Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

Construction of LWCNN Framework and its Application to Pedestrian Detection with Segmentation Process

R. Kanthavel

Department of Computer Engineering, King Khalid University, Abha, Kingdom of Saudi Arabia

E-mail: kanthavel2005@gmail.com

Abstract

To solve the challenges in traffic object identification, fuzzification, and simplification in a real traffic environment, it is highly required to develop an automatic detection and classification technique for roads, automobiles, and pedestrians with multiple traffic objects inside the same framework. The proposed method has been evaluated on a database with complicated poses, motions, backgrounds, and lighting conditions for an urban scenario where pedestrians are not obstructed. The suggested CNN classifier has an FPR of less than that of the SVM classifier. Confirming the significance of automatically optimized features, the SVM classifier's accuracy is equal to that of the CNN. The proposed framework is integrated with the additional adaptive segmentation method to identify pedestrians more precisely than the conventional techniques. Additionally, the proposed lightweight feature mapping leads to faster calculation times and it has also been verified and tabulated in the results and discussion section.

Keywords: Object Detection, CNN

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

1. Introduction

In recent times, there has been a lot of progress in the field of computer vision. The object tracking method is the most often used approach for identifying the moving objects in video sequences after time passed. The major objective of object tracking is to connect objects, their form or features, and their position in successive video sequences. As a result, it will be critical for computer vision systems to monitor the item classification and identification [1-5].

In addition, finding any moving item in the frame is the first step towards tracking it. Finally, the identified items may be categorized as trees swaying in the wind, birds flapping their wings, moving people, and so on. It is a difficult job to track objects in videos by using image processing techniques. Additional problems are object occlusion; complicated object motion, real-time requirement, and an incorrect or distorted form of the object are also observed [6-8].

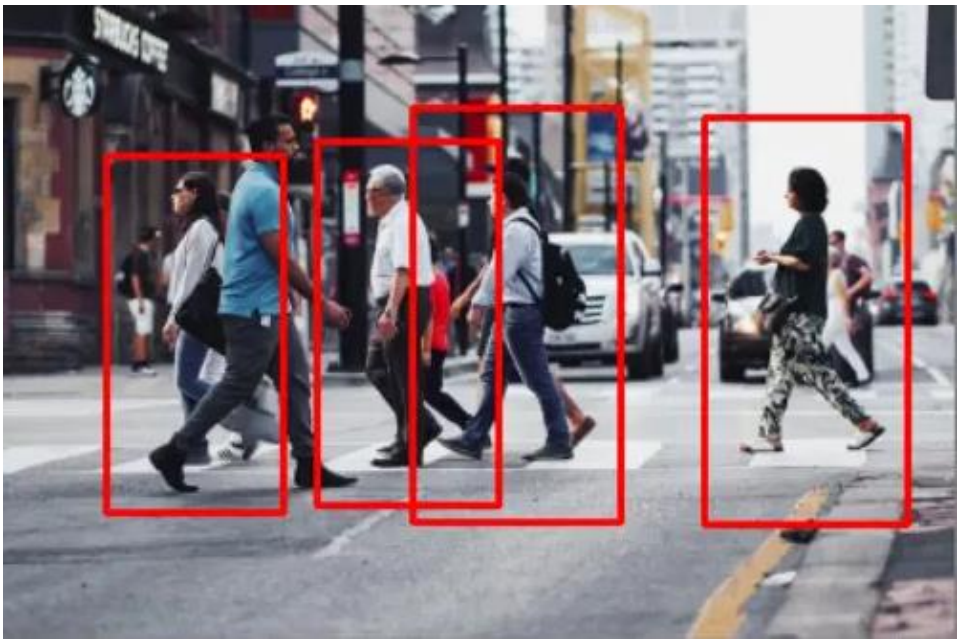

Figure 1. Pedestrian Detection in the Road-Cross

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

Nonetheless, this type of surveillance has grown increasingly prevalent in recent years. As observed, these seven technologies are largely utilised for a variety of reasons, including traffic monitoring, robot vision, surveillance, security, and video communication in public places such as subway stations, airports, stadiums, and amusement parks. Using this technique, the application must discover the best combination of processing, communication, and accuracy throughout the network. Revenue in the computer and communication sector is connected to the quantity and the type of collaboration between cameras in data analysis [9-13]. Further, a simpler way to think about it is to consider the process of obtaining the orientation of an item from the moment it appears in the image scene until the end of the processing.

