DEEP LEARNING FOR CROWD IMAGE CLASSIFICATION FOR IMAGES CAPTURED UNDER VARYING CLIMATIC AND LIGHTING CONDITION

Hemant T. Ingale
Research Scholar,
S.S.B.T.C.O.E&T, Bambhori,
Jalgaon, India.
hetui@rediffmail.com

Shekhar S. Suralkar
Professor, S.S.B.T.C.O.E&T,
Bambhori, Jalgaon, India.
shekhar_srs@rediffmail.com

Anil J. Patil
Professor, SGOI COE&FOM, Belhe,
Pune, India. anilj48@gmail.com

Abstract
Most of recent events have attracted a lot of attention towards importance of automatic crowd classification and management. COVID-19 is the most setback for the entire world. During these events proper breakout and public crowd management leads to the requirement of managing, counting, securing as well as tracking the crowd. But automatic analysis of the crowd is very challenging task because of varying climatic and lighting conditions, varying postures etc. During this paper we have developed PYTHON based system for automatic crowd images classification using Deep learning. This paper is the first attempt for automatic classification of crowd images. We have prepared the dataset of crowd classification consisting of three categories. The proposed methodology of crowd classification starts with preprocessing during which we have used median filtering for noise removal. Deep learning models are developed using 70% training images. The performance of the system is evaluated for various deep learning algorithms including one block VGG, two block VGG and three block VGG. We have also evaluated the performance of three block VGG using dropout. VGG16 transfer learning based crowd classification is developed using PYTHON. Using VGG16 transfer learning we achieved the accuracy of 69.44.% which is highest among all deep learning classification models during this study

Keywords: **deep learning, one block VGG, two block VGG, three blocks VGG, transfer learning**

1. INTRODUCTION

Deep learning plays major role in solving artificial intelligence problems related with speech recognition, object recognition, machine translation, moving object detection and tracking etc. Deep learning algorithms have been characterized by their success in discovering intricate structures in high-dimensional data. This makes it particularly useful for handling complex problems including natural language processing, vision etc. But key challenge during deep learning representations of complex high dimensional data is the requirement of large labelled data. But in many cases the labelled dataset is not available. Humans have to manually label it before. During this work we have first time prepared the dataset for crowd classification based on standard crowd counting datasets.

Crowd analysis and management involves the analysis of large group of people sharing some common area. Crowd analysis involves numerous tasks including crown counting, crowd density estimation, crowd tracking, crowd behavior analysis etc. Hence there are lots of applications of automatic crowd analysis and classification. It is useful in controlling the COVID19 spread. We can assure about the social distance between individuals in the crowd. Automatic crowd scene analysis helps in securing public places including cinema halls, malls, stores parks etc. It also helps in controlling the public events including sport championships, new year celebrations, carnivals etc. Crowd scene analysis involves extracting anomalous behaviors from huge number of people.

During this paper we have developed a system for automatic crowd classification using deep learning. During this work, we have focused:

- Preparing crowd density image classification dataset from multiple crowd dataset sources
- Evaluating the performance of various deep learning approaches in crowd classification.

The rest of this paper is organized as follows: Section 2 deals with review of related work in crowd analysis and crowd counting. Section 3 describes the dataset preparation and gathering. Section 4 presents implementation details about different deep learning models for crowd image classification. Section 5 deals with methodology for crowd classification. Performance of different deep learning models is been evaluated in Section 6 deals. Section 7 discuss the conclusion along with future scope of the work.

2. DEEP LEARNING FOR CROWD SCENE ANALYSIS

As discussed earlier, crowd scenes analysis plays important role in huma lives. Lot of research work has been carried out in crowd analysis. During this section we have reviewed different works carried out in the field of crowd scene analysis. The survey deals the main two aspects of crowd analysis: (1) crowd counting and (2) crowd action recognition.

2.1 CROWD COUNTING

Crowd counting deals with estimation of number of people in the region. In this section we have reviewed various methods for computing number of persons in the crowd.

Authors in [1] have proposed an end-to-end convolutional neural network (CNN) architecture for crowd counting. After taking whole image as its input the model outputs the count directly. Computations are shared over overlapping regions. Contextual

978-1-6654-9291-1/22/$31.00 ©2022 IEEE

information us used in prediction of both local and global count. Image is fed to pre trained CNN and a set of high level features is obtained. By mapping the features to local counting numbers, counting is performed. Results of the proposed method is challenging.

