

An Implicit Approach on Classification of Gender using CNN Methods with Crowd Analysis

Abhishek Nazare¹ and Sunita Padamannavar²

^{1,2}KLS Gogte Institute of Technology/Master of Computer Applications, Belagavi, India
Email: abhinazare@gmail.com, sunitapdm@gmail.com

Abstract—As the urban population continues to rise, intelligent identification systems will become increasingly important to meet the computing analysis. According to the discussions on gender and age identification in public for the development an Intelligent transportation systems need to understand the relativity of identification in public places for providing appropriate facilities. Face recognition refers to the process of recognizing or tracking a person's identification by using his face. Individuals can be classified in photographs and videos. An image's content can include places, actions, items, people, and so on. CNN is used to generate a variety of complicated face appearances. Researcher surveyed through the previous works for identification of gender through CNN methods. Among the region based architectures researcher proposes to use the combination of models for fetching an image details with accurate pattern and right outcome. ImageNet is used for collecting dataset containing photos with varying resolutions. Network can be trained using the pixels' (centred) raw RGB values. Researcher uses combination of methods to provide better solutions with more accuracy, the approach of combinations can be put forth in a way to understand the limitations from previous work done and make an approach to combine a good method and yield better results. This research work will be utilized for the analysis of crowd at public places for public transportation with a blend of smart systems in the future development of transportation system. The most benefited sectors from this technique are real-time surveillance mechanisms and access management systems.

Index Terms— Urbanized people, ImageNet, crowd analysis, Face recognition, Convolutional Neural Network(CNN), RCNN, Public transportation.

I. INTRODUCTION

Over the last few decades, there has been an increase in the number of several research communities which have looked into the issue of object counting. As a result, object counting is their primary research focus. To count the number of works published, there have been many ways across a wide range of domains of objects in images or videos to count which leads to crowd counting. This process helps in finding the number of individuals at a crowded place. This leads to good identification system with a variety of mobility alternatives can have a substantial impact on people's quality of life by allowing them to get to desired destinations [2]. The quality of transportation networks influences each person's decisions and how they organize various activities like work, leisure, social gatherings, shopping, education, and health. Inequalities in society result from a lack of access to the transportation network. Ethnic exclusion occurs only when the chances are not identified, but rather due to a lack of access to them [3]. Difficulties in acquiring employment, as well as key public infrastructure and services

such as health and education, are examples of social barriers, as well as participation in social and leisure activities [4]. Because mobility is so important in development, cities and increases individual quality of life [5], access to appropriate transportation solutions is tied to each individual's well-being [6]. Access to opportunity for all, elimination of negative externalities of transportation for all, and participation of representatives in decision-making, with a focus on urbanized populations, are critical for the provision of [1] inclusive transportation schemes.

The work attempts to contribute to the solution of gender biases in the provision of transportation services. By evaluating current gender-oriented approaches in transportation planning, A knowledge "library" of gendered mobility behaviour and gendered transportation service requirements is established [8]. This research seeks to analyze existing gender concerns in transportation planning methods and provide the results by investigating gender perceptions and attitudes about mobility services[1].

Hence to identify the gender races the work proposes the best Convolutional Neural Network(CNN) methods for predicting the gender and age which is the major scope to be achieved in this platform. The various methods of CNN are being experimented to identify the gender by using the still images, live images or real time images using the cameras or sensors which helps in projecting the picture. The focus is towards the crowd located near the bus stop or railway station or places where we can locate maximum people and identify the crowd with individualistic view.

The idea behind this study is that face detection offers a wide range of potential applications. From digital cameras that concentrate on people's faces automatically to security cameras that can match a person's face to their identity. Webcams are widely utilized as a security measure while protecting a computer. Only if the webcam's facial recognition system detects a person's face may they gain access to the computer. This technique can be reduced down to eye tracking and recognition.

Face Recognition has recently come in to popularity. In the learning stage, CNNs have a high computational cost in terms of memory and speed, but they are capable of some shift and deformation invariance. Furthermore, in order to tackle gender diversity, we employ an inter-sectional perspective, which is founded on demographical grounds, such as ethnicity, gender, age and disability, which must be regarded as mutually influencing and closely interwoven [11,12]. As a new area of machine learning research, the Deep Learning framework was introduced. Deep learning evolved from artificial neural networks (ANNs), is also known as Hierarchical learning [9] is a powerful technique in comparison to other machine learning algorithms.

