

Object Detection and Classification from Thermal Images Using Region based Convolutional Neural Network

Usha Mittal, Sonal Srivastava and Dr. Priyanka Chawla

Department of Computer Science and Engineering, Lovely Professional University, Punjab, India

Article history

Received: 04-04-2019

Revised: 19-06-2019

Accepted: 16-07-2019

Corresponding Author:

Dr. Priyanka Chawla

Department of Computer

Science and Engineering,

Lovely Professional University,

Punjab, India

Email: priyanka.22046@lpu.co.in

Abstract: In recent years, object detection and classification has gained so much popularity in different application areas like face detection, self-driving cars, pedestrian detection, security surveillance systems etc. The traditional detection methods like background subtraction, Gaussian Mixture Model (GMM), Support Vector Machine (SVM) have certain drawbacks like overlapping of objects, distortion due to smoke, fog, lightening conditions etc. In this paper, thermal images are used as thermal cameras capture the image by using the heat generated by the objects. Thermal camera images are not influenced by smoke and bad weather conditions which makes them a built-up apparatus in inquiry and safeguards or fire-fighting applications. These days, deep learning techniques are extensively used for detection and classification. In this paper, a comparative analysis has been done by applying Faster region based convolutional neural network on thermal images and visual spectrum images. The experimental results show that thermal camera images are better as compared to visible spectrum images.

Keywords: Object Detection, Classification, Faster R-CNN, Thermal Images, Visible Spectrum Images

Introduction

In computer vision, the process of scanning and searching for an object in an image or a video is known as detection of objects. People can easily recognize and distinguish objects present in a picture. The human visual framework is quick and exact and can perform complex undertakings like distinguishing different objects and identify obstructions with minimal aware ideas (Kaur and Talwar, 2016). With the accessibility of a lot of information, faster GPUs and better calculations, we can now effortlessly prepare systems to identify and classify various objects inside an image with high precision. Images taken with cell phones are normally complicated and contain various objects. Thus, assigning labels with image classification models can end up being complicated and questionable. Hence, in an individual picture numerous significant objects can be recognized by utilizing various models of object detection. Another significance of object detection is that the localization of the objects is given as compared to image classification. Nonetheless, because of huge varieties of perspectives, positions, obstacles and lighting conditions, it's hard to splendidly achieve object

detection with an extra object localization work. The main objective of object detection is to decide where objects are situated in a given picture (object localization) (Javier, 2017) and then classifying the categories for each detected objects. So the task of object detection models can be categorized into three phases.

Selection of Region

As various objects may show up in many places of the picture and had different resolutions or sizes, it is an individual decision to filter the entire picture with a multi-scale sliding window (Harsha and Anne, 2016). Because of countless windows, it is computationally costly and creates an excessive number of repetitive windows. But if just a limited number of sliding window formats is used, inadmissible locales might be created.

Extraction of Features

It is a process to extract visual features to identify different objects by providing correct and powerful descriptions about these detected objects. There are various feature extraction technique like HOG, Haar-like features and SIFT (Harsha and Anne, 2016).

Due to various issues like a variety of appearance, background and lighting conditions, it is difficult for humans to manually predict the features for all the objects that are detected in an image.

Classification of Objects

In addition, classification (Sokolova and Lapalme, 2009) is required to classify the target object into different classes and to make the visual identification of an object more hierarchical, correct and knowledgeable.

Basically, object detection is based on two main approaches: Either deep learning or machine learning approaches. In machine learning-based approaches, it detects the features using Haar, SIFT and HOG then for classification it can use Support Vector Machine (SVM) (Sun *et al.*, 2006) whereas in deep learning approaches it generally uses the Convolutional neural network for the object detection without requiring any knowledge about the features. The various deep learning approaches for object detection are: Region proposal Convolutional neural network (Girshick *et al.*, 2015), fast Region

proposal Convolutional neural network (Girshick, 2015), Faster Region proposal Convolutional neural network (Ren *et al.*, 2015) and You Only Look Once (YOLO) (Redmon *et al.*, 2016).

Estimation models supported by deep learning are made out of various hidden layers which help in learning data representation along with consideration at each level. Deep learning basically deals with the deep neural network algorithms where deep (Zhang *et al.*, 2017) refers to the number of hidden layers and its main objective is to resolve the learning problems by copy the functioning of the human brain. The first ever idea of deep learning based neural network model was implemented in 1940's (Zhang *et al.*, 2017) but due to various reasons like the absence of large training datasets, over-fitting problem (Zhou *et al.*, 2016) related with training set, poor performance and restricted calculation power is contrasted against other machine learning models thus making these models exceptionally invaluable.