Authors in [2] have proposed a study about crowd counting. Crowd counting has numerous applications including health care, safety, disaster management etc. Authors have classified crowd counting techniques as supervised learning based and unsupervised learning based techniques. During the work authors have discussed traditional as well as CNN based techniques for crowd counting. Decrease in crowd density increases the calculation complexity of the crowd counting.

Authors in [3] have deal with the method of crowd counting using an automatic scale-adaptive approach. The proposed work is been divided into two tasks. The classification sub network followed by main network. The classification network is responsible for estimating scales of the crowd. The main network performs the actual density map estimation. Given the image as input to the classification network as well as main network, crowd scales are estimated. Based on estimated crowd scales, structure of the main network is adjusted to perform crowd counting task. Crowd spatial information mask is created in the main network. Finally crowd density maps and crowd counting results are obtained.

Most of the existing research works of crowd counting are based on regression models. These models are mapping the features to the corresponding class labels directly. Authors in [4] have proposed a conditional marked point process (CMPP)-based approach for counting the persons in the crowd image. Stochastic model is used here which is the estimation from training set. Shape and size of the shape in distribution is used for counting number of persons in the image. The performance is evaluated PETS2009 dataset.

Occlusion is very common problem in crowd counting. Authors in [5] have deal with this issue and proposed a new method for crowd counting by using a RGB plus depth camera. First step is to detect each head-shoulder of the passing person. Then each detected head shoulder is tracked and count using RGB information bidirectionally. The crowd counting method proposed here is fast and robust. By using this approach, we have built a practical system for robust and fast crowd counting. The method can also be applicable for real time applications of crowd counting.

Authors in [6] have introduces a weakly supervised crowd wise attention network. Accuracy and performance is increased in the proposed model due to weakly supervised crowd segmentation and more attention of spatial information. Motion based region growth is used for generating segmentation label. Spatial attention map is generated according to active crowd segmentation. This map is used for reweighing the appearance feature and density estimation is achieved.

Use of CNN in crowd counting is not possible in today's small scale crowd counting datasets. To deal with this problem authors in [7] have constructed a large scale crowd counting dataset. The dataset NWPU-Crowd consists of 5109 images along with 2,133,375 annotations. The dataset contains varying illuminations as compared with other datasets. Authors have developed a benchmark website for evaluating the performance of various algorithm on the dataset.

Controlling the crowd manually is difficult due to human error and time consuming. Authors in [8] have focused on L&T Smart World AI-based crowd management system. The system was developed during Kumbh Mela 2019 gathering in Prayagraj. There are many situations that go beyond human capability during this event. Data is gathered for providing as core for framework. Crowd count and management can be performed using deep learning

2.2 CROWD ACTION RECOGNITION

Crowd analysis involves recognition of different activities for the individual or crowd. It is very important for the safety of crowd. This section deals with review of different works carried out in crowd action recognition.

Authors in [9] have addressed the task of crowd action recognition in videos. In this paper, we investigate the task of human action detection in crowded videos. The proposed methodology requires no human segmentation or tracking techniques. Cluttered and crowded backgrounds are deal with the help of shape and motion templates. Features are refined using shape templates. Action training is performed by involving only the moving body parts. This helps in solving the partial occlusion problem. Performance is evaluated on the CMU dataset with encouraging results.

Action recognition in moving background is a challenging problem due to contamination of motion field by background motions. Authors in [10] have proposed hierarchical filtered motion (HFM) approach for recognition of action in crowded videos. Authors have used motion history images for representing motions. First interest points are detected with recent motion as two dimensional Harris corners. Then isolated unreliable or noisy motions are obtained by applying global spatial motion smoothing. A local motion field filter is applied at each interest point. Performance is evaluated on KTH dataset. Cross dataset crowd action recognition is also performed in which KTH dataset is used for training and MSR dataset is used for testing.

Authors in [11] have addressed the issue of human action detection in dense crowds in the form of locally consistent state. Similarity in the scale in the local neighborhoods is computed and its smooth variation over the image is considered. Scale and confidence priors are inferred based on scale and confidence using Markov random field. The proposed method is able to detect various combinations of body parts without the need of any annotations.

