Integrated Video Based Crowdedness Forecasting Framework with a Review of Crowd Counting Models

L.B.I.P. Thilakasiri, D.M.P.M. Alwis, R.T. Nanayakkara,
G.M.R.I. Godaliyadda, M.P.B. Ekanayake, H.M.V.R. Herath, J.B. Ekanayake

Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Peradeniya

***Abstract*—Crowd counting and forecasting is an important problem amidst Covid 19 circumstances. A unified system to automate crowd monitoring, collect data about crowdedness and predict future crowds is presented in this paper. An evaluation of existing state-of-the-art crowd counting algorithms on a novel dataset is conducted in the first part of the paper, which demonstrates the shortcomings of these algorithms. Several novel algorithms, including a densely connected neural network, convolutional neural network, and a long short term memory based recurrent neural network, for predicting crowd counts in the near and distant future are presented afterwards in the second half of the paper.**

***Index Terms*—Crowd counting, Crowd population forecasting, Evaluation**

I. Introduction

Crowd counting and forecasting using computer vision techniques have become important in recent years for a wide variety of applications such as smart transportation systems [1] [2], video surveillance and it has become even more important in the recent after COVID-19 spread. Crowd counting and having datasets about the population information can help in estimating the scale of the crowd at social events and all sorts of public spaces which can be used for better controlling of crowds and better division of resources overall. This can help the authorities in the process of taking necessary steps to control the spread of the virus and also to minimize resurgences of coronavirus cases.

The objective of this paper was to propose a unified system to automate crowd monitoring, collect data about the crowd count, and predict the crowd count in both the near and distant future. To accomplish this, Closed Circuit Television Video (CCTV) footage was used, as they are a common sight around the world and are highly suited for observing crowd patterns [3]. The first step was to obtain the crowd count, i.e. the number of people present in an area covered by a CCTV camera, by using deep learning based models. Several such models, which have been trained on large datasets, were evaluated in this study using a dataset of our own.

The second half of this paper focuses on forecasting the crowd present at a given location. This was done for two cases, predicting the crowdedness at any given date and time and predicting the crowd variation in the next 15 hours. A Dense Neural Network (DNN) and a Convolutional Neural Network (CNN) were discussed and compared for the first case. For the second case a Long Short-Term Memory (LSTM) network was used, and it was compared to a baseline model.

II. Related Work

A. Crowd Counting

There have been various approaches for crowd counting such as object-level crowd counting [4] [5], regression-based crowd counting [6], pixel-based crowd counting [7] [8] [9], density contribution probability estimation [10]. Most of the preceding crowd counting methods were mainly based on the detection of people's bodies and faces [11], [12] and object detectors were sometimes used for such purposes and other attempts for object identification, enhancement under dynamic backgrounds had also been made [13]. But these types of detectors failed to detect the crowds accurately due to occlusion and the many different arrangements of the origins of images. Furthermore, bounding-box annotations were needed to train these models, which was very labour-intensive. In order to avoid this, a direct count regression approach was introduced. Here, a regression model was trained to learn a direct mapping from image features to crowd count. In most of the crowd counting datasets, each person is annotated by a pixel at the center of the head. Therefore, the pixel level point supervisions can't be fully utilized by the direct count regression-based models. To overcome this issue, the density map approach was proposed by Lempitsky and Zisserman [14]. Here, the pixel-level annotations are transformed into a density map using a Gaussian kernel. In the present, density estimation techniques have been utilized to surpass the more traditional approaches based on classical regressors using CNN based models. A CNN with the two target outputs crowd count and crowd density maps, was proposed by Zhang et al. [7]. Further developments using CNNs were proposed in [8] and [9] using MCNN (Multi-Column Convolutional Neural Network) and Switching Convolutional Neural Network (SCNN) respectively. Complex loss functions were used in these models and the performance was determined by the quality of the generated ground-truth density maps.

In the recent, better crowd counting models & loss functions [10] [15] have been introduced. As an example, to minimize the complexity of the loss function and improve the performance, a loss function that constructs a density contribution probability model [10] using the point annotations was introduced. Therefore, more definitive supervision on the expected value of the count at each annotated point was acquired. But, if these models were to be used for crowd counting at any place, these models will be predicting counts for images from different distributions than what they have been trained on. Therefore, the performance of such models against a new dataset is tested in section III to validate the use of these models in real world applications.