Fig. 1: (a) Visible spectrum image and (b) thermal camera image during the day time (c) visible spectrum image and (d) thermal camera image at night time

There are various factors which helped in the improvement of deep learning like: the development of large training dataset like ImageNet (Deng *et al.*, 2009) which is able to completely express its large learning capability, the development of parallel computing frameworks to boost the performance like GPU's (Kwaśniewska *et al.*, 2017) and the important approaches in the design of models architecture and training methodologies. There are various techniques developed to overcome the problem of over-fitting like data augmentation and regularization and with Batch Normalization (BN) technique. Hence, neural network models training process becomes easier and proficient. There are some more network structures which are used to enhance the performance like AlexNet (Krizhevsky *et al.*, 2012), Resnet (He *et al.*, 2016), Overfeat (Sermanet *et al.*, 2014), Google Net (Szegedy *et al.*, 2015) and SqueezeNet (Hamida *et al.*, 2018).

In the past, gathered labeled or unlabeled training data (Pan and Yang, 2010) were used to infer the data by utilizing the conventional machine learning and data mining models. But this process is altered by transfer learning, as it makes use of information experienced by at least one source assignment and implementing it over the target assignment for improving experience (Torrey and Shavlik, 2009; Zhang *et al.*, 2018). The NIPS-95 (Pan and Yang, 2010) is the workshop on "Learning to learn" where the basic inspiration to implement transfer learning in the areas of machine learning was developed. The main concentration of this workshop is on the requirements of machine-learning methods that preserve and reiterate earlier experienced information.

Thermal Images vs. Visible Spectrum Images

Thermal imaging is used to extend the human vision to an Infrared region by utilizing the lights emitted by warm objects (Gade and Moeslund, 2014). The working of thermal imaging can be defined as it detects and after that, the relative contrast between the intensities of the infrared energies are being reflected or produced from an object of a series of other objects are displayed (Berg *et al.*, 2015). Thermal imaging isn't only an alternate sort of night vision – it sees the heat generated and do not focus on the light, so it does these 24 h every day. Since thermal imagers make pictures from contrasts in heat energy, anything that typically generates heat can be identified and imaged. For instance, creatures (humans and animals), plants, electro-mechanical systems and industrial processes all had individual heat signs that will be visible through thermal cameras. Figure 1 shows the different images taken by visible spectrum cameras and thermal cameras.

A thermal imaging camera can examine the whole region and segments at the same time, failing to miss any overheating perils, regardless of how little. In some of the cameras the detector would begin recognizing its very own radiation (Berg, 2016) which is surely exceptionally unfortunate. In this manner, such detectors should be cooled to worthy levels.

In this paper, for the evaluation of results, confusion matrix is used. From the confusion-matrix, two evaluation metrics was computed i.e. recall and precision. Precision and recall (Davis and Goadrich, 2006) can be evaluated using equation 1 and 2 respectively:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \dots\dots (1)$$

$$Recall = \frac{Precision \times Recall}{True\ Positive + False\ Positive} \dots (2)$$

To compute the accuracy of the model, F1 score is calculated. It is calculated by:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots\dots (3)$$

Related Work

Zhang *et al.* (2017) had worked on the detection and classification of the vehicles by implementing deep neural network. The aim of the research was to extract high-level features from lower level features. The experimental result shows that the deep neural network is better for vehicle classification with 3.34% error rate as compared to a traditional neural network for which the error rate is 6.67%. Author (Harsha and Anne, 2016) had proposed enhanced Gaussian mixture model and background subtraction for vehicle detection. Then applied the AlexNet and SIFT (scale invariant feature transform) for feature extraction further on which PCA and LDA, was applied for dimensionality reduction and lastly with the help of SVM (support vector machine) classification was done. The experimental results depicted that enhanced GMM with AlexNet DNN at FC6 and FC7 was more accurate for detecting and classifying vehicles. Zhou *et al.* (2016) had worked on the deep neural network in which for detection YOLO method and for classification AlexNet DNN was used. To extend the capabilities of DNN, vehicle classification on dark images by applying the scene transformation, late fusion technique and color transformation method was used.

Gao and Lee (2015) had proposed a framework for detecting a moving car with the help of frame difference. Then symmetrical filter on the front view of the car was used to get the binary frontal view, on these 3 layers restricted Boltzmann machine with deep learning was used to detect the model of the car.