It makes use of a step-by-step paradigm for achieving amazing progress toward data abstraction It is becoming increasingly popular as a result of the exponential increase in data access and remarkable advancements in hardware technology. When compared to standard neural networks, deep learning outperforms them significantly. Deep learning provides a kind of autonomous extracting features with little to no human interaction [9]. Deep learning, in general, necessitates two key characteristics: layered nonlinear dispensation and learning (supervised or unsupervised)[10]. Layers form a hierarchical structure in which the output of next layer helps in generating the input for next layer and so on multiple layers are worked out.. Automatic further improvisation of extraction can be found in a wide range of utilization.

Section 2 provides the different types of CNN with the study of models. In next section 3 we highlight on the use of the CNNs methods stated in the previous work for the different set of sample dataset used for this article for our initial phases of understanding the work with the images or pictures which were considered previously. The study of performance analysis of CNN for various models for the future development in Transportation Systems in section 4. In section 5 we summarize the study and propose how the comparison of CNN's performance to those of other ways and have firsthand knowledge with its effectiveness. This review focuses on CNN's fashionable paradigm, its increasing advantages, and current applications.

II. CONVOLUTIONAL NEURAL NETWORK MODELS

Convolutional Neural Network may be a framework composed of various layers and stages, every of that is created from a group of arrays referred to as feature maps. These arrays function the inputs and outputs for every stage. A CNN is created from 3 divisions:

- Convolutional Layer - Extraction of Characteristics- The functioning of non-linearity- Layer of pooling or subsampling- Layer that is fully connected

Essentially, based on the study of CNN stages, one is created from these 3 parts. Multiple convolutional layers could be natural ability in multiple phases. Layers that area unit pooled area unit followed by layers that area unit entirely connected. Confined relationships, weight sharing, pooling, and therefore the use of various layers is one among the four key ideas of CNN that creates use of natural signal options for convenience[13]. The Fig. 1 CNN

architecture describes the process of handling the images in a Non linearity, Max pooling methods which helps in detecting the image to identify the minute particles and generate the clarification of the single image.

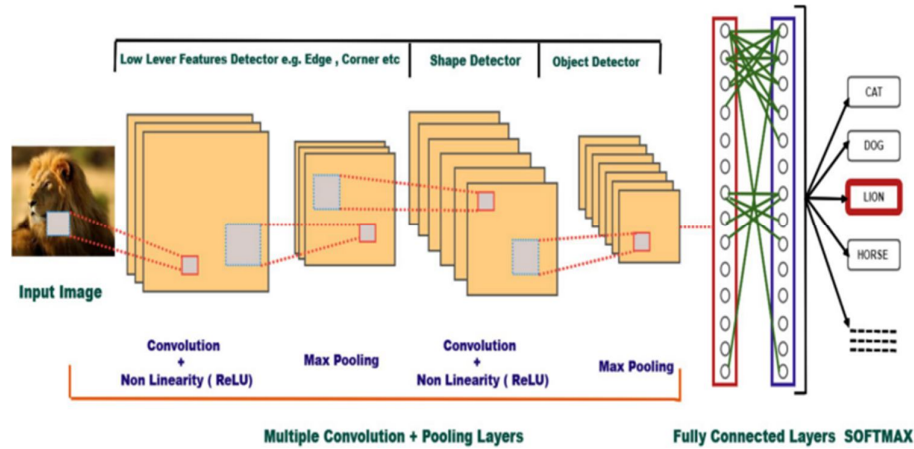


Fig. 1 CNN Architecture

The three activation functions are depicted graphically in Fig 2. Among all activation functions, ReLU [27] is the best. The mathematical definition of ReLU is $y = \max(0, x)$. It works on the non positive values and assigns zero.