B. Crowd Count Forecasting

Crowd forecasting is a frequently addressed problem in the field of computer vision and social computing. One approach for forecasting crowd behaviour has been to use big public data for prediction [16]. This approach analyses content published on web sources including social media to make predictions about crowds at public events. Another approach taken has been to use location based social networks to predict the number of visitors to a set of given locations [17] and attempts to analyze & cluster motion patterns in videos [18] has also been made. Mobile crowd-sensing has also become increasingly popular with the increased use of smartphones and this has been done by using either GPS data or mobile location data of users [19]. Another very popular approach to mobile crowd-sensing is to use the number of public WiFi users as a measure of the crowd present at a given location, and use this data in an LSTM to predict future variations of the crowd [20]. The third approach to forecasting crowd density is to use CCTV surveillance footage, to obtain crowd data [21]. All of these work focus on either forecasting crowd presence in the near future, or predicting crowd count at a given date, but a framework to combine these crowd counting methods to create datasets on the crowd population & forecasting models for both the near & distant future predictions have not been proposed. Datasets can be created by converting CCTV footage into images with an appropriate sampling rate along the time axis and these images can be used to create data logs about the human occupancy at any location to be used for forecasting as discussed in section IV. Such attempts [22] [23] to create datasets by stacking images along the frequency axis have also been made. Therefore, an integrated framework that incorporates these sections is introduced in this paper.

III. Experiments - Crowd Counting Model Evaluation

A. Evaluated Models

Bayesian Loss for Crowd Count Estimation with Point Supervision [10]. Here, the VGG19 [24] architecture except for the last max-pooling and fully connected layers, has been utilized as the backbone of the model. The output of the backbone has been fed to a regression header with 3 convolutional layers. The Bayesian loss function with background pixel modeling which acquires more authentic supervision on the expected value of the crowd count at each annotated point has been used in this model.

SS-DCNet [15]. An encoder and a decoder, based on VGG16 [24] and UNet [25] respectively have been used in this model to address the open set and closed set problem natures of the crowd counting problem. A supervised approach has been taken by this model by dividing the image and taking counts in those sub-regions for better generalization of the problem.

B. Evaluation Metrics

The metrics, Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE), given by equations 1 and 2, were used in the evaluation stage of chosen crowd counting models [10], [15].

$$MAE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} |y_{pred(i)} - y_{true(i)}| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} |y_{pred(i)} - y_{true(i)}|^2} \tag{2}$$

and n_{test}, $y_{true(i)}$ & $y_{pred(i)}$ represent the number of datapoints in the test set, predicted value of the i^{th} test item & the actual/ground truth value of the i^{th} test item respectively.

C. Datasets

In this section, the models from SS-DCNet [15] & BAYESIAN+ [10] that were previously trained on ShanghaiTech [26] and UCF_QNRF [27] were evaluated using the VET_Hospital_Dataset.

ShanghaiTech [26] consists of two parts, A and B. Part A (SH_A) has 300 images for training and 182 images for testing and Part B (SH_B) has 400 images for training and 316 images for testing.

UCF_QNRF [27] consists of high quality 1,535 jpeg images collected from various sources. The training and the test sets consist of 1201 and 334 images respectively and these images cover over 1.25 million point annotations.

VET_Hospital_Dataset was created by us for the evaluation of the crowd counting models listed above, using footage collected from a single CCTV camera from the Veterinary Teaching Hospital, Faculty of Veterinary Medicine & Animal Science, University of Peradeniya. The dataset consists of 1570 images of (2304×1296) resolution with a total of 16718 annotations. The images were collected with 30-second intervals from 0700 hours to 1800 hours and all the images were collected from the same camera angle & orientation. Both sparse & dense crowd scenes and a small portion of images of the empty premises are also included in the dataset. A majority of these images contain dogs & this helped to evaluate how the models handle animals during crowd counting. The faces of individuals in Fig. 1 were censored to maintain anonymity.

The dataset was collected amidst post-covid circumstances. Thus, all the people in these images were wearing masks, face shields and even Personal Protective Equipment (PPE) in some occasions as shown in Fig. (1.i.a), (1.ii.a), (1.iii.a) & (1.iv.a). Such footage was useful in evaluating the performance of the crowd counting models in detecting people against the new post-Covid-19 health regulations.

