I.J. Image, Graphics and Signal Processing, **2021, 1, 40-49**
Published Online February 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2021.01.04

Pedestrian Detection in Thermal Images Using Deep Saliency Map and Instance Segmentation

A. K. M. Fahim Rahman, **Mostofa Rakib Raihan**, **S.M. Mohidul Islam**
Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh
Email: {fahim1615, rakib1620}@cseku.ac.bd, mohid@cse.ku.ac.bd

Received: 19 February 2020; Revised: 23 April 2020; Accepted: 02 July 2020; Published: 08 February 2021

Abstract: Pedestrian detection is an established instance of computer vision task. Pedestrian detection from the color images has achieved robust performance but in the night time or in bad light conditions it has low detection accuracy. Thermal images are used for detecting people at night time, foggy weather or in bad lighting situations when color images have a lower vision. But in the daytime where the surroundings are warm or warmer than pedestrians then the thermal image has lower accuracy. Hence thermal and color image pair can be a solution but it is expensive to capture color-thermal pair and misaligned imagery can cause low detection accuracy. We proposed a network that achieved better accuracy by extending the prior works which introduced the use of the saliency map in pedestrian detection tasks from the thermal images into instance-level segmentation. We worked on a subdivision of KAIST Multispectral Pedestrian Detection Dataset [8] which has pixel-level annotations. We have trained Mask-RCNN for pedestrian detection task and report the added effect of saliency maps generated using PiCA-Net. We have achieved an accuracy of 88.14% over day and 91.84% over night images. . So, our model has reduced the miss rate by 24.1% and 23% over the existing state-of-the-art method in day and night images.

Index Terms: Thermal image, saliency map, deep saliency network, instance segmentation, mask-RCNN

1. Introduction

A pedestrian is someone who travels on foot. Pedestrian detection is an instance of object detection which detects the pedestrian in image. It is a useful technique that provides safety for both drivers and pedestrians on the road. Pedestrian detection is also a crucial task for major applications like self-driving vehicles [1, 27], Security [16], monitoring traffic flow [26], on-site vehicle drivers assisting system [28], etc.

Pedestrian detection is a laborious task and time-consuming. For our daily essential need like safety, security and for other purposes, an accurate and fast pedestrian detection technique is very essential. Pedestrian detection task in both color image and videos [29-31] has achieved an outstanding accuracy in recent times. There are various machine learning methods like - Support Vector Machine [32], AdaBoost [33], Decision Tree [34], Neural Networks [46] are used for detecting the pedestrian in color images. But there are few challenges like occlusion, low resolution and bad light conditions etc. that lacks accuracy for detecting pedestrian in color image. In night time or low light condition, the use of thermal image is a very effective technique [8, 47] for detecting pedestrian. So using thermal images, eliminates the limitations of color image for situations like low light, bad weather conditions. Thermal camera detects the infrared radiation from surroundings. Then images are formed based on that information. Humans can easily be distinguished from the surroundings at night time because then the surroundings are less warm than the pedestrian. But, on sunny day humans are less distinguishable from the surrounding objects which are almost as warm as human. Therefore, thermal images are more effective in the night time, low or bad light condition than luminous day time images.

So, there comes the concept of color and thermal image pair. It resolves the drawbacks of both color and thermal image in terms of detecting pedestrian at both day and night time. In recent time, various detecting methods that works on the color and thermal image pair architectures [2-3] has achieved inspiring result. But collecting the color-thermal pair [35] are expensive because it needs two separate cameras to capture both thermal and color images. Moreover if the color thermal pair is somehow misaligned then it can reduce performance accuracy.

To overcome the challenges of detecting pedestrians in a thermal image during the daytime, Ghose et al. [8] introduced the impact of a saliency map [36] for pedestrian detection tasks in a thermal image. Saliency is defined at a given location as how much the object is differentiated from its surroundings. Moreover saliency is a visual attention technique that highlights the noticeable objects in image which is called as salient object. We have used deep saliency network (PiCA-Net [6]) to determine the salient part in thermal image.

The performance obtains by using only thermal image [8] may have difficulties to achieve a higher detecting accuracy. Thus a thermal image integrated with its saliency map can improve the detection performance in daytime.

This saliency map highlights the salient part of the image thus it becomes easy for the detector to look at the import part. where a pedestrian instance might present. They [8] proved the impact of the saliency in pedestrian detection task but their main limitation is Faster RCNN generates a significant amount of false positive and thus it reduces its accuracy. Arnab et al. [50] introduced the impact of the instance segmentation in object detection task.

