been significant advances in deep learning in detecting drowsiness/fatigue in drivers. Authors in [66] proposed a non-intrusive method for detecting fatigue and drowsiness in a driver. The authors acquired data from 12 subjects to further study and estimate the blood perfusion level when they are drowsy. The observer rating of drowsiness (ORD) approach was used to measure the individuals' sleepiness levels separately. A four-step method was used to find and monitor facial blood vessels in each image. The research also observed that the facial arteries' temperature decreased from full wakefulness to drowsiness (0.54 °C, 0.33 °C, and 0.32 °C). A series of data was created by the average value of the image intensity of the face blood patches.

Further, the same authors proposed another non-intrusive approach to detect drowsiness in thermal imaging by monitoring the variation in driver respiration rate from wakefulness to drowsiness [67]. According to the findings, the rate of breathing decreases as the subject goes from fully alert to entirely asleep. Physiological aspects of the face were used to pinpoint the zone around the nostrils. The respiration signal was created by adding the average temperature of the nasal area over all frames. Herrera-Granda [68], proposed real-time drowsiness detection by processing the human eye. Using the Viola-Jones algorithm for image processing, the authors advised utilizing the AdaBoost training algorithm, in which a cascade classifier detects the location and region of the driver's eye in each frame. Once the driver's eyes have been identified, colour segmentation and thresholding based on the sclera binarized region are used to determine whether they are closed or open. Furthermore, an audible alert activates when detected driver is drowsy.

Knapik and Cyganek [69] proposed a unique approach to detect fatigue in drivers based on yawn detection in thermal images. Firstly, face alignment begins with the identification of eye corners. Then the suggested yawning thermal model is then used to detect yawns using the Viola-Jones algorithm with a cascade classifier. An annotated image database was established for quantitative assessment and made publicly available [69]. Kajiwara [70] examined and reviewed if there's the possibility of monitoring drivers using thermal infrared imaging. Also included in the article were proper baselines, the independent nature of thermal imprints, experimental data, methodological considerations, and limitations. Authors of [66], [67] further proposed an upgraded version of analyzing respiration to detect drowsiness in thermal imaging. The authors proposed Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) classifiers to be used to detect fatigue. The results indicated a good accuracy of 90% with a precision of 91% [71].

Authors in [72] proposed a new benchmark dataset for driver fatigue research. It contains thermal images, depth maps as well as visible images. Kajiwara [73] attempted to improve the real-time estimation method of low alertness rate in drowsiness detection. The author tested the model on full bright and low lighting condition images, the detection rate was found to be weak, and facial landmark detection was misaligned, the ear aspect ratio and mouth aspect ratio could not be determined reliably. Therefore, it was discovered that there is an issue that leads to the identification of a low wakefulness state and the incorrect detection of yawning. As a result, utilizing thermal images of the face produced from an infrared thermal camera that can be utilized in backlight and night-time conditions, the author developed a low-alertness state estimate method.

Tashakori et al. [74] proposed yet another good approach to detecting drowsiness. The authors located the forehead and the cheek skin temperature in thermal images and trained it on the Support Vector Machine, the K-Nearest Neighbor, and the regression tree classifiers. The drowsiness was detected with an accuracy of 82%, the sensitivity of 85%, specificity of 90%, and precision of 84%, according to the research. The authors in [75] proposed drowsiness detection using multimodal data. The proposed methodology suggested an analysis of the effects of early fusion on the classification of the driver's state using multiple physiological and thermal channels. The research outlined that it is better to detect drowsiness using a multimodal approach as it gives two separate factors and a clear picture of which is influencing the driver, drowsiness, or distraction [75].

V. OUT-CABIN THERMAL MONITORING SYSTEMS

Sensors are becoming increasingly important in advanced driver assistance systems, vehicle automation, vehicle networking, and new mobility services as technology advances. With highly automated driving levels, not only are the interior support systems improving, but the cars' exteriors are redesigning themselves to create a complete out-vehicle experience as well. To avoid any mishaps, a vehicle must be aware of its surroundings. Many out-of-cabin applications, namely object identification, and image segmentation have

recently been developed. This section will go through these applications and current advances in depth.

A. OBJECT DETECTION IN THE THERMAL SPECTRUM

Detecting objects in the thermal spectrum has numerous advantages. Thermal imaging produces better and more practical outcomes in difficult conditions, including poor light scenarios, and weather conditions, and is resistant to optical limitations in general. Recent advances in deep learning and a shift away from traditional machine learning have had a significant influence on thermal object detection.

1) MULTIMODAL APPROACHES

Devaguptapu et al. [76], proposed a pseudo-multimodal object detector. The proposed network borrows features from rich domains like visual RGB. They generated pseudo-RGB versions of a given thermal image using well-known image-to-image translation architectures and then utilized a multi-modal framework to detect objects in the thermal image.

Various applications in driver assistance such as tracking, monitoring, and multispectral pedestrian detection have become increasingly relevant in the field of computer vision. The authors in [77] suggested a deep learning-based brightness estimation model for pedestrian detection. The proposed unique brightness estimation technique presents various illumination circumstances to predict in both day and nighttime. The suggested technique performed well on the FLIR-ADAS Thermal dataset, with an mAP of 81.27%.

Agrawal and Subramanian [78], proposed a study that will examine the effectiveness of object detection with a fusion of visible and thermal images in a publicly available dataset. They provided a comparison of object detection in night images and showed that thermal images boost detection accuracy considerably. Object detection, such as persons and vehicles, is critical for autonomous driving. ThermalDet is a DNN-based, one-stage detector suggested by the authors in [79]. ThermalDet's fundamental notion is that, since thermal images lack many precise visual qualities (such as color and texture), features from low and high levels are equally essential when conducting detection tasks on images. The suggested detector builds on RefineDet's design and enhances it. To begin, they build a dual-pass fusion block (DFB) that allows them to immediately merge features from all layers. Later they introduced a channel-wise enhance module (CEM) to iteratively allocate relative weights to channels of feature maps [79].

The resolution of objects in the thermal spectrum is frequently poor. For real-time object detection in embedded applications, Talluri and Dua [80] proposed a modified tiny Yolo v3 trained on FLIR thermal dataset. Thermal cameras are also an essential component of advanced video surveillance systems due to the inability to use RGB cameras adequately at night and in adverse weather circumstances. Kristo et al. [81] proposed utilizing convolutional neural network models initially designed for detection in RGB images to detect people in thermal images. They examined the performance of state-of-the-art object detectors, such as Faster R-CNN, SSD, Cascade R-CNN, and YOLOv3, and retrained on a dataset of thermal images acquired from recordings. It was also observed that YOLOv3 performed significantly better than other detectors. Again, to focus on the small objects in thermal images, authors in [90] proposed a DDSSD (dilation and deconvolution single shot multi-box detector). This network is a modified or enhanced version of SSD with a novel feature fusion module for small object detection. The network achieves a mAP of 79.7 % with an FPS of 41 with a 300×300 input image.

2) DOMAIN ADAPTATION

The authors in [82] proposed a survey to examine the current state of the art in deep-domain adaptive object detection techniques in both the optical and thermal domains. They began by outlining the fundamental notions of deep domain adaptation. Next, the deep domain adaptive detectors are divided into five groups, with full explanations of typical approaches for each group. Furthermore, recommendations for future study trends are provided.

For object detection, Yadav et al. [83] suggested a CNN-based fusion architecture. They evaluated the KAIST multispectral pedestrian dataset and the FLIR thermal object detection dataset. They trained a baseline FasterRCNN model for detection in the daytime, the Color model exceeded the Thermal model, while in the nighttime, the Thermal model performed better than the Color model, demonstrating their complementary nature. Further, they built a basic mid-level CNN fusion architecture that outperforms the baseline models considerably. When compared to conventional approaches, they found a 0.62% reduction in the miss-detection rate. For safe autonomous vehicles, underexposure zones are critical for constructing a full sense of the environment.

Thermal cameras have become an important alternative for exploring areas where conventional optical sensors fail to capture interpretable information. In [84], the authors suggested a domain adaptation system that uses a style transfer approach to transfer learning from visual to thermal images. Using style consistency, the method incorporates a generative adversarial network (GAN) to move low-level characteristics from the visible spectrum domain to the thermal domain. Dai et al. [85] presented TIRNet, a novel object detection technique based on convolutional neural networks (CNN), for robust and sustainable object recognition in thermal infrared (TIR) images. The lightweight feature extractor (VGG) is used rather than the deep-CNN backbone (ResNet, ResNeXt), which has lower bandwidth and significant computational cost. This approach attained state-of-the-art detection accuracy while maintaining a high detection efficiency at the time.

Another unique approach of domain adaption for thermal object detection was proposed by authors in [86]. Using the self-supervised contrastive learning technique, the authors investigated thermal object detection to model a view-invariant model representation. They proposed a self-supervised thermal network (SSTN) for learning features to optimize data between both visual and thermal spectrums. The proposed network was trained on FLIR and KAIST Multi-Spectral datasets.

3) OTHER APPROACHES

Bongini et al. [87] proposed a different technique to improve object detection in the thermal domain. They suggested integrating synthetic 3D objects into actual scenes as a novel data augmentation strategy for visual content domains with sparse training datasets. They analyzed other augmenting methods, such as state-of-the-art approaches acquired through reinforcement learning (RL) techniques, the infusion of simulated data, and the use of a generative model, and investigated how to integrate their suggested augmentation with these other techniques as effectively as possible. On the FLIR ADAS dataset, their single-modality detector produces state-of-the-art results, demonstrating the effectiveness of this technique.

