



# Intelligent System Utilizing HOG and CNN for Thermal Image-Based Detection of Wild Animals in Nocturnal Periods for Vehicle Safety

Yuvaraj Munian, Antonio Martinez-Molina, Dimitrios Miserlis, Hermilo Hernandez & Miltiadis Alamaniotis

To cite this article: Yuvaraj Munian, Antonio Martinez-Molina, Dimitrios Miserlis, Hermilo Hernandez & Miltiadis Alamaniotis (2022) Intelligent System Utilizing HOG and CNN for Thermal Image-Based Detection of Wild Animals in Nocturnal Periods for Vehicle Safety, Applied Artificial Intelligence, 36:1, 2031825, DOI: [10.1080/08839514.2022.2031825](https://doi.org/10.1080/08839514.2022.2031825)

To link to this article: <https://doi.org/10.1080/08839514.2022.2031825>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 08 Feb 2022.



Submit your article to this journal [↗](#)



Article views: 1744



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

# Intelligent System Utilizing HOG and CNN for Thermal Image-Based Detection of Wild Animals in Nocturnal Periods for Vehicle Safety

Yuvaraj Munian<sup>a</sup>, Antonio Martinez-Molina<sup>b</sup>, Dimitrios Miserlis<sup>c</sup>,  
Hermilo Hernandez<sup>d</sup>, and Miltiadis Alamaniotis<sup>a</sup>

<sup>a</sup>Department of Electrical and Computer Engineering, The University of Texas at San Antonio (Utsa), San Antonio, Texas, USA; <sup>b</sup>School of Architecture and Design, the University of Texas at San Antonio (Utsa), San Antonio, Texas, USA; <sup>c</sup>Department of Surgery – Vascular Surgery, The University of Texas – Health Sciences Center, Texas, USA; <sup>d</sup>25:2 Solutions Corporation, Pocatello, Idaho USA

## ABSTRACT



Animal Vehicle Collision, commonly called *roadkill*, is an emerging threat to drivers and wild animals, increasing fatalities every year. Currently, prevalent methods using visible light cameras are efficient for animal detection in daylight time. This paper focuses on locating wildlife close to roads during nocturnal hours by utilizing thermographic obtained images, thus enhancing vehicle safety. In particular, it proposes an intelligent system for animal detection during nighttime that combines the technique of Histogram of Oriented Gradients (HOG) with a Convolutional Neural Network (CNN). The proposed intelligent system is benchmarked against a variety of CNN's like basic CNN and VGG16-based CNN and also with the machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree Algorithm (DT), Linear Regression (LR), and Gaussian Naïve Bayes (GNB). The proposed detection system was tested on a set of real-world data acquired with a thermal camera on the move in the city of San Antonio, TX, USA that includes images of wild deer. Obtained results exhibit that the HOG-CNN combination achieved approximately 91% correct detection accuracy of wild deer on roadsides, while it outperformed the rest of the tested machine learning algorithms.

## ARTICLE HISTORY

Received 25 October 2021  
Revised 14 December 2021  
Accepted 18 January 2022

## Introduction

Acquaintance and experiments associated with animal-vehicle collision mortalities are becoming an anticipated prerequisite of rapid globalization. Carcass/Roadkill by animal-vehicle collision is a constant threat for both humans and wild animals, which leads to significant loss of wildlife, human death, and injury every year. Amid vehicular crashes, animal actions (i.e., deer) are unpredictable and erratic on roadways. Due to the hideaway of several

**CONTACT** Miltiadis Alamaniotis  [miltos.alamaniotis@utsa.edu](mailto:miltos.alamaniotis@utsa.edu)  Department of Electrical and Computer Engineering, The University of Texas at San Antonio (Utsa), One UTSA Circle, San Antonio, Texas 78249 USA, Tel.: +1-210-458-6237

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

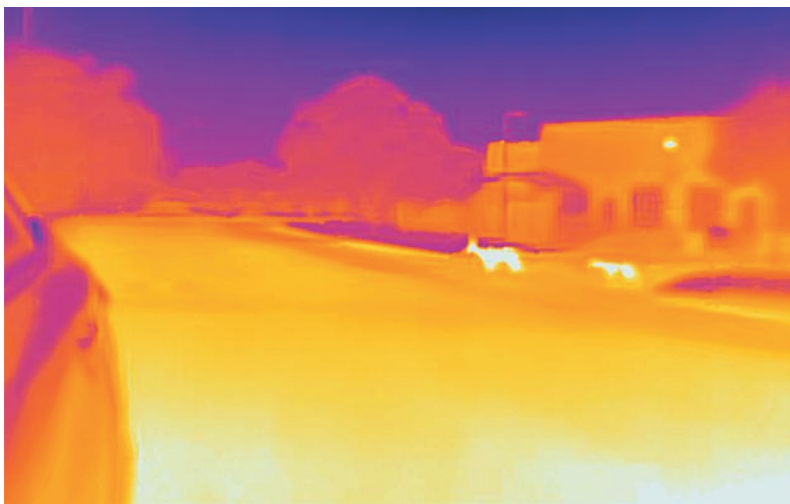
accidents, the real data around collisions that vehicles and wildlife are involved in is not determined precisely. However, the estimated wildlife crashes are over 35,000 – which means 5% of the yearly cases – in the US. The fatality rate of humankind is around 200/year (U.S. Department of Transportation, Federal Highway Administration 2020) (San Antonio News 2020). The insurance industry (Insurance Journal, Oct 2012) estimated American spending is over \$4 billion annually for property damage, and the mortality rate is around 200 per year due to Deer-Vehicle collisions. More are seriously injured, with deer being a significant part of all critical and overall collisions, particularly in Texas, USA. Furthermore, the number of property damage case filings is around 4000/year, and the cost is approximately \$1,000 per case (U.S. Department of Transportation, Federal Highway Administration 2020) (Wilkins, Kockelman, and Jiang 2019). A five-state study circumscribes that 15,000 reptiles and amphibians, 48,000 mammals, and 77,000 birds have lost their lives due to collisions with vehicles in a single month.

This paper focuses on deer detection, one of the most common forms of wildlife found by roadsides in nocturnal periods. For instance, in San Antonio, TX, every two accidents out of three are counted for wildlife crashes and are the highest reported type of car accidents (San Antonio News 2020). As per the Texas Department of Transportation, between 8 pm and midnight is the most dangerous time of the day for automobile drivers. High beams may be used to spot the deer lurking along the roadside at dark hours, but as per the state law, high beams should not be used on highways when a vehicle approaches in the opposite lane. A useful aspect of deer is that they always move in herds and groups. This primary peril leads the way to contrive effective image detection by identifying animals' presence during the dark hours. Notably, various methods have been proposed to distract and alert the deer off the road using warning reflectors, audio alarms to seize the motion of the deer, long tunnel, over bridge construction across the highways and signboards (Wilkins, Kockelman, and Jiang 2019) (Sawyer, Rodgers, and Hart 2016). However, all the above-cited approaches produced results toward the animal-vehicle collision, which are not up to the expected mark. Hence, it is unsuccessful in eliminating the collision collectively.

In the last decade, image processing and machine learning tools in automobiles have led to several advanced systems like self-driving cars, pedestrians' autodetection, and cruise control, reducing fatalities. This research focuses on detection application by combining an image processing technique with a deep learning tool, and more particularly, the Histogram of Oriented Gradients (HOG) with a convolutional neural network (CNN), as it has been preliminarily introduced in (Munian, Martinez-Molina, and Alamaniotis 2020). Notably, deep learning models are the most effective in image processing applications since they distinguish the high-level features from the low-level features (Wu, Sahoo, and Hoi 2020).

In the current work, we aim at utilizing thermal images to detect wild deer on the roadsides. Thermal images are obtained by a thermal camera that receives heat emission from the object and generates the composite image, displaying an object's approximate temperature. The thermal camera plays an important role in different applications like water leakage detection in the hydrogeological department (Massaro, Panarese, and Galiano 2021). The basic need of the human is food, this thermal camera is used to identify fresh food and alert the user that the quality is maintained or not, by alerting the system (Massaro, Panarese, and Galiano 2021). (Wang et al. 2020) showed that all over the world so many crashes happened due to gas leakage, the thermal camera detects the gas leakage or emission from the tank, which stops the big accidents. For instance, [Figure 1](#) depicts a thermal image taken during a nocturnal period using a thermal camera. Visual information is the best way to understand the prerequisites. The comprehensive parent of photographic computation utilizes imagery extraction for systematic evaluation. Object classification and identification are the central building blocks of computer vision. Before entering the core process, the dataset is preprocessed to remove irrelevant and noisy images by the basic image processing technique called "filtering the dataset." These techniques are used to extricate the information from the given visual data as per the requirements. In general, computer vision algorithms have high computational complexity, which must be coupled with fast and low complexity processing algorithms to make their real-time application to automobile systems plausible.

It is challenging to know what precisely happens in the moments during roadkill. Therefore, the vehicle's dashcam helps the drivers monitor the ambient scenery and take preventive measures. To that end, research efforts turned their



**Figure 1.** Sample thermal image captured after 8 pm from dataset collection-were used for experiments.

direction toward designing systems that predict or detect the object before any action-taking. This research follows that direction: it adopts detecting the animal objects in acquired thermal images by a camera mounted on the vehicle.