D. Model Evaluation

The publicly available models & weights of both SS_DCNet [1] [15] and BAYESIAN+ [2] [10] were used in the evaluation process. The experimental results of all six models against the VET_Hospital_Dataset are listed in Table I where the best overall result is highlighted in **bold**. The published results of the models for the datasets that they were initially trained on are shown in Table II, where the best for each dataset is highlighted in **bold**.

Comparing the results in Table I with the results in Table II, all these models except the BAYESIAN+ model which had been trained on the SH_A dataset, outperformed other models in generalizing to the situation and recognizing the crowds against the VET_Hospital_Dataset.

Despite the low MAE and RMSE values, it was observed that the models consistently miscalculated some aspects of the images in the VET_Hospital_Dataset. Those points are discussed with the aid of density maps generated for five chosen images shown in the Fig. 1 as (i.a), (ii.a), (iii.a), (iv.a) & (v.a).

Commending results were not achieved by the publicly available BAYESIAN+ model trained on SH_A. As illustrated by the images in column (b) of Fig. 1, the entropy map generated by this model has noise all over the entropy map which has resulted in an overestimation of the crowd count.

People wearing masks/ face shields have been identified by the models successfully as shown best by the density maps of (1.ii.a). Workers wearing white color PPE covering their heads can be seen in sub-figures (i.a), (ii.a), (iii.a), (iv.a) of Fig. 1, and the BAYESIAN+ model & SS-DC model trained on the SH_B dataset have failed to identify all occurrences of such workers as shown in the images in columns (c) and (f) of Fig. 1. However, BAYESIAN+ model & SS-DC model trained on the the SH_A dataset have detected those workers in some occasions [(1.i.b), (1.i.e), (1.iii.e)] and some have missed out [(1.ii.b), (1.iii.b), (1.iv.b), (1.ii.e), (1.iv.e)]. Out of the three datasets, the models which had been trained on UCF_QNRF have detected many of such workers. It can be seen that training on a dataset [27] with a huge variety and a large number of annotations has made the models better at recognizing people with unusual clothing than those that were trained on SH_A and SH_B. This is also evident as the SS-DC [15] model that has been trained on the UCF_QNRF [27] has achieved the best overall MAE & RMSE both as shown in Table I.

[1]https://github.com/xhp-hust-2018-2011/SS-DCNet
[2]https://github.com/ZhihengCV/Bayesian-Crowd-Counting

TABLE I: Comparison of the performance of evaluated models against the VET_Hospital_Dataset

Initially trained on dataset	SH_A		SH_B		UCF-QNRF	
Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE
SS-DC	2.930	5.492	3.937	5.241	**2.425**	**3.197**
BAYESIAN+	68.853	69.764	3.039	3.879	4.362	5.335

TABLE II: Comparison of the performance against their respective trained datasets

Datasets	SH_A		SH_B		UCF-QNRF	
Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE
SS-DC	**56.1**	**88.9**	**6.6**	**10.8**	**81.9**	**143.8**
BAYESIAN+	62.8	101.8	7.7	12.7	88.7	154.8

When the dogs present in the range of [0-40] of the x-axis and [60-100] of the y-axis in (i.a), (ii.a), (iii.a), (iv.a) of Fig. 1 are considered, every model has identified and added the dogs into the crowd count at least on one occasion. Especially, the dogs have been detected by the BAYESIAN+ model trained on UCF_QNRF at every instant. Considering the dark fur of the dogs in the images, there is a high chance of the dogs being identified by these models as they have been trained to identify the heads of people. For example, the SS-DC model trained on SH_B has resulted in two false-positive spots for the dog in (1.i.a).

Clusters have been identified in most of the images and a count more than the actual amount has been predicted from those regions. In the Fig. (1.iv.e) the SS-DC model trained on SH_A has identified a cluster in the mid-section of the image. The model had predicted 82.994 as the output of this image while the ground truth was only 17, making an error of 65.994. The same miscalculations can be seen in images with people far away from the camera. Despite these errors, the error metrics have achieved low values as the overall score for the entire dataset due to this compensation of missing out people in some places and predicting a larger value than the actual count in the cluster areas.