Authors in [88], proposed Bayesian fusion object detection in visible and thermal spectrum based on multimodal data to increase safety-critical perception. They further explore and investigate different strategies for fusing data with different modalities. Kumar and Gaur [89] proposed YOLO-v3 and Spatial Pyramid Pooling (SPP) approaches to detect objects in thermal images. The YOLO-v3 technique for object detection is unique as it obtains the bounding box coordinates and a confidence score for the image using a single CNN. They then built a Spatial Pyramid Pooling (SPP) layer on top of the CNN in YOLO-v3. Cropping the image in the fully connected layer is no longer necessary because of the SPP layer. This resulted in a 100-fold increase in speed. The network was trained on the FLIR dataset, the proposed technique achieves above 80% precision.

Lu et al. [91] proposed a pedestrian detection method based on centre, temperature, scale, and ratio prediction in thermal imagery. The suggested technique is divided into two parts: (1) extraction of features and (2) predictions of centroid, temperature, scale, and ratio. The feature extraction section extracts high-level semantic characteristics from thermal imaging data as input feed data using the ResNet-101 network. The temperature, scale, and ratio prediction section determine if a target centroid exists in each place of the feature map, which is a binary classification challenge. The temperature branch determines whether the centroid is a heat-radiating pedestrian target or the background. The scale and aspects ratios prediction branches are used to establish the target size, which is a regression challenge. The suggested approach and its detection performance are better than the benchmark of night-time pedestrian detection systems, according to experimental results.

B. SEMANTIC SEGMENTATION

One of the most difficult challenges in computer vision is semantic segmentation. In the great scale of things, segmentation is one of the high-level tasks that leads to comprehensive scene interpretation. Semantic Segmentation aims to apply an object class to each pixel in an image. In self-driving surroundings, these classes may be ''pedestrians, automobiles, buildings, trees, poles, etc.'' Semantic segmentation, for example, supports self-driving cars in determining which regions of an image are safe to drive.

Yoon et al. [92], proposed a pixel-level matching object segmentation using a convolutional neural network. Based on pixel-level similarity between two object units, the network seeks to differentiate the target region from the background. To take use of both geographical details and classification semantic information, the proposed network depicts a target object utilizing characteristics from distinct depth layers. In addition, they offer a feature compression strategy that dramatically decreases memory needs while preserving feature representation capabilities. This network was primarily built on visual data, but the authors extended the network's transferability to other domains and tested it on thermal data too. In addition, the network performed better in terms of precision, efficiency, and stability.

Authors in [93] proposed segmentation of pedestrians in the thermal spectrum. The authors described the testing of the thresholding-based segmentation procedure in FIR images. The usefulness of two types of thresholding procedures is demonstrated by an evaluation of the acquired results: Otsu global thresholding vs. single threshold. Wang and Bai [94] proposed a different approach to tackle thermal pedestrian segmentation. The authors suggested a new conditional generative adversarial network-based thermal infrared pedestrian segmentation technique (IPS-cGAN) [94]. According to the results, the suggested approach outperformed variously supervised and unsupervised segmentation techniques in terms of accuracy and robustness, particularly in complex images. Thermal imaging is a very unique way of segmenting the road which is not visible in RGB images.

Authors in [95] proposed a real-time unmanned aerial vehicle semantic segmentation in thermal images. The suggested model included an encoder-decoder architecture, as well as a convolutional layer extracted features and a constrained Boltzmann machine in the network. The algorithm was also put to the test and evaluated with five state-of-the-art segmentation approaches. The proposed model was shown to be a robust model with an average accuracy of 0.97 in the results obtained.

Li et al. [96] suggested an edge conditional convolutional neural network for segmenting objects in thermal images at different times of day and night. The authors also present ''Segmenting Objects in Day And Night'' (SODA), a new benchmark data set for extensive assessments in thermal image semantic segmentation. SODA has approximately 7168 carefully annotated and synthetically produced thermal images from a variety of angles and scene complexity, each

with 20 semantic area labels. Xiong et al. [97] proposed a MCNet-multi-level correction network for semantic segmentation in thermal images for night-time, foggy, and snowy driving scenarios. It can generate a more accurate correlation matrix and step-by-step adjust the feature development process. A large thermal dataset was also introduced, called SCUT- Seg. SCUT-Seg comprises 10 manually annotated semantic area labels for 2010 thermal images taken from diverse road scenarios. Comprehensive tests on SCUT-Seg and the public MFNet dataset showed that the suggested approach outperforms the state-of-the-art methods.

Other significant advancements in the thermal semantic segmentation domain have mostly focused on narrowing the gap between RGB and thermal images, by creating a fusion between them. Vertens et al. [98] proposed a multimodal semantic segmentation model and utilize thermal images in addition to RGB images, making the network substantially more resilient. Moreover, this article also offers a unique two-stage training technique that uses a transfer learning mechanism to align the learned feature spaces across contexts. A new dataset with over 20,000 time-synchronized and coordinated RGB-thermal image sets to overcome the shortage of thermal data for self-driving cars was also introduced.

The authors of [99] utilize thermal images to build a unique deep neural network that fuses both RGB and thermal information. The proposed network uses ResNet to extract features, and a new decoder is being built to restore feature map resolution so that visual and thermal images can be fused. The results of the experiments show that their network outperforms the state of the art. Further authors in [100] proposed FuseSeg to achieve superior performance of semantic segmentation in urban scenes. The proposed network is made up of an end-to-end deep neural network that receives a pair of visual and thermal images as input and produces pixel-by-pixel semantic labels as outputs. The network interprets urban scenes, which is an important part of many self-driving operations including environment modelling, obstacle detection, mobility prediction, and planning.

John et al. [101] proposed BVTNet, a multi-label multi-class fusion of visible and thermal images to determine free space and person segmentation. The BVTNet calculates the number of pedestrians and available space in each multi-class output. In a post-processing phase, the boundaries semantics segmentation is incorporated into the overall semantic segmentation framework. The proposed model has been evaluated on the public MFNet dataset. The authors of [102] proposed a sensor fusion system that does both semantic forecasting and optimum semantic segmentation. The network predicts the available or free space, and pedestrian crossing labels, including their spatial and motion behaviour. The presented system was tested using the publicly available KAIST dataset and the framework can not only properly predict but also update the semantic segmentation map very accurately.

Kim et al. [103] proposed a framework that overcomes the data limitation issues and enhances segmentation results in the thermal. The approach further improves the classification performance of the thermal segmentation network in day and night thermal images with pixel-level domain adaptation. By utilizing sequential multi-spectral knowledge transfers, such as RGB -to- RGB, RGB -to-thermal, and thermal-to-thermal adaptations, a thermal image segmentation network achieved exceptional performance without any ground-truth labels. In addition, the authors include a real-world RGB-Thermal segmentation dataset that includes 950 manually labelled Cityscapes-style ground-truth labels in Nineteen classes.

Further authors in [104] proposed an RGB-thermal segmentation for snowy road scenarios. This research compared some of the most advanced semantic segmentation approaches for categorizing snow road surfaces from RGB images. The authors also proposed a completely new dataset for feature classification in various lighting circumstances (day and night). Zhou et al. [105] proposed a graded feature multi-label learning network (GMNet) with two RGB-thermal fusion modules, namely a shallow feature fusion and deep feature fusion to determine the segmentations in urban thermal image scenes. The network outperformed the state-of-the-art methods for urban scene semantic segmentation. Guo et al. [106] proposed the state-of-the-art MLFNet to determine the robust semantic segmentation based on the fusion between RGB-thermal images in the variable lighting scenes. Furthermore, the results show that this network is precise and resilient in a variety of illumination situations and that it exceeds state-of-the-art networks in terms of overall performance.

C. OTHER APPLICATIONS

This section will detail the use of thermal technology for other critical applications which includes but not limited to adaptive cruise control, lane keeping and lane changing. During poor visibility scenarios, a driver-assistive system that makes use of a precision based differentially corrected GPS (DGPS) system, highly accurate geospatial data, radar, and driver interfaces can assist the driver in maintaining lane position and avoiding crashes. The intelligent transportation system institute presented a similar system with the integration of infrared imaging for head-up display lane keeping and collision avoidance [129]. This system was further deployed in Minnesota and Alaska. Further, there are few studies with thermal/ infrared imaging to monitor roads using unmanned aerial vehicles to patrol the roads to reduce road fatalities [130], [131]. Moreover, focusing on drowsy drivers, authors in [132] presented a lane departure system using IR cameras for nighttime road conditions. Authors in [133] proposed a road detection algorithm with 3D information. The proposed method is incorporating a thermal stereo system-based methodology, which is supplemented by the depth data from the disparity map, to assess the thermal features of the road. The research community have tremendous potential to do further research in this area in the coming years.

VI. OUR CONTRIBUTIONS TOWARDS SAFE AUTONOMOUS SYSTEMS USING THERMAL INFRARED IMAGING

This section will summarize our core contributions toward the effective use of thermal imaging technology for both in-cabin and out-cabin vehicular applications. The entire experimental work was carried out under the Heliaus project [107]. The project tends to develop a smart breakthrough thermal perception system using an uncooled thermal camera based on a microbolometer sensing array. An indigenous prototype VGA camera was used specially designed for data recording and validation purposes. The camera embeds a Lynred Long Wave Infrared (LWIR) sensor with a focal length of 7.5 mm and an F-number of 1.2. Figure 3 shows the images of the 640×480 LWIR thermal camera used in this project. The first phase of this section will highlight experimental results for in-cabin vehicular applications whereas the second phase will describe the experimental findings for out-cabin vehicular applications carried out under the Heliaus project [107].

FIGURE 3. Uncooled Prototype LWIR 640 × 480 thermal imaging developed under the Heliaus project.

A. HELIAUS IN-CABIN CONTRIBUTIONS USING THERMAL INFRARED IMAGING

The in-cabin applications development targeted in the context of the Heliaus project aims at prototyping new smart thermal systems enabling the monitoring of driver activities by specifying the person's soft biometrics, vital sign monitoring, and drowsiness detection. The main contributions are listed below.