Applications like autonomous vehicles, video surveillance, security, robotics, and the driver-assist vehicle utilize the latest technologies which support the long-awaited social and economic search. Artificial Intelligence and Computer Vision systems are remarkable realizations in intensifying human success. A machine learning (ML) governed intelligent system model is the kernel conceptualization of the transport systems. In Intelligent Transportation Systems (ITS), deep learning and artificial intelligence are adventitious and entrenched. Necessarily, analyzing wild animal behavior is more complicated than training a network model. The traditional models like overpass, underpass, and double fencing (Sawyer, Rodgers, and Hart 2016) are replaced by the contemporary machine learning techniques (Pons, Jaen, and Catala 2017). In summary, the objective of this paper is to

- (1) Design and develop the HOG-CNN Intelligent System for the prediction/detection of wild animals,
- (2) Train, Validate, and Test over the own generated thermal dataset,
- (3) Benchmark the Proposed System with other machine learning classifiers.

The rest of the paper is organized as follows: Section II explains the literature review of different ideas and techniques, stimulating the current research's basic concepts. Section III provides background knowledge on HOG and CNN, while section IV states the whole intelligent system design and other algorithms. Section V discusses the results, comparisons, and findings of the proposed system in charts and tabulation, and section VI concludes the paper.

## **Literature review**

Researchers are working arduously to meet the target of reducing animal-vehicle collisions in urban highway areas. The published work reveals that the current trend is researchers waving toward machine learning, deep learning algorithms, and artificial intelligence to optimize the reduction in the number of accidents.

### ***Roadway design cost-effective methods***

Hundreds of humans and thousands of wild animals meet their end by animal-vehicle collision (U.S. Department of Transportation, Federal Highway Administration 2020). To reduce wild animal crossing, warn humans, and separate traffic, the government tried some traditional methods, including

constructing bridges in the forest areas and installing reflectors, signboards, and fencing. It is a socio-economical model, whereas it is not an effective way to reduce roadkill (Benten et al. 2018).

By constructing the overpass and underpass in the highways, the number of accidents is reduced, but it comes at a high cost. For instance, in US191, the average cost for constructing these underpasses and overpasses is approximately US\$400 K and US\$2 M, respectively (Sawyer, Rodgers, and Hart 2016). Wilkins, Kockelman, and Jiang (2019) compared all the existing traditional methods to the machine learning algorithm and regression model to detect animals (Sawyer, Rodgers, and Hart 2016). The outcome of this work motivates and paves the path for our proposed system.

### ***Image preprocessing***

In the last few decades, computer-based technology (image processing) has played a prominent role in alleviating image quality, feature detection, manipulation, interpretation, and classification. The HOG transformation, which is used for feature detection, has been improved by integrating contour-based or gradient-oriented-based methods that have shown effectiveness in thermal image analysis (Zhou, Wang, and Wang 2012).

The CENSUS TRansform hISTogram (CENTRIST) is one of the transforms used to extract thermal image features. In particular, this transform is used to detect the human presence (Riaz, Jingchun Piao, and Shin 2013). Both HOG and CENTRIST are highly effective algorithms concerning thermal image preprocessing. A segment of the system proposed by Su et al. (2013) and Zhu et al. (2006) is the cornerstone upon which our proposed method has been built (i.e., contour-based and cascaded gradient-oriented-based preprocessing, respectively).

### ***Thermal imaging***

Predominantly in object detection or identification techniques, the input data form is essential for accommodating feature selection and processing. Thermal imaging measures the environment's temperature and accommodates objects' detection by identifying temperature differences (Sibanda et al. 2019).

The main characteristic of thermal images is that the pixel's intensity expresses the amount of temperature emitted from the object (Christiansen et al. 2014). Based on thermal imaging, the shrubs' concealed deer are unveiled by their body temperature (Zhou, Wang, and Wang 2012). Furthermore, Riaz, Jingchun Piao, and Shin (2013) utilized the thermal image dataset, which has been captured accurately at 21°C. The human body temperature will always be constant on a concave surface, by which the detection is high-speed (Santhi 2017). Zilkha and Spanier (2019) proposed a classification and detection

method using a set of both regular and thermal images. HOG and cascaded HARR are used to detect the human in the thermal image on both day and nighttime (Emine and Ahmet 2017).

### ***Machine learning for detection applications in transportation***

Machine learning is expected to act as the hardcore to vindicate the transportation systems and improve road safety. Of significance is the technology growth that must be diverted toward safeguarding animals to improve animal welfare. Christiansen et al. (2014) describe the detection of various animal stances (sitting, jumping, turning, walking, semi-sitting, and standing) by depth-based tracking systems. All classifiers like decision tree, random tree, random forest, rule induction, support vector machine, K-nearest neighbor, Naive Bayes, and logistic regression are applied and compared for different animals' postures. Hence, it is proven that for different stances, accuracy is attained at its maximum through the decision tree, SVM, KNN, NB, and rule induction (Pons, Jaen, and Catala 2017).

In Guo et al. (2012), the AdaBoost algorithm and SVM are used to detect humans, while results encourage the combination of image processing with learning algorithms, while in Sibanda et al. (2019), active and passive infrared sensors are used in the vehicle. The Bayesian dynamic logistic regression model predicts and updates the real-time crash risk evaluation (Yang, Wang, and Yu 2018). Because of the growth of people and vehicles in the city, traffic, and accidents are increasing. An intelligent transportation system rectifies these with a deep learning network model. Affonso et al. (2017) compared machine learning techniques like KNN, SVM, DT, NN with the texture descriptor to the deep learning model of CNN. The CNN performs more efficiently when compared to all other techniques. The confusion matrix shows all the models' real predicted values, while the importance of labeling the dataset is highlighted.

The comparison of HOG-SVM and CNN for human detection in occlusion and non-occlusion regions is presented by Islam et al. (2017). Even though the HOG-SVM is more efficient for the non-Occlusion images, the CNN produced adequate accuracy in real-time world application. This sets the grounds for using HOG with CNN to improvise the results (Emine and Ahmet 2017) (Aslan et al. 2020). Wang et al. (2019) show the comparison of deep learning networks for enhancing transportation systems – the four network models are Deep Neural Network (DNN), CNN, Recurrent Neural Network (RNN), and Deep Q Networks (DQN).

In a divergence from studies described above to the preeminent of the author's understanding, this work's uniqueness does not experiment in any of the current research work. This work's individuality is real-time data captured during driving is used for animal detection during the nocturnal period. The following sections describe the promising intelligent system model for advanced

transportation systems. Deviating away from these methods, it is increasingly necessary to compare the machine learning and deep learning algorithms in combination with the image processing technique. Before inflowing into the methodology and proposed system, the discussion on basic knowledge about image processing, machine learning, and deep learning is highly required.

### Background knowledge

This section gives an overview of the two techniques, namely, Histogram of Oriented Gradients (HOG) and Convolutional Neural Network (CNN), that are used to develop the proposed intelligent system. The following information will encourage beginners to leverage their quest of image processing and kindles their interest in research in these areas.

#### Histogram of Oriented Gradients (HOG)

The features of the image are determined by HOG and are known as HOG descriptors. HOG is the technique used to define the object in the image by applying computer vision approaches, while it consists of significant preprocessing techniques for images (Aslan et al. 2020).

Figure 2 depicts the gradient orientation calculation and detection of the localized portions in the image. The HOG descriptor shows the image’s filter measurement, which helps in object or target image identification. A filter, whose size is 2x2, as shown in Figure 2, in the form of a cell block is moved horizontally from cell 1 to cell n. The gradient orientation shows the object’s continuity, while the arrow direction clearly shows the object’s presence in the image. HOG’s widespread significance made it utilized in applications like animal detection, vehicle identification, concrete cracks, etc., in static imagery (Christiansen et al. 2014) (Wei et al. 2019) (Zhu et al. 2006) (Lowe 2004).