In this paper, we have proposed a model, which compact both two method generally. Our proposed model will use - instance segmentation instead of semantic segmentation in pedestrian detection task. We hope for achieving more accuracy and lower miss rate than the current state-of-the-art methods by using instance segmentation. The aim of this research is finding out the impact of instance segmentation with saliency map in pedestrian detection and also achieved a higher accuracy and lower miss rate in pedestrian detection task. We have trained Mask RCNN, a pedestrian detector on the thermal image, fused with their saliency map using PiCA-Net. We worked on a subdivision of KAIST multispectral pedestrian dataset [8] for detecting pedestrian.

2. Related Works

Pedestrian detection is one of the most important task in computer vision. It is a useful technique that detects pedestrian which is useful for surveillance [48] and security [16] purpose. Detecting pedestrian from the thermal image has achieved increasing attention in recent times. Ghose at el. [8] first established the impact of saliency maps in thermal images and the pedestrian detection accuracy improved. They use both static and deep saliency network and fused the saliency maps with the thermal image. Then they used Faster R-CNN [11] to detect pedestrian in thermal image. They have used a subdivision of KAIST Multispectral Pedestrian Dataset. As they introduced the impact of saliency map in pedestrian detection, at first they created baseline using only thermal image. They achieved baseline of 55.8% in day time and 59.6% in night time using only thermal image. Then they used both static-saliency and deep saliency network for generating saliency map. Using static saliency in thermal images, accuracy of detecting pedestrian from day time image have been improved. Meanwhile at night time, accuracy does not changed significantly. Then they bring out the concept of using deep saliency networks. They achieved 67.8% for day time images and 78.3% for night images using PiCA-Net and achieved 69.6% for day images and 79% for night time images using R^3-Net. By using deep saliency networks, detection performance in both day and night time image have increased than using only thermal image. The drawback of their model is, Faster R-CNN generates a significant number of false positives. J. Kim [19] proposed a technique for detecting pedestrian in a low light environment and calculated the approximate distance. They used the smartphone-based thermal camera. The pedestrian detection task is performed using a multi-stage cascade learning device. They plotted the Two-Dimensional thermal image in Three-Dimensional space to calculate the distance. This is done by the inverse perspective transformation method. This method has achieved a 91% accurate object detection rate and takes only 0.34s to detect and the distance estimation accuracy is 95%. The method only tested at the indoor environment and no deteriorated phase. In paper [20], J. W. Davis at el. proposed a method for extracting salient objects from thermal image in different environmental condition. At first, they used statistical background subtraction to identify the local RoI. Then they got the background gradient information from that background subtraction. Then they integrated both region input and the background gradient information to generate a saliency Map. Finally, the contour image is filled to produce the mask. Experiments with their method and six challenging thermal video sequences of pedestrians recorded at very different environmental conditions showed promising results. The main constrain of this method is that they only tested on six challenging thermal video sequences of the pedestrian as the area of the testing is very small for this method.

3. Proposed Method

There are a few methods described in the previous section. One of them [8] used saliency map for detecting pedestrians. As they proved the positive impact of the saliency map, we proposed to use instance segmentation with the saliency map for detecting pedestrians. Our model takes a thermal image as input, then the pixel level mask is generated by the deep saliency network (PiCA-Net [6]). After that, we fused the saliency map with the corresponding thermal image. Then the combined image is fed into the pedestrian detector. Fig.1 illustrates system architecture. At first it takes thermal images as input. Then the thermal images are fed into a Deep Saliency network (i.e. PiCA-Net) to detect the salient part of the images. Then the salient images are combined with the corresponding thermal images. Thus it outputs images as it is in the input images additionally it identifies the salient part as well. Then the fused images are feed into the object Detector for detecting the pedestrian instances from the images.

Elaboration of the points of our proposed method are discussed below.

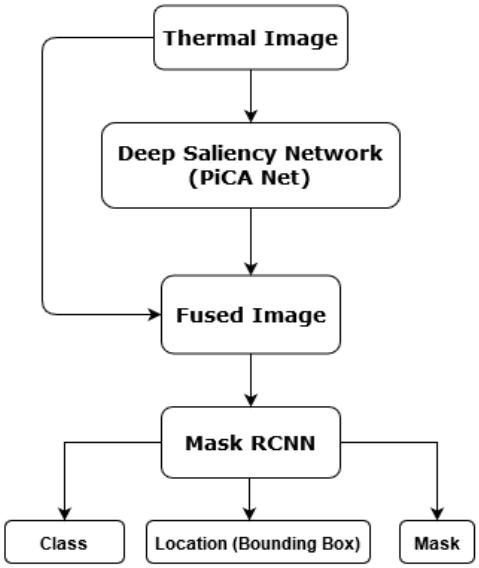

Fig.1. System architecture of the proposed method.