In the first phase, we proposed a composite mechanism to generate a large-scale synthetic thermal dataset using various computer and deep learning methods. These include data augmentation/ data transformation, synthetic data generation using SoA styleGAN, and lastly 2D-3D thermal face reconstruction using end-to-end Position map Regression Network (PRN) architecture [108]. The generated synthetic data along with the real-world thermal data gathered from the prototype thermal camera and other public thermal datasets is further used for various experimental work. The main goal of synthetic data is to further use it for robust training of deep learning models and the development of an autonomous driver gender classification system. The complete working methodology along with detailed experimental results are published in [109], [110].

In the next phase, we have developed a thermal driver gender classification framework [111] for human-machine interface applications. For this, we have trained nine state-of-the-art pre-trained networks from scratch (by unfreezing all the network layers) on a large-scale casia facial dataset. These models include AlexNet, VGG-19, MobileNet-v2, Inception-v3, ResNet-52, ResNet-50, ResNet-101, DenseNet-121, Dense-201 and EfficientNet-B4. The trained architectures are further fine-tuned using Tufts public thermal dataset [1], [7], [112]. In addition to utilizing the pretrained architectures, the main contribution of this work is designing a novel CNN architecture 'GENNet' [111] for the thermal gender classification task, and further its performance is compared against all the pre-trained state-of-the-art architectures. For rigorous validation tests of all the trained architectures including the newly proposed GENNET architecture on thermal data nine different quantitative metrics have been employed. These include accuracy, sensitivity, specificity, precision, negative predictive value, False Positive Rate (FPR), False Negative Rate (FNR), Matthews Correlation Coefficient (MCC), and F1-score.

In our study, the EfficientNet-B4 model achieved the highest test accuracy of 93.3% followed by the DenseNet-201 and the proposed GENNet network which has achieved an overall testing accuracy of 92.2 and 91.1% however, GENNet architecture is good for a compute-constrained thermal gender classification use-case as it performs significantly better than other low-parameter models. The complete experimental results are published and available at [111].

The third phase of our experimental work contributes toward driver stress and drowsiness detection using thermal infrared imaging technology. Cardone et al. [113] from NEXT2U [114] proposed a driver stress evaluation based on ECG signals. To estimate the ''stress index'' (SI) using thermal features derived from the face region of interest (i.e., nose tip, nostrils, glabella), a non-linear support vector regression (SVR) method was employed. The predicted, ''stress index'' ($r = 0.61$, $p = 0$) had a strong relationship with the actual SI. Based on the anticipated SI, a two-level categorization of the stress condition (STRESS, SI 150, vs NO STRESS, SI 150) was performed. Considering an AUC of 0.80, a sensitivity of 77%, and a specificity of 78%, the ROC analysis revealed that the classification results were improved drastically.

Further, NEXT2U [114] proposed the drowsiness classification work [115] using a low-cost and high-resolution prototype thermal camera developed under the Heliaus project. A total of 10 subjects participated in this study among which six were male subjects between the age range of 23-44. The data was recorded at a 30hz frame rate for performing further experiment work. The facial skin temperature was acquired using LWIR thermal camera along with visible facial videos which were recorded using an Intel RealSense D415 camera. The purpose of recording the visible facial data was to

transfer the visible facial landmark features tracked to the thermal imagery, thus estimating the geometrical transformation between the two imaging optical devices. The authors utilized PERCLOS (percentage of eyelid closure over the pupil across time) which is one of the most accurate parameters to assess the drowsiness state on the visible data whereas the facial feature (i.e., nose tip, glabella) estimation was performed using the acquired thermal data. The recorded data was then used to train a conventional machine learning-based support vector classifier with a polynomial kernel to classify the data into three different classes which include AWAKE, SLEEP, and FATIGUE. The ROC curve showed satisfactory performance of the classifier with an average AUC of 0.65, a sensitivity of 72.52%, and an overall specificity of 67.69% [115].

A method for enhancing the quality of thermal image data using deep learning-based multi-image super-resolution was also proposed as part of the Heliaus project in [119]. In this paper, a novel architecture for a fully convolutional recurrent neural network was presented and was trained for 4x super-resolution on a custom thermal dataset of 30 unique subjects in a driving simulator. The trained network significantly outperformed traditional bicubic interpolation both quantitatively and qualitatively. Further work has since been done on optimizing this neural network model for real-time inference on an embedded platform.

In one of the recent studies by Cardone et al. [120] carried out under the Heliaus project [107], the authors have proposed the evaluation of mental workload (MW) for Advanced Driver-Assistance Systems, since it is correlated with traffic accident risk. In this work, two different cognitive tests which include Digit Span Test (DST) and Ray Auditory Verbal Learning Test (RAVLT) were monitored and examined on participants while driving in a simulated environment. The authors utilized infrared sensing technology along with heart rate variability (HRV) data to collect features related to the psychophysiology of the subjects, which were then used for training machine learning (ML) classifiers. The authors achieved the best classifier performances with a maximum accuracy of 73.1%, the sensitivity of 0.71, and specificity of 0.69 for the Digit Span Test and Ray Auditory Verbal Learning Test the systems achieve overall accuracy of 75.0%, average sensitivity of 0.75, and an average specificity of 0.87.

B. HELIAUS OUT-CABIN CONTRIBUTIONS USING THERMAL INFRARED IMAGING

In this project, we have mainly focused on supervised learning methodology and used different types of CNN architectures for out-cabin driver assistance which include thermal object detection and classification framework. Moreover, the further stage of this research work focuses on the deployment of trained/ fine-tuned networks on single-board edge-GPU devices for onboard real-time feasibility testing. The first phase of experimental work contributes toward a novel roadside thermal object detection dataset collection named 'C3I Thermal Automotive Dataset'. The main purpose of this dataset is to further use it for out-of-cabin applications which include the development of the SoA thermal object detection framework that should be effective in all weather and environmental conditions. The further details of the newly acquired thermal dataset from the prototype LWIR thermal camera are summarized in Table 4. The complete dataset is open-sourced and available on IEEE Dataport [116].

TABLE 4. Novel thermal datasets attributes.

Dataset Name	Dataset Details	Processing Methods	Camera settings
C3I Thermal Automotive Dataset	1. Total No of distinct frames: 39,770 2. Class annotations: bike, bicycle, bus, car, person & pole 3. Environmental conditions: roadside, industrial town, downtown 4. Time: daytime, evening time and nighttime 5. Weather conditions: clear/ sunny, cloudy, windy, and foggy conditions	1. Shutterless calibration, 2. Automatic gain correction (AGC) 3. Bad pixel removal (BPR) 4. Temporal denoising (TD)	Resolution : 640x480, FPS: 30, DPI: 96

Figure 4 depicts various thermal frames acquired in different environmental and weather conditions selected from the C3I thermal automotive dataset.

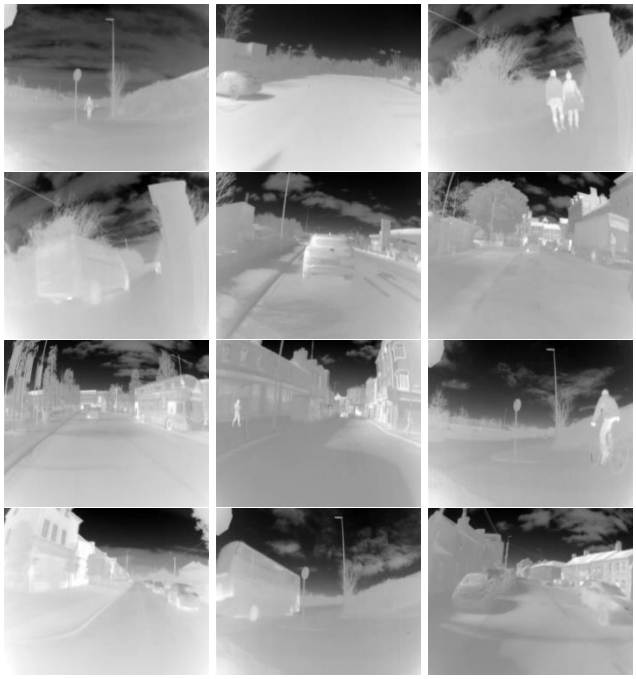


FIGURE 4. Newly acquired C3I sample thermal frames acquired in the daytime, evening time, and nighttime showing different classes.

The newly acquired is further used for training and deployment of SoA end-to-end YOLO-v5 object detector models on GPU and edge GPU devices. The main purpose for choosing the YOLO-v5 framework for thermal object detection as compared to all the previous versions of YOLO released is that YOLO-v5 is different, as this is a PyTorch implementation rather than a fork from the original Darknet library. Moreover, the YOLO-v5 has a Cross-Stage-Partial (CSP) backbone and PA-NET neck. The foremost improvements include mosaic data augmentation and auto-learning bounding box anchors thus saving the efforts for manual tuning of anchors for performing optimal training on detector models. The complete study is published and available online [117]. Three alternative test methodologies, including test-time with no augmentation (TTNA), test-time augmentation (TTA), and test-time with model ensembling, are used to validate the performance evaluation of all trained models (ME). Model ensembling, also known as an ensembling engine, is the process of combining several trained networks concurrently to create the best possible predictive inference model.

Figure 5 displays the results of the qualitative inference on nine challenging thermal frames with complex circumstances, such as numerous objects with overlapping classes, object scale and viewpoint fluctuations, and various weather conditions. These frames are selected from public test data as well as locally acquired C3I thermal automotive datasets. During the validation phase, the large network variant comprising 47.4 million parameters achieved the best qualitative results thus achieving the mean average precision (mAP) score of 84.1% using TTNA and 86.6% using the TTA method. The additional optimization and deployment of neural networks on GPUs and resource-constrained edge devices, including Nvidia Jetson Nano and Nvidia Jetson Xavier development boards, are part of the Heliaus project's major contribution [107].