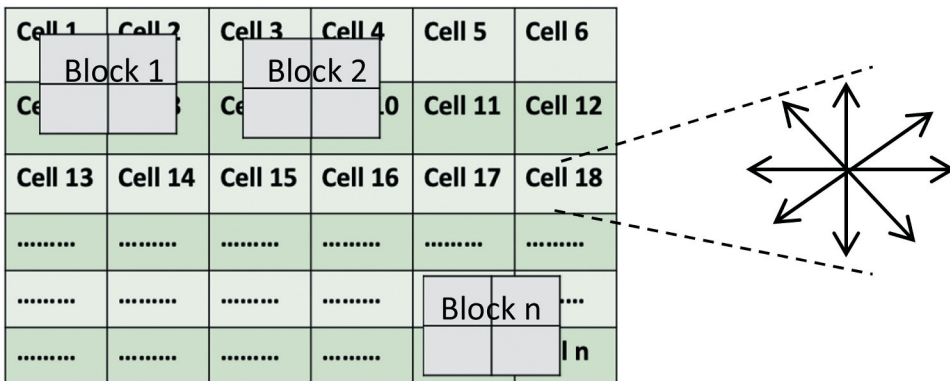
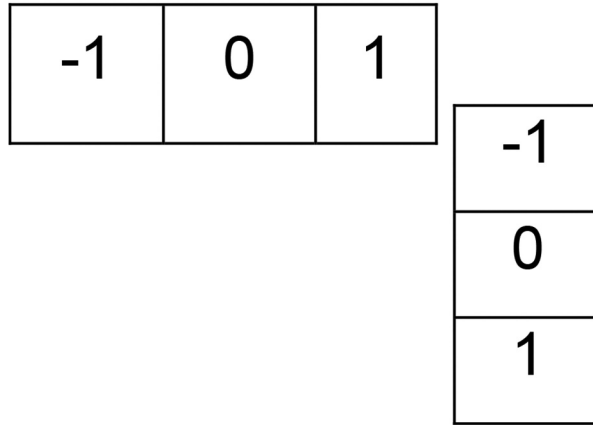


Figure 2. Hog descriptor with pixel gradient orientation.





**Figure 3.** Sobel operator values for gradient computation in the x and y directions of the image.

Algorithmic steps in HOG transform include normalization, magnitude, orientation calculation, block normalization, and feature vector calculation. The general normalization defines the whole image's division by 255 and transforms the original image into a normalized image. The third stage's normalized image is forwarded to the gradient computation step using the Sobel operator, whose values are shown in [Figure 3](#). This is used to calculate the gradients  $g_x$  and  $g_y$  of the image, respectively. To calculate and emphasize how the high spatial frequency corresponds to edges, Sobel operators are moved all over the image.

The magnitude and direction of the spatial continuity pixel in the image are calculated using the equations (1) and (2) given below. Computation is based on the blocks, pixel per cell, cells per block, and the number of blocks per image.

$$g = \sqrt{g_x^2 + g_y^2} \quad (1)$$

$$\theta = \arctan\left(\frac{g_x}{g_y}\right) \quad (2)$$

The notation 'g' in equation (2) is the Sobel operator's magnitude ( $g_x, g_y$ ). The respective magnitude is determined by taking the square root of vertical (y) and horizontal (x) operators. The notation ' $\theta$ ' denotes the concerning operator's angle or direction ( $g_x, g_y$ ). Here the conversion of cartesian to polar coordinates is needed. This calculation mostly depends on the edges of the object.

$$f = \frac{v}{(\|v\|_1 + e)} \quad (3)$$

L1 normalization is used as the block normalization, eliminating the image's extra details using equation (3). In (3), 'v' is the non-normalized vector values holding all histograms in each block,  $\|v\|_1$  the L1 – norm value, and 'e' the negligible small constant. The features are extracted by the rescaling intensity within a range from 0 to 255.

Feature extraction accuracy is enhanced by the normalization of the image with a dense grid of blocks, as discussed in Dalal and Triggs (2005). The two main block geometrics are square/rectangle HOG (R-HOG) and circular HOG (C-HOG) (Zhu et al. 2006) (Su et al. 2013). The fundamental parameters are i) the number of pixels, ii) the number of cells per block, and iii) the number of channels per cell. These normalization methods motivated the HOG usage in our proposed intelligent system.

### Convolutional neural network

An abstract structure of a convolutional neural network is shown in Figure 4. It consists of convolution layers (filters and kernel size), which select the specific characteristics of every input to one output. The max-pooling layers extracted the feature vector, it only maintains the feature considered by the convolutional layers in the previous step. The fully connected layer is the multilayer neurons, which get the input from the previous layer output, and it has the binary representation used for the classification output. The output matrix of CNN is generated only when the dimension of the images are primarily 2D. The extraction of features is determined as the product of the height, width, and channels (i.e., colors) of the image at hand. The first layer's output feature is fed into the next convolution layer, while the kernel size will be defined as the sliding window used for the convolution across the data.

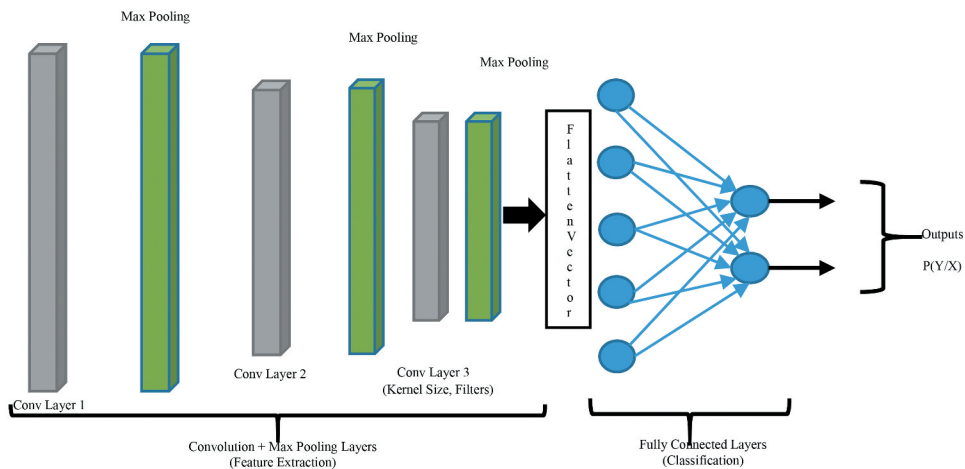


Figure 4. Convolution neural network.

The filters serve as the input parameters which determine the number of used sliding windows. Also, the max-pooling layer is used to prevent over-fitting by considering the maximum value of several features, while the height and kernel size determines the number of neurons. The fully connected layer (shown in [Figure 4](#)) uses an activation function to generate the probability distribution over the number of classes as discussed in ([Aslan et al. 2020](#)), ([Gomez Villa, Salazar, and Vargas 2017](#)), ([Zhao et al. 2020](#)), and ([Ferreira and Giraldi 2017](#)). Notably, this defined 1D CNN will not consider the feature's location in the vector segment ([Affonso et al. 2017](#)). It should be noted that detected feature sets in 1D, 2D, and 3D CNN versions are the same. However, there are significant differences in the feature detector slider movement in the segment, as shown in ([Islam, Raj, and Al-Murad 2018](#)) and ([Bai 2017](#)).

## **Methodology**

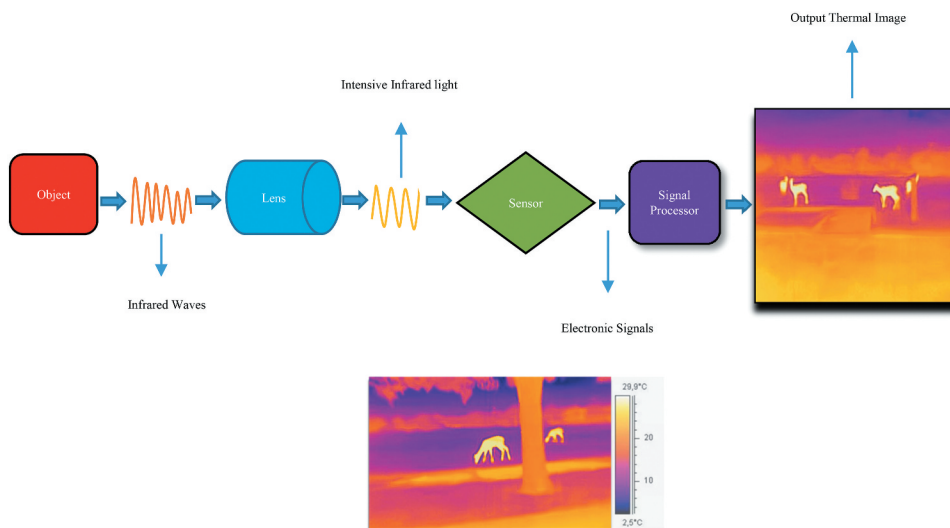
This section vividly depicts the methods, implementation, and synergism of comparing machine learning classifiers, HOG, and CNN to detect wild animals in automobile applications. During nocturnal periods, the proposed system aims to avoid road crashes resulting from the increase of roadside wild animal detection accuracy. From a researcher's perspective, the whole system encompasses the original thermal image, HOG transformation, feature extraction, HOG image, convolutional neural network, and confusion matrix. It should be noted that the confusion matrix tabulates the detection and misdetection rates as applied to our test dataset.

### ***Thermal camera, specifications, and thermal images***

Thermogram images (also called thermal images) express the measured heat emitted from an object in the form of infrared radiation. The hottest regions will generate high radiation, and, in turn, the image will produce more intensity or clarity. The big difference between a night vision camera and a thermal camera is how it captures infrared radiation.

The night vision camera captures the shorter wavelength radiation of the object. In contrast, the thermal camera captures the longer wavelength of infrared radiation, allowing detection of the object precisely and accurately ([Peeters et al. 2018](#)) ([Kwasniewska et al. 2020](#)). The mid and long wavelengths are referred to as thermal infrared. The atmosphere only emits radiation with a specific wavelength due to various objects' presence and proximity. Generally, the invisible infrared radiation is captured and converted into visible images.