Authors in [12] have discussed the task of human action recognition in crowded videos. Human action is recognized using a novel mask based shape matching method. The method is not based on human detection and segmentation. During features extraction both shape and flow based features are combined. The performance of the proposed algorithm is evaluated on the CMU dataset.

Authors in [13] have addressed the issue of human action recognition in a crowded and cluttered environment. This task is achieved by using a holistic 4D scan of the cluttered scene. A new method is proposed for tracking people in 4D environment. Then a deep neural network is built for recognizing the action of tracked individual. Adaptive 3D convolution layer is also designed for improving the performance of the system.

Lot of research has been carried out in crowd motion segmentation. Authors in [14] haves used flow dynamics for modelling the motion of the crowd. Authors have proposed crowd motion segmentation approach according to streak flow. Streak flow deals with the dynamical property of crowd motion. Segmentation accuracy of the proposed method is better on the benchmark datasets.

Authors in [15] have concentrated on high density crowd scenes and proposed the approach for recognizing unexpected behavior in image sequence. Occlusions and low-resolution images are challenging in determining abnormal behavior in images. During the proposed model authors have used features based on motion information for abnormality detection in the crowd.

Crowd scenes are always extremely cluttered. Hence event detection form videos is a challenging task. Authors in [16] have designed a video content analysis task. Various MPEG-7 descriptors are used including kinetic energy of the crowd, motion detections histograms, motion directions are used in the beginning. SVM is then used for training and testing these descriptors.

3. CROWD CLASSIFICATION DATASET GATHERING AND PREPARATION

The classification dataset for crowd is prepared by using two datasets as below:

- JHU-CROWD++: Large-Scale Crowd Counting Dataset[17]
- ShanghaiTech[18]

3.1 JHU-CROWD++: LARGE-SCALE CROWD COUNTING DATASET

JHU-CROWD ++ is a large scale crowd dataset containing 4372 images with 1.51 million annotations. The dataset contains images which are captured under varying scenarios and environmental conditions. To be more specific the images in the dataset are including weather based degradations and illumination variations etc. Due to these the dataset is a challenging. Figure 1 shows some of the images from JHO CROWD dataset.

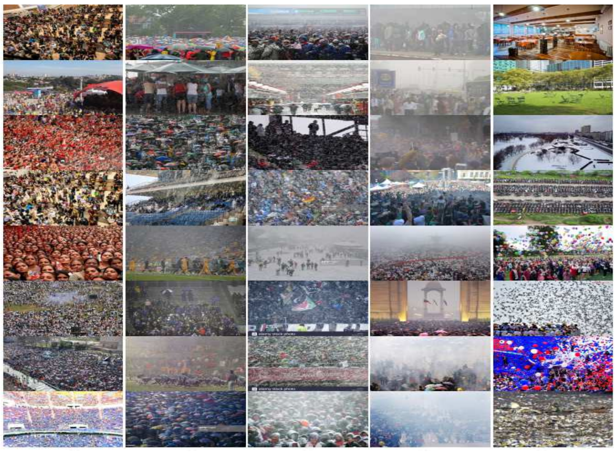


Fig. 1. Examples of the images in the JHU-CROWD++ dataset. (a) Overall (b) Rain (c) Snow (d) Haze (e) Distractors.

3.2 SHANGHAITECH DATASET:

ShanghaiTech dataset is a crowd counting dataset. It contains 1198 images with 330165 annotations. The dataset images are organized into two parts namely part A and part B. Images form part A are random images taken from the internet. Part B images are captured from street view.

3.3 CROWD CLASSIFICATION DATASET:

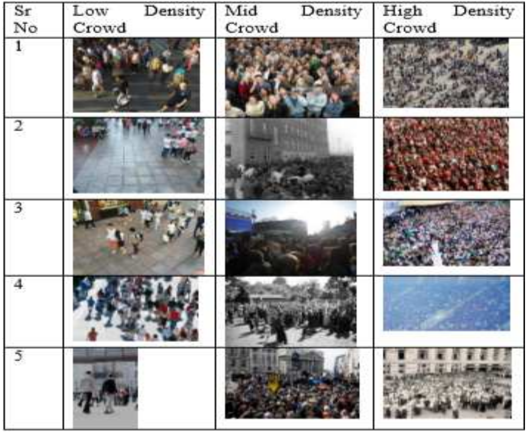
We have prepared the dataset of crowd classification by using both datasets of crowd counting. Images are divided into three categories according to number of persons in the image. Details of the crowd classification dataset are shown in table 1.