A. Thermal Image

Thermal imaging is a special equipment that can detect the heat produced by the people or objects and it uses these heat signature to produce images of them. Thermal images are the visual representation of the degree of infrared radiation emitted by any object. The thermal camera consists of sensors that can detect the electromagnetic radiation. Human eye cannot visualize electromagnetic radiation. A thermal camera can form images from the radiation that an objects emits. The infrared energy that an object radiates is its distinct property. In general, the hotter object radiates more infrared radiation than the cooler one. A thermal camera can detect small scale differences in temperature. Thermal camera converges the radiation emitted from object and from those information it creates an electronic image. In general different object have different temperature. So a thermal camera can easily detect object distinctly. There are three most commonly used night vision technologies. Among the three methods, only thermal imaging works perfectly in low or bad light condition.

Fig.2. Sample color Image [15]

In Fig.2 we can see a man with a plastic bag in his arm. The person's left hand is not visible in this image for the plastic bag.

Fig.3. Thermal Image [15]

We can see in Fig.3 that the hand is warm and the plastic bag contains almost zero heat. So, the human hand is visible in the Thermal image.

So, it is easily understandable from these two examples is the main difference of thermal image with the color image is that color image is constructed by sensing the visible light and the thermal image is the image constructed by measuring the heat of the surroundings of a frame.

Thermal images are basically represented in grayscale (i.e. black objects refers to be cold, white objects refers to be hot and the depth of gray shows the heat differences between objects.) Some thermal cameras, can make RGB images by coloring images for identifying objects of different temperatures.

B. Saliency Detection

Vision is the foremost and vital senses that human owns. Saliency is a visionary attention mechanism that narrow down to the specific part of any object that is important to see. Saliency highlights the most important part of an image. Koch and Ulman [9] define saliency at a given location by how different this location is from its surroundings in color, orientation, motion, and depth. Itti et al. [21] determine saliency by considering the feature difference between each and every pixel and its surrounding region.

Detecting saliency in thermal images helps to detect pedestrians precisely in our work. Thermal images are useful to differentiate a warm object from less warm surroundings. In day time, the objects in the surrounding becomes warm or warmer than the pedestrians. It can make pedestrians less distinguishable from surroundings. To face the challenge of detecting pedestrian in thermal images during daytime, we use the saliency maps extracted from thermal images. Saliency maps would help to perform better during the daytime where a pedestrian is less distinguishable from the surroundings.

In Fig.4. We see that the left two images are color image which contains two objects with a colorful background. Right side images contain the salient part of the corresponding images of the left side.

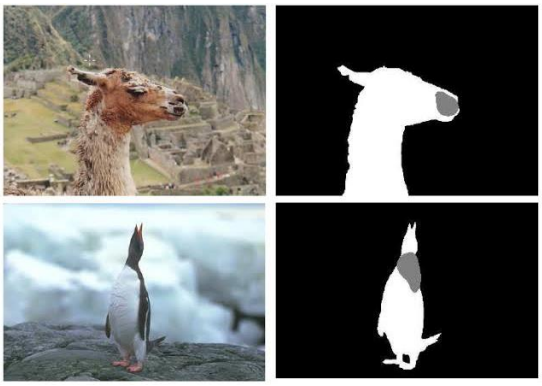

Fig.4. Saliency Detection

Prior works [9, 38-40] has unveiled the uses of saliency detection from images. Some authors [41-42] used the local method which is the neighbor pixel or region difference in image.

Others [43-44] introduced a global methods that depends on the color differences in terms of statistics. In recent years few deep saliency networks [6, 45] achieved a state-of-the-art performance in terms of detecting saliency in image.

We used PiCA-Net [6] which creates an attention map for each pixel of its corresponding location. It Uses Bidirectional LSTM for scanning the image both horizontally and vertically to obtain its global context. The attention mechanism is applied on a local neighbor region using deep convolutional layers for finding local context. Finally, a U-Net architecture unites the PiCA-Net hierarchically for detecting salient object.

C. Fused Image

Thermal image is used as the input in deep saliency network [6]. Fig.5. is a sample from the dataset we have used. Few pedestrian walking in the road in Fig.5. There are vehicle and trees in background.

