Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

Speedy Image Crowd Counting by Light Weight Convolutional Neural Network

B. Vivekanandam

Senior Lecturer, Faculty of Computer Science and Multimedia, Lincoln University College, Malaysia

E-mail: vivekanandam@lincoln.edu.my

Abstract

In image/video analysis, crowds are actively researched, and their numbers are counted. In the last two decades, many crowd counting algorithms have been developed for a wide range of applications in crisis management systems, large-scale events, workplace safety, and other areas. The precision of neural network research for estimating points is outstanding in computer vision domain. However, the degree of uncertainty in the estimate is rarely indicated. Point estimate is beneficial for measuring uncertainty since it can improve the quality of decisions and predictions. The proposed framework integrates Light weight CNN (LW-CNN) for implementing crowd computing in any public place for delivering higher accuracy in counting. Further, the proposed framework has been trained through various scene analysis such as the full and partial vision of heads in counting. Based on the various scaling sets in the proposed neural network framework, it can easily categorize the partial vision of heads count and it is being counted accurately than other pre-trained neural network models. The proposed framework provides higher accuracy in estimating the headcounts in public places during COVID-19 by consuming less amount of time.

Keywords: CNN, Crowd counting, COVID 19

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

1. Introduction

The Computer Vision (CV) applications include accurately predicting the number of objects and human beings in a picture. While object recognition and counting is used in many contexts, it is most often employed in the areas of security and development. Likewise, the counting of crowds in images may also be used in other situations, including surveys and traffic control. Accurate crowd counts assist in situations such as stampedes and fires since these types of incidents may benefit from knowing the number of people present in the particular area [1-3]. Because of this, researchers are most likely to investigate the use of image-based object counting in different areas. Recently, gathering of people in public places are prohibited due to COVID-19. Figure 1 shows the picture with dense crowd at public place.

Figure 1. Crowd at Public Place

Additionally, a significant portion of the research literature deals with remarkable achievements in the aforementioned areas, which are discussed extensively. Crowd counting is

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

required for crowd gathering activities such as religious and political rallies, parades, marathons, and concerts to assist in their administration and security [4-6]. Finally, crowd counting helps in determining the significance of demonstrations. It's unusual for different political parties to estimate the number of individuals, who attend rallies in such dramatically different ways. Even with the deployment of security cameras, crowd monitoring is challenging [7, 8].

Figure 2. Head Count with Segmented Area

Figure 2 shows crowd counting on a segmented area. Deep learning algorithms and CNNs have led to great advancements in object and crowd counting applications. Improvements in one application tend to lead to improvements in other related applications. The significance of crowd counting techniques may be used to analyze the crowds and track their behavior, too [9, 10].

CNNs (convolutional neural networks) have recently proven to be highly capable of performing a wide range of computer vision tasks, including object identification, image recognition, face recognition, and image segmentation. As a result of these recent achievements, several crowd counting techniques based on CNN technology have been suggested [11-13].

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

2. Organization of the Research

The rest of the research article is organized as follows; section 3 contains the description of existing research works on crowd counting techniques. Section 4 describes the proposed crowd counting methodology. Section 5 discusses the results obtained through the proposed framework. Section 6 concludes the proposed research work by including the future possible enhancements.

3. Preliminaries

The first large-scale crowd counting projects were implemented by using different detection techniques. Head detection methods use a sliding window over an image to track the movement. There have been many new approaches, such as Region-based Convolutional Neural networks (RCNN), YOLO, and Single-Shot multi-box Detector (SSD), which provide excellent accuracy in sparse situations and representation. Moreover, they fail to deliver enough performance in the crowded areas. Features such as pixels or areas are used in the density-based techniques. Regression-based methods have the drawback of losing the location information over time [14, 15].

Lei et al. built a fundamental crowd-counting model by using the weaker type of supervision. Annotations that are less complete will simply demand to know the total number of items. The best results have been obtained by using the multiple density map estimation method [16].

Crowd counting has been made simpler with the use of a smart camera created by Tong et al. Multi-task learning was used to achieve density-level categorization by using the suggested method. Transposed convolutional layers were used to compensate the loss of image description. Using the suggested approach, it was estimated that the crowd density was between a certain limit and an unknown one [17].