Table 1: Crowd Classification Dataset

Crowd Type	Crowd Count	Number of images
Low Density Crowd	0 to 100	2462
Mid Density Crowd	101 to 500	2317
High DensityCrowd	501 to 1000	790
Total		5569

Table 2 shows sample images from crowd classification dataset for various crowd category.

Table 2: Crowd Classification Dataset Images



Sr No	Low Density Crowd	Mid Density Crowd	High Density Crowd
1			
2			
3			
4			
5			

4. DEEP LEARNING FOR CROWD CLASSIFICATION

In this paper we have proposed deep learning based crowd classification. General architecture of VGG model consists of stacking of convolutional layers having 3x3 filters and max pooling layer. A block is formed using these layers. Repetition of these blocks can be done by increasing the depth of the network like 32,64,128,256 for the first four blocks of the model. Padding in the convolutional layers ensures the height and weight shapes of the output feature maps matches the input.

In this study we have explored this model on crowd classification dataset. We have evaluated the performance on 1,2 and 3 blocks. In each layer ReLU activation function and He weight initialization is used.

4.1 ONE BLOCK VGG MODEL

The one-block VGG model is composed of a single convolutional layer having 32 filters along with max pooling layer.

The dataset is loaded and split into train test data separately. The one block VGG model is fit using training data and evaluated. As depicted in figure 2 the model has achieved the training accuracy of 90% and evaluation accuracy of 71%. A model is run for 30 iterations. Figure shows the accuracy and loss of the model for both train and evaluation data.

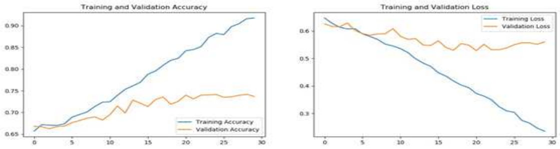


Figure 2: Performance of One Block VGG Model

4.2 TWO BLOCK VGG MODEL

The two-block VGG model is composed of two convolutional layer having 32 and 64 filters along with max pooling layer.

The dataset is loaded and split into train test data separately. The two block VGG model is fit using training data and evaluated. As depicted in figure 3 the model has achieved the training accuracy of 82.5%% and evaluation accuracy of 72.5%. A model is run for 30 iterations. Figure shows the accuracy and loss of the model for both train and evaluation data. As compared with one block VGG model, the accuracy is increased from 71% to 72.5% for evaluation data.

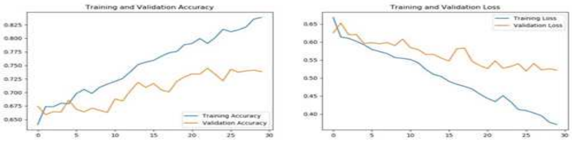


Figure 3: Performance of Two block VGG Model

4.3 THREE BLOCK VGG MODEL

The three-block VGG model extends the two block model by adding third block of 128 filters. The two-block VGG model is composed of three convolutional layer having 32, 64 and 128 filters along with max pooling layer.

The dataset is loaded and split into train test data separately. The three block VGG model is fit using training data and evaluated. As depicted in figure 4 the model has achieved the training accuracy of 80% and evaluation accuracy of 75%. A model is run for 30 iterations. Figure shows the accuracy and loss of the model for both train and evaluation data. As compared with one block VGG model, the accuracy is increased from 71% to 75% for evaluation data

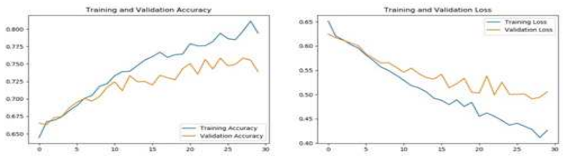


Figure 4: Performance of Three block VGG Model

During this study we have evaluated the performance of three VGG models for crowd classification. The performance of evaluation dataset are summarized below.