Fig.5. Input Thermal Image

The deep saliency network highlights the salient part and discard the background. The salient object is explained in section 2.B. Fig.6 shows the output of the PiCA-Net. Here only the pedestrian instances are present. All other backgrounds are being discarded by the saliency detector.

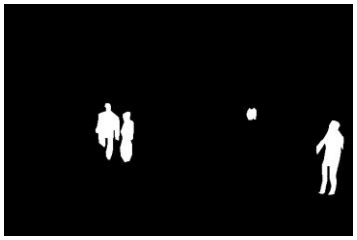

Fig.6. Output of PiCA-Net

It can be hard sometimes to decide whether the salient part is a pedestrian or not only looking at the mask. This is why we have to fuse the mask with the corresponding thermal image. Fig.7 is the fused image which contains both salient object and background. This will give the thermal image with the highlighted salient part. As a result, we get the input image as it is along with this, we also get the salient objects like pedestrians, cars etc. Then we feed this image to the detector for detecting the pedestrian from the input images.

Fig.7. Fused Image

D. Mask RCNN

In computer vision task, object detection is a technique that distinguishes between objects in an image or video. As it is associated with classification, it is more specific in what it identifies. It applies classification to distinct objects in an image or video and using bounding boxes to tell us where each object is. Image segmentation partitions an image into various segments (i.e. image objects). Image segmentation make image simple and change the image into a meaningful and easier image to analyze [23-24].

Segmenting an image allows separating the foreground from background and identifies the precise location of every object in that image.

Fig.8. illustrate the difference between instance segmentation and semantic segmentation. Here we can see that in semantic segmentation there are balloon present but in instance segmentation there are seven balloon instances present in the image. In semantic segmentation, each and every image pixel consists of a specific class. Instance segmentation goes one step further and separates distinct objects belonging to the same class [7]. Image segmentation is associated with both object classification and detection. So, both these techniques should take place before segmentation. After the object of an image is restrained with a bounding box, the pixel by pixel outline of that object which consists in the image can be done. In case of detecting, there can be many bounding boxes which represents different objects of interest within the image and we would not know how many beforehand.

Prior works [2-4] have tried to solve the problem using a different schema. R. Girshick [2] proposed to use selective search and extract 2000 region proposal from each image then these region proposals are feeding into the CNN. But the classification task of that vast number of region proposals takes a huge time to train. Additionally, as the selective search is a fixed algorithm it could lead to the generation of bad candidate region proposals. The limitations are removed in Fast R-CNN [3]. It proposed to feed the input image to the CNN to develop a feature map. Then both proposed region and bounding boxes are generated. It reduces the training time but the region proposals slow down the algorithm.

Then Shaoqing Ren et al. [4] propose to eliminate the selective search. They propose a different system to anticipate the region proposal. The ROIPool in Faster R-CNN [14] caused slightly misalignment from the regions of the original image but it is faster than the prior models it can even be used for object detection. K. He et al [5] introduced Mask R-CNN which extends Faster RCNN [14] by adding a prediction mask. It also expanded Faster R-CNN for accomplishing pixel-level segmentation. Mask R-CNN [5] does this by adding an additional branch to Faster R-CNN that produces a binary mask. This mask indicates if a pixel is part of an object or not. Mask RCNN [7] has achieved the

new state-of-the-art technique in case of instance segmentation. Mask RCNN is a deep CNN which tries to explain the instance segmentation problem in computer vision.

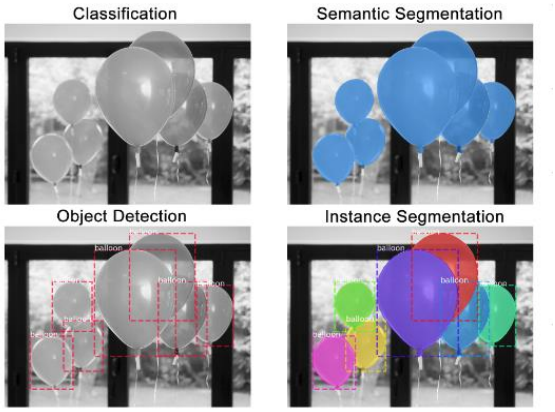

Fig.8. Example of Image Segmentation

Mask Representation: A mask is the objects spatial layout in an image. The spatial structure of masks can be addressed by the correspondent of pixel to pixel convolutions.