Chan *et al.* (2012) had worked on a vision-based system for detecting preceding vehicles considering various scenarios like poor lighting and weather conditions on a highway. To detect the various properties of vehicle appearance, author had fused four things like underneath shadow of vehicle, vertical edge, symmetry and taillight of the vehicle. Various vehicles had been detected using clustering along with the single particle filter and data-driven initial sampling technique was used in the identification of objects and it prevented from collapsing the multi-modal distribution to the local maxima. Berg *et al.* (2015) the author had proposed a method for the tracking of short-term single objects by utilizing thermal images. They had compared various tracking methods by utilizing visible spectrum images and thermal images and concluded that the best tracking methods for normal images and thermal images are (ASLA, SCM and DSST) on the basis of its dimensional structure and/ or scattered representation and (EDFT) on the basis of pixel value distribution respectively.

Rodin *et al.* (2018) had worked on thermal images captured by Unmanned Ariel Systems (UAS) for the sea surface detection and classification of objects. It was beneficial for searching and finding the marine objects. Experimental results obtained shows 92.5% accuracy over a testing dataset. Nam and Nam (2018) had proposed a surveillance system for detecting and classifying of vehicles during the day and at night time. The various parameters used for the feature extraction was textures, entropy, homogeneity, energy and contrast.

Moranduzzo and Melgani (2014) had worked on UAV (unmanned aerial vehicle) images for detecting cars using catalog based approach. Author had discussed the existing system that made use of screening operations in which asphalted-areas were recognised for making the car detection speed up and robust. Then HOG features were extracted by filtering operations for finding the exact points of the car, finally obtained the 36-directions for higher value of similarity. Since, UAV image had a high resolution because of which single car can be detected more than once. So, to avoid redundancy, points were merged which belongs to the same car. Finally, SVM was used for classification. Similarly, for vehicle detection, localization and tracking of the vehicle the author (Sivaraman and Trivedi, 2013) had worked. The research solved the problems of movement of vehicles in lane for driver assistance. There was an issue that it feasible only on the 11 frame per second videos. Hence, cars with higher speed could be easily skipped. Author (Chen *et al.*, 2014) had worked on developing an intelligent urban video surveillance system for automated vehicle detection and tracking in clouds and the installation of the digital surveillance system in video surveillance system camera was done to obtain the image or vehicle data containing the vehicles.

Tuermer *et al.* (2013) had worked on the HOG based Vehicle Detection technique in which dense traffic areas was considered. Similar areas were excluded by region growing algorithm and then the classification of remaining parts of input data was done based on HOG features. Author (Prabha and Shah, 2016) had worked on the detection and classification using the hybrid deep neural network. In this, the Non-Negative Matrix Factorization (NMF) was used for extracting features and compression. This research helped the administration to build the adaptable traffic shaping strategies for the crowded highways in the urban areas.

Chen *et al.* (2012) had used background Gaussian Mixture Model and shadow removal method for sudden illumination changes and vibration from the camera for detecting, tracking and classifying where vehicles were classified into four categories i.e. car, van, bus and motorcycle. For tracking, Kalman filter was used. Three modules namely background-subtraction, foreground-extraction and vehicle detection was used. He *et al.* (2015) had worked on a vehicle detection and classification using deep Convolutional activation feature (DeCAF). Visual features were extracted and compared the accuracy of various techniques like large-scale sparse learning and deep convolutional neural network.

In a paper, the author (Chen *et al.*, 2011) proposed a method to classify road vehicles by utilizing the structure i.e. size and shape derived from view dependent binary figures with the help of CCTV camera images, in which manual segmentation was performed to get the boundaries of the vehicles in the images and then features are extracted from each binary figure. Gupte *et al.* (2002) had proposed an algorithm for vision based detection and classification of vehicles by utilizing a monocular images captured by a CCTV camera. Images were processed into three categories like vehicle level images, locale level images and untrained images and to support the research, the experimental result from highway scenes was used. They had developed a camera calibration tool that recovers the camera parameter by utilizing features selected by the user in the image. By utilizing minimal scene specific knowledge, their system was able to detect, track and classify vehicles. For each visible vehicle the system provides location and velocity information. For classification of the larger number of categories of vehicles, they had utilized non-rigid model-based technique. The parameterized 3-D models will be used for each category of vehicles.