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

Using a density-based method with a linear mapping between local characteristics and density maps, Lemptisky et al. have uncovered a hitherto unknown fauna. Rather than using a linear technique to solve its problems, Random Forest was suggested, and it included crowdedness before and after training two distinct forests. Moreover, the forest-storing technique performs better than the linear method while using less memory. However, it is challenging to use conventional features to extract low-level information that cannot be counted on a high-quality density map with the proposed approach [18].

For crowd counting, the adaptable CNN was developed by Sang et al. To calculate the estimated headcount, CNN was utilized to generate the crowd density map, which was further processed. To evaluate the suggested method, the researchers have carried out different experiments on the Shanghai Tech dataset and proved the system to be effective on sparse representation and web camera surveillance for detecting crime scenes [19].

The researchers Zhang et al. have conducted a research study that use CNN to count the number of passengers on subway stations. Additional datasets include a collection of 627 images and 9243 annotated heads. During weekday peak and weekend off-peak periods, the images were taken. The authors have utilized the first 13 levels of VGG-16, as shown in the illustration. Compared to state-of-the-art techniques, the typical datasets have exhibited lower MAE and MSE [20].

Research Gap

Despite the excellent performances shown by density estimates and CNN-based methods for counting crowds, wherein less emphasis has been devoted to determining the level of uncertainty in the predicted outcomes. Because of overfitting and assessing uncertainty in crowd positions, many traditional models are not reaching appropriate results. Both lacks of

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

comprehension of model outputs and neural networks' susceptibility to overfitting require probabilistic explanations.

4. Methodologies

4.1 Light weight CNN (LW-CNN)

The proposed frame work has constructed a LW - CNN architecture of multi-scale, which has used VGG for crowd counting, and that application is tested in real-world settings. Figure 3 illustrates how the VGG-16 provides 10 convolution layers and 3 max-pooling layers at the front end. With a dilated convolution depth of 6 and a dilation rate of 2, it is possible to accurately count the number of people present in a crowd. At all times, the kernel size is kept at 3X3. Figure 3 shows the block diagram of proposed architecture.

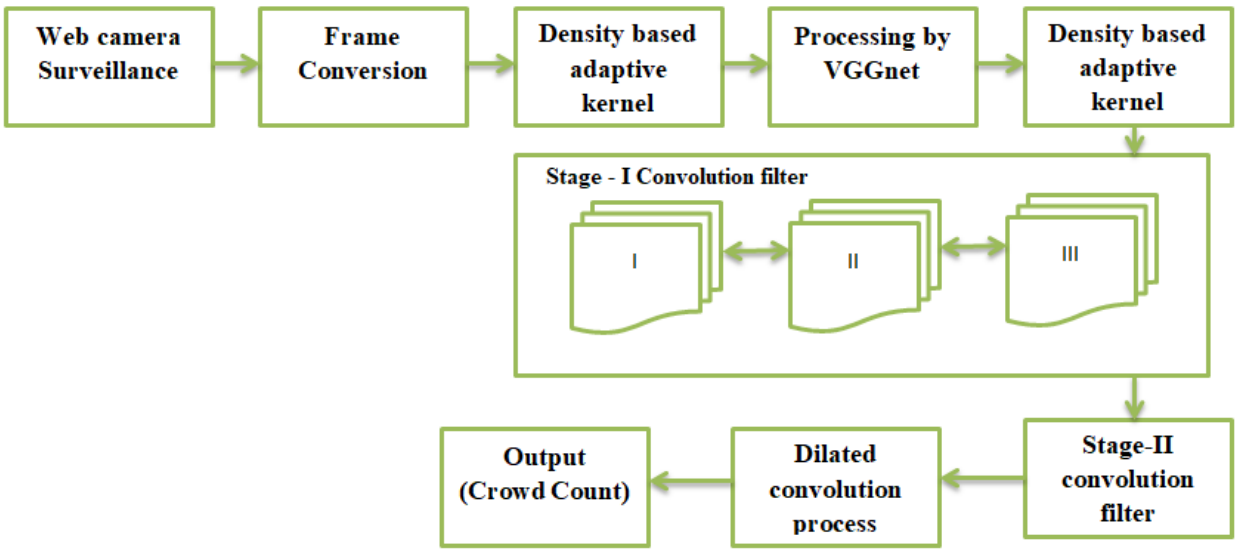

Figure 3. Proposed Architecture

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

4.1.1 Density Estimation

Density estimation is used to get an estimate by utilizing data on a probability density function that is unobservable. By using spatial information, this method has enabled a novel way to solve the issue of occlusion and clutter by utilizing the density estimation [21].