RoIAlign: RoIPool [4] creates a feature map from each Region of Interest. In Faster RCNN [14] RoI pooling adjusts the RPN (Region Proposal Networks). Though RoIAlign does not adjust the RPN. Instead it takes the object proposal and then divides those proposals into a certain number of bins. In each bin, there present a certain number of points which are sampled. After that it used bilinear interpolation to compute the value at those points. Usually, the size of the feature map that RoI Align layer produces, is a hyper parameter. So instead of using RoIPool [4] here RoIAlign is introduced. Input features are calculated by using bilinear interpolation [37] and regularly sampled location and then aggregate the result.

E. Prediction Output

The Mask RCNN [7] output the object class, Location with bounding box and mask. Here in Fig.9. Illustrate the pedestrian class, locating certain locations on the image.

In the output images the background remain as it was in the input images. Additionally it has illuminated salient part with the prediction score of the detector along with bounding boxes and masks.

Fig.9. Output with object class, Location with bounding box and mask.

4. Results and Discussion

We applied Mask-RCNN instead of Faster-RCNN, which has less false positive. Thus, it increases the detection rate.

A. Dataset Description

KAIST Multispectral pedestrian dataset contains around 95k images (50k for training and 45k for testing set). We take every 15th image from the day image sets and take every 10th image from the night image sets. Thus we selected 1702 images from the training set of the KAIST Multispectral Pedestrian dataset [5]. So, the training set contains 913

day and 789-night images contain 4646 pedestrian instances. We [8] took 362 images from the test set of the KAIST Multispectral dataset [5]. Which contains 193 day and 169-night images, containing 1230 pedestrian instances? To train the deep saliency network i.e. PiCA-Net, we need a thermal image with a ground truth saliency map. As there is no publicly available thermal image datasets with ground truth i.e. mask. We manually annotate these images using VGG Image Annotator [8] and although the pixel level annotations are not precise. But this dataset [8] works well for pedestrian detection tasks.

B. Implementation Details

We use PiCA-Net [9] for determining the salient part of the thermal image. For PiCA-Net we use an open-source implementation [10] and keep the same architectural design as described in the paper. We use SGD optimizer with momentum 0.9, weight decay 0.0005, and batch size 4. The learning rate of the decoder is trained with learning rate 0.01 and for the encoder is 0.001 and learning decay 0.1. The training is done on a single Tesla P4 with 32 GB memory using Google's colaboratory. The generated Saliency maps are size 224x224. And the image resolution in the dataset [6] is 640x512. So, we resize the saliency map to the original image size and merge with the corresponding input thermal images thus we obtain the fused images.

We then use Mask R-CNN for our pedestrian detection task. We use Detectron2 [13] library provided by Facebook AI Research to fine-tune Mask R-CNN with our rendered training set. We annotate the images with Labellmg in COCO format. We use to train the model learning rate 0.02 with Batch size 4 with maximum iteration 300. We use 0.7 as the threshold of the detector. The entire setup is done on a single Tesla P4 with 32 GB memory on Google's Collaboratory.

C. Results Evaluation

Our proposed method has got an accuracy of 88.14% over day and 91.84% night images. Section IV (B) described the implementation details of the Mask-RCNN.

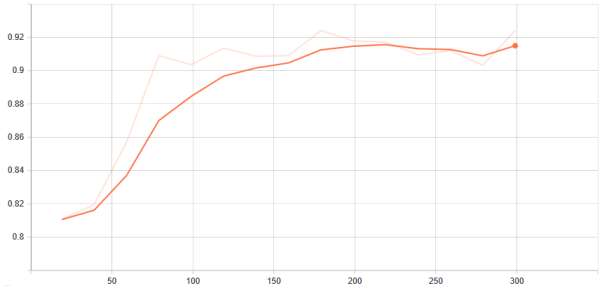

Fig.10. Accuracy Graph from Tensorboard for night set

From Fig.10 we can graphically observe that our model has 91.84% night time pedestrian detection accuracy and 88.14% for daytime pedestrian detection.

Table 1. Bounding box AP on COCO metrics

	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Night	44.592	87.734	41.557	38.636	53.644	50.957
Day	43.244	90.746	33.197	38.941	50.569	75.455

Table 2. Instance segmentation mask AP on COCO metrics

	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Night	44.887	88.189	42.460	38.493	54.880	43.660
Day	47.522	90.649	43.470	42.430	56.202	73.465

In table 1 and 2 we reported the COCO metrics of the bounding box and instance segmentation along with AP, AP_{50}, AP_{75}, AP_S, AP_M, AP_L. Mask IoU is used to evaluate AP. The IoU is an evaluation metric which is the ratio of the ground truth bounding box and the predicted bounding box. The average precisions are calculated at different IoU value. From this COCO metrics, in table 2, we can see that our model has achieved 0.882 and 0.906 mean average precision at IoU=0.5. In the next section we will compare our results with the existing state-of-the-art methods.