Fig. (1.v.a) is an image of the empty room and the density maps generated for this image as shown the imperfections of these models clearly. A majority of these models have identified the cleaning equipment & sink, ceiling/fans and the dustbin respectively in the x-axis & y-axis ranges of [(80-100), (40-60)], [(60-80), (0-20)], [(40-60), (80-100)] of Fig. (1.v.a). The SS-DC model trained on SH_A was the only model that has not identified the sink or the fans, but it has identified the aluminium block present in [(0-20),(0-40)] Fig. (1.v.a) as illustrated by the maps in the column (e). Therefore, it was evident that these models tended to miscalculate common objects as humans quite often.

IV. EXPERIMENTS - CROWD COUNT FORECASTING

In this section, a novel DNN & CNN, mainly for distant future predictions, and a LSTM network for near future predictions are proposed and tested for crowd count forecasting.

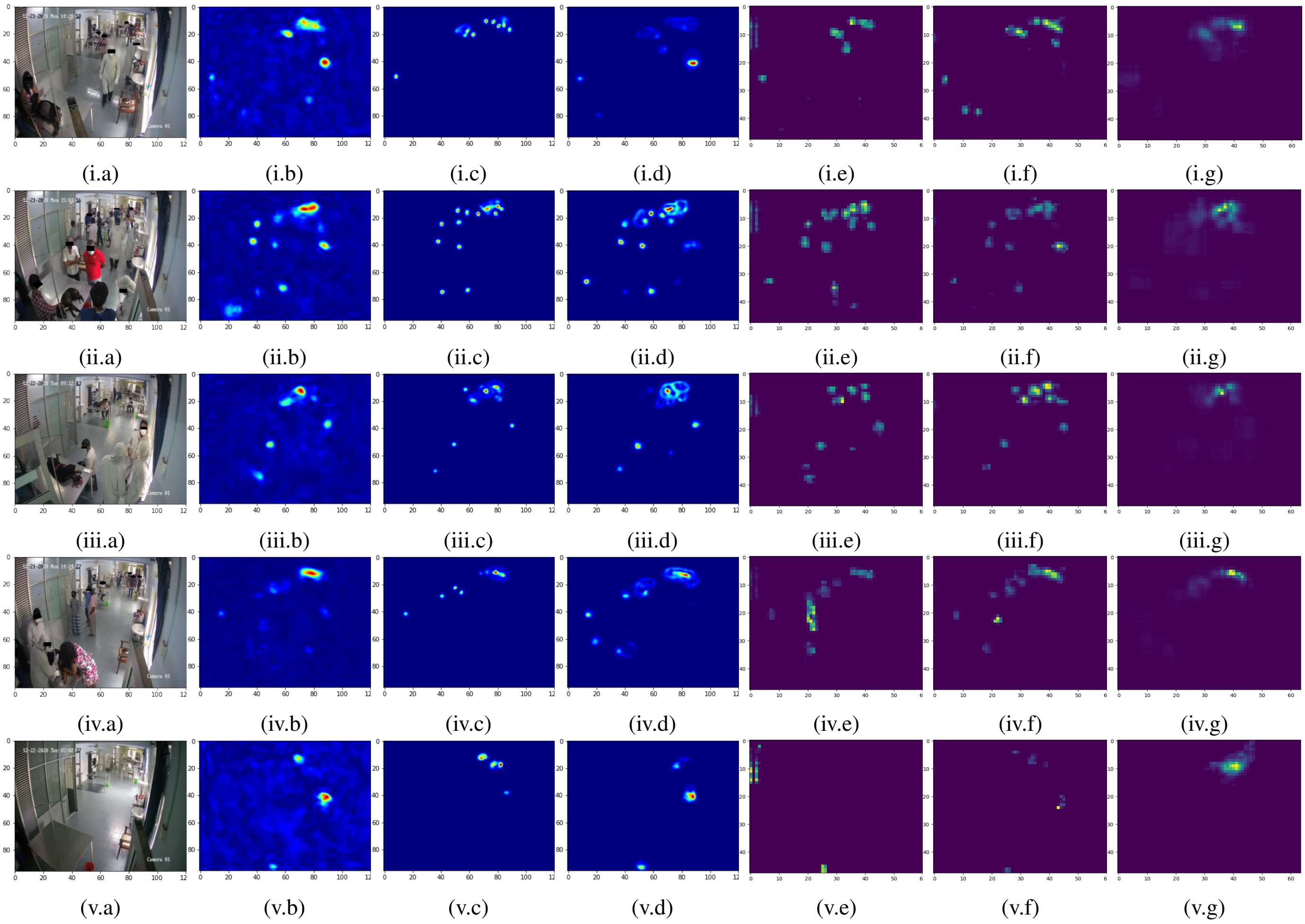

Fig. 1: Density maps generated for five chosen images from VET_Hospital_Dataset. Warmer the color, higher the density. The images in the row (i) in Fig. 1 correspond to the results obtained for the image (i.a). The results of the BAYESIAN+ models trained on SH_A, SH_B, UCF_QNRF are illustrated by the images in the columns (b), (c), (d) respectively and the results of the SS-DC models trained on SH_A, SH_B, UCF_QNRF are illustrated by the images in the rows (e), (f), (g) respectively. The same pattern is applied for the images (ii.a), (iii.a), (iv.a), (v.a).

A. Evaluation Metrics

For the Evaluation of the DNN, CNN & LSTM Networks, Mean Absolute Error (MAE) given in the equation (1) and Root Mean Squared Error (RMSE) in the equation (2) were used.

B. Datasets

Crowdedness_at_the_Campus_Gym [28] was used for the evaluation of the forecasting models.

Crowdedness_at_the_Campus_Gym [28] dataset consists of the data about the number of people that were present inside a campus gym over the course of 19 months and the data had been collected in rough intervals of 10 minutes. The maximum crowd count was 145 and the mean crowd count of the preprocessed dataset was 38.3192.

C. Data Preprocessing for DNN and CNN

The focus of this section was to predict the crowd population during working hours and to get a better idea about its crowd population patterns. Therefore, the data collected during the night from 2300 hours to 0800 hours were removed from the dataset used for the training and testing. Another reason for this removal was the tendency of the networks to always predict very low values since the majority of the crowd count recorded at night had been either zero or one. For the DNN, the day_of_week label was replaced with additional labels for each of the hours, days & months, and then the dataset was normalized. The preprocessed dataset had 41 features. This inclusion of extra features gave an advantage with the results and the new labels were: number of people, timestamp, day of week (categorical), hour (categorical), month (categorical), weekend (binary), holiday (binary), temperature, start of semester (binary), during semester (binary).