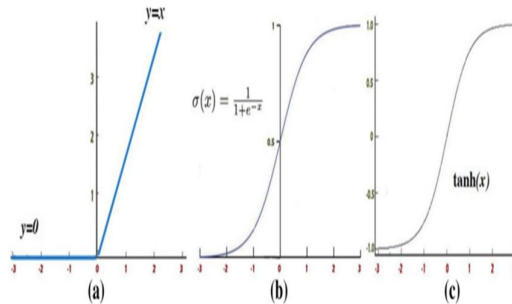


Fig. 2 Functions used are a. ReLU b. Sigmoid c. tanh[27]

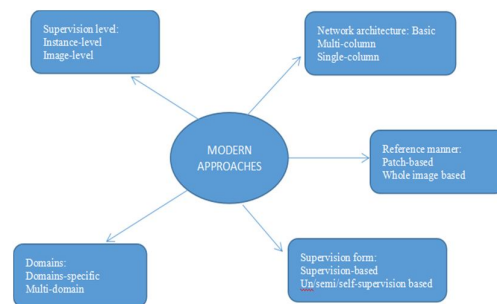


Fig. 3 Modern Approaches of CNN

It balances the CNN by tightly controlling the minimum and maximum thresholds. In this work we utilize the following concept in a crowd analysis where there is huge crowd and identifying the individuals is the key factor. The Fig. 3 below states the modern methods in CNN architecture paves us to think on the possibility of utilizing the approaches based on the requirement with accuracy and efficiency. The approaches are organized around network topologies, reference modes, supervision forms, domain specifics, and supervision levels. We classified network architecture types as basic CNN-based approaches, multi-column and single-column methods.[14]. The basic CNN adopts here the layered feature as shown in Fig. 1. The multi-column method employs distinct columns to record multi-scale information based on the various receptive fields, resulting in improved performance. Often single column network study with the purpose of not expanding network complexity utilizes single and complex CNNs instead of bulky structure of multi-column network study. Based on their training methodologies, CNN-based crowd counting systems are classified into two types: patch-based inference and completely based inference.

1) Patch-based: To train utilizing randomly cropped patches from an image, this inference approach is necessary. During the testing process, a sliding window is dispersed throughout the full test image, and the estimations of each window are gathered and then combined to obtain the image's final total count. 2) Methods based entirely on images: Because of the sliding window method, patch-based techniques do not focus on overall data and incur a high computational cost. Full image-based algorithms, on the other hand, usually use the entire image as input and produce a density map or total number of crowds, which is more efficient but may occasionally omit internal data.

2) In Table I an overview of the required architecture and their description has been approximately provided in an order.[22] For representative network architectures for crowd counting it is broadly divided into basic CNN based, multi column and single column based methods.

TABLE II. SUMMARY OF THE STATE-OF-THE-ART

Methods	Year & Venue	Network architecture	Reference manner	Supervision form	Learning paradigm	Supervision type
Fu et al. [16]	2015, EAAI	Basic	Patch-based	Fully	STL	Instance -level
Wang et al. [17]	2015 ,ACMMM	Basic	Patch-based	Fully	STL	Instance-level
MCNN [18]	2016 ,CVPR	Multi-column	Whole image-based	Fully	STL	Instance-level
Crowdnet [1]	2016, ACMMM	Multi-column	Whole image-based	Fully	STL	Instance -level
CNN-Boosting [19]	2016, ECCV	Multi-column	Whole image-based	Fully	STL	Instance-level

The below table II summarizes the different literature review for the work on the crowd density estimation through systematic survey the title provides the areas wherein they have focused in the year and the venue where they have have conducted it. The description is very crucial as it states the a study on mass density calculation and rate-counting methods utilized for visible investigation in view of computer visual sense research, another reference describes the statistical evidence from existing research and recommends focusing on broad elements of approaches rather than any specific algorithm. Crowd analysis is been divided into two areas like statistics and behavior understanding. Recently the new trends have been highlighted about the CNN based single image crowd counting.

TABLE III. ANALYSIS OF LITERATURE REVIEWS

Sl. No	Title	Year	Place	Brief Description
1	Recent study on mass density computation and counting for visual investigation [18]	2015	EAAI	This research paper introduces a study on mass density calculation and rate-counting methods utilized for visible investigation in view of computer visual sense research.
2	Enhancements and modes in visible multitude study: A systematic study and valuation of mass modelling techniques [19]	2016	Neurocomputing	This study aims to contribute deriving significant applied math data from the recent literary study and present towards focusing on the universal elements of approaches rather than any peculiar methodology.
3	Crowd scene understanding from video: a survey [1]	2017	TOMM	It studies on the two major keys for further study first is crowd computation and next is behaviour comprehension.
4	A survey of recent advances in cnn-based single image crowd counting and density estimation [20]	2018	PRL	A new approach of CNN based methods have to be identified based on the earlier study of single based image.
5	Convolutional NN for mass behaviour analysis: a survey [21]	2019	VC	Analysis of crowd using CNN based architecture.