D. Comparison

Prior works [2, 11, 17] have introduced the image classification and detection task with semantic segmentation. They use bounding boxes to locate the pedestrian. Mask RCNN [7] extends the detection task into instance segmentation. This indicates which pixel belongs to which instance. Another drawback for Faster RCNN [2] is, it produces a significant number of false positives. Thus, reduces the accuracy rate. This problem is reduced later by adding an additional CNN to the faster RCNN [18]. Again RoIPool [25] in faster RCNN causes misalignment between RoI and extracted features. This negatively impacts predicting pixel-accurate masks although this may not have any effect on the classification tasks. Thus, the RoIAlign is used instead of RoIPool but our proposed network uses Mask RCNN which created from the Faster RCNN and it has removed the false positive drawbacks.

Table 3. Comparison between the accuracy of existing and proposed methods

	Day	Night
Baseline [8]	55.80%	59.60%
Ghose et al [8]	67.80%	78.30%
Ours	88.14%	91.84%

Table 3 compares the accuracy of the proposed model and the existing models.

Table 4. Comparison between the mAP value of existing and proposed methods

	Day	Night
Ghose et al [8]	0.640	0.676
Ours	0.882	0.906

Ghose et al [8] calculate mAP at IoU=0.5 we have also calculated the mAP at IoU=0.5 of our model using the COCO metrics. Table 4 is illustrated that our model has achieved a very good mean average precession with respect to previous one.

5. Conclusion

The aim of our model is to achieve higher pedestrian detection accuracy using deep saliency networks. Pedestrian detection task can be very effective for many cases such as autonomous driving or self-driving car or security purposes. We have used deep saliency networks to detecting the salient part and Mask RCNN as deep learning object detection techniques. We make an important contribution by using the Mask RCNN along with the deep saliency network (i.e. PiCA-Net) and improved the accuracy of the pedestrian detection task in thermal images. We also proved the positive impact of the instance segmentation in pedestrian detection. Additionally, the instance segmentation with saliency map is also expect to work for color images.

In this paper we have used OpenCV library to augment the thermal images with their corresponding saliency map and then feed into the model. In future if we can combine the extraction of the saliency information and the pedestrian detection task like SDS RCNN that can improve detection accuracy along with jointly learning the saliency and pedestrian detection task. The SDS RCNN can improve detection accuracy along with jointly learning the saliency and pedestrian detection task. The experimental result shows that the proposed method is robust than the existing method in case of detecting pedestrians in both day and night time.

References

[1] Geiger A, Lenz P and Urtasun R. “Are we ready for autonomous driving? the kitti vision benchmark suite.” In 2012 IEEE Conference on Computer Vision and Pattern Recognition 2012 Jun 16 (pp. 3354-3361). IEEE. doi: 10.1109/CVPR.2012.6248074

[2] Li C, Song D, Tong R and Tang M. “Illumination-aware faster R-CNN for robust multispectral pedestrian detection.” Pattern Recognition. 2019 Jan 1;85:161-71. doi:10.1016/j.patcog.2018.08.005

[3] Liu J, Zhang S, Wang S and Metaxas DN. “Multispectral deep neural networks for pedestrian detection.” arXiv preprint arXiv:1611.02644. 2016 Nov 8.

[4] Klein DA and Frintrop S. “Center-surround divergence of feature statistics for salient object detection.” In2011 International Conference on Computer Vision 2011 Nov 6 (pp. 2214-2219). IEEE. doi: 10.1109/ICCV.2011.6126499

[5] Hwang S, Park J, Kim N, Choi Y and So Kweon I. Multispectral pedestrian detection: Benchmark dataset and baseline. In Proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 1037-1045).

[6] Liu N, Han J and Yang MH. “Picanet: Learning pixel-wise contextual attention for saliency detection.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018 (pp. 3089-3098).

[7] He K, Gkioxari G, Dollár P and Girshick R. "Mask R-CNN." In Proceedings of the IEEE international conference on computer vision 2017 (pp. 2961-2969). arXiv:1703.06870

[8] Ghose D et al. "Pedestrian Detection in Thermal Images using Saliency Maps." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 2019 (pp. 0-0).

[9] Koch C and Ullman S. "Shifts in selective visual attention: towards the underlying neural circuitry." In Matters of intelligence 1987 (pp. 115-141). Springer, Dordrecht.