FIGURE 5. Inference results on nine different thermal frames using small, medium, large and x-large model variants of the Yolo v5 framework.

The SoA inference accelerator Tensor RT is employed to further accelerate the thermally tailored YOLO architectures, resulting in higher frames per second (FPS) and shorter inference times. When deploying the models on edge hardware for the automobile sensor suite, the major goal of the quantization process is to demonstrate the viability of thermally tuned object detection models for real-time onboard testing.

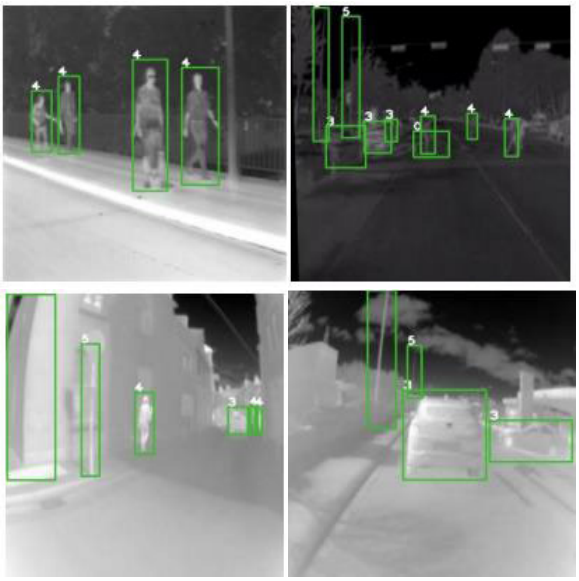
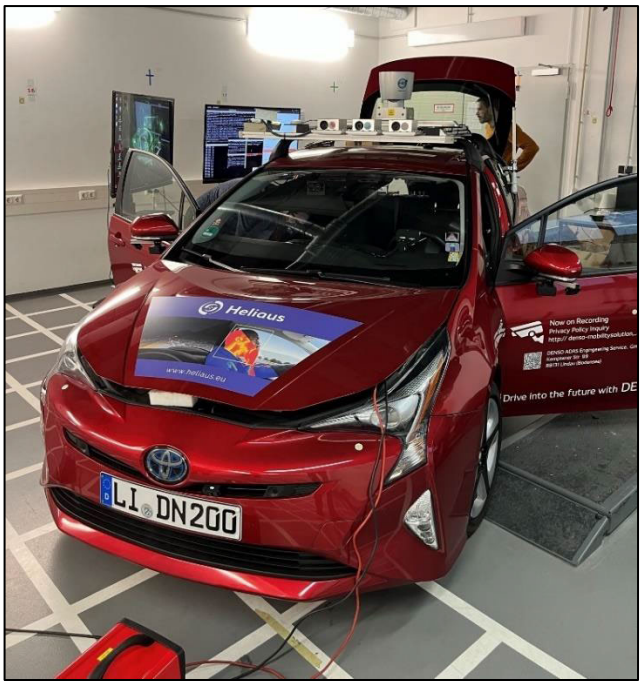


FIGURE 6. Inference results on four different thermal frames using TensortRT optimizer.

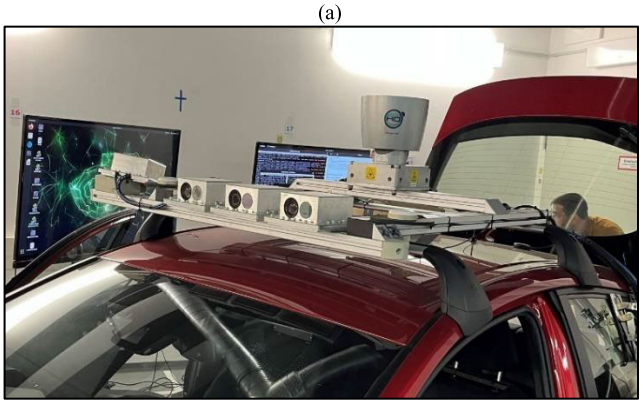
Figure 6 shows the inference results on 4 different thermal frames using the TensorRT inference accelerator engine. The complete study along with detailed experimental outcomes is published in IEEE Transactions on Intelligent Vehicles titled "Evaluation of Thermal Imaging on Embedded GPU Platforms for Application in Vehicular Assistance Systems" [118]. The optimized version of the smaller network variant achieved 60 FPS on the Nvidia Jetson Xavier development board and 11 FPS on the Nvidia Jetson Nano board.

C. FINAL PROJECT OUTCOMES

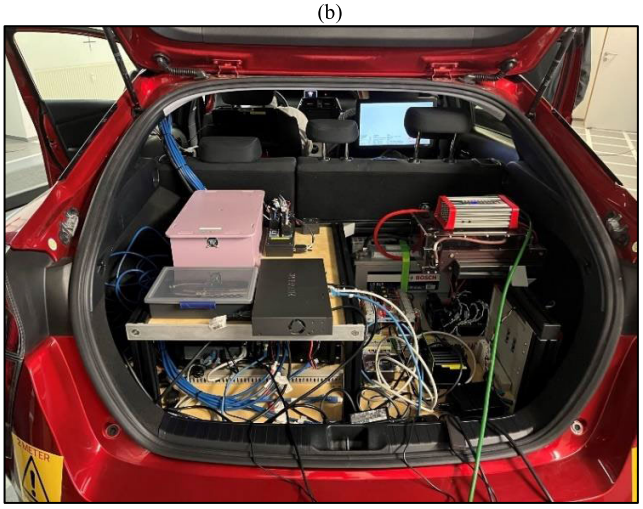
At the final phase of this project all the developed in-cabin and out-cabin thermal perceptions systems are discussed in section VI-A and VI-B were installed on a prototype vehicle. The prototype vehicle used for complete onboard deployment is the Toyota Prius car to validate the real-time performance of developed systems. Figure 7 shows the complete deployment of all the developed AI-based thermal systems on a prototype car demonstrated at Denso Germany Headquarters. The first columns in Figure 7 show the prototype car used under the Heliaus project with all the camera and sensors which includes thermal LWIR, NIR, visible RGB, lidars, and global positioning systems installed on it. Further, it displays the embedded power supply and control system which integrates all the AI-based imaging and sensor pipelines with low-power edge computing boards installed in the trunk of



(a)

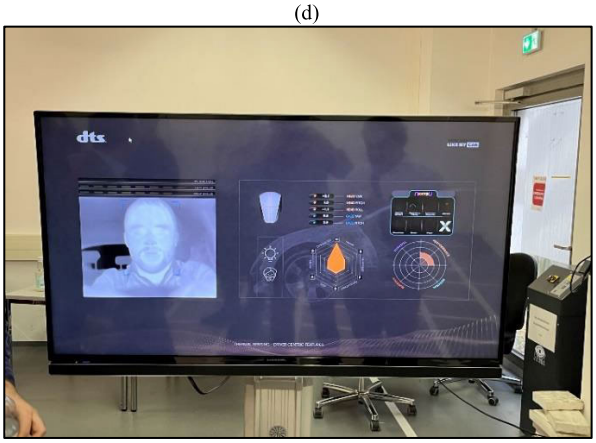


(b)

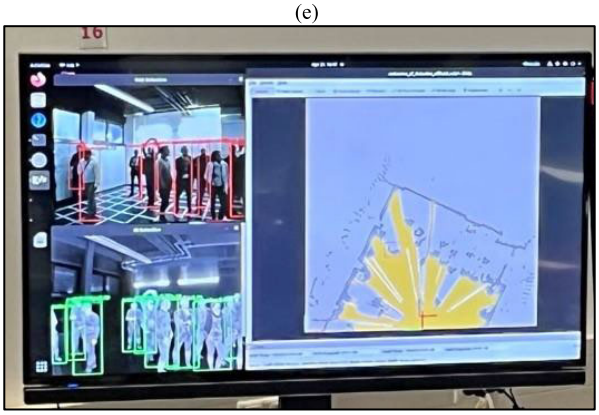


(c)

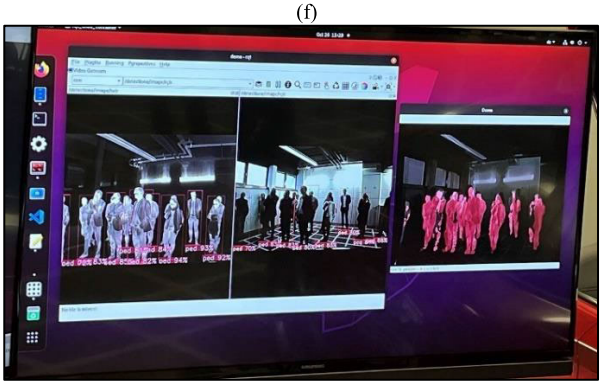
(d)



(e)



(f)



(g)

FIGURE 7. In-Cabin and Out-Cabin AI-thermal perceptions systems deployed on the HELIAUS project demo car. (a) vehicle, (b) roof-mounted sensor array, (c) power, computation & data storage, (d) In-cabin camera setup, (e) in-cabin driver monitoring data & UI, (f) out-cabin forward perception system with lidar mapping & UI, (g) out-cabin thermal object detection and semantic segmentation & UI.

the car. The second column shows the results of both in-cabin driver monitoring systems (DMS) and out-cabin advanced driver assistance systems (ADAS). The DMS systems incorporate the results of thermal super image resolution, stress monitoring, head pose and emotion estimation. On the other side, the ADAS incorporates lidar map estimation, thermal multi-class object detection, and semantics segmentation outputs.