III. DATASET

ImageNet is a collection of more than 10 million tagged perfectly framed photos organized into over 20,000 collection. The pictures were gathered over the online store and categorized by person via Amazon's Mechanical Turk crowd-sourcing programme. An annual competition named the "ImageNet Large-Scale Visual Recognition Challenge" (ILSVRC)[22] has been designed as part of the "Pascal Visual Object Challenge" since 2010.

ILSVRC use a subset of ImageNet, with approximately 1000 photos in each of the 1000 categories. The amount rises up to 1.2 million training images, 50,000 validation images, and 150,000 testing images in all. ImageNet is made up of photos with varying resolutions, however our algorithm requires a consistent input dimensionality. As a result, we down-scaled the photos to a fixed resolution of (256x256). Given a rectangular image, we first re-scaled it so that the shorter side was 256 pixels long, and then cropped the resulting image of removing the centre (256x256) patch. We did not preprocess the photographs in any way other than subtracting the mean activity throughout the training set from each pixel. As a result, we trained our network using the raw RGB values of the pixels (centre).

Table III sums-up some representational datasets, including counting of crowd datasets using scenario from real world and one with synthetically generated data. To get the accuracy, we additionally study many datasets used in other fields to assess the abstraction potential of the constructed algorithmic program. Figure 4 depicts several samples from the represented datasets. Reference[23], have stated that convolutional neural network(CNN) is a sort of multi-layer network that comprises of numerous alternating convolutional and pooling (subsampling) layers, followed by a sequence of full-connected layers, as shown in Fig. 1. A convolutional layer network's input is often an image ($m \times m \times r$), where m is the picture's height and width combined and r is the number of channels. Convolutional layers have l kernels (or filters) with dimensions of ($n \times n \times q$), where (n and q) are determined by the designer (where generally q is equal to r). Each filter is $(m - n + 1) \times (m - n + 1) \times q$ in size. This is produced by a convolution map of features or characteristics, where p denotes the number of filters to be used. Table III lists the specific statistics for the datasets, which are arranged by chronology order in year.



Fig 4. Some samples of datasets for Gender, age and crowd analysis

The datasets samples used are from the Shanghai Tech for the initial work for understanding the CNN method for identification of gender, age and crowd with more specific computations and accuracy. Shanghai Tech [2] is with 1198 pictures and 330,165 annotations, being the most popular and having maximum amount of dataset available in it. Part A and Part B of the dataset are divided based on different density distributions. Part A has photographs chosen across the network and picked up randomly, Part B features pictures picked up from a busy place of a city Shanghai resident area. In contrast of Part B has a much lower concentration than Part A. The store of images delivers a tough amount of data spanning a wide range of view types and concentration. However, because the number of images in different density sets is unequal, the training and test sets are frequently light concentrated sets. Regardless, the size alterations and viewpoint damage in the given store of data present situation which are new and potentially on the development of multiple Convolutional Neural Network-based approach.

TABLE III: OBJECT COUNTING DATASET STATISTICS, INCLUDING CROWD COUNTING AND OTHER FIELDS. TOTAL-THE TOTAL NUMBER OF OCCURRENCES IN THE DATASETS, AS WELL AS THE NUMBER OF INSTANCES

Dataset	Year	Attributes	Number of Images	Training/Test	Average Resolution	Count Statistics Total
CrowdFlow [24]	2018	Synthetic	-	-	300~450	-
UCF-QNRF [29]	2018	Real-World	1535	1201/334	2013 x 2902	1,251,642
CHSR [25]	2019	Real- World	1000	-	1080 x 720	-
Indoor[26]	2019	Real- World	148,243	-	352 x 288 or 704 x 576	1,834,770
ShanghaiTechRGBD [27]	2019	Real- World	2193	1193/1000	1080 x 1920	144,512
DroneCrowd [28]	2019	Drone-based	33,600	-	1920 x 1080	4,864,280

IV. A REVIEW OF PERFORMANCE ANALYSIS OF CNN MODELS

According to the discussions on gender and age identification in public for the development in Intelligent transportation systems we need to understand the relativity of identification in public places for providing

appropriate facilities [41]. The different CNN models for identifying the objects like vehicle, human beings with RCNN, Fast RCNN, YOLO, Mask RCNN for region based. These models provide the platform to study the various aspects defining them with the proper parameters to find the objects. Lets check the architecture of the following models:

A. RCNN(Region based CNN)

The RCNN helps in identifying the region for the CNN to work but it has some disadvantages like it take too much of time to train the network as we have to classify roughly 2k to 3k region proposal per image, cannot be taken for real time as it takes seconds to scan on one image. Selective search algorithm is used. The below fig. 5 gives the depiction of its working.