A thermal image will have both cooler and warmer objects. The cooler can take up purple, blue, or green color, and the warmer area is seen as red, yellow, or orange as shown in [Figure 5](#). Infrared lays between the visible and



**Figure 5.** Thermal camera operational process.

microwave light spectrum (Gade and Moeslund 2014) (Kwasniewska et al. 2020). Its acquisition includes a set of camera lens that points at an object and focus on the infrared light emitted by all the surrounding objects. An array of detectors scans the focused light, and a precise temperature pattern is generated, known as a thermogram image. Subsequently, the thermogram is translated into impulses sent to the circuit board/signal processor. Furthermore, the chip inside the camera converts the impulses into display images. The whole thermal image acquisition process is shown in Figure 5.

In this research, the iOS device-based Forward-Looking Infrared (FLIR) thermal camera FLIR ONE Pro is used to acquire the animal image (deer) during nocturnal hours. The FLIR can capture the images between 0 to 35 degrees centigrade operating temperature. The scene dynamic range of the thermal camera is  $-20$  to  $120$  degrees Centigrade. The file formats are MPEG, and MOV with the lens focus is fixed from 15 cm to infinity. The thermal sensitivity of this pro model is 100mK; both video recording and image capture are possible. The video and still image resolution are four times higher than any FLIR ONE model.

### **Data collection**

This section describes the data collection processes. The dataset is 1500 images long; only 1068 images are used for the proposed system. The blurred images are removed from the collection. The dataset includes images without animals also captured from the same places and consists of various dimensional and behavioral images that captured different movements of the deer with various paces. Overall, the dataset collection

**Table 1.** Dataset details.

Descriptions	No. of Images
Total Images	1500 (Before Filtering)
Images used for Research	1068 (After Filtering)
Images with animal	714
Images without animal	354
No. of images used for training	854
No. of images used for testing	214

contains diversified images with different sizes and have various resolutions like 640x480, 480x480,  $640 \times 520$  pixels are used for the proposed system model training and testing. In Summary,

Data collection shown in [Table 1](#) was executed every night around from 6 pm to 10 pm for three consecutive weeks. The challenges in data collection are i) Unavailability of animals during rainy nights, ii) High speeding of vehicles resulting in divided animal attraction and presence, and finally iii) Extended hibernation due to natural calamity. The full dataset is comprised of a set of complete JPEG images. The time from 6 to 10 pm is selected because more animal movement and accidents occur during these timings, as explained in the introduction section.

Natural climate condition is an essential deliberation for the whole proposed system. If the temperature during the daytime is very high, then the road and trees' temperature will be very high in the evening time. This increases the complexity of the system. The images are captured at different places around the San Antonio area in Texas, the USA, in May 2019. The data is collected with different backgrounds (mostly from the forest environment), while all was taken from a highway. The animal's temperature is captured in the thermal camera despite being hidden behind trees and bushes or crossing the roads.

### ***Histogram of orientation gradients***

The computer vision preprocessing technique, Histogram of Oriented Gradients, is generally used to find the localized feature of the image or object in the form of information data structure and generate a feature vector. The local intensity or the edge direction are identified irrespective of the gradient and edge detection information.

### ***General setup***

The feature vectors of the image use descriptors, including circular, rectangular, or dense grid to define the so-called HOG transformation. The simplified image information is known as descriptors by eliminating the given image's extra information and effectively utilizing the image patch or representation. The general block diagram of the HOG transformation is shown in [Figure 6](#), where the HOG transformation's steps are presented.



**Figure 6.** HOG transform – Overview.

The descriptor detects the image's features based on color, dimension, and mainly on the objects' shape or edge. The HOG descriptor is the joint venture or histogram vector values of all the blocks of the image. Before the transformation, the image must be preprocessed for its object detection readiness, that is, resizing and cropping. These types of preprocessing techniques in image processing accommodate the increase in the quality of the dataset (Santhi 2017). The output of the descriptor is the HOG features of the image, as shown in Figure 6.

### **Block diagram and algorithm description**

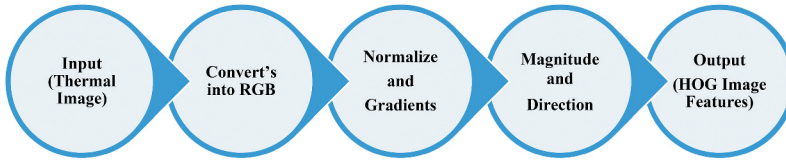
The HOG transform includes gradient computation, orientation binning, descriptor block formation, and block normalization (Santhi 2017). The previous section provides a conceptual description of the HOG transform; however, the HOG transform's detailed description is given below (Mallick and Learn Open 2016). The number of pixels per cell, the number of cells per block, and the number of channels per cell histogram are the HOG transform parameters (Su et al. 2013) (Zhu et al. 2006).

The separation of the image into cells and the histogram of the gradient calculation on directions within each cell is called gradient computation. The discretization of each cell into angular bins depends on the orientation known as orientation binning. It also contributes a weighted gradient to its angular bin. Descriptor block formation is defined as the grouping of cells as per the constraint based on the adjacent cell in the blocks. Block normalization is defined as the normalizing group of histograms, represented as the block histogram, and will serve as the block descriptor. The Gaussian window is used inside each block of the image to minimize pixels weighing all over the blocks' edges.

The algorithm used for HOG transform is to identify the edges of the object by measuring the continuity of the pixel values of the image is given below:

- Step 1: Preprocessing of the Images by Normalization.
- Step 2: Calculate the gradient magnitude images – Sobel Operator.
- Step 3: Calculate the Histogram of Oriented Gradients in  $8 \times 8$  cells.
- Step 4: Block Normalization (16x16) – L1 norm.
- Step 5: Calculate the HOG feature vector.

Figure 7 presents the entire process of the proposed feature detector HOG transform. Firstly, the input thermal image with the dimension of  $480 \times 640$  is given to the feature detection system. Before that, the



**Figure 7.** HOG transform – Block diagram.

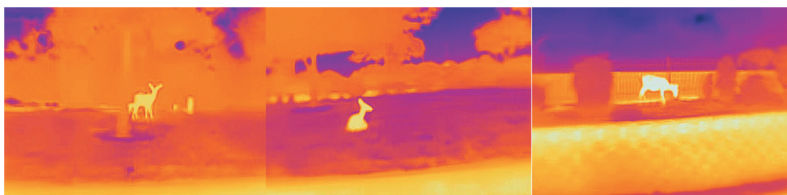
preprocessing techniques of resizing (to reduce the number of features without information loss) and cropping (to remove the brand camera logo from the image) are the prerequisites for preparing the input as the preferred one for the next-level processing. The given image is converted into an RGB image to be compliant with the remaining process's flexibility. The features are detected, and these images are utilized as an input for CNN for optimal results.

### ***Proposed system setup***

This section describes the second part of the proposed detection system. In conjunction with the significant HOG transform output, the HOG image is given to the deep learning algorithms to detect wild animals.

**Figure 8** shows a sample of input thermal images from the dataset directory. These images with a dimension of either  $1440 \times 1080$  or  $640 \times 480$  are collected by the high-resolution FLIR brand thermal camera. In the acquired images, the deer are in different positions, which makes their detection more complicated. The quality or clarity of the image is still low because of the nocturnal period. The thermal images shown in **Figure 8** are captured in the nighttime, specifically from 7 pm to 10 pm near the highway roads. These thermal images have high color density due to the high-temperature property.

During the nighttime, the temperature of the animals' (deer) blood is more visible to the human eye through thermal cameras, which helps the expert system or human in object detection. Due to the lower temperature range, the background objects like the sky, road, curb, tree, and branches lose their image detection prominence. Sometimes the objects in the background have a higher temperature than the animal, which leads to misdetection. The HOG



**Figure 8.** Input thermal images.

transform can easily detect the edges of the animals and the continuity of the pixels. Furthermore, the continuity in the direction of orientation plays a vital role in the feature analysis and animal detection using HOG.

The following steps are the general methodology/algorithm used in the proposed system:

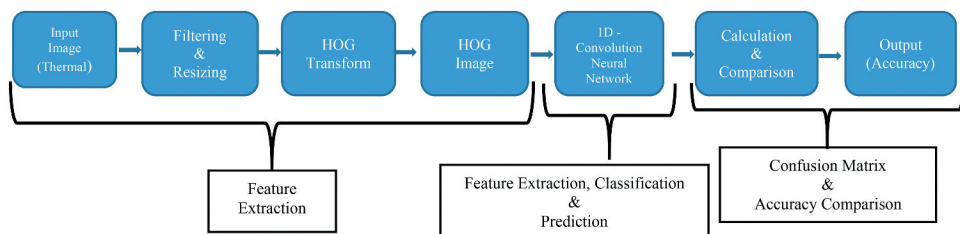
- Step 1: Image acquisition.
- Step 2: Load the image files as a folder directory.
- Step 3: Images are preprocessed by image processing and HOG, and features are taken into consideration.
- Step 4: Append all the calculated features into vector values.
- Step 5: Split the data for training and testing as 80% and 20% or 70% and 30%.
- Step 6: Train data for different epochs with different activation functions, filter, kernel, dropout, dense and max pooling.
- Step 7: Predict the results as a confusion matrix with the true positive, true negative, false positive, and false negative.
- Step 8: The testing accuracy and losses are calculated and shown in the confusion matrix.