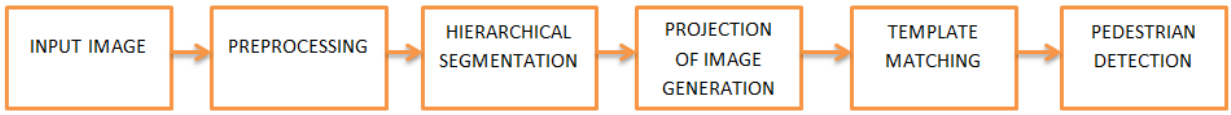

Figure 2. Simplified Block Diagram of Pedestrian Detection

Object identification is a basic computer vision issue that has broad application in video surveillance, robotics, and smart transportation domain. Both intelligent video surveillance systems and driving assistance systems use pedestrian detection. Generally, pedestrian detection is used to provide basic information from video surveillance, object identification, and fast crowd counting and it is considered as an important element for semantic comprehension of the environment [14, 15].

Feature extraction is performed by using a CNN (Convolutional Neural Network). The feature representation capabilities of CNN are superior to those of hand-crafted features. Given that CNN feature maps can be effectively used for the semantic segmentation problem, this discovery may be credited to CNN. To enhance the performance of a CNN-based pedestrian detector, this research work provides a framework based on a semantic network [16].

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

2. Organization of the Research

The rest of the research article is structured as follows: Section 3 discusses about previous research works on pedestrian detection by using a variety of techniques. Section 4 explains the suggested approach via the use of a diagram. Section 5 outlines the proposed research findings. The conclusion and future work of this research study are summarized in the final section.

3. Preliminaries

The approach described by Ronneberger et al. benefits in classifier network development by expanding the receptive field while decreasing spatial resolution. The "encoder-decoder" architecture was proposed in a semantic segmentation approach that aims to restore spatial dimension [17].

According to the study proposed by Badrinarayanan et al, pooling layers in the encoder would reduce the dimension of the space the code is working in, while the decoder will assist to restore the dimension of the space. SegNet is a highly sophisticated pixel-wise semantic labelling encoder-decoder [18].

Long et al. describe the network, which is composed of a convolutional network and an upscaling network followed by a classification layer. In upsampling, the feature maps acquired are incomplete. SegNet uses the closest neighbor method to transform dense feature maps into sparse ones for dense image labeling applications. SegNet reportedly outperforms other state-of-the-art deep semantic segmentation techniques while requiring less memory [19].

The researchers Liu and chan suggest semantic picture segmentation often referred to as semantic image labelling, in which each pixel in an image gets a specific item class label. Prior to DNN's widespread use of semantic picture segmentation, the most often used methodologies were random forest (RF) classifiers [20].

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

The previous DNN-based semantic segmentation methods proposed by Ciresan et al. may be used to classify picture patches. Each pixel was classified into one of eight distinct groups using a fixed-size image patch around each pixel. Since deep neural networks are generally composed of fully linked layers, they require inputs of a specific size [21].

Research Gap

With the proposed framework, there is a shortage of CNN designs for dense predictions like crowd counting or object identification in dense areas. Furthermore, it will gain popularity when used with lightweight Convolutional Networks to accelerate the process. By providing this capacity, segmentation can be performed on any image size and is considerably quicker than traditional patch-based techniques.

4. Proposed Method

The LW-CNN framework with SegNet has been developed to count the detected objects or pedestrians. With a dilated convolution depth of 6 and a dilation rate of 2, it is possible to accurately segment pedestrians in roadside crowds. The kernel size is always 3x3 as a matter of policy. The proposed design is depicted in Figure 3.