Research Methodology

In this research, a comparative analysis has been done using thermal images and visual spectrum images. The motivation behind the experiment is to evaluate the results and performance of deep learning algorithms on thermal images compared to visual spectrum images. As

it is already discussed that thermal images are providing good quality images even where lighting conditions are inappropriate. While visual spectrum cameras may fail to give efficient results at night time or in bad weather conditions like fog or rain. Detecting objects and recognizing them is the popular research area now days. In this paper, deep learning based faster R-CNN algorithm is used on the thermal images and visible spectrum images. Convolutional Neural Networks expect and preserve the spatial relationship between pixels by learning internal feature representations using small squares of input data. Feature are learned and used across the whole image, allowing for the objects in the images to be shifted or translated in the scene and still detectable by the network. CNN provides following major advantages:

- They use fewer parameters (weights) to learn than a fully connected network
- They are designed to be invariant to object position and distortion in the scene
- They automatically learn and generalize features from the input domain

To perform experiment, dataset is collected from the open source library i.e. FLIR (released in July 2018). The dataset contains the thermal images as well as visible spectrum images from which 2000 images (1000 visible spectrum image and 1000 thermal images) were taken. Annotation of images was done using labeling software and four classes of objects are considered for detection, i.e. human, 2-Wheeler, 4-Wheeler and Traffic-Light. Figure 2 shows the proposed methodology.

Ren *et al.* (2015) had developed a new method for the detection of objects, prediction of bounding boxes and region proposal generation which is known as region proposal network (RPN) to overcome the cost issue in the traditional method which utilizes the selective search methods for the generation of region proposals. Hence, the combination of Region Proposal Network and fast region proposal Convolutional neural network models

computes the faster region proposal Convolutional neural network.

Algorithm of Faster R-CNN

- a. The entire input image is passed through convolutional layer and feature maps are extracted
- b. A sliding window is used in region proposal network for each location over the feature map
- c. For each location, k ($k=9$) anchor boxes are used (3 scales of 128, 256 and 512 and 3 aspect ratios of 1:1, 1:2, 2:1) for generating region proposals
- d. The *classifier* layer outputs $2k$ scores whether there is object or not for k boxes
- e. The *regressor* layer outputs $4k$ for the coordinates (box center coordinates, width and height) of k boxes
- f. With a size of $W \times H$ feature map, there are $W * H * k$ anchors in total
- g. Non-maximum suppression is used to reduce the number of proposals

For increasing training and testing speed and enhancing the performance, RPN is used by Faster R-CNN instead of using the method of selective search. For classification, RPN is applied over the ImageNet dataset which was pre-trained and on the PASCAL VOC dataset it was fine-tuned. Finally, to train fast R-CNN, region proposal are generated along with anchor boxes. Hence, it is a repetitive process.

Results and Discussion

In this experiment, the dataset contains the 1000 visible spectrum images and 1000 thermal images which is divided into two parts i.e. for training 800 images and for testing 200 images respectively. The various categories of the objects that are detected are 4-wheeler, 2-wheeler, traffic light and human. Implementation was done in python using Tensorflow API and code was executed on NVIDIA GPU with 4 GB capacity.

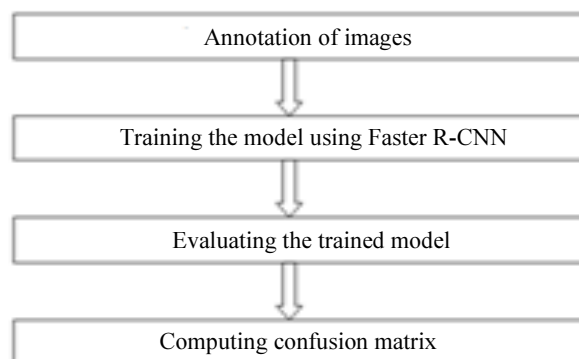


Fig. 2: The proposed research methodology

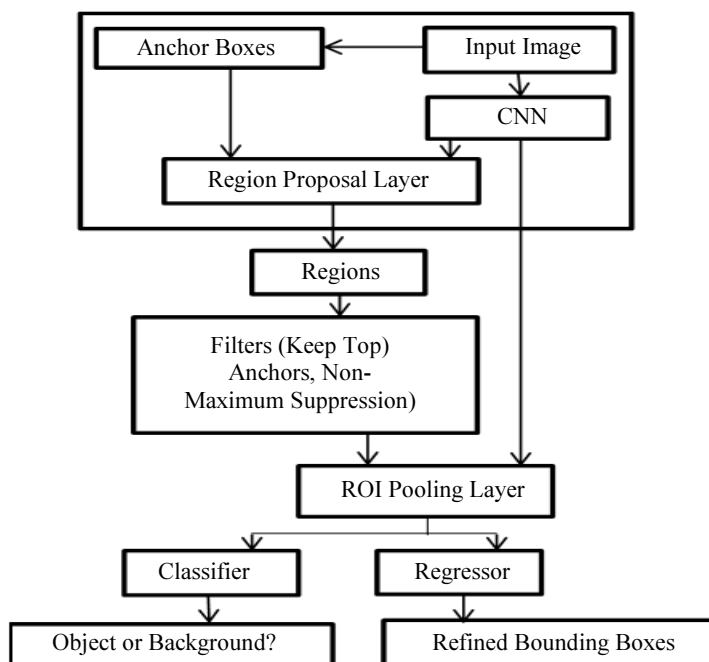


Fig. 3: The basic architecture of Faster Region-based Convolutional neural network