[10] Y. Jaehoon. Pytorch implementation of Pica-Net: Learning pixel-wise contextual attention for saliency detection. URL https://github.com/Ugness/PiCANetImplementation, 2018.

[11] Girshick R. "Fast R-CNN." In Proceedings of the IEEE international conference on computer vision 2015 (pp. 1440-1448).

[12] Dollar P, Wojek C, Schiele B and Perona P. "Pedestrian detection: An evaluation of the state of the art." IEEE transactions on pattern analysis and machine intelligence. 2011 Aug 4;34(4):743-61.

[13] Yuxin Wu and Alexander Kirillov and Francisco Massa and Wan-Yen Lo and Ross Girshick. Detectron2: https://github.com/facebookresearch/detectron2, 2019.

[14] Ren S, He K, Girshick R and Sun J. "Faster R-CNN: Towards real-time object detection with region proposal networks." In Advances in neural information processing systems 2015 (pp. 91-99).

[15] "Thermography" [Online]. Available: https://en.wikipedia.org/wiki/Thermography Collected 26 January 2020.

[16] Wang X, Wang M and Li W. "Scene-specific pedestrian detection for static video surveillance." IEEE transactions on pattern analysis and machine intelligence. 2013 Jun 25;36(2):361-74. doi: 10.1109/TPAMI.2013.124

[17] Girshick R, Donahue J, Darrell T and Malik J. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition 2014 (pp. 580-587).

[18] Ren S, He K, Girshick R and Sun J. "Faster R-CNN: Towards real-time object detection with region proposal networks." In Advances in neural information processing systems 2015 (pp. 91-99).

[19] Kim J. "Pedestrian Detection and Distance Estimation Using Thermal Camera in Night Time." In2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) 2019 Feb 11 (pp. 463-466). IEEE.

[20] Davis JW and Sharma V. "Background-subtraction in thermal imagery using contour saliency." International Journal of Computer Vision. 2007 Feb 1;71(2):161-81.

[21] Itti L, Koch C and Niebur E. "A model of saliency-based visual attention for rapid scene analysis." IEEE Transactions on pattern analysis and machine intelligence. 1998 Nov;20(11):1254-9. doi: 10.1109/34.730558

[22] Turkowski K. "Filters for common resampling tasks." In Graphics gems 1990 Aug 1 (pp. 147-165). Academic Press Professional, Inc..

[23] Shapiro L, Stockman G. Computer Vision Prentice Hall. Inc., New Jersey. 2001.

[24] Lauren B and Lee LW. "Perceptual information processing system." Paravue Inc. US Patent Application. 2003 Jul;10(618,543).

[25] Hariharan B, Arbeláez P, Girshick R and Malik J. "Simultaneous detection and segmentation." In European Conference on Computer Vision 2014 Sep 6 (pp. 297-312). Springer, Cham.

[26] Tai JC and Song KT. "Background segmentation and its application to traffic monitoring using modified histogram." In IEEE International Conference on Networking, Sensing and Control, 2004 2004 Mar 21 (Vol. 1, pp. 13-18). IEEE. doi: 10.1109/ICNSC.2004.1297401

[27] Göktürk K and Jönsson A. "Developing a Resource-Efficient Sensor Cleaning System for Autonomous Heavy Vehicles (2019)."

[28] Zhao D, Hanson EJ, Nix MC, Chin S, inventors; Uber Technologies Inc, assignee. Systems and Methods for On-Site Recovery of Autonomous Vehicles. United States patent application US 15/884,852. 2019 Aug 1.

[29] Li C, Song D, Tong R and Tang M. "Illumination-aware faster R-CNN for robust multispectral pedestrian detection." Pattern Recognition. 2019 Jan 1;85:161-71.

[30] Zhang L et al. "Cross-modality interactive attention network for multispectral pedestrian detection." Information Fusion. 2019 Oct 1;50:20-9.

[31] Lahmyed R, El Ansari M and Ellahyani A. "A new thermal infrared and visible spectrum images-based pedestrian detection system." Multimedia Tools and Applications. 2019 Jun 30;78(12):15861-85.

[32] Bilal M and Hanif MS. "High performance real-time pedestrian detection using light weight features and fast cascaded kernel SVM classification." Journal of Signal Processing Systems. 2019 Feb 1;91(2):117-29.

[33] Bastian BT and Jiji CV. "Pedestrian detection using first-and second-order aggregate channel features." International Journal of Multimedia Information Retrieval. 2019 Jun 1;8(2):127-33.