The proposed system block diagram in [Figure 9](#) clearly explains the correlation between the HOG transform feature detection output (HOG Image) and the machine learning algorithm.

This section is divided into three subsections,

- HOG Transformation – Feature Extraction.
- Convolution Neural Network (CNN) – Feature Extraction, Classification, and Prediction.
- Comparison – To identify the efficient algorithm for the detection.

The input thermal image dataset with the dimension  $640 \times 480$  is collected and given for the HOG transform for the feature detection. The size of the image directory contains 1068 images. Due to the nature of thermal images, all of them are similar in color and resolution. This similarity makes the proposed



**Figure 9.** Block diagram of the proposed system.

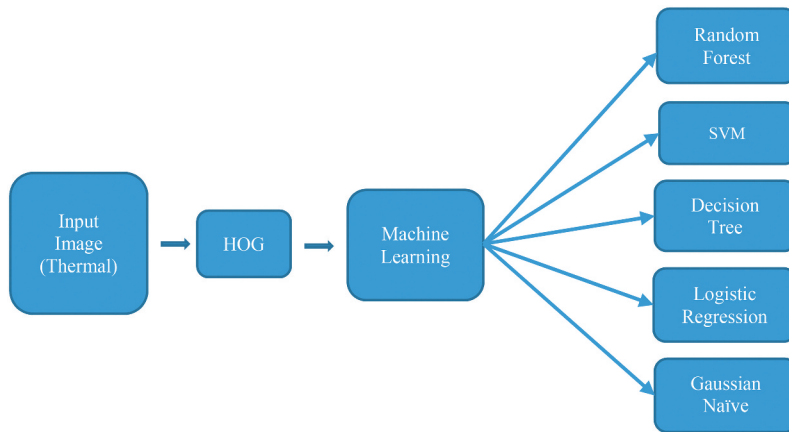


system more complex in detection. Assign the labels (0&1) based on the deer's presence for all the images in the dataset. The preprocessing tools are used to generate a 'filtered' dataset by cropping and resizing the images, which increases the dataset's accuracy. The challenges that occurred during the thermal image detection are, the similarity of the image will increase the misdetection of the animal in the given image, the high-temperature background creates more problems in the detection. The other vehicle that is parallel to the proposed vehicle also creates problems in the system. Sometimes, the high-temperature objects like stone, wood, green leaves on the roadside also shake the stability and performance of the proposed system. If a metallic object is present in the image, the system will lead to misdetection. The emissivity scale ranges from 0.01 to 0.99. By using emissivity the heat radiation will be measured.

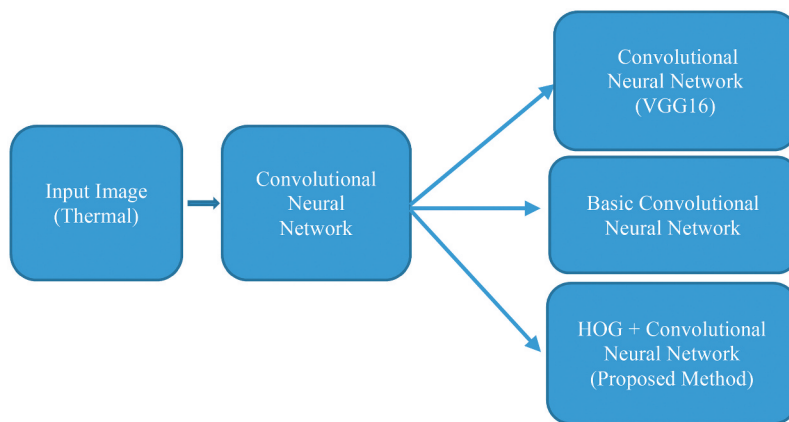
The original data directory is fed into the HOG transformation for detecting the features from the images. The HOG transform works based on the orientation, gradients, and magnitude of the images. The HOG parameters like window size, block size, cell size, and bin size determines the structure of HOG descriptors. The orientation value is 9, pixels per cell (8, 8), cells per block (2, 2), and the L1 block normalization is used as the parameters for the feature detection. The HOG transformation ('HOG features') is then saved in another directory in the list of arrays. The second section of the proposed system is the machine learning algorithm, which is used to detect the animal's features and the classification and prediction. The output images from the HOG transform are fed into the machine learning. The CNN architecture has the feature parameters as filter size 256, kernel size 5, and the relu activation function is used, with the binary cross-entropy loss function and adam optimizer.

Figure 10 shows the machine learning classifiers used to compare the proposed methodology. The classifiers RF, SVM, DT, LR, and GNB, are used for feature extraction and object (deer) detection. The extracted features are stored as vector-matrix values in an array list. The Jupyter notebook<sup>1</sup> is used for all the algorithms other than 1D - CNN. The keras<sup>2</sup> inbuilt libraries are used for the binary classification prediction. The dataset directory is split into two, with deer ('1') and without deer ('0') for all algorithms except 1D-CNN.

Figure 11 shows the comparison between the CNN with vgg16 as the backbone, CNN, and the proposed method (CNN + HOG), which uses the thermal image as input. For CNN technique, python v3.6 package<sup>3</sup> is used for the simulation to detect the animal in the image. The precision, weighted and macro average are calculated for the assigned labels. The training and validation of the dataset will take 80% or 90% of the original dataset. The testing will have the remaining 20% or 10% of the whole dataset. Around 682 images for the training, 172 images for the validation, and 214 images for the testing are used for the proposed intelligent expert system.



**Figure 10.** Benchmark – Machine learning algorithms.



**Figure 11.** Convolutional Neural Network comparison with the proposed method.

The third section of the proposed system in [Figure 9](#) gives all algorithms' output, calculated by the test data's confusion matrix and accuracy. The confusion matrix provides the detection and misdetection of the given test dataset. The precision, recall, and average values give the accuracy of the algorithms. By increasing or decreasing the convolutional layers, which have filters, kernel size, and normalization, the accuracy will be optimized based on the previous step's features.

Also, by changing the activation functions, the accuracy will be optimized. The activation functions are selected based on the applications like medical, transportation, image processing, etc. The similarity of all the images for the diversified images in the dataset will produce the overfitting problem. The probability and binary cross-entropy detect the object features. All the classifier's output (accuracy) is compared to identify the efficient algorithm for detecting the animal.

## Results and discussion

This section explains the experimental results and the subsequent solution based on the influence analysis performed. The experimental results show the classification performance and the parameters used in the proposed system implementation. The bar chart, accuracy graph, tabulation, confusion matrix comparison for all training, validating, and testing results of all the machine learning classifiers, HOG, and CNN, are intensively detailed below. This section is further divided into data analysis and test setup, which describes when and where the dataset collection occurred and the environmental conditions. The device setup explains the simulation's device and is followed by the model parameters and classification reports, which describe the trainable and non-trainable parameters. The description and result analysis discuss the inputs, stagewise results, and outputs of the proposed system.

### Device setup

Regarding the collection instrument, a FLIR brand thermal camera is used for capturing the images. The device setup used for the complete training, validation, and testing is MacBook Pro with specifications as follows,

- Processor 2.3 GHz Dual-Core Intel Core i5.
- Memory 8 GB 2133 MHz LPDDR3.
- Graphics Intel Iris Plus Graphics 640 1536 MB.

### Detection evaluation

In our work, a set of metrics was adopted to evaluate the system's performance that contains a precision, recall, f1-score, and macro average. By definition, i) Precision is the fraction of relevant instances among the retrieved instances in (4) ii) Recall is the fraction of relevant instances that have been retrieved over pertinent total instances in the image as shown in (5) iii) F1-Score is a measure of a test's accuracy, and it can be interpreted as a weighted average of the precision and recall (0 & 1) as shown in Equation (6) and, iv) Macro-averaging : Collect decisions for all classes, compute contingency table, evaluate.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (5)$$

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

**Table 2.** Model parameters with the specification for CNN.

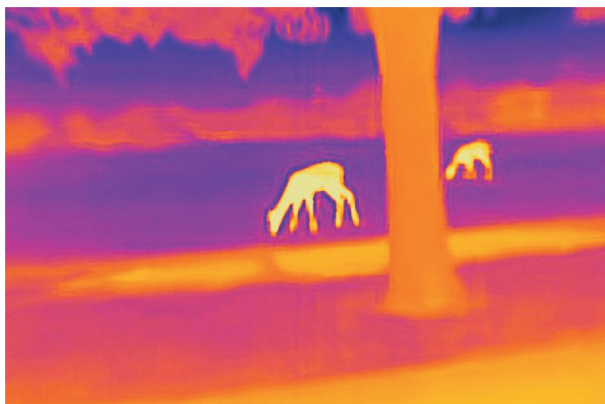
Layer (type)	Output shape	Parameters
Conv1d_7 (Conv1D)	(None, 2696, 125)	2000
Conv1d_8 (Conv1D)	(None, 2692, 125)	78250
Global_max_pooling1d_4(Global)	(None, 125)	0
Dense_4(Dense)	(None, 1)	126
Total Parameters: 80,376		
Trainable Parameters: 80,376		
Non-Trainable Parameters: 0		