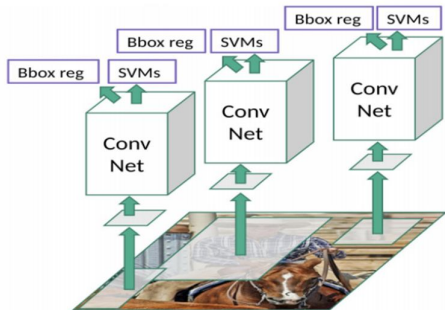


Fig 5 R-CNN Architecture

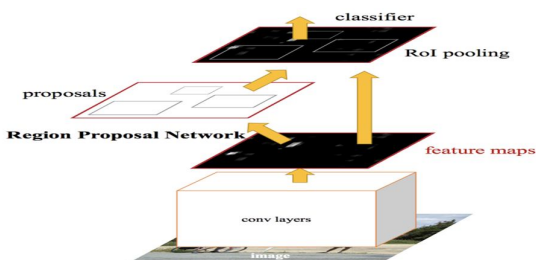


Fig 6 Fast R-CNN Architecture

B. Fast RCNN

In Fast RCNN the method used is to compute the given image faster than RCNN method, the RCNN method has to check the whole image with a multi region wise approach whereas in Fast RCNN dose it at once for a given image which produces a feature map as a result. The below figure 6 provides the details about its working.

C. YOLO

All previous object detection techniques used regions to locate the object within the image. YOLO is one of the many methods known as You Only Look Once used by CNN for detecting entities. The approach is of a network is not just interpret a whole picture but also each part of it and from there to pick the right portion of it. A single neural network generates the conjugated boxes as well as their class probabilities. The figure 7 depicts the working model of YOLO with proper approach. Other object detection techniques are orders of magnitude slower than YOLO the rate of frames per second is 45. The YOLO formula efforts has issue with small particles in the given image.

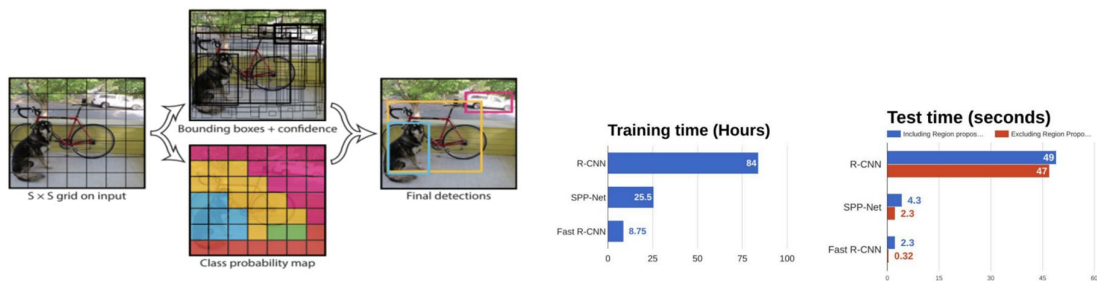


Fig 7 YOLO working style Fig 8 Train and Test

The above figure 8 provides the time taken to train and test the objects in various object detecting algorithm.[40] The below figures of 9 (a)(b)(c)(d) represents the outcome of CNN based state-of-the-art. The eight methods with four approaches are depicted in a chart representative traditional approaches over the crowd counting work. The valuation is done using the MAE and RMSE metrics for evaluating the models' reliability and precision. The charts below state the results of the earlier study.[14] The approach to utilize the methods in combinations would help on getting a more accurate result for identification of gender in crowd.