4.1.2 Crowd Net

"CrowdNet" uses 13 of its convolution layers, which uses it to produce a density map and then adds one 1x1 convolution layer as an output to increase the overall density of the images.

Step 1

On the other hand, multi-scale uses VGG 16, which is a classifier for density levels to label the images and feed them through one of its columns. VGG 16 is still the primary component but its impact on network accuracy is unnoticeable since it is used in almost all the publications. The training set loss is calculated as follows;

$$L(\theta) = \frac{1}{2M}\sum_{i=1}^{M}\left\|output(x_i;\theta - y_i^{GT})\right\|_2^2$$

Step 2

While Google utilizes VGG 16's convolution layers with the classification module not included, Microsoft uses VGG 16's convolution layers by removing the classification (i.e., fully connected layers). A picture with an input size of $1/8^{th}$ of its output size would be created. The geometric adaptive kernel can be computed as follows;

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

$$F(y) = \sum_{i=0}^{M} Z(y - y_i) * G_T(y)$$

Where, y_i is final object, $G_T(y)$is targeted final ground truth factors.

4.1.3 Dilated Process

This process helps to preserve resolution and produce high-quality density maps from the output. This is the output of the dilated convolution network, which is sent to the back-end and it is defined as follows;

$$y(x, y) = \sum_{i=0}^{x} \sum_{j=1}^{y} output(x + r * i, y + r * j)$$

$y(x, y)$ is the output dilated convolution, where r is dilated rate.

4.1.4 Spatial Information

Local characteristics such as landmarks are linearly mapped to the estimated-density (ED) maps, which yield spatial information. Since, image density may be calculated for any area of a picture, individual objects can no longer be located and detected.

4.2 Crowd Analysis

In cutting-plane optimization, risk-based quadratic cost functions are used to solve convex optimization problems. Despite the use of handcrafted characteristics, other types of CNN-based algorithms (including crowd analysis, motion analysis, and the 3D construction of body parts) have also been proposed for crowdsourcing [22]. The benchmark dataset is utilized by including 1,000

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

pictures. This has been suggested to be used in finding accurate counting in extremely crowded settings by using geometric adaptable kernels in 4.1.2 section and step 2.

5. Results & Discussion

Here, this section show some of the algorithm's outputs on the testing set of pictures that also includes the count. Figure 4 shows count results obtained by various algorithms.

Figure 4. Results obtained by various algorithms

It is calculated by using an adaptive kernel function after the framed picture that has been taken from the web camera surveillance video and stored in memory [23]. The headcount is

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

calculated from the input image by using the suggested framework, and the result is shown in the image that is displayed. Table 1 shows the comparison of computed metrics.

Table 1. Comparison of computed metrics

S.No	Model	Accuracy	MAE	MSE	Computation time
1	Image Object detection	89.34%	21.6	2.3	High
2	Pre-trained CNN	80.98%	26.4	1.75	Moderate
3	Proposed LW-CNN Approach	95.34%	10.32	0.81	Very small

When compared to previous methods, the suggested approach has accurately counted the number of heads in the crowd image while requiring the shortest amount of processing time. Aside from that, the suggested method has identified the heads in a variety of different rotations that the heads might take [24]. On the other hand, another conventional method was unable to identify the changes in the headcount that were present however the ground truth count is near to the proposed algorithm. Figure 5 shows the overall performance of proposed framework.