4.1 Pre Processing

When the proposed LW-CNN is initialized with low weight, a kernel of 3x3 in size and a stride of 1 have been applied to the picture to obtain the result. It attempts to get rid of everything except face and backdrop, and then the item gets cropped. After down-sampling, the next preprocessing step is to crop the picture to 32x32, 64x64, or 128x128. To measure the performance of the CNN model, it should be tested with varying input sizes. Finally, we are adding salt, pepper, and speckle to the dataset [22]. To illustrate that, despite the negative remarks, CNN's performance

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

stays consistent. Further, normalization is completed via GCN, local normalization, and histogram equalization.

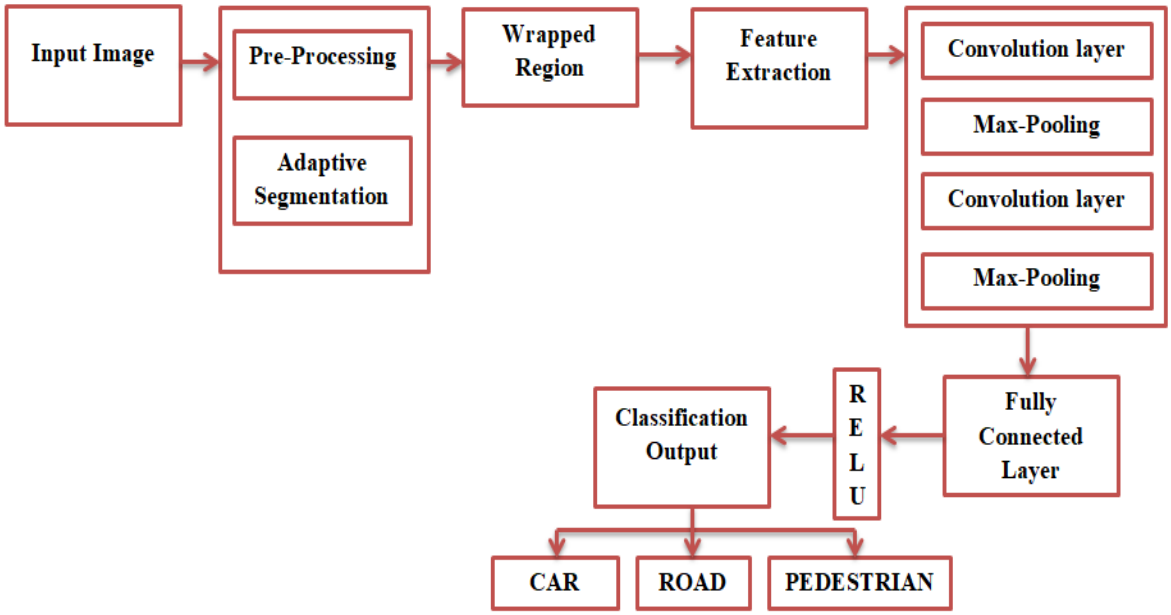

Figure 3. Overall Proposed Framework

4.2 Adaptive Segmentation

In this phase, finding the pedestrian in the image necessitates spending close attention to the available information. The objects have well-defined boundaries with this segmentation, allowing the pedestrian to be accurately recognized. Using a pre-defined list of sensitivities and internal thresholds, the system will set the internal variables to calculated values depending on the input parameters, and the internal variables will be set to their calculated values [23, 24]. In contrast, for the variables present in this system, it is much easier to deal with more diverse picture data by utilizing the method of adaptive computation for internal variables.

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

4.3 Detection

If the outer scale of objects in the scene varies over a wide range, a third external parameter may be specified. This can be delineated as the approximate fraction of background, or the minimum acceptable region size is considered as an alternative invocation. This parameter allows the non-uniform application of the outer scale across the image. It is used when the outer scale of some regions (e.g. background) is much larger than the outer scale of other important regions in the scene [25, 26]. Since the background itself needs to be classified, it is not desirable to limit the maximum size.