D. ADDITIONAL CONTRIBUTIONS

In addition to all the contributions under the Heliaus project, we have also started focusing on other low-powered and economical edge computing boards. Recently, we have trained a multi-class object classifier using a nano variant of the YOLO-v5 framework. The tuned thermal model was further optimized using TensorFlow lite and deployed on a Raspberry Pi-4 computing board. The optimized model achieves the maximum FPS of 2 [134]. In addition to that we are also focusing on other imaging modalities which mainly include neuromorphic event cameras. In contrast to visible camera sensors that produce pixels, neuromorphic vision or event vision is a more advanced vision technology that produces events whenever a brightness change in the field of view (FoV) exceeds a predetermined threshold [135]. We have conducted a proof-of-concept study with event modality. Two networks, small and large variants of YOLO-v5 were trained. The trained model achieved a maximum of 201 FPS [136].

VII. CONCLUSION AND FUTURE WORK

As a result of the tremendous advancements in imaging physics over the previous few decades, the infrared thermal imaging modality has witnessed numerous technological advancements. Further integrating this with AI-based imaging pipelines we can develop smart thermal perception systems for advanced vehicular applications. The same has been highlighted in the proposed research study by presenting state-of-the-art studies for in-cabin and out-cabin applications for automotive sensor suites. The most important reason for selecting thermal modality over conventional CMOS imaging is that it is unaffected by light or any other environmental conditions, making it ideal for all-weather and day conditions, thus providing redundancy. Further, this study lists large-scale thermal datasets and highlights the technique for generating synthetic thermal datasets. This is required to overcome the shortcomings of thermal datasets for optimal training of deep neural networks. Further, this study highlights key contributions of the EU-funded Heliaus project and presents the state-of-the-art studies and datasets which are published and open-sourced as the result of the dissemination of this project.

The possible future directions of this research would be to study multi-imaging modalities which can be integrated with the automotive sensor suite. Such as, recently event camera has gained more popularity in the research community because they can measure per-pixel brightness changes asynchronously. As an outcome, a stream of events is generated, each of which encodes the time, location, and signal of the brightness changes.

ACKNOWLEDGMENT

The authors would like to acknowledge the Xperi-Ireland, Lynred-France, Denso-Germany, and Next2U-Italy for all the technical assistance and administrative management. They would also like to acknowledge the individual experts, including Christopher Dainty, Carles Person, Quentin Noir, and Joe Lemley for providing their guidance in completing this project and for giving their feedback. They would also like to acknowledge the contributors of all the public datasets for providing the image resources to carry out this research work.

REFERENCES

[1] J. Izquierdo-Reyes, R. A. Ramirez-Mendoza, M. R. Bustamante-Bello, S. Navarro-Tuch, and R. Avila-Vazquez, "Advanced driver monitoring for assistance system (ADMAS)," *Int. J. Interact. Des. Manuf. (IJIDeM)*, vol. 12, no. 1, pp. 187–197, Feb. 2018.

[2] S. Cafiso and A. Di Graziano, "Evaluation of the effectiveness of ADAS in reducing multi-vehicle collisions," *Int. J. Heavy Veh. Syst.*, vol. 19, no. 2, pp. 188–206, 2012, doi: 10.1504/IJHVS.2012.046834.

[3] Z. Liu, Y. Cai, H. Wang, and L. Chen, "Surrounding objects detection and tracking for autonomous driving using LiDAR and radar fusion," *Chin. J. Mech. Eng.*, vol. 34, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/s10033-021-00630-y.

[4] Y. Zhu, X. Lei, K. X. Wang, and Z. Yu, "Compact CMOS spectral sensor for the visible spectrum," *Photon. Res.*, vol. 7, no. 9, p. 961, Sep. 2019, doi: 10.1364/prj.7.000961.

[5] H. Tian, *Noise Analysis in CMOS Image Sensors*. Stanford, CA, USA: Stanford Univ., 2000.

[6] K. Panetta, Q. Wan, S. Agaian, S. Rajeev, S. Kamath, R. Rajendran, S. P. Rao, A. Kaszowska, H. A. Taylor, A. Samani, and X. Yuan, "A comprehensive database for benchmarking imaging systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 3, pp. 509–520, Mar. 2020, doi: 10.1109/TPAMI.2018.2884458.

[7] K. M. S. Kamath, R. Rajendran, Q. Wan, K. Panetta, and S. S. Agaian, "TERNet: A deep learning approach for thermal face emotion recognition," in *Proc. SPIE*, vol. 10993, May 2019, Art. no. 1099309, doi: 10.1117/12.2518708.

[8] R. G. Shoja, "Face recognition using infrared vision," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Laval, Quebec City, QC, Canada, 2014.

[9] *University of Notre Dame (UND) Thermal Dataset*. Accessed: Nov. 20, 2021. [Online]. Available: https://cvrl.nd.edu/projects/data/

[10] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska, "A new face database simultaneously acquired in visible, near-infrared and thermal spectrums," *Cognit. Comput.*, vol. 5, no. 1, pp. 119–135, Mar. 2013, doi: 10.1007/s12559-012-9163-2.

[11] V. Espinosa-Duró, M. Faundez-Zanuy, J. Mekyska, and E. Monte-Moreno, "A criterion for analysis of different sensor combinations with an application to face biometrics," *Cognit. Comput.*, vol. 2, no. 3, pp. 135–141, Sep. 2010, doi: 10.1007/s12559-010-9060-5.

[12] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *Proc. 7th IEEE Workshops Appl. Comput. Vis.*, Jan. 2005, pp. 364–369.

[13] *CVC-09 FIR Sequence Pedestrian Dataset*. Accessed: Jun. 22, 2021. [Online]. Available: http://adas.cvc.uab.es/elektra/enigma-portfolio/item-1/

[14] A. Torabi, G. Massé, and G.-A. Bilodeau, "An iterative integrated framework for thermal-visible image registration, sensor fusion, and people tracking for video surveillance applications," *Comput. Vis. Image Understand.*, vol. 116, no. 2, pp. 210–221, Feb. 2012.

[15] Z. Wu, N. Fuller, D. Theriault, and M. Betke, "A thermal infrared video benchmark for visual analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 201–208.

[16] *FLIR Thermal Dataset*. Accessed: Dec. 22, 2021. [Online]. Available: https://www.flir.com/oem/adas/adas-dataset-form/

[17] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1037–1045.