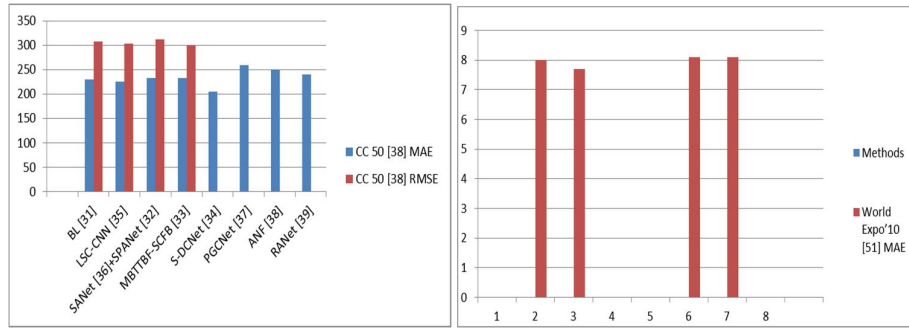


Fig 9(a) Methods and CC Approach

Fig 9(b) Methods and World expo'10 Approach

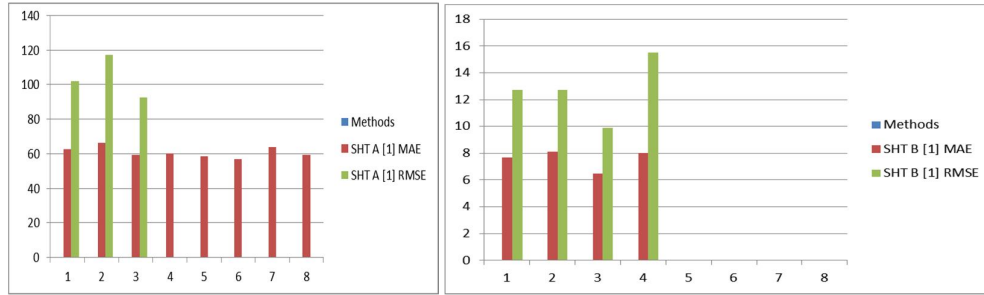


Fig 9(c) Methods and CC Approach

Fig 9(d) Methods and CC Approach

The above chart provides an idea of the valuation metrics of MAE and RMSE for the the CC 50 dataset for the place ICCV in 2019. The above chart provides an idea of the valuation metrics of MAE and RMSE for the the world Expo'10 dataset for the place ICCV in 2019. The chart above 9(c) states the statistics of evaluation metrics of MAE and RMSE for the SHT A dataset and 9(d) provides the data of the calculations carried out for SHT B dataset. The comparison of the above charts helps in understanding the evaluation metrics for the methods and the representing a proper picture about accuracy and precision from each method for the different dataset.

V. PROPOSED WORK

The above points have helped in understanding the overall idea about crowd analysis, gender identification using various CNN methods. Among them we can use combination of methods to provide better solutions with more accuracy, the approach of combinations can be put forth in a way to understand the limitations from previous references and make an approach to combine a good method and yield better results. Below listed references provide us some inputs for the different methods utilized, with few considered parameters and the limitations which have been observed.

Mean absolute error (MAE) for the evaluation of age and is defined as the mean error between predicted age and actual age for the input or test image. Mathematically, it is expressed in (1) as:

$$\text{Mean absolute error (MAE)} = \frac{1}{N} \sum_{i=1}^N |\hat{A}_i - A_i| \quad (1)$$

The table below states the boundary of the approach mentioned in earlier literature.

TABLE V. RELATED WORK OF VARIOUS METHODS ON GENDER IDENTIFICATION APPROACH

References	Method	Evaluation Parameters Used	Observations
Yang et al. [42]	The methodology used was ScatNet, PCA and fully connected layers	MAE score on MORPH 2 dataset was 3.49, LIFESPAN was 5.19, and for FACES was 7.04	Required more database for training to enhance the performance further
Yi et al. [43]	Multi-scale CNN was used with sub-network per patch	MAE on MORPH 2 is 3.63	For gender classification achieving higher accuracy can be taken as future work
Wang et al. [44]	CNN with dimensionality reduction and classification (SVR, PLS, CCA)	MAE on MORPH 2 is 4.77 and for FG-NET is 4.26	-
Rothe et al. [45]	VGG16 architecture was used	MAE on MORPH 2 was 2.68 and for FGNET was 3.09	The training dataset could be further enlarged

The combination of methods can help us in increasing the accuracy so we suggest to consider the test analysis of dataset in large scale and small scale. The approach of CNN methods with dimensionality reduction and classification with feature extractions methods. Another approach is to work with dimensionality and fully connected layers. These two combined approaches can help in generating better results.