- One-block VGG : 71%
- Two block VGG: 72.5%
- Three-block VGG: 75%

Improvement in performance is observed as we increase the capacity of the network.

4.4 THREE BLOCK VGG WITH DROPOUT

Deep neural networks can be regularized using dropout regularization. Inputs to a layer which are the input variables in the data samples or activations from the previous layers are removed or dropped out probabilistically in dropout regularization. A large number of networks are simulated with different network structures. These are more robust to inputs. After each VGG block a small amount of dropout can be applied. For the fully connected layers near the output layer of the model more dropout is applied. A small dropout of 20% is applied after each VGG block followed by a larger dropout of 50% after fully connected layer.

Performance of three block VGG model with addition of dropout is slightly increased from 75% to 77% for validation data. As shown in figure 5 the accuracy of training is increasing as we increase the number of iterations. Hence increase in the training epochs will increase the performance of the model.

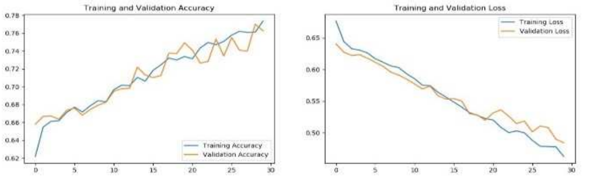


Figure 5: Performance of Three block VGG model with Dropout

4.5 EXPLORE TRANSFER LEARNING

Transfer learning deals with use of all parts of a model which is trained on a related task. Various pre trained models provided by Keras can be used. One of the widely used model in transfer learning is VGG16 which is the VGG model with 16 layers. The model has performed best on ImageNet photo classification challenge. VGG16 model is consists of two parts mainly. Feature extractor part of VGG 16 is made up of VGG blocks. The classifier part of VGG16 consists of fully connected layers along with final output layer.

During this work we have used the feature extractor part of the VGG 16 model. New classification part of the model is tailored for crowd classification dataset. To be more specific, we have hold all weights of all of the convolutional layers fixed during training. Only new fully connected layers are trained for classifying the crowd dataset. Steps for performing this task are as below

- Load VGG 16 model
- Remove fully connected layers from the output-end of the model,
- Add new fully connected layers to interpret the model output and make a prediction.

- “include-top” argument is set to “False” for removing the classifier model automatically.

.After defining the model, we train the model on the training dataset. Number of training epochs is set to 30. Performance of three block VGG model using transfer learning is slightly increased 79% for validation data. As shown in figure 6 the accuracy of training is increasing as we increase the number of iterations. Hence increase in the training epochs will increase the performance of the model.

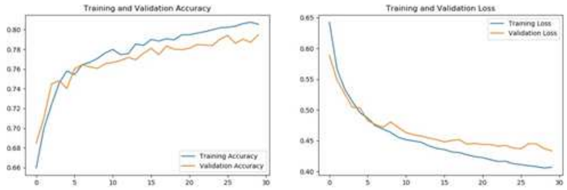


Figure 6: Performance of Transfer Learning

5. CROWD CLASSIFICATION USING DEEP LEARNING

The proposed methodology for crowd classification using deep learning is depicted in figure 7.

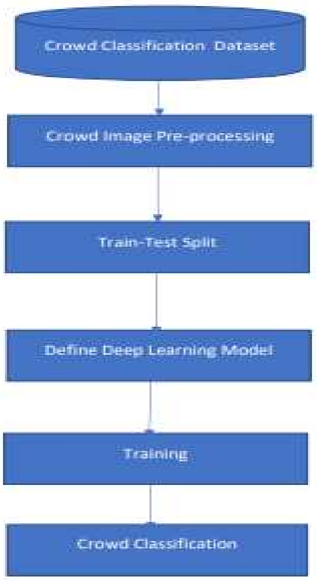

Crowd Classification Dataset
Crowd Image Pre-processing
Train-Test Split
Define Deep Learning Model
Training
Crowd Classification

Figure 7: Crowd Classification Using Deep Learning

5.1 PRE-PROCESSING

For all images in the dataset, we have resized the images to 256x256 pixels. Median filtering is used for noise removal. Crowd image is first split into three separate R, G, and B channels. We have applied the filter to separate R, G and B components of image and then after merged them. The denoised images are further used for segmentation, features extraction and classification.