[18] C. Pramerdorfer, J. Strohmayer, and M. Kampel, "SDT: A synthetic multi-modal dataset for person detection and pose classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 1611–1615, doi: 10.1109/ICIP40778.2020.9191284.
[19] C. Chen and A. Ross, "Evaluation of gender classification methods on thermal and near-infrared face images," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–8, doi: 10.1109/IJCB.2011.6117544.
[20] S. Wang, Z. Gao, S. He, M. He, and Q. Ji, "Gender recognition from visible and thermal infrared facial images," *Multimedia Tools Appl.*, vol. 75, no. 14, pp. 8419–8442, Jul. 2016, doi: 10.1007/s11042-015-2756-5.
[21] D. Nguyen and K. Park, "Body-based gender recognition using images from visible and thermal cameras," *Sensors*, vol. 16, no. 2, p. 156, Jan. 2016, doi: 10.3390/s16020156.
[22] D. Nguyen, K. Kim, H. Hong, J. Koo, M. Kim, and K. Park, "Gender recognition from human-body images using visible-light and thermal camera videos based on a convolutional neural network for image feature extraction," *Sensors*, vol. 17, no. 3, p. 637, Mar. 2017, doi: 10.3390/s17030637.
[23] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Multimodal gender detection," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, Nov. 2017, pp. 302–311, doi: 10.1145/3136755.3136770.
[24] T. Chaudhuri, D. Zhai, Y. C. Soh, H. Li, and L. Xie, "Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology," *Energy Buildings*, vol. 166, pp. 391–406, May 2018, doi: 10.1016/j.enbuild.2018.02.035.
[25] J. Tapia and C. C. Aravena, "Gender classification from periocular NIR images using fusion of CNNs models," in *Proc. IEEE 4th Int. Conf. Identity, Secur., Behav. Anal. (ISBA)*, Jan. 2018, pp. 1–6, doi: 10.1109/ISBA.2018.8311465.
[26] N. Dwivedi and D. K. Singh, "Review of deep learning techniques for gender classification in images," in *Harmony Search and Nature Inspired Optimization Algorithms*. Berlin, Germany: Springer, 2019, pp. 1089–1099.
[27] G. Koukiou and V. Anastassopoulos, "Gender discrimination based on the thermal signature of the face and the external ear," *Signal Image Process., Int. J.*, vol. 11, no. 4, pp. 13–23, Aug. 2020, doi: 10.5121/sipij.2020.11402.
[28] G. Koukiou and V. Anastassopoulos, "Simple face thermal features for gender discrimination," in *Proc. CS IT Conf.*, 2020, pp. 23–31, doi: 10.5121/csit.2020.100803.
[29] K. Prihodova and J. Jech, "Gender recognition using thermal images from UAV," in *Proc. Int. Conf. Inf. Digit. Technol. (IDT)*, Jun. 2021, pp. 83–88, doi: 10.1109/IDT52577.2021.9497627.
[30] A. Basu, A. Routray, S. Shit, and A. K. Deb, "Human emotion recognition from facial thermal image based on fused statistical feature and multi-class SVM," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2015, pp. 1–5, doi: 10.1109/INDICON.2015.7443712.
[31] I. A. Cruz-Albarran, J. P. Benitez-Rangel, R. A. Osornio-Rios, and L. A. Morales-Hernandez, "Human emotions detection based on a smart-thermal system of thermographic images," *Infr. Phys. Technol.*, vol. 81, pp. 250–261, Mar. 2017.
[32] C. Goulart, C. Valadao, D. Delisle-Rodriguez, E. Caldeira, and T. Bastos, "Emotion analysis in children through facial emissivity of infrared thermal imaging," *PLoS ONE*, vol. 14, no. 3, pp. 1–17, 2019, doi: 10.1371/journal.pone.0212928.
[33] B. Ilikci, L. Chen, H. Cho, and Q. Liu, "Heat-map based emotion and face recognition from thermal images," in *Proc. Comput., Commun. IoT Appl. (ComComAp)*, Oct. 2019, pp. 449–453, doi: 10.1109/ComComAp46287.2019.9018786.
[34] C. K. Kyal, H. Poddar, and M. Reza, "Human emotion recognition from spontaneous thermal image sequence using GPU accelerated emotion landmark localization and parallel deep emotion net," in *Proc. Int. Conf. Innov. Comput. Commun.*, 2020, pp. 931–943.
[35] S. Sarath, "Human emotions recognition from thermal images using YOLO algorithm," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Jul. 2020, pp. 1139–1142, doi: 10.1109/ICCSP48568.2020.9182148.
[36] S. Wirjopawiro, "Super-resolution to enhance low-resolution thermal facial expression images for thermal facial emotion recognition," M.S thesis, Dept. Elect. Eng., Math. Comput. Sci., Delft Univ. Technol. (TU Delft), Delft, The Netherlands, 2021.
[37] A. K. Prabhakaran, J. J. Nair, and S. Sarath, "Thermal facial expression recognition using modified ResNet152," in *Advances in Computing and Network Communications*. Singapore: Springer, 2021, pp. 389–396. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-33-6987-0_32#citeas
[38] M. M. M. Al Qudah, A. S. A. Mohamed, and S. L. Lutfi, "Affective state recognition using thermal-based imaging: A survey," *Comput. Syst. Sci. Eng.*, vol. 37, no. 1, pp. 47–62, 2021, doi: 10.32604/CSSE.2021.015222.
[39] Y. H. Lai, Y. C. Chang, C. W. Tsai, C. H. Lin, and M. Y. Chen, "Data fusion analysis for attention-deficit hyperactivity disorder emotion recognition with thermal image and Internet of Things devices," *Softw., Pract. Exper.*, vol. 51, no. 3, pp. 595–606, Mar. 2021.
[40] K. Ganesh, S. Umapathy, and P. T. Krishnan, "Deep learning techniques for automated detection of autism spectrum disorder based on thermal imaging," *Proc. Inst. Mech. Eng., H, J. Eng. Med.*, vol. 235, no. 10, pp. 1113–1127, Jun. 2021, doi: 10.1177/09544119211024778.
[41] D. Bhattacharjee, A. Seal, S. Ganguly, M. Nasipuri, and D. K. Basu, "A comparative study of human thermal face recognition based on Haar wavelet transform and local binary pattern," *Comput. Intell. Neurosci.*, vol. 2012, pp. 1–12, Jan. 2012, doi: 10.1155/2012/261089.
[42] M. Kristo and M. Ivasic-Kos, "An overview of thermal face recognition methods," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 1098–1103, doi: 10.23919/MIPRO.2018.8400200.
[43] Z. Wu, M. Peng, and T. Chen, "Thermal face recognition using convolutional neural network," in *Proc. Int. Conf. Optoelectronics Image Process. (ICOIP)*, Jun. 2016, pp. 6–9, doi: 10.1109/OPTIP.2016.7528489.
[44] A. Kwasniewska, J. Ruminski, and P. Rad, "Deep features class activation map for thermal face detection and tracking," in *Proc. 10th Int. Conf. Human Syst. Interact. (HSI)*, Jul. 2017, pp. 41–47, doi: 10.1109/HSI.2017.8004993.
[45] T. Zhang, A. Wiliem, S. Yang, and B. Lovell, "TV-GAN: Generative adversarial network based thermal to visible face recognition," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 174–181, doi: 10.1109/ICB2018.2018.00035.
[46] S. M. J. Tamboli and P. K. R. Desai, "Approach of thermal imaging as a facial recognition," *JournalNX*, vol. 2, no. 4, pp. 1–4, 2016.
[47] M. K. Bhowmik, K. Saha, S. Majumder, G. Majumder, A. Saha, A. N. Sarma, D. Bhattacharjee, D. K. Basu, and M. Nasipuri, "Thermal infrared face recognition—A biometric identification technique for robust security system," *Rev., Refinements New Ideas Face Recognit.*, vol. 7, pp. 113–138, Jul. 2011.
[48] A. Seal, D. Bhattacharjee, M. Nasipuri, and D. K. Basu, "Minutiae based thermal face recognition using blood perfusion data," in *Proc. Int. Conf. Image Inf. Process.*, Nov. 2011, pp. 1–4, doi: 10.1109/ICIIP.2011.6108928.
[49] P. Buddharaju, I. T. Pavlidis, and P. Tsiamyrtzis, "Pose-invariant physiological face recognition in the thermal infrared spectrum," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, 2006, p. 53, doi: 10.1109/CVPRW.2006.160.
[50] V. A. Weidlich, "Thermal infrared face recognition," *Cureus*, vol. 13, pp. 1–7, Mar. 2021, doi: 10.7759/cureus.13736.
[51] S. D. Lin, L. Chen, and W. Chen, "Thermal face recognition under different conditions," *BMC Bioinf.*, vol. 22, no. S5, pp. 1–17, Nov. 2021, doi: 10.1186/s12859-021-04228-y.
[52] D. D. Poster, S. Hu, N. J. Short, B. S. Riggan, and N. M. Nasrabadi, "Visible-to-thermal transfer learning for facial landmark detection," *IEEE Access*, vol. 9, pp. 52759–52772, 2021, doi: 10.1109/ACCESS.2021.3070233.
[53] W.-T. Chu and P.-S. Huang, "Thermal face recognition based on multi-scale image synthesis," in *Proc. MultiMedia Model., 27th Int. Conf.*, 2021, pp. 99–110.
[54] G. Hermosilla, D.-I.-H. Tapia, H. Allende-Cid, G. F. Castro, and E. Vera, "Thermal face generation using StyleGAN," *IEEE Access*, vol. 9, pp. 80511–80523, 2021, doi: 10.1109/ACCESS.2021.3085423.
[55] T. Jin, C. Shouming, X. Xiuzhen, and J. Gu, "Eyes localization in an infrared image," in *Proc. IEEE Int. Conf. Autom. Logistics*, Aug. 2009, pp. 217–222, doi: 10.1109/ICAL.2009.5262927.
[56] S. Budzan and R. Wyzgolik, "Face and eyes localization algorithm in thermal images for temperature measurement of the inner canthus of the eyes," *Infr. Phys. Technol.*, vol. 60, pp. 225–234, Sep. 2013.
[57] J. Ahlberg, N. Markus, and A. Berg, "Multi-person fever screening using a thermal and a visual camera," in *Proc. Swedish Symp. Image Anal.*, Mar. 2015, pp. 17–18.
[58] Q. Wang, L. Boccanfuso, B. Li, A. Y.-J. Ahn, C. E. Foster, M. P. Orr, B. Scassellati, and F. Shic, "Thermographic eye tracking," in *Proc. 9th Biennial ACM Symp. Eye Tracking Res. Appl.*, Mar. 2016, pp. 307–310, doi: 10.1145/2857491.2857543.

[59] M. Marzec, A. Lamża, Z. Wróbel, and A. Dziech, "Fast eye localization from thermal images using neural networks," *Multimedia Tools Appl.*, pp. 1–14, 2016.

[60] M. N. Hussien, M.-H. Lye, M. F. A. Fauzi, T. C. Seong, and S. Mansor, "Comparative analysis of eyes detection on face thermal images," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2017, pp. 385–389, doi: 10.1109/ICSIPA.2017.8120641.

[61] M. Knapik and B. Cyganek, "Fast eyes detection in thermal images," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 3601–3621, Jan. 2021, doi: 10.1007/s11042-020-09403-6.

[62] M. Knapik and B. Cyganek, "Comparison of sparse image descriptors for eyes detection in thermal images," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 638–644, doi: 10.5220/0007576506380644.

[63] C. Ferrari, L. Berlincioni, M. Bertini, and A. Del Bimbo, "Inner eye canthus localization for human body temperature screening," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 8833–8840, doi: 10.1109/ICPR48806.2021.9412015.

[64] N. Padmapriya, N. Venkateswaran, R. Ravikumar, and R. Chelliah, "Localization of eye region in infrared thermal images using deep neural network," in *Proc. 6th Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2021, pp. 446–450, doi: 10.1109/WiSPNET51692.2021.9419446.

[65] M. A. Jaddoa, L. Gonzalez, H. Cuthbertson, and A. Al-Jumaily, "Multiview eye localisation to measure cattle body temperature based on automated thermal image processing and computer vision," *Infr. Phys. Technol.*, vol. 119, Dec. 2021, Art. no. 103932, doi: 10.1016/j.infrared.2021.103932.

[66] M. Tashakori, A. Nahvi, A. Shahiidian, S. E. H. Kiashari, and H. Bakhoda, "Estimation of driver drowsiness using blood perfusion analysis of facial thermal images in a driving simulator," *J. Sleep Sci.*, vol. 3, nos. 3–4, pp. 45–52, 2018.

[67] S. E. H. Kiashari, A. Nahvi, A. Homayounfard, and H. Bakhoda, "Monitoring the variation in driver respiration rate from wakefulness to drowsiness: A non-intrusive method for drowsiness detection using thermal imaging," *J. Sleep Sci.*, vol. 3, no. 2, pp. 1–9, 2018.

[68] E. P. Herrera-Granda, "Drowsiness detection in drivers through real-time image processing of the human eye," in *Intelligent Information and Database Systems*. Cham, Switzerland: Springer, 2019, pp. 626–637.

[69] M. Knapik and B. Cyganek, "Driver's fatigue recognition based on yawn detection in thermal images," *Neurocomputing*, vol. 338, pp. 274–292, Apr. 2019, doi: 10.1016/j.neucom.2019.02.014.