VI. CONCLUSION

Face recognition refers to the process of recognizing or tracking a person's identification by using his face. Individuals can be classified in photographs and videos. The most benefited sectors from this technique are real-time surveillance mechanisms and access management systems. Deep learning-based algorithms are capable of extracting more complex face features. CNN can generate a variety of complicated face appearances, and it performs rather well in comparison to humans. It is a sub-field in computer vision. An image's content can include places, actions, items, people, and so on. Convolutional Neural Networks outperform other algorithms on large datasets due to its combination of feature extraction and classification mechanisms. These are the outcomes we have surveyed through the previous works for identification of gender through CNN methods with their architectures, factors and models stated earlier. Among the region based architectures we need to use the combination of models for fetching a image details with accurate pattern and right outcome. This work will be utilized for the analysis of crowd at public places for public transportation with a blend of smart systems in the future development of transportation system.

REFERENCES

- [1] Pirra, M. Kalakou, S. Carboni, A. Costa, M. Diana, M. Lynce, "A Preliminary Analysis on Gender Aspects in Transport Systems and Mobility Services", Presentation of a Survey Design. Sustainability 2021, 13, 2676. <https://doi.org/10.3390/su13052676>
- [2] Boisjoly G, Yengoh, G.T. , "Opening the door to social equity: Local and participatory approaches to transportation planning in Montreal", Eur. Transp. Res. Rev. 2017, 9, 1–21
- [3] Preston, J. Rajé, F, "Accessibility, mobility and transport-related social exclusion", J. Transp. Geogr. 2007, 15, 151–160.
- [4] Herwangi, Y, Pradono, P, Syabri, I. Kustiwan, I, "Assessing transport disadvantage and transport related social exclusion for inclusive transportation planning: A review of methodology", In Proceedings of the Planocosmo International Conference, Bandung, Indonesia, 21–22 October 2013.
- [5] Thynell M, "The quest for gender-sensitive and inclusive transport policies in growing Asian cities", Soc. Incl. 2016, 4, 72–82.
- [6] Delbosc, A., Currie, G., "Exploring the relative influences of transport disadvantage and social exclusion on well-being:", Transp. Policy 2011, 18, 555–562.
- [7] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in CVPR2016, pp. 589–597.
- [8] Church, A., Frost, M., Sullivan, K., "Transport and social exclusion in London", Transp. Policy 2000, 7, 195–295.
- [9] Najafabadi M M, Villanustre F, Khoshgoftar TM, Seliya N, Wald R, Muharemagic E (2015) , "Deep learning applications and challenges in big data analytics", J Big Data 2(1):1–21, <https://doi.org/10.1186/s40537-014-0007-7>
- [10] Bengio Y , "Learning deep architectures for AI", Foundation Trends Machine Learn 2:1–127. <https://doi.org/10.1561/2200000006>, 2009
- [11] Salakhutdinov R, Tenenbaum JB, Torralba A, "Learning with hierarchical-deep models". IEEE Trans Pattern Anal Mach Intell 35(8):1958–1971. <https://doi.org/10.1109/TPAMI.2012.269>, 2013
- [12] Xiangang Li and Xihong Wu, "Constructing long shortterm memory based deep recurrent neural networks for large vocabulary speech recognition", In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp 4520–4524. <https://doi.org/10.1109/ICASSP.2015.7178826>, 2015.
- [13] LeCun Y, Bengio Y, Hinton G, "Deep learning". Nature 521(7553):436–444. <https://doi.org/10.1038/nature1453>, 2015
- [14] Gao, Guangshuai, Junyu Gao, Qingjie Liu, Qi Wang, and Yunhong Wang, "Cnn-based density estimation and crowd counting: A survey", arXiv preprint arXiv:2003.12783 (2020).
- [15] A. Garain et al., "GRA_Net: Deep Learning Model for Classification of Age and Gender From Facial Images", 10.1109/ACCESS.2021.3085971
- [16] M. Fu, P. Xu, X. Li, Q. Liu, M. Ye, and C. Zhu, "Fast crowd density estimation with convolutional neural networks", EAAI, vol. 43, pp. 81–88, 2015.
- [17] C. Wang, H. Zhang, L. Yang, S. Liu, and X. Cao, "Deep people counting in extremely dense crowds", in ACM MM. ACM, 2015, pp. 1299–1302.
- [18] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in CVPR2016, pp. 589–597.
- [19] E. Walach and L. Wolf, "Learning to count with cnn boosting", in ECCV. Springer, 2016, pp. 660–676.