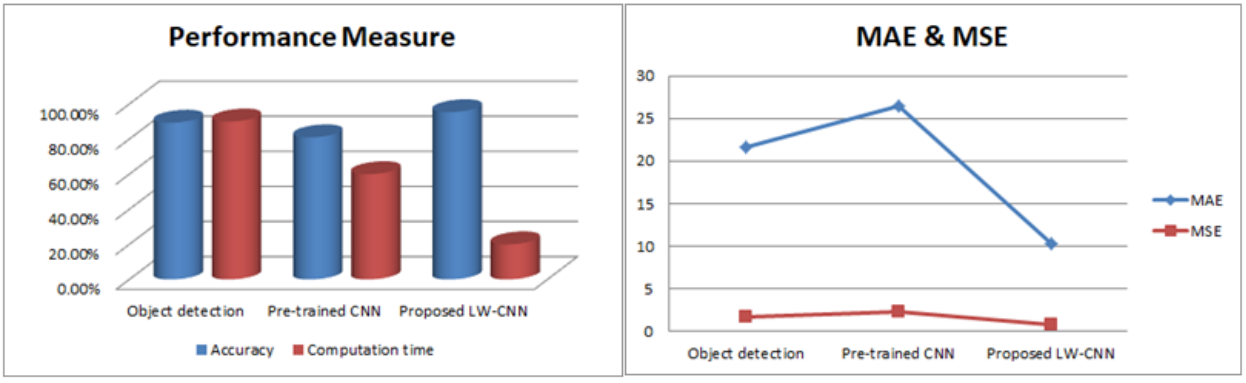

Figure 5. Overall Performance of Proposed Framework

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

The density or crowd has shown in the figure 4, where the network does well with dense images. Sparse pictures rely heavily on the dataset benchmark to perform well. In sparsely attended areas, the visibility of people's heads is enhanced [25]. The Gaussian kernel is initialized to the average human head size rather than utilizing a geometry adaptive kernel for ground truth production. These dataset images must be manually counted in order to get an exact count. As a result, the actual count is not completely accurate due to human error. The output of the crowd-count approach must be used to demonstrate that the estimated findings are appropriate. To measure the accuracy, perform these steps:

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

The obtained results are evaluated through our network on the basis of performance metrics, such as MAE (Mean Absolute Error) and MSE (Mean Square Error) by the equation,

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$$

6. Conclusion

Thus, the proposed algorithm has been examined with various benchmark datasets and proved superior to other pre-trained algorithms. Based on the multi-scale function in LW-CNN, the crowd density is computed very accurately. The graph and table have shown the computation time and performance of the proposed algorithm. The ultimate goal of the proposed research work is to get robust assessments by using much better and more advanced machine learning-based methods for crowd counting. Studies are required to address the difficulties such as shadows and

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

the lack of a clear 'leave' lookalike in the crowd count method. This project has an enormous potential and it may one day be carried out through satellite imagery. The concept may be modified to track the movement of people, which might assist in controlling mass gatherings like riots, protests, and other public disturbances. Also, a density map for automobiles may be utilized for real-time traffic monitoring by allowing vehicles to operate within a certain range.

References

[1] 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.

[2] Kaur, Preetjot, and Roopali Garg. "Towards Convolution Neural Networks (CNNs): A Brief Overview of AI and Deep Learning." Inventive Communication and Computational Technologies (2020): 399-407.

[3] 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.

[4] Sunitha, P. J., and K. R. Joy. "Deep CNN-Based Fire Alert System in Video Surveillance Networks." In Computational Vision and Bio-Inspired Computing, pp. 599-615. Springer, Singapore, 2021.

[5] Chen, Joy Iong Zong, and P. Hengjinda. "Early Prediction of Coronary Artery Disease (CAD) by Machine Learning Method-A Comparative Study." Journal of Artificial Intelligence 3, no. 01 (2021): 17-33.

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

[6] Smys, S., and Jennifer S. Raj. "Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study." Journal of trends in Computer Science and Smart technology (TCSST) 3, no. 01 (2021): 24-39.

[7] Murugeswari, P., and S. Vijayalakshmi. "A New Method of Interval Type-2 Fuzzy-Based CNN for Image Classification." In Computational Vision and Bio-Inspired Computing, pp. 733-746. Springer, Singapore, 2021.