[70] S. Kajiwara, "Evaluation of driver status in autonomous vehicles: Using thermal infrared imaging and other physiological measurements," *Int. J. Veh. Inf. Commun. Syst.*, vol. 4, no. 3, pp. 232–241, 2019.

[71] S. E. H. Kiashari, A. Nahvi, H. Bakhoda, A. Homayounfard, and M. Tashakori, "Evaluation of driver drowsiness using respiration analysis by thermal imaging on a driving simulator," *Multimedia Tools Appl.*, vol. 79, pp. 17793–17815, Jul. 2020, doi: 10.1007/s11042-020-08696-x.

[72] K. Malecki, P. Forczmanski, A. Nowosielski, A. Smolinski, and D. Ozga, "A new benchmark collection for driver fatigue research based on thermal, depth map and visible light imagery," in *Progress in Computer Recognition Systems*. Cham, Switzerland: Springer, 2020, pp. 295–304.

[73] S. Kajiwara, "Driver-condition detection using a thermal imaging camera and neural networks," *Int. J. Automot. Technol.*, vol. 22, no. 6, pp. 1505–1515, Dec. 2021.

[74] M. Tashakori, A. Nahvi, and S. E. H. Kiashari, "Driver drowsiness detection using facial thermal imaging in a driving simulator," *Proc. Inst. Mech. Eng., H, J. Eng. Med.*, vol. 236, no. 1, pp. 43–55, Sep. 2021, doi: 10.1177/09544119211044232.

[75] K. Das, S. Sharak, K. Riani, M. Abouelenien, M. Burzo, and M. Papakostas, "Multimodal detection of drivers drowsiness and distraction," *Assoc. Comput. Machinery*, vol. 1, no. 1, pp. 416–424, 2021.

[76] C. Devaguptapu, N. Akolekar, M. M. Sharma, and V. N. Balasubramanian, "Borrow from anywhere: Pseudo multi-modal object detection in thermal imagery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1029–1038, doi: 10.1109/CVPRW.2019.00135.

[77] K. N. R. Chebrolu and P. N. Kumar, "Deep learning based pedestrian detection at all light conditions," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2019, pp. 838–842, doi: 10.1109/ICCSP.2019.8698101.

[78] K. Agrawal and A. Subramanian, "Enhancing object detection in adverse conditions using thermal imaging," 2019, *arXiv:1909.13551*.

[79] Y. Cao, T. Zhou, X. Zhu, and Y. Su, "Every feature counts: An improved one-stage detector in thermal imagery," in *Proc. IEEE 5th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2019, pp. 1965–1969, doi: 10.1109/ICCC47050.2019.9064036.

[80] P. Talluri and M. Dua, "Low-resolution human identification in thermal imagery," in *Proc. 5th Int. Conf. Commun. Electron. Syst. (ICCES)*, vol. 2020, pp. 1283–1287, doi: 10.1109/ICCES48766.2020.9138039.

[81] M. Kristo, M. Ivasic-Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using YOLO," *IEEE Access*, vol. 8, pp. 125459–125476, 2020, doi: 10.1109/ACCESS.2020.3007481.

[82] W. Li, F. Li, Y. Luo, P. Wang, and J. Sun, "Deep domain adaptive object detection: A survey," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2020, pp. 1808–1813, doi: 10.1109/SSCI47803.2020.9308604.

[83] R. Yadav, A. Samir, H. Rashed, S. Yogamani, and R. Dahyot, "CNN based color and thermal image fusion for object detection in automated driving," in *Proc. Irish Machine Vis. Image Process. Conf.*, Jul. 2020, pp. 69–76.

[84] F. Munir, S. Azam, M. A. Rafique, A. M. Sheri, M. Jeon, and W. Pedrycz, "Exploring thermal images for object detection in underexposure regions for autonomous driving," 2020, *arXiv:2006.00821*.

[85] X. Dai, X. Yuan, and X. Wei, "TIRNet: Object detection in thermal infrared images for autonomous driving," *Int. J. Speech Technol.*, vol. 51, no. 3, pp. 1244–1261, Mar. 2021, doi: 10.1007/s10489-020-01882-2.

[86] F. Munir, S. Azam, and M. Jeon, "SSTN: Self-supervised domain adaptation thermal object detection for autonomous driving," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2021, pp. 206–213, doi: 10.1109/iros51168.2021.9636353.

[87] F. Bongini, L. Berlincioni, M. Bertini, and A. Del Bimbo, "Partially fake it till you make it: Mixing real and fake thermal images for improved object detection," in *Proc. 29th ACM Int. Conf. Multimedia*, vol. 1, 2021, pp. 5482–5490.

[88] Y.-T. Chen, J. Shi, Z. Ye, C. Mertz, D. Ramanan, and S. Kong, "Multimodal object detection via probabilistic ensembling," 2021, *arXiv:2104.02904*.

[89] S. Kumar and D. Gaur, "Thermal object detection using YOLOv3 and spatial pyramid pooling," in *Proc. Int. Conf. Mach. Intell. Data Sci. Appl.*, 2021, pp. 553–565.

[90] H. Zhang, X.-G. Hong, and L. Zhu, "Detecting small objects in thermal images using single-shot detector," *Autom. Control Comput. Sci.*, vol. 55, no. 2, pp. 202–211, Mar. 2021, doi: 10.3103/S0146411621020097.

[91] C. Lu, S. Zhang, and M. Liu, "Pedestrian detection based on center, temperature, scale and ratio prediction in thermal imagery," in *Proc. 40th Chin. Control Conf. (CCC)*, Jul. 2021, pp. 7288–7293, doi: 10.23919/CCC52363.2021.9550656.

[92] J. S. Yoon, F. Rameau, J. Kim, S. Lee, S. Shin, and I. S. Kweon, "Pixel-level matching for video object segmentation using convolutional neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2186–2195, doi: 10.1109/ICCV.2017.238.

[93] K. Piniarski and P. Pawlowski, "Segmentation of pedestrians in thermal imaging," in *Proc. Baltic URSI Symp. (URSI)*, May 2018, pp. 210–211, doi: 10.23919/URSI.2018.8406765.

[94] P. Wang and X. Bai, "Thermal infrared pedestrian segmentation based on conditional GAN," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6007–6021, Dec. 2019.

[95] M. K. Masouleh and R. Shah-Hosseini, "Development and evaluation of a deep learning model for real-time ground vehicle semantic segmentation from UAV-based thermal infrared imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 155, pp. 172–186, Sep. 2019, doi: 10.1016/j.isprsjprs.2019.07.009.

[96] C. Li, W. Xia, Y. Yan, B. Luo, and J. Tang, "Segmenting objects in day and night: Edge-conditioned CNN for thermal image semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3069–3082, Jul. 2020, doi: 10.1109/TNNLS.2020.3009373.

[97] H. Xiong, W. Cai, and Q. Liu, "MCNet: Multi-level correction network for thermal image semantic segmentation of nighttime driving scene," *Infr. Phys. Technol.*, vol. 113, Mar. 2021, Art. no. 103628, doi: 10.1016/j.infrared.2020.103628.

[98] J. Vertens, J. Zurn, and W. Burgard, "HeatNet: Bridging the day-night domain gap in semantic segmentation with thermal images," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 8461–8468, doi: 10.1109/IROS45743.2020.9341192.

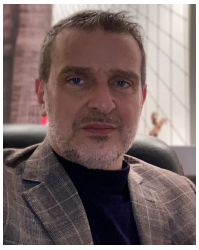
[99] Y. Sun, W. Zuo, and M. Liu, "RTFNet: RGB-thermal fusion network for semantic segmentation of urban scenes," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2576–2583, Jul. 2019, doi: 10.1109/LRA.2019.2904733.