[34] Kim S, Kwak S and Ko BC. "Fast pedestrian detection in surveillance video based on soft target training of shallow random forest." IEEE Access. 2019 Jan 11;7:12415-26. doi: 10.1109/ACCESS.2019.2892425

[35] Zhang L et al. "Weakly aligned cross-modal learning for multispectral pedestrian detection." In Proceedings of the IEEE International Conference on Computer Vision 2019 (pp. 5127-5137).

[36] Van de Weijer J, Gevers T and Bagdanov AD. "Boosting color saliency in image feature detection." IEEE transactions on pattern analysis and machine intelligence. 2005 Nov 21;28(1):150-6. doi: 10.1109/TPAMI.2006.3

[37] Li Y, Qi H, Dai J, Ji X and Wei Y. "Fully convolutional instance-aware semantic segmentation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017 (pp. 2359-2367).

[38] Goferman S, Zelnik-Manor L and Tal A. "Context-aware saliency detection." IEEE transactions on pattern analysis and machine intelligence. 2011 Dec 27;34(10):1915-26. doi: 10.1109/TPAMI.2011.272

[39] Yan Q, Xu L, Shi J and Jia J. "Hierarchical saliency detection." In Proceedings of the IEEE conference on computer vision and pattern recognition 2013 (pp. 1155-1162).

[40] Hou X and Zhang L. "Saliency detection: A spectral residual approach." In2007 IEEE Conference on computer vision and pattern recognition 2007 Jun 17 (pp. 1-8). Ieee. doi: 10.1109/CVPR.2007.383267

[41] Harel J, Koch C and Perona P. "Graph-based visual saliency." In Advances in neural information processing systems 2007 (pp. 545-552).

[42] Achanta R, Estrada F, Wils P and Süsstrunk S. "Salient region detection and segmentation." In International conference on computer vision systems 2008 May 12 (pp. 66-75). Springer, Berlin, Heidelberg.
[43] Achanta R, Hemami S, Estrada F and Susstrunk S. "Frequency-tuned salient region detection." In2009 IEEE conference on computer vision and pattern recognition 2009 Jun 20 (pp. 1597-1604). IEEE. doi: 10.1109/CVPR.2009.5206596
[44] Cheng MM, Mitra NJ, Huang X, Torr PH and Hu SM. "Global contrast based salient region detection." IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014 Aug 5;37(3):569-82. doi: 10.1109/TPAMI.2014.2345401
[45] Deng Z. et al. "R3net: Recurrent residual refinement network for saliency detection." In Proceedings of the 27th International Joint Conference on Artificial Intelligence 2018 Jul 13 (pp. 684-690). AAAI Press.
[46] Szarvas, Mate, Akira Yoshizawa, Munetaka Yamamoto, and Jun Ogata. "Pedestrian detection with convolutional neural networks." In IEEE Proceedings. Intelligent Vehicles Symposium, 2005., pp. 224-229. IEEE, 2005.
[47] Setjo, Christian Herdianto, and Balza Achmad. "Thermal image human detection using Haar-cascade classifier." In 2017 7th International Annual Engineering Seminar (InAES), pp. 1-6. IEEE, 2017.
[48] Paul, Manoranjan, Shah ME Haque, and Subrata Chakraborty. "Human detection in surveillance videos and its applications-a review." EURASIP Journal on Advances in Signal Processing 2013, no. 1 (2013): 176.

Authors' Profiles

A. K. M. Fahim Rahman is a student at Computer Science and Engineering Discipline, Khulna University. He was born on 18th September 1998. He involves in programming contests. His research interests include Machine Learning, Deep Learning and Image Processing.

Mostofa Rakib Raihan is a student at Computer Science and Engineering Discipline, Khulna University. He was born on 1st January 1998. His research interests include Machine Learning, Pattern Recognition and Artificial Intelligence.

S.M. Mohidul Islam is an Associate Professor at the Computer Science and Engineering Discipline, Khulna University, Bangladesh. He received his B.Sc. Engg. And M.Sc. Engg. Degree from Khulna University. His research interests include Machine learning, Data Mining, Pattern Recognition, and Digital Image Processing.

How to cite this paper: A. K. M. Fahim Rahman, Mostofa Rakib Raihan, S.M. Mohidul Islam, " Pedestrian Detection in Thermal Images Using Deep Saliency Map and Instance Segmentation", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.13, No.1, pp. 40-49, 2021.DOI: 10.5815/ijigsp.2021.01.04