5.2 TRAIN-TEST SPLIT

Pre-processed images from crowd dataset are split randomly into 70% data for training and 30% data for evaluation.

5.3 DEFINE DEEP LEARNING MODEL

During this work we have considered various models of deep learning for crowd classification.

Following models are defined using deep learning

- One Block VGG
- Two block VGG
- Three Block VGG
- Three Block VGG with Dropout
- Deep Learning using transfer learning (VGG-16)

5.4 TRAINING

Different deep learning models are trained using 70% training data.

5.5 CROWD CLASSIFICATION

During the stage of crowd classification, we have divided the data in 70% training data and 30% testing data. We have evaluated the performance for different deep learning models.

6. EXPERIMENTAL RESULTS AND PERFORMANCE MEASURE

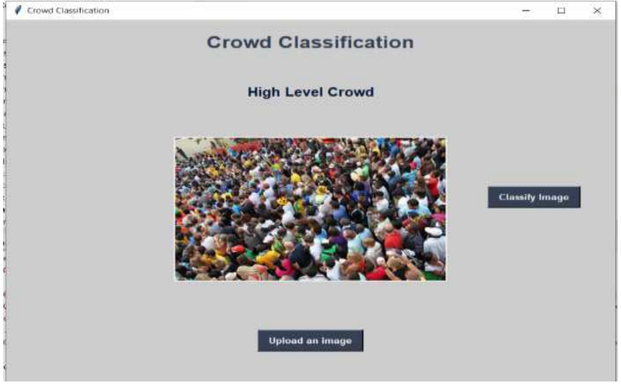
During this work of crowd images classification we have evaluated the performance of various deep learning models. Various parameters used for performance evaluation are accuracy, precision, recall and fmeasure.

- Accuracy: Accuracy is the measure of correctly classified crowd images.
- Precision is the fraction of retrieved crowd images that are similar to the query
- Recall is the fraction of the crowd images that are successfully classified.
- F Measure considers both precision and recall. It is the harmonic mean(average) of the precision and recall.

6.1 CROWD CLASSIFICATION USING PYTHON

We have used tkinter for developing the GUI for crowd classification as shown in figure 8. The crowd classification dataset is prepared using two datasets [17, 18] Steps involved in crowd classification are

- Pre-processing crowd images using median filtering for noise removal.
- Train-test split of the dataset
- Building various deep learning models
- Training the models with 70% training data
- Crowd image classification and performance evaluation using 30% evaluation data.


Crowd Classification

High Level Crowd

Figure 8 : PYTHON GUI for crowd Classification

Accuracy of the proposed crowd classification is shown in table 3 in terms of accuracy, precision, recall and f measure.

Table 3: Accuracy for Crowd Classification

	One Block VGG	Two Block VGG	Three Block VGG	Three Block VGG with Dropout	VGG16 with transfer Learning
Accuracy	0.608333	0.6125	0.631944	0.638889	0.694444
Precision	0.614194	0.610969	0.663043	0.633564	0.691726
Recall	0.608333	0.6125	0.631944	0.638889	0.694444
F Measure	0.608386	0.609583	0.625413	0.630204	0.68883

As shown in table 3, performance of crowd classification is increased to 69.44% using VGG16 with transfer learning.

Figure 9 depicts the performance of the proposed system of crowd images classification.

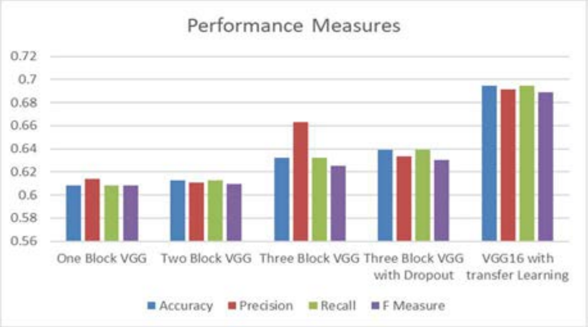


Figure 9: Performance Evaluation

As shown in figure 9 VGG16 with transfer learning performs better for crowd classification.