4.3.1 Feature extraction

To get a result of 64 feature maps, a first convolution layer of size 28 x 28 with an input picture size of 32 x 32 is used. The output picture of 28x28 is transferred to a 2x2 kernel pooling layer of stride 2 for each dimension with a max-pooling layer of 2x2 kernels with stride 2. This results in an output image of 14x14. Max pooling has examined every dimension to uncover the hidden structure of an image. Before the second convolution layer, the results of the first convolution layer were convoluted by a 5 x 5 kernel and a stride of 1. After the second convolution layer, the results were 10 x 10 pixels.

The last layer contains 6 output nodes that correspond to the right item. Rectified linear unit (ReLu) is the activation function used for all layers. ReLu is defined as a value that is higher than or equal to zero but less than the maximum value. Training is used to optimize the CNN model for non-convex circumstances and adjust the gradient exponentially. Cross-entropy is a measure of the performance of an algorithm as defined by its error function.

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

4.3.2 Road Map

In the perspective, straight-line, uphill, and downhill sections form the basic straight forms that provide visual contrast. To represent the road information in a better way, we should use the methods other than the conventional Hough transform and RANSAC.

4.3.3 Vehicle Detection

Vehicle movement takes place on the road by occupying the whole road space. Vehicle detection on the road relies on the collection of road contours. Vehicle detection is suggested to be implemented on the road by using ground-based testing for vehicle detection. The vehicle threshold for this experiment is set at 5, while the vehicle aspect ratio range is set between 0.5 and 2.

4.3.4 Pedestrian detection

That is, the general form of the road area defines the human source position on the road. This research work suggests using a multi-feature fusion technique to accurately identify and segment the pedestrian target region, as well as extracting the pedestrian image. To generate a multi-feature fusion target, feature fusion approaches use properties such as aspect ratio, perspective ratio, and area ratio.

5. Results & Discussion

Many experiments are used to investigate the classifier system as suggested in this article. The proposed experiment includes a variety of diverse picture sequences including various city streets, suburban roads, and varied weather situations. The road map will be notified in the output image, which is shown in the resultant image figure 4.

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

Figure 4. Input and Output Images

The Daimler dataset is a real-life stereo picture collection of approximately 21k photos, of which around 7,000 are actual stereo pairs, which is recorded for a total duration of 27 minutes in the continuous metropolitan road traffic [27, 28]. Traffic items that exist in the collection include both continuous and complicated road situations, such as roadways, cars, and people. With an average accuracy of 78.5%, the basic segmentation process will be able to distinguish between roadways, people, and cars. Also, the computation time is very higher than other classification

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

methods. SVM classifier shows its superiority than traditional segmentation method. Table 1 shows the computed overall performance metrics.

Table 1. Computed Overall Performance Metrics

S.No	Method	Accuracy	Precision	Recall	Total computation time (Sec)
1	Traditional Segmentation process	78.5%	75%	72.56%	65
2	Support Vector Machine classifier	87.34%	80%	79.12%	64
3	Proposed LW-CNN	94.8%	90%	93.2%	62

Several performance metrics have been obtained by using the proposed framework, and further it accurately recognizes the items referred to in the table as road, vehicle, and pedestrian while requiring less calculation time than previous approaches. It is necessary to calculate the observed value in order to determine accuracy and recall index. In order to calculate, the following formulae are used:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

This graph depicted in Figure 5 represents the overall performance of the framework, which was calculated to demonstrate recall, precision, and accuracy.