[8] Bindhu, V., and G. Ranganathan. "Hyperspectral Image Processing in Internet of Things model using Clustering Algorithm." Journal of ISMAC 3, no. 02 (2021): 163-175.

[9] Chen, Joy Iong Zong, and Joy Iong Zong. "Automatic Vehicle License Plate Detection using K-Means Clustering Algorithm and CNN." Journal of Electrical Engineering and Automation 3, no. 1 (2021): 15-23.

[10] Ahuja, Komal R., and Nadir N. Charniya. "Design of Near Optimal Convolutional Neural Network Based Crowd Density Classifier." In International Conference on Innovative Data Communication Technologies and Application, pp. 204-212. Springer, Cham, 2019.

[11] Adam, Edriss Eisa Babikir. "Deep Learning based NLP Techniques In Text to Speech Synthesis for Communication Recognition." Journal of Soft Computing Paradigm (JSCP) 2, no. 04 (2020): 209-215.

[12] Quitian, Oscar Iván Torralba, Jenny Paola Lis-Gutiérrez, and Amelec Viloria. "Supervised and unsupervised learning applied to crowdfunding." In International Conference on Computational Vision and Bio Inspired Computing, pp. 90-97. Springer, Cham, 2019.

[13] 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.

[14] Shraddha S., Gulshan Pathak, Simran Panchal, and Monika Patil. "RFID-Based Railway Crowd Prediction and Revenue Analysis." In Data Intelligence and Cognitive Informatics, pp. 97-109. Springer, Singapore, 2021.

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

[15] 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.

[16] Lei, Y.; Liu, Y.; Zhang, P.; Liu, L. Towards using count-level weak supervision for crowd counting. Pattern Recognit. 2021, 109, 107616.

[17] Tong, M.; Fan, L.; Nan, H.; Zhao, Y. Smart Camera Aware Crowd Counting via Multiple Task Fractional Stride Deep Learning. Sensors 2019, 19, 1346.

[18] Lempitsky, V.; Zisserman, A. Learning to count objects in images. In Proceedings of the Neural Information Processing Systems, Hyatt Regency, Vancouver, BC, Canada, 6–11 December 2010; pp. 1324–1332.

[19] Sang, J.; Wu, W.; Luo, H.; Xiang, H.; Zhang, Q.; Hu, H.; Xia, X. Improved crowd counting method based on scale-adaptive convolutional neural network. IEEE Access 2019, 1, 24411–24419.

[20] Zhang, Y.; Zhou, D.; Chen, S.; Gao, S.; Yi, M. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 589–597.

[21] 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.

[22] Kang, D.; Ma, Z.; Chan, A.B. Beyond counting: Comparisons of density maps for crowd analysis tasks—Counting, detection, and tracking. IEEE Trans. Circuits Syst. Video Technol. 2018, 29, 1408–1422.

[23] Walach, E.;Wolf, L. Learning to count with cnn boosting. In European Conference on Computer Vision; Springer: Berlin/Heidelberg, Germany, 2016; pp. 660–676.

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021

Journal of Innovative Image Processing (JIIP) (2021)
Vol.03/ No. 03
Pages: 208-222
https://www.irojournals.com/iroiip/
DOI: https://doi.org/10.36548/jiip.2021.3.004

[24] Kumagai, S.; Hotta, K.; Kurita, T. Mixture of counting cnns: Adaptive integration of cnns specialized to specific appearance for crowd counting. arXiv 2017, arXiv:1703.09393.

[25] Marsden, M.; McGuinness, K.; Little, S.; O'Connor, N.E. Fully convolutional crowd counting on highly congested scenes. arXiv 2016, arXiv:1612.00220.

Author's biography

B. Vivekanadam is a senior lecturer in the Department of Computer Science and Multimedia at Lincoln University College, in Malaysia. His major area of research are machine learning, neural network algorithms, image processing, video and signal processing, cloud computing, deep learning, artificial intelligence, object recognition, complex feature extraction and vision graphics.

ISSN: 2582-4252 (online)
Submitted: 29.06.2021
Revised: 21.08.2021
Accepted: 28.08.2021
Published: 08.09.2021