Fig. 4: Visible spectrum image (original) and Visible spectrum image (objects detected)

Figure 4-7 shows the experimental results performed on both types of images. Figure 4 shows that the accuracy of visible spectrum image during the night time is poor as compared to the Fig. 5 which contains the thermal image of the same scene at the same time. While Fig. 6 and 7, contains the day time images and the quality of visible spectrum images is somewhat better than thermal images. The numbers of objects detected during night time from thermal images are more accurate as compared to visible spectrum images while accuracy is almost same in day time scene in both the images.

The confusion-matrix for visible spectrum and thermal camera images is given in Table 2 and 3 respectively.

Experimental results are compared on the basis of precision, recall and accuracy and given in Table 4.

The higher value of recall means the class is correctly classified and the higher value of precision means positively labeled classes are certainly positive. Hence, two cases can be concluded from Table 5.

Case 1: High recall, low precision- means there is a huge number of false positive even though most of the positive examples are correctly recognized (low FN).

Case 2: Low recall, high precision- means even though examples that are predicted positive are certainly positive (low FP) but we missed a huge number of positive examples as the case with visible spectrum images in this research.



Fig. 5: Thermal image (original) and thermal image (objects detected)



Fig. 6: Visible spectrum image (original) and Visible spectrum image (objects detected)



Fig. 7: Thermal image (original) and thermal image (objects detected)

Table 1: Literature review summary

Year	Authors	Image Type	Approach	Results and Discussion
2002	Gupte <i>et al.</i>	Monocular images.	Developed camera calibrations tool.	The tool requires minimum scene related information for detecting tracking and classifying vehicles.
2011	Chen <i>et al.</i>	Images using CCTV cameras.	View-dependent binary figures were used and compared the two classifiers: SVM and random forest.	It achieved the classification accuracy of 96.26% for SVM which was much better as compared to the random forest.
2011	Chan <i>et al.</i>	CCD camera images.	Clustering along with the single particle filter and data-driven initial sampling technique was used. Provided the contrast between AdaBoost classifier and their proposed system.	Obtained the accuracy of 92.84% over seven videos for detection under poor illumination conditions.
2012	Chen <i>et al.</i>	CCTV camera images	Background Gaussian Mixture Model (GMM) and shadow removal method was used for detection. Kalman filter and SVM was used for classification.	It conclude that training the SVM using MBF+IPHOG features extracted by background subtraction gives better results, classifying foreground blobs using majority voting.
2013	Sivaraman and Trivedi	RGB images.	Vehicle tracking for driver assistance.	Solved the problems of lane movement
2013	Chen <i>et al.</i>	Surveillance camera	Developed urban video surveillance system for detecting vehicles.	Higher detection accuracy as well as higher positioning accuracy.
2013	Tuermer <i>et al.</i>	Disparity images.	HOG based Vehicle Detection technique was used.	Faster and accurate detection of vehicles.
2014	Moranduzzo and Melgani	UAV images.	Catalog-based approach and SVM was used.	Higher accuracy for a higher number of possible directions of movements.
2015	Gao and Lee	Binary images.	Frame difference and symmetrical filters was used.	Achieves 100% accuracy on their own dataset.
2015	He <i>et al.</i>	RGB images.	Deep Convolutional activation feature (DeCAF) and SVM was used.	The accuracy of DeCAF is highest as compared to other traditional methods.
2015	Berg <i>et al.</i>	Thermal images.	For the tracking of short-term single-objects by utilizing thermal images.	It concluded that the best tracking methods for normal images and thermal images are (ASLA, SCM and DSST) on the basis of its dimensional structure and/or scattered representation and (EDFT) on the basis of pixel value distribution respectively.
2016	Zhang <i>et al.</i>	RGB images	Deep neural network was used.	Deep neural network gives 3.34% error rate as compared to a give traditional neural network which 6.67%.
2016	Harsha and Anne	RGB images.	AlexNet, SIFT, Enhanced GMM and background subtraction used for vehicle detection. SVM for classification.	Combination of enhanced GMM with AlexNet DNN at layers FC6 and FC7 is more accurate for detection and classification.
2016	Zhou <i>et al.</i>	RGB images.	DNN trained for a smaller dataset. It uses YOLO model.	Extended the capability for classifying the vehicles using dark images.
2016	Kaur and Talwar	RGB images.	Probabilistic neural network was used.	The overall improvement in the existing system.
2016	Prabha and Shah	RGB images.	Hybrid deep neural network and Non-negative matrix factorization (NMF) was used.	The administration was able to build the adaptable traffic shaping strategies for the crowded highways in the urban areas.
2017	Bhartee <i>et al.</i>	Thermal images.	Utilizes heat signatures generated by the objects.	Beneficial for the security purpose and for finding and locating objects during the time of natural calamities. It is useful in both day and night vision.
2018	Rodin <i>et al.</i>	Thermal images.	GMM and CNN for sea surface detection and classification of objects was used.	The accuracy of 92.5% over a testing dataset.
2018	Nam and Nam	Both visible light and thermal images are used.	For detection and classification GMM was used.	The accuracy for visible spectrum images and thermal images were 92.7% and 65.8% respectively.