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

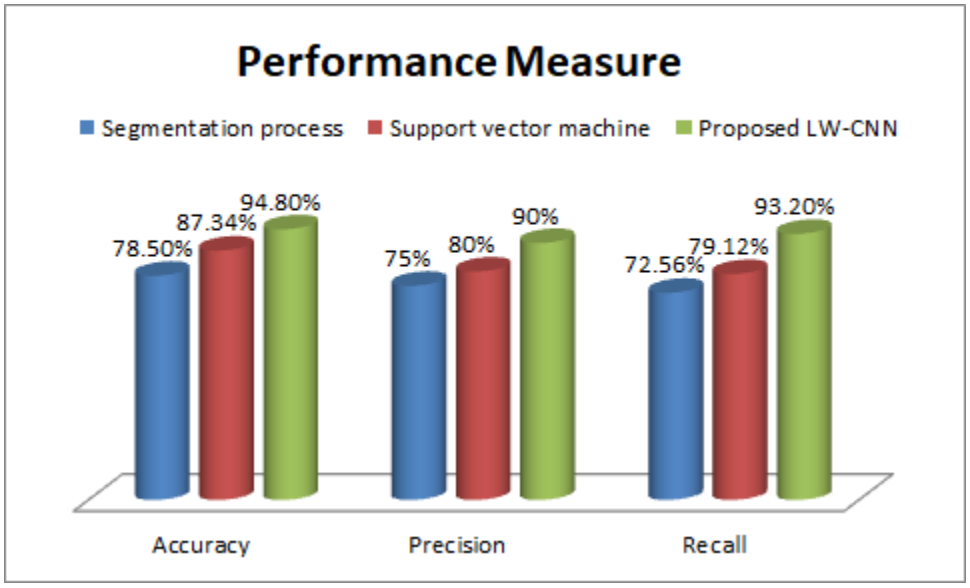

Figure 5. Overall Performance Graph

The major goal accomplished by the proposed lightweight CNN is to decrease the amount of time taken to compute information, which is shown in the table 1.

6. Conclusion

As a result, the proposed novel LW-CNN with SegNet method offers high accuracy with less calculation time. The suggested categorization detection system classifies roads, cars, and people while maintaining unity and addressing ambiguity in conventional detection techniques. The proposed work includes the detection of r and pedestrian detection. In the future, more research into the proposed technique by employing more complex deep neural networks, such as the residual network, for enhanced segmentation and detection is expected. Nonetheless, the algorithm has not completely satisfied the end-user requirements.The following are the major reasons:

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

1. The picture shows a lot of noise and regions that does not exist. These big regions will impair the detection results if they are not corrected completely.

2. Inaccurate ROI areas and missing identification of road vehicles occur due to inaccuracies in road detection results [29].

3. Since the color and texture of the road around the car are quite similar to the vehicle, it may be difficult to accurately separate the vehicle from the road.

References

[1] Balasubramaniam, Vivekanadam. "Artificial Intelligence Algorithm with SVM Classification using Dermascopic Images for Melanoma Diagnosis." Journal of Artificial Intelligence and Capsule Networks 3, no. 1: 34-42.

[2] D. M. Gavrila and J. Giebel, “Shape-based pedestrian detection and tracking,” in Proc. IEEE Intelligent Vehicle Symposium, IV 2002, Versailles, France, June 2002.

[3] Adam, Edriss Eisa Babikir. "Evaluation of Fingerprint Liveness Detection by Machine Learning Approach-A Systematic View." Journal of ISMAC 3, no. 01 (2021): 16-30.

[4] Manoharan, J. Samuel. "Capsule Network Algorithm for Performance Optimization of Text Classification." Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 1-9.

[5] K. Levi and Y. Weiss, “Learning object detection from a small number of examples: the importance of good features,” in Proc. International Conference on Computer Vision ICCV 2003, Nice, France, Oct. 2003.

[6] Adam, Edriss Eisa Babikir. "Survey on Medical Imaging of Electrical Impedance Tomography (EIT) by Variable Current Pattern Methods." Journal of ISMAC 3, no. 02 (2021): 82-95.

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

[7] J. Louie, “A biological model of object recognition with feature learning,” Master’s thesis, Massachusetts Institute of Technology, Cambridge, 2003.

[8] Karuppusamy, P. "Building Detection using Two-Layered Novel Convolutional Neural Networks." Journal of Soft Computing Paradigm (JSCP) 3, no. 01 (2021): 29-37.

[9] H. Schneiderman, “Learning a restricted bayesian network for object detection,” in Proc. Computer Vision and Pattern Recognition CVPR’04, Washington, DC, USA, June 2004.

[10] Vijayakumar, T., Mr R. Vinothkanna, and M. Duraipandian. "Fusion based Feature Extraction Analysis of ECG Signal Interpretation–A Systematic Approach." Journal of Artificial Intelligence 3, no. 01 (2021): 1-16.