- [20] J. Shao, K. Kang, C. Change Loy, and X. Wang, "Deeply learned attributes for crowded scene understanding", in CVPR, 2015, pp. 4657–4666.
- [21] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking", in CVPR. IEEE, 2008, pp. 1–7.
- [22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database", in CVPR. IEEE, 2009, pp. 248–255
- [23] Ren, S., He, K., Girshick, R. Y Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks", Advances in neural information processing systems. S.I.: s.n., pp. 91-99, 2015
- [24] G. Schröder, T. Senst, E. Bochinski, and T. Sikora, "Optical flow dataset and benchmark for visual crowd analysis", in AVSS. IEEE, 2018, pp. 1–6.
- [25] M. Xu, Z. Ge, X. Jiang, G. Cui, B. Zhou, C. Xu et al., "Depth information guided crowd counting for complex crowd scenes," PRL, 2019.
- [26] M. Ling and X. Geng, "Indoor crowd counting by mixture of gaussians label distribution learning", TIP, vol. 28, no. 11, pp. 5691–5701, 2019
- [27] D. Lian, J. Li, J. Zheng, W. Luo, and S. Gao, "Density map regression guided detection network for rgb-d crowd counting and localization", in CVPR, 2019, pp. 1821–1830
- [28] L. Wen, D. Du, P. Zhu, Q. Hu, Q. Wang, L. Bo, and S. Lyu, "Dronebased joint density map estimation, localization and tracking with space-time multi-scale attention network", arXiv:1912.01811, 2019.
- [29] 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 ECCV, 2018, pp. 532–546
- [30] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli et al., "Image quality assessment: from error visibility to structural similarity", TIP, vol. 13, no. 4, pp. 600–612, 2004.
- [31] Z. Ma, X. Wei, X. Hong, and Y. Gong, "Bayesian loss for crowd count estimation with point supervision", in ICCV, 2019, pp. 6142–6151.
- [32] Z.-Q. Cheng, J.-X. Li, Q. Dai, X. Wu, and A. G. Hauptmann, "Learning spatial awareness to improve crowd counting", ICCV, 2019.
- [33] V. A. Sindagi and V. M. Patel, "Multi-level bottom-top and top-bottom feature fusion for crowd counting", ICCV, 2019.
- [34] H. Xiong, H. Lu, C. Liu, L. Liang, Z. Cao, and C. Shen, "From open set to closed set: Counting objects by spatial divide-and-conquer", in ICCV, 2019.
- [35] D. B. Sam, S. V. Peri, A. Kamath, R. V. Babu et al., "Locate, size and count: Accurately resolving people in dense crowds via detection", arXiv preprint arXiv:1906.07538, 2019.
- [36] X. Cao, Z. Wang, Y. Zhao, and F. Su, "Scale aggregation network for accurate and efficient crowd counting", in ECCV, 2018, pp. 734–750.
- [37] Z. Yan, Y. Yuan, W. Zuo, T. Xiao, Y. Wang, S. Wen, and E. Ding, "Perspective-guided convolution networks for crowd counting", in ICCV, 2019.
- [38] A. Zhang, L. Yue, J. Shen, F. Zhu, X. Zhen, X. Cao, and L. Shao, "Attentional neural fields for crowd counting", in ICCV, 2019, pp. 5714–5713.
- [39] A. Zhang, J. Shen, Z. Xiao, F. Zhu, X. Zhen, X. Cao, and L. Shao, "Relational attention network for crowd counting", in ICCV, 2019, pp. 6788–6797.
- [40] <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>
- [41] Olatunbosun AgboAjala, Serestina Viriri, "Deep learning approach for facial age classification: a survey of the state of the art", published online 19 June 2020.
- [42] Yang, H. F., Lin, B.Y., Chang, K.Y., & Chen, C.S., "Automatic age estimation from face images via deep ranking networks", 35(8), 1872–1886, 2013
- [43] Yi, D., Lei, Z., & Li, S. Z., "Age estimation by multi-scale convolutional network", In Asian Conference on Computer Vision, 144–158. Springer, Cham, 2014
- [44] Wang, X., Guo, R., & Kambhamettu, C., "Deeply-learned feature for age estimation", In 2015 IEEE winter conference on applications of computer vision, 534–541. IEEE, 2015
- [45] Rothe, R., Timofte, R., & Gool, L.V., "Deep Expectation of real and apparent age from single image without facial landmarks", International Journals of Computer Vision, 144–157, 2016.