Table 2: Confusion-matrix for visible spectrum image

Categories	Actual 4wheeler	Actual 2wheeler	Actual Traffic-light	Actual Human	Others
Predicted 4wheeler	103	0	3	0	620
Predicted 2wheeler	2	17	2	7	879
Predicted Traffic-light	1	0	7	3	534
Predicted Human	2	0	13	85	1052
Others	3	3	1	4	0

Table 3: Confusion-matrix for thermal camera images

Categories	Actual 4wheeler	Actual 2wheeler	Actual Traffic-light	Actual Human	Others
Predicted 4wheeler	538	4	2	17	165
Predicted 2wheeler	82	438	2	232	153
Predicted Traffic-light	1	1	289	5	249
Predicted Human	29	16	2	951	154
Others	39	128	95	123	0

Table 4: Comparative results of precision and recall for visible spectrum and thermal camera images

	Precision for visible-spectrum images	Precision for thermal images	Recall for visible-spectrum images	Recall for thermal images	F1 score of visible-spectrum images	F1 Score of thermal images
4wheeler	0.9300	0.7800	0.1400	0.7400	0.2430	0.7590
2wheeler	0.8500	0.7500	0.0200	0.4800	0.0390	0.5850
Traffic-light	0.2700	0.7400	0.0100	0.5300	0.0190	0.6170
Human	0.8600	0.7200	0.0700	0.8300	0.1290	0.7710

It can be clearly seen from Table 4 that the value of recall is higher in case of thermal images, as well as accuracy for thermal images in case of 4-Wheeler is 75.9% whereas in visible spectrum images it is 24.3% only. Similarly, in other cases like 2-wheeler, traffic-light and human the value of accuracy for thermal images are 58.5%, 61.7% and 77.1% respectively whereas, in case of visible spectrum images the values of accuracy are 3.9%, 1.9% and 12.9% respectively.

Conclusion and Future Work

Traffic monitoring is a very important application of object detection and classification for managing and controlling the traffic on roads and at intersections. At intersections, traffic light controllers can be optimized by estimating the vehicle density present at a particular lane. To compute the vehicle density, images from cameras are taken and processed. So, at night time and in odd weather conditions, due to lack of lighting, cameras are not able to provide the quality images which leads to inefficient results. In such applications, thermal cameras can provide efficient results. The experimental result shows that the accuracy of thermal images is better than visible spectrum images during the night time whereas the accuracy of visible spectrum images is almost same as thermal camera images during the day time but overall thermal images give better results. Although at day time, thermal cameras may not provide the same image quality because of excessive heat but overall efficiency of the system will be improved if different time, seasons and weather conditions are considered as whole. In future, other deep learning techniques such as YOLO, Mask R-CNN, SSD, etc. would be implemented over thermal images and normal images.

Author's Contributions

Ms. Usha Mittal: Participated in data collection from online repository, data preprocessing, literature survey, design the methodology, result analysis and writing the manuscript.

Ms. Sonal Srivastava: participated in data preprocessing, data annotations, literature survey, design the methodology, result analysis and writing the manuscript.