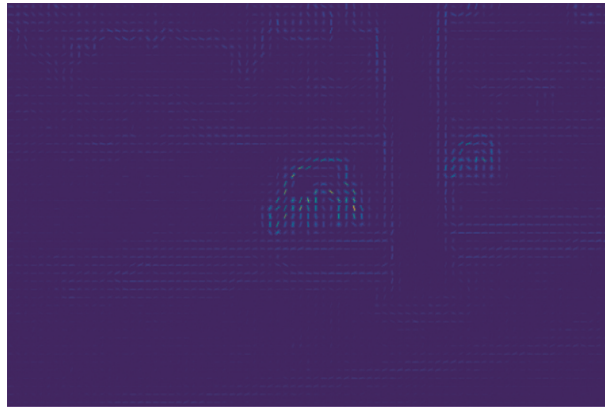
### **Model parameters**

Table 2 depicts the parameters calculated from the deep learning algorithm along with source code execution. The trainable and non-trainable parameters are generated based on the number of filters and kernels chosen in the convolution layer. Overall, only two 1D convolution layers, one global max-pooling layer, and one dense layer are used, as shown in Table 2.

### **Result analysis**

The result analysis shows all possible inputs and outputs of the proposed Intelligent System. Figures 12 and 13 are the resized and cropped original thermal image and the HOG image from the full dataset. These figures provide the input thermal image of dimension  $640 \times 480$  and HOG image of  $128 \times 128$  or  $256 \times 256$  dimension, respectively. The images shown are resized and cropped accordingly. This image processing excludes the undesired information like trees, roads, trucks, and curb from the image without losing the required information. The original image size can be in a different dimension, but in the proposed system, all the images are converted into  $128 \times 128$ . This image size

**Figure 12.** Original thermal input image from the dataset.



**Figure 13.** The pre-processed HOG output image.

reduction will increase the processing speed by making the training, validating, and testing times very short. It takes around 10 to 20 sec for each training iteration or epoch.

Table 3 depicts the confusion matrix comparison for the proposed system. From this table, the accuracy percentage value shows that the convolutional neural network synergistically with HOG provides a higher percentage than the rest-tested machine learning algorithms. Thus, the proposed system is the most accurate way to detect wild animals from the given thermal image. In Table 3, the true positive defines the animal presence in the image, and the true negative states that there is no animal in the image. Equation (7), is used to calculate the machine learning algorithm’s accuracy from the confusion matrix.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}} \tag{7}$$

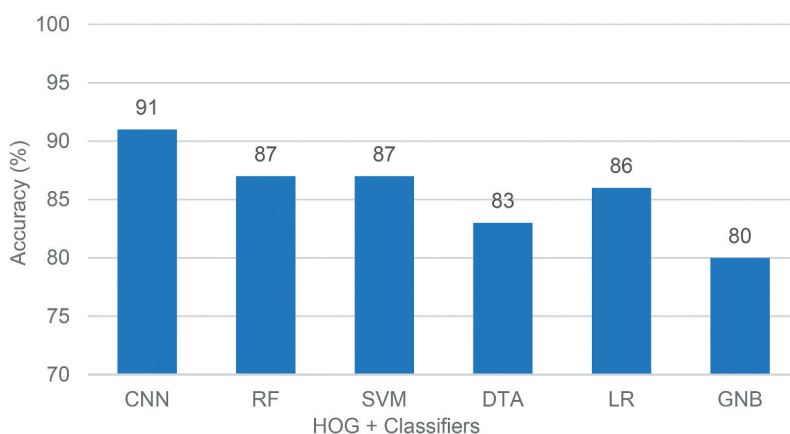
The false-negative and false-positive are the missed detection rate of the whole proposed system. The Gaussian Naïve Bayesian produces a lower percentage compared with all other algorithms. The SVM and RF algorithms are mostly

**Table 3.** Confusion matrix comparison of the classifiers for 214 images.

Serial Number	HOG + Classifiers	Confusion Matrix	True Positive	True Negative	False Positive	False Negative	Accuracy (%)
1	Proposed System – HOG + CNN	[[54 8] [12 140]]	140	54	8	12	91
2	Ensemble – Random Forest (RF)	[[51 12] [15 136]]	136	51	12	15	87
3	Support Vector Machine (SVM)	[[46 17] [10 141]]	141	46	17	10	87
4	Decision Tree Algorithm (DTA)	[[48 15] [22 129]]	129	48	15	22	83
5	Logistic Regression (LR)	[[39 24] [6 145]]	145	39	24	6	86
6	Gaussian Naïve Bayes (GNB)	[[42 21] [21 130]]	130	42	21	21	80

close to the proposed system's accuracy, but the false positive is very low in our proposed method compared with other algorithms. From Table 3, we can say that the proposed approach is the most efficient. Therefore, more experiments are conducted for the CNN network with different convolutional layers, activation functions, number of epochs, max-pooling layers, and changing the dense layers. Figure 14 shows the pictorial representation of all the classifiers' accuracy concerning correct detection percentage. Simultaneously, it is apparent that the HOG-CNN combination upon which our proposed system is based exhibits the highest performance.

As mentioned in the previous section, the precision, recall, and F1 score are calculated and shown in a detailed manner for all the classifiers. Table 4 shows the accuracy of the classifier based on the labels 0 and 1. From these tables, both the random forest algorithm and support vector machine generated the same accuracy of 87% for the thermal input images. The decision tree algorithm produces 83%, and the logistic regression has 86%. The Gaussian Naïve Bayes classifier produces moderate classification in binary classification.



**Figure 14.** Bar chart of the accuracy comparison of the proposed system.

**Table 4.** Weighted average comparison of HOG + RFC, SVM, DT, LR, and GNB classifiers reports.

Labels	Type of Classifiers	Precision	Recall	F1 Score
0	RF	0.77	0.81	0.79
	SVM	0.82	0.73	0.77
	DTA	0.69	0.76	0.72
	LR	0.87	0.62	0.72
	GNB	0.67	0.67	0.67
1	RF	0.92	0.90	0.91
	SVM	0.89	0.93	0.91
	DTA	0.90	0.85	0.87
	LR	0.86	0.96	0.91
	GNB	0.86	0.86	0.86

**Table 5.** CNN Validation accuracy comparison.

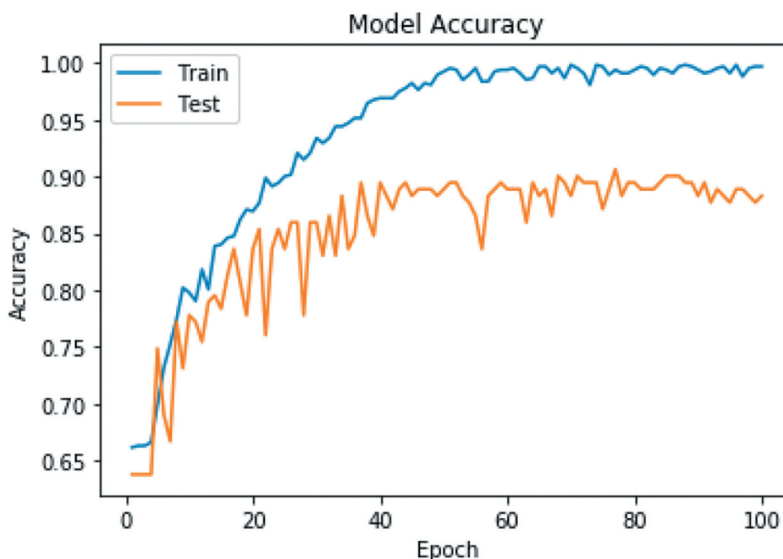
Methods	Combinations (Conv. Layer, Dense Layer)	Activation Layer	No. of Epochs	Average Time in sec./run (approx.)	Validation Accuracy (Round-off)
Convolutional Neural Network (VGG16)	(12,3)	ReLu, SoftMax	50	3–4	69%
	(9,2)	ReLu, SoftMax	100		69%
	(6,2)	ReLu, SoftMax	50		67%
	(3,2)	ReLu, Sigmoid	100		67%
Convolutional Neural Network (CNN)	(5,1)	ReLu, SoftMax	50	1–2	69%
	(3,1)	ReLu, SoftMax	100		71%
	(2,1)	ReLu, Sigmoid	100		71%
Proposed Method (HOG + CNN)	(5,1)	ReLu, Sigmoid	50	1–2	89%
	(3,1)	ReLu, Sigmoid	100		90%
	(2,1)	ReLu, Sigmoid	150		91%

The results demonstrated in Table 5 show the comparative study of the various frameworks in CNN. The result proved that the proposed system produced the most effective and optimal result compared with other techniques. In this experimentation, the same dataset has been fed to three different convolutional neural network frameworks.