D. Data Preprocessing for LSTM

A similar preprocessing technique was used for constructing the sequential data for training the LSTM. Data collected during the night from 2300 hours to 0800 hours was removed to balance the dataset. The dataset contained samples taken at roughly 10-minute intervals. The time was converted to a sine cosine pair to capture its cyclic nature. The final labels were: number of people, year, month, day, time (sin & cos format), day of week, temperature. Sequences consisting of 15 hours of inputs and 15 hours of labels were constructed.

E. Novel DNN Architecture

Our proposed DNN consists of 12 layers. The input is a (1×41) vector. It is followed by two 128 node layers, two 256 node layers, two 512 node layers, two 256 node layers, and three 128 node layers before the output layer which outputs the expected crowd count. Activation functions and hyperparameters were chosen by experimentation. Exponential Linear Unit (ELU) was chosen as the activation function of the first hidden layer, last hidden layer, and the output layer while Rectified Linear Unit (ReLU) was chosen for the remaining nine hidden layers. RMSprop was chosen as the optimization function with a learning rate of 0.001 and a mini batch size of 64.

F. Novel CNN Architecture

The CNN architecture in Fig. 2 was used to forecast the crowd count. To extract features, eight one-dimensional convolutional layers with a filter size of 64 followed by a flatten layer were used. The extracted feature vector was fed into a neural network of 3 dense layers. All the layers utilized a Rectified linear unit (ReLU) activation function. Here, max-pooling layers were not used since the down-sampling effect of max-pooling layers tended to miss out on information and reduced the performance of the model.

The adam optimizer with a mini batch size of 32 gave the best outcome for the dataset, as extensive mini batch sizes and RMSprop lessened the performance of the model.

G. Novel LSTM Network

An LSTM network with several densely connected layers showed the best results in forecasting the crowd count. As shown in Fig. 3 the 8 input features were fed into an LSTM layer to generate 64 features. This was followed by another LSTM layer with 64 outputs which then fed into two consecutive 64 node densely connected layers, each utilizing a ReLU activation function. The output was obtained through another densely connected layer of size 54, which generated the expected crowd count variation for the next 15 hours. During the training process, the adam optimizer was used with a learning rate of 0.001 and a mini-batch size of 32.