Dr. Priyanka Chawla: Coordinated the experiments, result analysis and writing the manuscript.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Berg, A., 2016. Detection and tracking in thermal infrared imagery. Sweden: Linkoping: Linkoping University Electronic Press.
 DOI: 10.3384/lic.diva-126955
- Berg, A., J. Ahlberg and M. Felsberg, 2015. A thermal object tracking benchmark. Proceedings of the 12th IEEE International Conference on Advanced Video and Signal Based Surveillance, Aug. 25-28, IEEE Xplore Press: Karlsruhe, Germany, pp: 1-7.
 DOI: 10.1109/AVSS.2015.7301772
- Bhartee, A.K., K.M. Srivastava and T. Sharma, 2017. Object Identification using thermal image processing. Int. J. Eng. Sci. Computing,
- Chan, Y.M., S.S. Huang, L.C. Fu, P.Y. Hsiao and M.F. Lo, 2012. Vehicle detection and tracking under various lighting. IET Intelligent Transport Systems, 6: 1-8. DOI: 10.1049/iet-its.2011.0019
- Chen, Y.L., T.S. Chen, T.W. Huang, L.C. Yin and S.Y. Wang, 2013. Intelligent urban video surveillance system for automatic vehicle detection and tracking in clouds. Proceedings of the IEEE 27th International Conference on Advanced Information Networking and Applications, March 25-28, IEEE Xplore Press: Barcelona, Spain, pp: 814-821.
 DOI: 10.1109/AINA.2013.23

- Chen, Z., T. Ellis and S.A. Velastin, 2011. Vehicle Type Categorization: A comparison of classification schemes. Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems Washington, DC, USA, pp: 74-79.
DOI: 10.1109/ITSC.2011.6083075
- Chen, Z., T. Ellis and S.A. Velastin, 2012. Vehicle detection, tracking and classification in urban traffic. Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems, Sep. 16-19, IEEE Xplore Press: Anchorage, Alaska, USA, pp: 951-956.
DOI: 10.1109/ITSC.2012.6338852
- Davis, J. and M. Goadrich, 2006. The Relationship Between Precision-Recall and ROC Curves. Int. Conf. Mach. Learning, pp: 1-8.
DOI: 10.1145/1143844.1143874
- Deng, J., W. Dong, R. Socher, L.J. Li and K. Li, 2009. ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, IEEE Xplore Press: Miami, FL, USA, pp: 1-8.
DOI: 10.1109/CVPR.2009.5206848
- Gade, R. and T.B. Moeslund, 2014. Thermal cameras and applications: A survey. Machine Vision Applications, 25: 245-262.
DOI: 10.1007/s00138-013-0570-5
- Gao, Y. and H.J. Lee, 2015. Moving Car Detection and Model Recognition based on Deep Learning. Advanc. Sci. Technol. Lett.
DOI: 10.14257/astl.2015.90.13
- Girshick, R., 2015. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV) Santiago, Chile, pp: 1-9.
- Girshick, R., J. Donahue, T. Darrell and J. Malik, 2015. Region-based convolutional networks for accurate object detection and segmentation. Proce. IEEE Trans. Pattern Analysis Machine Intelligence, 38: 142-158. DOI: 10.1109/TPAMI.2015.2437384
- Gupte, S., O. Masoud, R.F. Martin and N.P. Papanikolopoulos, 2002. Detection and classification of vehicles. Proce. IEEE Trans. Intelligent Trans. Syst., 3: 37-47.
DOI: 10.1109/6979.994794
- Hamida, A.B., A. Benoit, P. Lambert and C.B. Amar, 2018. 3-D deep learning approach for remote sensing image classification. Proceedings IEEE Trans. Geosciences Remote Sensing, 56: 4420-4434.
DOI: 10.1109/TGRS.2018.2818945
- Harsha, S.S. and K.R. Anne, 2016. Gaussian mixture model and deep neural network based vehicle detection and classification. Int. J. Advanc. Computer Sci. Applica., 7: 17-25.
DOI: 10.14569/IJACSA.2016.070903
- He, D., C. Lang, S. Feng, X. Du and C. Zhang, 2015. Vehicle detection and classification based on convolutional neural network. Proceedings of the 7th International Conference on Internet Multimedia Computing and Service, Zhangjiajie, Hunan, China, pp: 1-5. DOI: 10.1145/2808492.2808495
- He, K., X. Zhang, S. Ren and J. Sun, 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, IEEE Xplore Press: USA, pp: 1-12. DOI: 10.1109/CVPR.2016.90
- Javier, 2017. Tryolabs. Retrieved march 05, 2019. <https://tryolabs.com/blog/2017/08/30/object-detection-an-overview-in-the-age-of-deep-learning/>
- Kaur, R. and M. Talwar, 2016. Automated vehicle detection and classification with probabilistic neural network. IJARIT.
- Krizhevsky, A., G.E. Hinton and I. Sutskever, 2012. ImageNet classification with deep convolutional neural networks. NIPS, 25: 1-9.
DOI: 10.1145/3065386
- Kwaśniewska, A., J. Rumiński and P. Rad, 2017. Deep features class activation map for thermal face detection and tracking. Proceedings of the IEEE 10th International Conference on Human System Interactions, July 17-19, IEEE Xplore Press: Ulsan, South Korea, pp: 41-47.
DOI: 10.1109/HSI.2017.8004993
- Moranduzzo, T. and F. Melgani, 2014. Detecting cars in UAV Images with a catalog-based approach. Proce. IEEE Trans. Geoscience Remote Sensing, 52: 6356-6367.
DOI: 10.1109/TGRS.2013.2296351
- Nam, Y. and Y.C. Nam, 2018. Vehicle classification based on images from visible light and thermal cameras. Eurasip J. Image Video Proc.
DOI: 10.1186/s13640-018-0245-2
- Pan, S.J. and Q. Yang, 2010. A Survey on Transfer Learning. Proce. IEEE Trans. Knowledge Data Eng., 22: 1345-1359.
DOI: 10.1109/TKDE.2009.191
- Prabha, C. and I. Shah, 2016. Study of vehicular traffic using hybrid deep neural network. Int. J. Innovative Res. Computer Communica. Eng., 4334-4338.
- Redmon, J., S. Divvala, R. Girshick and A. Farhadi, 2016. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp: 1-10.
- Ren, S., K. He, R. Girshick and J. Sun, 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, Neural Information Processing Systems Foundation, Inc. pp: 1-14. DOI: 10.1109/TPAMI.2016.2577031