[11] D. M. Gavrila and S. Munder, “Vision-based pedestrian protection: The PROTECTOR system,” in Proc. IEEE Intelligent Vehicle Symposium, IV 2004, Parma, Italy, June 2004.

[12] Chen, Joy Iong-Zong. "Design of Accurate Classification of COVID-19 Disease in X-Ray Images Using Deep Learning Approach." Journal of ISMAC 3, no. 02 (2021): 132-148.

[13] Sharma, Rajesh, and Akey Sungheetha. "An Efficient Dimension Reduction based Fusion of CNN and SVM Model for Detection of Abnormal Incident in Video Surveillance." Journal of Soft Computing Paradigm (JSCP) 3, no. 02 (2021): 55-69.

[14] M. Soga, T. Kato, M. Ohta, and Y. Ninomiya, “Pedestrian detection using stereo vision and tracking,” in Proc. World Congress on ITS, Nagoya, Japan, Oct. 2004.

[15] Sungheetha, Akey, and Rajesh Sharma. "Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural Network." Journal of Trends in Computer Science and Smart technology (TCSST) 3, no. 02 (2021): 81-94.

[16] Smys, S., and Wang Haoxiang. "Naïve Bayes and Entropy based Analysis and Classification of Humans and Chat Bots." Journal of ISMAC 3, no. 01 (2021): 40-49.

[17] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. arXiv:1505.04597.

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

[18] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2015). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561.

[19] Long, J., Shelhamer, E., and Darrell, T. (2015). “Fully convolutional networks for semantic segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Boston, MA).

[20] Liu, T. R., and Chan, S. C. (2015). “A hierarchical semantic image labelling method via randomforests,” in TENCON 2015-2015 IEEE Region 10 Conference (Macao), 1–5.

[21] Ciresan, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2012). “Deep neural networks segment neuronal membranes in electron microscopy images,” in Advances in Neural Information Processing Systems (Lake Tahoe), 2843–2851.

[22] He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Las Vegas, NV), 770–778.

[23] Dayana, A. Mary, and WR Sam Emmanuel. "A Patch-Based Analysis for Retinal Lesion Segmentation with Deep Neural Networks." In International conference on Computer Networks, Big data and IoT, pp. 677-685. Springer, Cham, 2019.

[24] Huang, J.-J., and Siu, W.-C. (2017). Learning hierarchical decision trees for single image super-resolution. IEEE Trans. Circ. Syst. Video Technol. 27, 937–950. doi: 10.1109/TCSVT.2015.2513661

[25] Hussain, J. "A Shape-Based Character Segmentation Using Artificial Neural Network for Mizo Script." In International Conference on Communication, Computing and Electronics Systems, pp. 231-239. Springer, Singapore, 2020.

[26] Huang, J. J., Siu, W. C., and Liu, T. R. (2015). Fast image interpolation via random forests. IEEE Trans. Image Process. 24, 3232–3245. doi: 10.1109/TIP.2015.2440751

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 269-283
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.008

[27] Swetha, O., and C. Ramachandran. "Counting and Tracking of Vehicles and Pedestrians in Real Time Using You Only Look Once V3." In Data Intelligence and Cognitive Informatics, pp. 873-886. Springer, Singapore, 2021.

[28] Kapuriya, B. R., Debasish Pradhan, and Reena Sharma. "Selective segmentation of piecewise homogeneous regions." In International Conference on Innovative Data Communication Technologies and Application, pp. 535-542. Springer, Cham, 2019.

[29] Mittal, Neetu, and Alexander Gelbukh. "Change detection in remote-sensed data by particle swarm optimized edge detection image segmentation technique." In Innovative Data Communication Technologies and Application, pp. 809-817. Springer, Singapore, 2021.

Author's biography

R. Kanthavel is a professor in the Department of Computer Engineering, in King Khalid University, Abha, in the Kingdom of Saudi Arabia. His research is mainly focused on the emerging smart computing technologies that includes Distributed Computing, quantum computers, Computer Graphics, Computer Networks, and Web Technologies.

ISSN: 2582-4252 (online)
Submitted: 08.07.2021
Revised: 01.09.2021
Accepted: 10.09.2021
Published: 20.09.2021