[100] Y. Sun, W. Zuo, P. Yun, H. Wang, and M. Liu, "FuseSeg: Semantic segmentation of urban scenes based on RGB and thermal data fusion," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 3, pp. 1000–1011, Jul. 2021, doi: 10.1109/TASE.2020.2993143.
[101] V. John, A. Boyali, S. Thompson, and S. Mita, "BVTNet: Multi-label multi-class fusion of visible and thermal camera for free space and pedestrian segmentation," in *Pattern Recognition, ICPR International Workshops and Challenges* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2021, pp. 277–288.
[102] V. John, S. Mita, A. Lakshmanan, A. Boyali, and S. Thompson, "Deep visible and thermal camera-based optimal semantic segmentation using semantic forecasting," *J. Auto. Vehicles Syst.*, vol. 1, pp. 1–10, Sep. 2021, doi: 10.1115/1.4052529.
[103] Y.-H. Kim, U. Shin, J. Park, and I. S. Kweon, "MS-UDA: Multi-spectral unsupervised domain adaptation for thermal image semantic segmentation," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 6497–6504, Oct. 2021, doi: 10.1109/LRA.2021.3093652.
[104] S. Vachmanus, A. A. Ravankar, T. Emaru, and Y. Kobayashi, "An evaluation of RGB-thermal image segmentation for snowy road environment," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2021, pp. 224–230, doi: 10.1109/ICMA52036.2021.9512708.
[105] W. Zhou, J. Liu, J. Lei, L. Yu, and J.-N. Hwang, "GMNet: Graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 7790–7802, 2021, doi: 10.1109/TIP.2021.3109518.
[106] Z. Guo, X. Li, Q. Xu, and Z. Sun, "Robust semantic segmentation based on RGB-thermal in variable lighting scenes," *Measurement*, vol. 186, Dec. 2021, Art. no. 110176, doi: 10.1016/j.measurement.2021.110176.
[107] (Aug. 2022). *Heliaus European Union Project*. [Online]. Available: https://www.heliaus.eu/
[108] Y. Feng, "Joint 3D face reconstruction and dense alignment with position map regression network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 534–551.
[109] M. A. Farooq and P. Corcoran, "Generating thermal image data samples using 3D facial modelling techniques and deep learning methodologies," in *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–5, doi: 10.1109/QoMEX48832.2020.9123079.
[110] M. A. Farooq and P. Corcoran, "Proof-of-concept techniques for generating synthetic thermal facial data for training of deep learning models," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2021, pp. 1–6, doi: 10.1109/ICCE50685.2021.9427690.
[111] M. A. Farooq, H. Javidnia, and P. Corcoran, "Performance estimation of the state-of-the-art convolution neural networks for thermal images-based gender classification system," *J. Electron. Imag.*, vol. 29, no. 6, Nov. 2020, Art. no. 063004.
[112] *Tufts Thermal Dataset*. Accessed: Aug. 29, 2022. [Online]. Available: http://tdface.ece.tufts.edu/
[113] D. Cardone, D. Perpetuini, C. Filippini, E. Spadolini, L. Mancini, A. M. Chiarelli, and A. Merla, "Driver stress state evaluation by means of thermal imaging: A supervised machine learning approach based on ECG signal," *Appl. Sci.*, vol. 10, no. 16, p. 5673, Aug. 2020.
[114] *Heliaus Consortium Partners Next2u Solutions*. Accessed: Aug. 17, 2022. [Online]. Available: https://www.next2u-solutions.com/
[115] D. Cardone, "Driver drowsiness evaluation by means of thermal infrared imaging: Preliminary results," in *Proc. SPIE*, vol. 11831, pp. 111–121, Aug. 2021.
[116] M. A. Farooq, W. Shariff, F. Khan, P. Corcoran, and C. Rotariu, "C3I thermal automotive dataset," IEEE Dataport, to be published. [Online]. Available: https://ieee-dataport.org/documents/c3i-thermal-automotive-dataset, doi: 10.21227/jf21-rt22.
[117] M. A. Farooq, P. Corcoran, C. Rotariu, and W. Shariff, "Object detection in thermal spectrum for advanced driver-assistance systems (ADAS)," *IEEE Access*, vol. 9, pp. 156465–156481, 2021, doi: 10.1109/ACCESS.2021.3129150.
[118] M. A. Farooq, W. Shariff, and P. Corcoran, "Evaluation of thermal imaging on embedded GPU platforms for application in vehicular assistance systems," *IEEE Trans. Intell. Vehicles*, early access, Mar. 9, 2022, doi: 10.1109/TIV.2022.3158094.
[119] D. O'Callaghan, C. Ryan, W. Shariff, M. Ali Farooq, J. Lemley, and P. Corcoran, "Recurrent super-resolution method for enhancing low quality thermal facial data," in *Proc. 24th Irish Mach. Vis. Image Process. Conf.*, Aug. 2022, pp. 193–196, doi: 10.56541/UAOV9084.
[120] D. Cardone, D. Perpetuini, C. Filippini, L. Mancini, S. Nocco, M. Tritto, S. Rinella, A. Giacobbe, G. Fallica, F. Ricci, S. Gallina, and A. Merla, "Classification of drivers' mental workload levels: Comparison of machine learning methods based on ECG and infrared thermal signals," *Sensors*, vol. 22, no. 19, p. 7300, Sep. 2022.
[121] R. Zhen, W. Song, Q. He, J. Cao, L. Shi, and J. Luo, "Human-computer interaction system: A survey of talking-head generation," *Electronics*, vol. 12, no. 1, p. 218, Jan. 2023.
[122] X. Sun, M. Wang, R. Lin, Y. Sun, and S. S. Cheng, "Deep-learned perceptual quality control for intelligent video communication," *IEEE Trans. Consum. Electron.*, vol. 68, no. 4, pp. 354–365, Nov. 2022, doi: 10.1109/TCE.2022.3206114.
[123] L. Tay, S. E. Woo, L. Hickman, B. M. Booth, and S. D'Mello, "A conceptual framework for investigating and mitigating machine-learning measurement bias (MLMB) in psychological assessment," *Adv. Methods Practices Psychol. Sci.*, vol. 5, no. 1, Jan. 2022, Art. no. 251524592110613.
[124] M. Jia, J.-H. Choi, H. Liu, and G. Susman, "Development of facial-skin temperature driven thermal comfort and sensation modeling for a futuristic application," *Building Environ.*, vol. 207, Jan. 2022, Art. no. 108479.
[125] P. Li, P. Dai, D. Cao, B. Liu, and Y. Lu, "Non-intrusive comfort sensing: Detecting age and gender from infrared images for personal thermal comfort," *Building Environ.*, vol. 219, Jul. 2022, Art. no. 109256.
[126] A. S. Ahire, "Night vision system in BMW," *Int. Rev. Appl. Eng. Res.*, vol. 4, no. 1, pp. 1–10, 2014.
[127] *BMW Night Vison Systems*. Accessed: Feb. 14, 2023. [Online]. Available: https://www.flir.eu/discover/oem/automotive/bmw-incorporates-thermal-imaging-cameras-in-its-cars/
[128] *Veoneer Thermal Camera*. Accessed: Feb. 14, 2023. [Online]. Available: productionhttps://europe.autonews.com/suppliers/veoneer-wins-thermal-camera-production-deal-self-driving-car
[129] P. M. Cheng and C. Shankwitz, "Integration of infrared imaging for a head up display lane keeping and collision avoidance system," Center Transp. Stud., Univ. Minnesota, Minneapolis, MN, USA, Tech. Rep. CTS 08-09, 2008.
[130] N. Sharma, "The role of infrared thermal imaging in road patrolling using unmanned aerial vehicles," in *Unmanned Aerial Vehicle: Applications in Agriculture and Environment*. Cham, Switzerland: Springer, 2020, pp. 143–157.
[131] L. F. Gonzalez, G. A. Montes, E. Puig, S. Johnson, K. Mengersen, and K. J. Gaston, "Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation," *Sensors*, vol. 16, no. 1, p. 97, 2016.
[132] M. Miman, O. A. Osman, and H. C. Korkmaz, "Lane departure system design using with IR camera for night-time road conditions," *Tem J.*, vol. 4, no. 1, p. 54, 2015.
[133] G. A. Pelaez, D. Bacara, A. De La Escalera, F. Garcia, and C. Olaverri-Monreal, "Road detection with thermal cameras through 3D information," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Seoul, Korea (South), Jun. 2015, pp. 255–260, doi: 10.1109/IVS.2015.7225695.
[134] M. Ali Farooq, W. Shariff, F. Khan, and P. Corcoran, "Development, optimization, and deployment of thermal forward vision systems for advance vehicular applications on edge devices," 2023, *arXiv:2301.07613*.
[135] M. S. Dilmaghani, W. Shariff, C. Ryan, J. Lemley, and P. Corcoran, "Control and evaluation of event cameras output sharpness via bias," 2022, *arXiv:2210.13929*.
[136] W. Shariff, M. Ali Farooq, J. Lemley, and P. Corcoran, "Event-based YOLO object detection: Proof of concept for forward perception system," 2022, *arXiv:2212.07181*.

MUHAMMAD ALI FAROOQ received the B.E. degree in electronic engineering from Iqra University, in 2012, the M.S. degree in electrical control engineering from the National University of Sciences and Technology (NUST), in 2017, and the Ph.D. degree from the National University of Ireland, Galway (NUIG). His Ph.D. research program was funded through the prestigious H2020 European Union (EU) Scholarship. He is currently a Postdoctoral Researcher with the University of Galway. Moreover, he is working as a machine learning research intern with Xperi corporation. His research interests include machine vision, computer vision, video analytics, machine learning, thermal imaging, and sensor fusion.

WASEEM SHARIFF received the B.E. degree in computer science from the Nagarjuna College of Engineering and Technology (NCET), in 2019, and the M.Sc. degree in computer science, specializing in artificial intelligence from the National University of Ireland, Galway (NUIG), in 2020. He is currently pursuing the Ph.D. degree with the University of Galway, under the IRC Employment Ph.D. Program. He is associated with the Heliaus (thermal vision augmented awareness) Project. He is also a Research and Development Engineer with Xperi Inc. His research interest includes machine learning for computer vision applications, with a particular emphasis on automotive in-cabin monitoring applications.

DAVID O'CALLAGHAN received the B.E. degree in electrical and electronic engineering from University College Cork (UCC), in 2017, and the M.Sc. degree in artificial intelligence from the School of Computer Science, National University of Ireland, Galway (NUIG), in 2020. He is currently a Research and Development Engineer with the Sensing Research Team, Xperi Inc., Galway, Ireland. At Xperi, he specializes in deep learning, computer vision, and signal processing for in-cabin monitoring applications.



ARCANGELO MERLA received the Ph.D. degree in biomedical technology from the University of Chieti–Pescara and the M.S. degree in physics from the University of Bologna. He is currently an Associate Professor of applied physics with the Department of Engineering and Geology, G. d'Annunzio University of Chieti–Pescara, Chieti, Italy. He has extensive scientific production with over 120 international articles. His research interests include biomedical imaging and modeling, with special emphasis on thermal infrared imaging and optical imaging and their applications in medicine and applied psychophysiology. His research interest includes the interaction of homo-machine.

PETER CORCORAN (Fellow, IEEE) currently holds a Personal Chair position in electronic engineering with the School of Engineering, University of Galway. He was the Co-Founder of several start-up companies, notably FotoNation, now the Imaging Division, Xperi Inc. He has more than 600 cited technical publications and patents, more than 120 peer-reviewed journal articles, and 160 international conference papers, and a co-inventor on more than 300 granted U.S. patents. He is an IEEE Fellow recognized for his contributions to digital camera technologies, notably in-camera red-eye correction and facial detection. He is a member of the IEEE Consumer Technology Society for more than 25 years. He is the Founding Editor of *IEEE Consumer Electronics Magazine*.