- Rodin, C.D., L.N. Lima, F.A. Andrade, D.B. Haddad and T.A. Johansen, 2018. Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems. International Joint Conference, Neural Networks, pp: 1-8.
DOI: 10.1109/IJCNN.2018.8489465
- Sermanet, P., D. Eigen, X. Zhang, M. Mathieu and R. Fergus, 2014. OverFeat: Integrated recognition, localization and detection using convolutional Networks. ICLR, pp: 1-16.
- Sivaraman, S. and M.M. Trivedi, 2013. Integrated lane and vehicle detection, localization and tracking: A Synergistic approach. IEEE Trans. Intelligent Transportation Syst., 14: 906-917.
DOI: 10.1109/TITS.2013.2246835
- Sokolova, M. and G. Lapalme, 2009. A systematic analysis of performance measures for classification tasks. Inform. Proc. Manage., 45: 427-437.
DOI: 10.1016/j.ipm.2009.03.002
- Sun, Z., G. Bebis and R. Miller, 2006. Monocular precrash vehicle detection: Features and classifiers. Proce. IEEE Trans. Image Processing, 15: 2019-2034.
DOI: 10.1109/TIP.2006.877062
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet and S. Reed *et al.*, 2015. Going deeper with convolutions. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, IEEE Xplore Press. Boston, MA, USA, pp: 1-12.
DOI: 10.1109/CVPR.2015.7298594
- Torrey, L. and J. Shavlik, 2009. Transfer Learning. In: Handbook of Research on Machine Learning Applications, Olivas, E.S., J.D. Guerrero, M.M. Sober, J.R. Benedito and A.J. Lopez, IGI Global. pp: 1-22.
- Tuermer, S., F. Kurz, P. Reinartz and U. Stilla, 2013. Airborne vehicle detection in dense urban areas using hog features and disparity maps. Proce. IEEE J. Selected Topics Applied Earth Observations Remote Sensing, 6: 2327-2337.
DOI: 10.1109/JSTARS.2013.2242846
- Zhang, J., W. Xing, M. Xing and G. Sun, 2018. Terahertz image detection with the improved fasterregion-based convolutional neural network. Sensors, 18: 1-19. DOI: 10.3390/s18072327
- Zhang, Z., C. Xu and W. Feng, 2017. Road vehicle detection and classification based on Deep Neural Network. Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science, Aug. 26-28, IEEE Xplore Press: Beijing, China, pp: 675-678.
DOI: 10.1109/ICSESS.2016.7883158
- Zhou, Y., H. Nejati, T.T. Do, N.M. Cheung and L. Cheah, 2016. Image-based vehicle analysis using deep neural network: A systematic study. Proceedings of the IEEE International Conference on Digital Signal Processing, Oct. 16-18, IEEE Xplore Press: Beijing, China. pp: 276-280.
DOI: 10.1109/ICDSP.2016.7868561