H. Preformance comparison of the models

The outputs of the models were rounded off to the nearest integer to acquire the crowd count forecast.

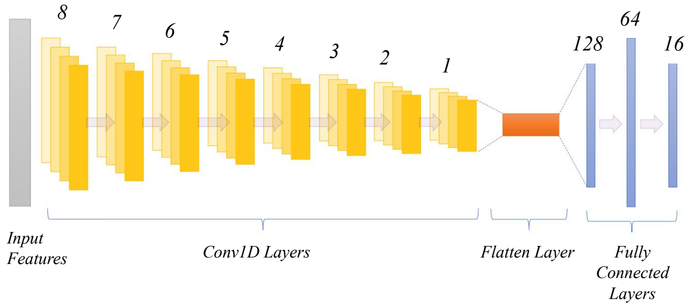

Fig. 2: Proposed CNN Architecture

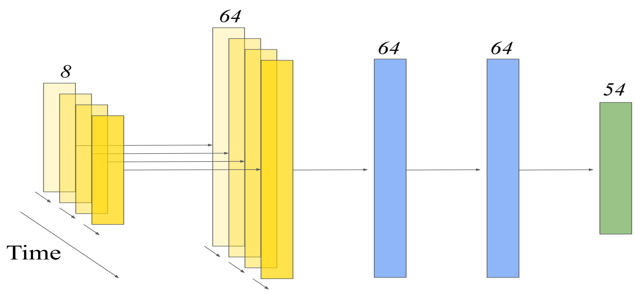

Fig. 3: Proposed LSTM Network Architecture

As depicted in Table III, it is possible to observe a better MAE and MSE for predictions with the CNN as opposed to the DNN. This is because CNNs are less sensitive to the amount of preprocessing done on the dataset compared to a DNN, as discussed by Jernelv et. al [29].

The LSTM was compared with a baseline model for predicting the near future crowd counts as a time series. The baseline model simply assumed that the crowd count for the next 15 hours was identical to the previous 15 hours, and gave a MAE of 24.31. In comparison, the LSTM was able to recognize and predict 15-hour variations in the data with a MAE of 8.50, which is considerably less.

V. Conclusion and Future Perspectives

The performance of crowd counting models and their shortcomings in crowd identification were discussed in the first part of this paper. It was clear that these models tend to mistake common objects as humans and training on bigger datasets containing images with more diverse scenes & more common day-to-day objects could result in better performances. The rest of the paper was focused on the types of models that can be used to create predictive models for both the near and distant

TABLE III: Comparison of the performance of Crowd Count Forecasting models with VET_Hospital_Dataset

Method	DNN	CNN	LSTM	Baseline
MAE	6.20	5.53	8.50	24.31
RMSE	8.88	7.57	11.31	28.48

future. Observing the results of the predictive models showed that the lack of information about special events and about the days the gym was closed has lessened the overall performance. As an example, when the premises were closed or filled with people because of a special event on an otherwise typical day, the training becomes less intuitive as the information about the closing of the gym or about the special event was not available. Therefore, including such information is essential for better results when creating such datasets in the future. As mentioned in the introduction, creating a unified model with (i) a crowd counting model, (ii) a format to log the information about the parameters that could affect the crowd count, (iii) creating models as discussed in the Section IV to predict the crowd population in near and distant future can be beneficial in many areas. Gathering of census and statistics, crowd pattern information collection & analyzing, and crowd controlling in a pandemic situation are such examples. Such a system can be further enhanced to collect information about other types of objects that can be identified using density mapping such as plantations, animals, vehicle population & occurrence patterns as well.

ACKNOWLEDGMENT

Authors are grateful to Dr. R. A. D. Eranda Indrajith, Head of the Department, Department of Veterinary Clinical Sciences, Faculty of Veterinary Medicine & Animal Science, University of Peradeniya for providing CCTV footage. In addition, authors would like to thank Mr. Gihan Jayatilake for helpful suggestions.

REFERENCES

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.

[2] X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang, and L. Yang, "Big data for social transportation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 3, pp. 620–630, 2015.

[3] W. Xia, J. Zhang, and U. Kruger, "Semisupervised pedestrian counting with temporal and spatial consistencies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1705–1715, 2014.