The first set of experiments is performed against the VGG16 neural network and the obtained accuracy is between 67% and 69%. Secondly, basic CNN is tested against the same data set and resulted in maximum accuracy of 71%. Finally, the HOG + CNN network is tested for the results and produced a maximum of 91%. Hence, the accuracy and prominence of the system are proved. Here different combinations like different activation layers, convolutional layers, and epochs are executed. The validation accuracy is compared for all three CNN's.

The experimental results and the comparison of the proposed system concerning time and accuracy are shown in the following sections. In Figure 15, the x-axis is the number of epochs, and the y-axis is the accuracy expressed in percentage. For every epoch, maximizing accuracy is shown vividly, inclusive of training and validation. The pictorial representation of model accuracy proves the efficiency of the proposed model.

Table 6 presents the testing accuracy generated of CNN for all 10 times the 100 epochs. The input state is fixed as random for all the training, validation, and testing data. Because of the random input to the CNN, it produces different accuracy for different combinations of images. The accuracy is varied for all the times, therefore experimental results are taken into consideration.



**Figure 15.** Model accuracy – training and validation (Blue – Training and Orange – Validation).

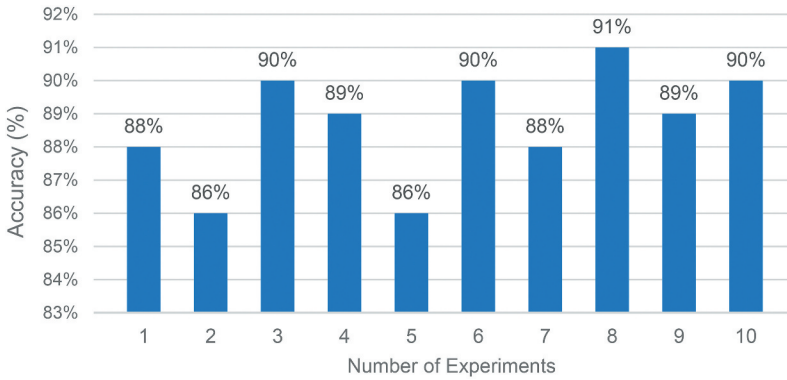
The testing accuracy is represented in percentage values, which are rounded off to the nearest integer. The average value will be 89%, the maximum and minimum will be 91% and 86% based on the detection. The time required for the testing iteration (i.e., experiment) is presented in [Table 6](#).

The time is denoted in seconds, which vary from max 3.15 secs to min 1.06 secs for 214 image frames. Because of the conversion of thermal into HOG image, the time taken for testing is a maximum of 0.0147 sec/frame and a minimum of 0.0049 sec/frame. To avoid the deep learning technique's complexity, the proposed technique (CNN + HOG) is considered to speed up the system testing time per frame. All experiments were performed on a CPU with specifications mentioned in the device setup section. It takes only a few minutes for the testing in the CPU; therefore, no GPU is required for the testing purpose.

**Table 6.** Testing accuracy and time taken per 214 frames in seconds.

Number of Experiments	Time/214 frames (sec)	Testing Accuracy (%)
1.	2.11	88
2.	2.01	86
3.	2.05	90
4.	1.06	89
5.	2.08	86
6.	2.06	90
7.	3.01	88
8.	2.05	91
9.	2.07	89
10.	1.05	90
Maximum		91





**Figure 16.** Bar chart testing accuracy for 10 times – 100 epochs for HOG-CNN.

Figure 16 illustrates the percentage variations for all 10 experiments where the training is set to have a max of 100 epochs. Based on the dataset fed for each experiment’s model, the accuracy is calculated and plotted as a bar chart graph for the (CNN + HOG) proposed intelligent system. The deep learning model will provide maximum accuracy of 91% for the eighth run. From these experimental results, the average and maximum percentage will be considered as the testing model accuracy.

Table 7 shows in detail the accuracy, confusion matrix, true positive (TP), true negative (TN), false positive (FP), and false-negative (FN) for each of the ten testing experiments. For each experiment, the input dataset is unordered and assembled randomly. The percentage variation is due to heterogeneous datasets. The false positives might increase due to background luminosity, with a brighter background showing higher false positives. Due to the

**Table 7.** Testing accuracy from 214 images for 10 times.

Serial Number	Accuracy (%)	Confusion Matrix	TP/TN	FN/FP
1	88	[[60 5] [21 128]]	128/60	21/5
2	86	[[49 9] [21 135]]	135/49	21/9
3	90	[[45 15] [11 142]]	142/45	11/15
4	89	[[51 13] [11 139]]	139/51	11/13
5	86	[[45 17] [12 140]]	140/45	12/17
6	90	[[48 10] [11 145]]	145/48	11/10
7	88	[50 13] [12 139]]	139/50	12/13
8	91	[[54 8] [12 140]]	140/54	12/8
9	89	[49 11] [12 142]]	142/49	12/11
10	90	[[51 11] [10 142]]	142/51	10/11

**Table 8.** Example of Confusion Matrix (CM).

		Predicted / Detected	
		N	P
Actual	N	51 True Negative	11 False Positive
	P	10 False Negative	142 True Positive

Confusion Matrix  
[[ 51 11]  
[ 10 142]]

problems in measuring the targets' range and velocity, the accuracy will sometimes slightly go down. Experiments 3, 6, 8, and 10 produce 90% accuracy, whereas 2 and 5, 1 and 7, and 4 and 9 produce 86%, 88%, and 89%, respectively. The confusion matrix for all 10 times of iterations is shown in Table 7, while the confusion matrix of testing number 10 is given explicitly in Table 8, clarifying the final output of the HOG + CNN of the proposed intelligent system. In particular, for experiment 10, the size of the testing dataset is 214 images. From the testing set of 214 images, there were 193 correct detected cases (true positive and true negative), and 21 were misses (false positive and false negatives). Overall, the average accuracy of the 10 cases is equal to 88.7%, while the maximum is 91%.

### Conclusion and future direction

In this paper, the HOG transformation-based machine learning classifier intelligent system is proposed and compared to identify the robust method for detecting wild animals on highways during nocturnal periods. This proposed system wherein is used to avoid damage and human kill by road accidents. As explained in the introduction section, the inefficacy of the traditional methods strikes the researchers to design the intelligent system. The dataset is gathered over different highways and other roads near the non-urban, forest region, and highways during the nocturnal hours from 6 pm to 10 pm. The stability of the system during (November – February) from evening 5 pm to morning 6 pm the system is stable. Suppose the temperature is high during 5 pm or during the daylight savings (March to November) the stability will have its limitation and the system works from 9 pm to 6 am.

To measure the proposed system's detection performance, experiments were executed using the dataset that was acquired from the real-world collection. The comparison of three types of CNN's (basic CNN, VGG16, HOG + CNN) and the machine learning classifiers like RF, SVM, DT, LR, and GNB

based on thermal images and HOG transformation, concerning obtained accuracy, leads to the selection of the efficient technique (HOG + CNN – 91%) for the detection of wild animals.

As shown in the result analysis section for the detection of deer, these generated results will improve the systems' general performance for the problem under research. The CNN machine learning technology, which relies entirely on learning features augmented with the HOG feature detection in imagery wildlife detection and classification, produces efficient and effective results. In this research exploration, the utmost accuracy has been obtained using CNN and HOG together. The sensitivity of the model is measured by conducting more experiments on the identified intelligent system. Although more time is required for the CNN model's training compared with other techniques, it produces high accuracy, which increases human survival during major road accidents caused by a wild animal.

This work demands time sensitivity, as the time taken for the testing, all the techniques require 1 and 3 seconds for all 214 image frames. The usage of two convolution layers gives good accuracy and approximately stable output for the proposed system. By increasing the convolutional layers, the stability of the accuracy is not disturbed. In another direction, by reducing the dataset's size or high filtration, the detection accuracy will further increase.

Considering the proposed intelligent system's outcome, the recommendation is given for the proposed method to use different datasets to avoid mitigating wild animal-vehicle collision in rural areas. The model's future work can be extended using the pre-trained model, changing the number of convolution layers for the different datasets, epochs and increasing the dataset size for high prediction to achieve the model's performance.

## Notes

1. <https://jupyter.org/>
2. <https://keras.io/api/>
3. <https://www.python.org/downloads/release/python-360/>

## Disclosure Statement

No potential conflict of interest was reported by the author(s).

## References

- Affonso, C., A. L. D. Rossi, F. H. A. Vieira, and A. C. P. D. L. F. de Carvalho. 2017. Deep learning for biological image classification. *Expert Systems with Applications* 85:114–22. doi:10.1016/j.eswa.2017.05.039.