7. **CONCLUSION AND FUTURE SCOPE**

Automatic crowd classification has been of prime importance due to our world witnessed major events. COVID19 breakout is one of the latest reason for improving the importance of crowd analysis. Automated system for crowd management, crowd counting, securing the crowd are getting the importance. Automatic classification of crowd images has been a challenge due to occlusion, varying climatic conditions while capturing images etc. In this paper we have addressed the problem of crowd classification in presence of varying lighting conditions, varying climatic conditions etc. We have prepared the crowd classification dataset consisting of 5569 crowd images. Crowd images are classified into three categories according crowd density. We have considered different models of deep learning for crowd classification. Deep learning with transfer learning using VGG 16 model is proposed here. The proposed model has achieved the accuracy of 69.44% which is greater than all machine learning models as well as deep learning models under consideration. Future aspects of the system will be towards features selection, features reduction techniques along with deep learning for improving the performance of the system.

REFERENCES

[1] C. Shang, H. Ai and B. Bai, "End-to-end crowd counting via joint learning local and global count," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 1215-1219.

[2] S. Thasveen M. and L. Mredhula, "Real Time Crowd Counting: A Review," 2020 International Conference on Futuristic Technologies in Control Systems & Renewable Energy (ICFCR), Malappuram, India, 2020, pp. 1-5.

[3] W. Kong, H. Li, G. Xing and F. Zhao, "An Automatic Scale-Adaptive Approach With Attention Mechanism-Based Crowd Spatial Information for Crowd Counting," in IEEE Access, vol. 7, pp. 66215-66225, 2019.

[4] Y. Yoon, J. Gwak, J. Song and M. Jeon, "Conditional marked point process-based crowd counting in sparsely and moderately crowded scenes," 2016 International Conference on Control, Automation and Information Sciences (ICCAIS), Ansan, Korea (South), 2016, pp. 215-220.

[5] H. Fu, H. Ma and H. Xiao, "Real-time accurate crowd counting based on RGB-D information," 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 2012, pp. 2685-2688.

[6] X. Kong, M. Zhao, H. Zhou and C. Zhang, "Weakly Supervised Crowd-Wise Attention For Robust Crowd Counting," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 2722-2726.

[7] Q. Wang, J. Gao, W. Lin and X. Li, "NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 6, pp. 2141-2149, 1 June 2021.

[8] L. Rajendran and R. Shyam Shankaran, "Bigdata Enabled Realtime Crowd Surveillance Using Artificial Intelligence And Deep Learning," 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea (South), 2021, pp. 129-132.

[9] P. Guo and Z. Miao, "Action Detection in Crowded Videos Using Masks," 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010, pp. 1767-1770.

[10] Y. Tian, L. Cao, Z. Liu and Z. Zhang, "Hierarchical Filtered Motion for Action Recognition in Crowded Videos," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 3, pp. 313-323, May 2012.

[11] H. Idrees, K. Soomro and M. Shah, "Detecting Humans in Dense Crowds Using Locally-Consistent Scale Prior and Global Occlusion Reasoning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 10, pp. 1986-1998, 1 Oct. 2015.

[12] P. Guo, Z. Miao and H. Cheng, "Masks based human action detection in crowded videos," 2010 IEEE International Conference on Image Processing, Hong Kong, China, 2010, pp. 693-696.

[13] Q. You and H. Jiang, "Action4D: Online Action Recognition in the Crowd and Clutter," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 11849-11858.

[14] M. Gao, Y. Wang, J. Jiang, J. Shen, G. Zou and L. Liu, "Crowd motion segmentation via streak flow and collectiveness," 2017 Chinese Automation Congress (CAC), Jinan, China, 2017, pp. 4067-4070.

[15] A. Zweng and M. Kampel, "Unexpected Human Behavior Recognition in Image Sequences Using Multiple Features," 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010, pp. 368-371.

[16] H. Liao, J. Xiang, W. Sun, Q. Feng and J. Dai, "An Abnormal Event Recognition in Crowd Scene," 2011 Sixth International Conference on Image and Graphics, Hefei, China, 2011, pp. 731-736.

[17] Sindagi, Vishwanath & Yasarla, Rajeev & Patel, Vishal. (2020). JHU-CROWD++: Large-Scale Crowd Counting Dataset and A Benchmark Method.

[18] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma. Singleimage crowd counting via multi-column convolutional neural network. In CVPR, pages 589–597, 2016.