[4] D. B. Sam, N. N. Sajjan, R. V. Babu, and M. Srinivasan, "Divide and grow: Capturing huge diversity in crowd images with incrementally growing cnn," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3618–3626.

[5] W. Ge and R. T. Collins, "Marked point processes for crowd counting," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 2913–2920.

[6] C. Wang, H. Zhang, L. Yang, S. Liu, and X. Cao, "Deep people counting in extremely dense crowds," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1299–1302.

[7] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 833–841.

[8] D. Ciregan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3642–3649.

[9] D. Babu Sam, S. Surya, and R. Venkatesh Babu, "Switching convolutional neural network for crowd counting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5744–5752.

[10] Z. Ma, X. Wei, X. Hong, and Y. Gong, "Bayesian loss for crowd count estimation with point supervision," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6142–6151.

[11] B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1. IEEE, 2005, pp. 90–97.

[12] Z. Lin and L. S. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 4, pp. 604–618, 2010.

[13] W. Fernando, H. Herath, P. Perera, M. Ekanayake, G. Godaliyadda, and J. Wijayakulasooriya, "Object identification, enhancement and tracking under dynamic background conditions," in *7th International Conference on Information and Automation for Sustainability*. IEEE, 2014, pp. 1–6.

[14] V. Lempitsky and A. Zisserman, "Learning to count objects in images," *Advances in neural information processing systems*, vol. 23, pp. 1324–1332, 2010.

[15] H. Xiong, H. Lu, C. Liu, L. Liu, C. Shen, and Z. Cao, "From open set to closed set: Supervised spatial divide-and-conquer for object counting," *arXiv preprint arXiv:2001.01886*, 2020.

[16] N. Kallus, "Predicting crowd behavior with big public data," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 625–630.

[17] M. A. Saleem, F. S. Da Costa, P. Dolog, P. Karras, T. B. Pedersen, and T. Calders, "Predicting visitors using location-based social networks," in *2018 19th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2018, pp. 245–250.

[18] R. Rupasinghe, S. Senanayake, D. Padmasiri, M. Ekanayake, G. Godaliyadda, and J. Wijayakulasooriya, "Modes of clustering for motion pattern analysis in video surveillance," in *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*. IEEE, 2016, pp. 1–6.

[19] G. Cardone, A. Cirri, A. Corradi, L. Foschini, R. Ianniello, and R. Montanari, "Crowdsensing in urban areas for city-scale mass gathering management: Geofencing and activity recognition," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4185–4195, 2014.

[20] U. Singh, J.-F. Determe, F. Horlin, and P. De Doncker, "Crowd forecasting based on wifi sensors and lstm neural networks," *IEEE transactions on instrumentation and measurement*, vol. 69, no. 9, pp. 6121–6131, 2020.

[21] C. Lijun and H. Kaiqi, "Video-based crowd density estimation and prediction system for wide-area surveillance," *China Communications*, vol. 10, no. 5, pp. 79–88, 2013.

[22] E. Ekanayake, E. Ekanayake, A. Rathnayake, S. Vithana, H. Herath, G. Godaliyadda, and M. Ekanayake, "A semi-supervised algorithm to map major vegetation zones using satellite hyperspectral data," in *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2018, pp. 1–5.

[23] E. Ekanayake, S. Vithana, E. Ekanayake, A. Rathnayake, A. Abeysekara, T. Oorloff, H. Herath, G. Godaliyadda, M. Ekanayake, and A. Senaratne, "Mapping ilmenite deposit in pulmudai, sri lanka using a hyperspectral imaging-based surface mineral mapping method," *Journal of the National Science Foundation of Sri Lanka*, vol. 47, no. 3, pp. 271–284, 2019.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[26] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 589–597.

[27] H. Idrees, M. Tayyab, K. Athrey, D. Zhang, S. Al-Maadeed, N. Rajpoot, and M. Shah, "Composition loss for counting, density map estimation and localization in dense crowds," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 532–546.

[28] "Crowdedness at the campus gym," https://www.kaggle.com/nsrose7224/crowdedness-at-the-campus-gym.

[29] I. L. Jernelv, D. R. Hjelme, Y. Matsuura, and A. Aksnes, "Convolutional neural networks for classification and regression analysis of one-dimensional spectral data," *arXiv preprint arXiv:2005.07530*, 2020.