- Aslan, M. F., A. Durdu, K. Sabanci, and M. A. Mutluer. 2020. CNN and HOG based comparison study for complete occlusion handling in human tracking. *Measurement: Journal of the International Measurement Confederation* 158. doi:10.1016/j.measurement.2020.107704.
- Bai, S. 2017. Growing random forest on deep convolutional neural networks for scene categorization. *Expert Systems with Applications* 71:279–87. doi:10.1016/j.eswa.2016.10.038.
- Benten, A., T. Hothorn, T. Vor, and C. Ammer. 2018. Wildlife warning reflectors do not mitigate wildlife–vehicle collisions on roads. *Accident Analysis and Prevention* 120:64–73. doi:10.1016/j.aap.2018.08.003.
- Christiansen, P., K. A. Steen, R. N. Jørgensen, and H. Karstoft. 2014. Automated detection and recognition of wildlife using thermal cameras. *Sensors (Switzerland)* 14 (8):13778–93. doi:10.3390/s140813778.
- Dalal, N., and B. Triggs (2005). Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 886–93 vol. 1. DOI:10.1109/CVPR.2005.177
- Emine, Cengil, and Cinars Ahmet. 2017. Comparison Of Hog (Histogram of Oriented Gradients) and Haar Cascade Algorithms with a Convolutional Neural Network Based Face Detection Approaches. *International Journal of Advance Research* 3: 244–55. www.IJARIIIT.com
- Ferreira, A., and G. Giraldi. 2017. Convolutional neural network approaches to granite tiles classification. *Expert Systems with Applications* 84:1–11. doi:10.1016/j.eswa.2017.04.053.
- Gade, R., and T. B. Moeslund. 2014. Thermal cameras and applications: A survey. *Machine Vision and Applications* 25 (1):245–62. doi:10.1007/s00138-013-0570-5.
- Gomez Villa, A., A. Salazar, and F. Vargas. 2017. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics* 41 (July):24–32. doi:10.1016/j.ecoinf.2017.07.004.
- Guo, L., P. S. Ge, M. H. Zhang, L. H. Li, and Y. B. Zhao. 2012. Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine. *Expert Systems with Applications* 39 (4):4274–86. doi:10.1016/j.eswa.2011.09.106.
- Islam, K. T., R. G. Raj, and A. Al-Murad (2018). Performance of SVM, CNN, and ANN with BoW, HOG, and image pixels in face recognition. 2nd International Conference on Electrical and Electronic Engineering, ICEEE 2017, 1–4, RUET, Rajshahi, Bangladesh. DOI:10.1109/CEEE.2017.8412925
- Kwasniewska, A., J. Ruminski, M. Szankin, and M. Kaczmarek. 2020. Super-resolved thermal imagery for high-accuracy facial areas detection and analysis. *Engineering Applications of Artificial Intelligence* 87:103263. doi:10.1016/j.engappai.2019.103263.
- Liu, S., and W. Deng, “Very deep convolutional neural network based image classification using small training sample size,” 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 2015, pp. 730–34, DOI: 10.1109/ACPR.2015.7486599.
- Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60 (2):91–110. doi:10.1023/B:VISI.0000029664.99615.94.
- Mallick, S., and C. V. Learn Open, 2016, Accessed 2020. <https://www.learnopencv.com/histogram-of-oriented-gradients/>
- Massaro, A., A. Panarese, and A. Galiano, “Infrared thermography applied on fresh food monitoring in automated alerting systems,” 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, 2020, pp. 554–58, doi: 10.1109/MetroInd4.0IoT48571.2020.9138207.

- Massaro, A., A. Panarese, and A. Galiano, “Technological platform for hydrogeological risk computation and water leakage detection based on a convolutional neural network,” 2021 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT), 2021, pp. 225–30, doi: [10.1109/MetroInd4.0IoT51437.2021.9488561](https://doi.org/10.1109/MetroInd4.0IoT51437.2021.9488561).
- Munian, Y., A. Martinez-Molina, and M. Alamaniotis (2020, July). “Intelligent system for detection of wild animals using HOG and CNN in automobile applications,” *11th International Conference on Information, Systems and Applications (IISA)*, Piraeus, Greece, pp. 8.
- Peeters, J., E. Louarroudi, B. Bogaerts, S. Sels, J. J. Dirckx, and G. Steenackers. 2018. Active thermography setup updating for NDE: A comparative study of regression techniques and optimisation routines with high contrast parameter influences for thermal problems. *Optimization and Engineering* 19 (1):163–85. doi:[10.1007/s11081-017-9368-z](https://doi.org/10.1007/s11081-017-9368-z).
- Pons, P., J. Jaen, and A. Catala. 2017. Assessing machine learning classifiers for the detection of animals’ behavior using depth-based tracking. *Expert Systems with Applications* 86:235–46. doi:[10.1016/j.eswa.2017.05.063](https://doi.org/10.1016/j.eswa.2017.05.063).
- Riaz, I., K. Jingchun Piao, and H. Shin. 2013. Human detection by using centrist features for thermal images. *IADIS International Journal on Computer Science and Information Systems* 8 (2):1–11.
- San Antonio News, Accessed 2020, <https://www.mysanantonio.com/news/local/article/Here-s-where-people-most-often-crash-into-animals-13266975.php#item-85307-tbla-1>
- Santhi, V. 2017. *Recent advances in applied thermal imaging for industrial applications*. 1st. Hershey:IGI Global. (Chapter 1 and 4) doi:[10.4018/978-1-5225-2423-6](https://doi.org/10.4018/978-1-5225-2423-6).
- Sawyer, H., P. A. Rodgers, and T. Hart. 2016. Pronghorn and mule deer use of underpasses and overpasses along U.S. Highway 191. *Wildlife Society Bulletin* 40 (2):211–16. doi:[10.1002/wsb.650](https://doi.org/10.1002/wsb.650).
- Sibanda, V., K. Mpofu, J. Trimble, and N. Zengeni. 2019. Design of an animal detection system for motor vehicle drivers. *Procedia CIRP* 84:755–60. doi:[10.1016/j.procir.2019.04.175](https://doi.org/10.1016/j.procir.2019.04.175).
- Su, X., W. Lin, X. Zheng, X. Han, H. Chu, and X. Zhang (2013). A new Local-Main-Gradient-Orientation HOG and contour differences based algorithm for object classification. Proceedings - IEEE International Symposium on Circuits and Systems, Beijing, 2892–95. DOI:[10.1109/ISCAS.2013.6572483](https://doi.org/10.1109/ISCAS.2013.6572483)
- U.S. Department of Transportation, Federal Highway Administration, Accessed 2020, <https://www.fhwa.dot.gov/publications/research/safety/08034/exec.cfm>
- Wang, J., L. P. Tchapmi, A. P. Ravikumar, M. McGuire, C. S. Bell, D. Zimmerle, S. Savarese, and A. R. Brandt. 2020. Machine vision for natural gas methane emissions detection using an infrared camera. *Applied Energy*. 257:113998. 0306-2619. doi:[10.1016/j.apenergy.2019.113998](https://doi.org/10.1016/j.apenergy.2019.113998).
- Wang, Y., D. Zhang, Y. Liu, B. Dai, and L. H. Lee. 2019 February 1. Enhancing transportation systems via deep learning: A survey. *Transportation Research Part C: Emerging Technologies*. 99:144–63. doi: [10.1016/j.trc.2018.12.004](https://doi.org/10.1016/j.trc.2018.12.004).
- Wei, X., Z. Yang, Y. Liu, D. Wei, L. Jia, and Y. Li. 2019. Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. *Engineering Applications of Artificial Intelligence* 80:66–81. doi:[10.1016/j.engappai.2019.01.008](https://doi.org/10.1016/j.engappai.2019.01.008).
- Wilkins, D. C., K. M. Kockelman, and N. Jiang. 2019. Animal-vehicle collisions in Texas: How to protect travelers and animals on roadways. *Accident Analysis and Prevention* 131:157–70. doi:[10.1016/j.aap.2019.05.030](https://doi.org/10.1016/j.aap.2019.05.030).
- Wu, X., D. Sahoo, and S. C. H. Hoi. 2020. Recent advances in deep learning for object detection. *Neurocomputing* 396 (d):39–64. doi:[10.1016/j.neucom.2020.01.085](https://doi.org/10.1016/j.neucom.2020.01.085).

- Yang, K., X. Wang, and R. Yu. 2018. A Bayesian dynamic updating approach for urban expressway real-time crash risk evaluation. *Transportation Research Part C: Emerging Technologies* 96:192–207. doi:10.1016/j.trc.2018.09.020.
- Zhao, Y., K. Hao, H. He, X. Tang, and B. Wei. 2020. A visual long-short-term memory based integrated CNN model for fabric defect image classification. *Neurocomputing* 380:259–70. doi:10.1016/j.neucom.2019.10.067.
- Zhou, D., J. Wang, and S. Wang (2012). Countour Based HOG Deer Detection in Thermal Images for Traffic Safety. Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2012 (vol.2, pp.969–74), Las Vegas, NV, United States.
- Zhu, Q., S. Avidan, M.-C. Yeh, and K.-T. Cheng (2006). Fast human detection using a cascade of histograms of oriented gradients. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on (CVPR'06), vol.2, no., pp. 1491–98, New York, NY, USA, DOI:10.1109/CVPR.2006.119
- Zilkha, M., and A. B. Spanier (2019). Real-time CNN-based object detection and classification for outdoor surveillance images: Daytime and thermal. 1116902(September 2019), 1. Proceedings volume 11169, Artificial intelligence and machine learning in defense applications, Strasbourg, France. DOI:10.1117/12.2532908