Recognition Aid for Visually Challenged to Make Out Indoor Environment

Dr. A. Pasumpon Pandian,
Professor, Department of CSE,
KGISL Institute of Technology,
Coimbatore, India.
Email: pasumponpandian32@gmail.com

Abstract: The applications that are capable of identifying the tangible products and the movement of those product has attained a predominant position in the field of robotics. The complexity in identifying the changes going around is very high on the indoor environments that are too messy. Segregating things and sorting objects in such messy environment becomes even more tedious and challenging for the people with visual disorders. To subdue these issues and enable the blind and the visually challenged to be aware of the changes or the tangible objects they come across in the indoor environment, the proffered method in the paper devise a recognition aid that is empowered with the deep learning neural networks. The usual conventional-CNN is refurbished by upgrading the components to achieve a better accuracy in recognizing. The images based on different scenes of the indoor environment gathered under different circumstance where used as training and the testing dataset for the proffered model and the accuracy and the recognition rate for the training as well as testing was examined.

Keywords: Recognition Aid, Deep Learning Neural Network, Conventional- Convolutional Neural Network (CNN), Indoor Environment, Visually Challenged People

1. Introduction

The recent development in the technology that in turn causes changes in the human life paves way for enormous generation of particulars and details that hold values. This huge progress has caused high complexities in computing as well as the tasks that are associated with the computer vision. Every image in the computer or any other devices are presented in the form of large number of pixels. To devise an aid with the capability of recognizing the circumstance in the messy environments such as the indoor scenarios a deep learning structure, by refurbishing the usual conventional-CNN with the upgraded the components is put forward. The structure is devised with the capacity of the sorting out the "geometric layout and the semantic clues". For instance in case of an "RGB-Image" the entire layout of the image is studied utilizing the AI that is vision centered. It necessary for the mechanism to be updated with the knowledge of the association that lays among the circumstance and the tangible things that are found in it. The

aforementioned are the basic and the primary requirements for a human to sort out and understand the scenario.

The complexity in making out and interpreting a "visual scenario" is very high in the field of AI and CV, so the issue is commonly segregated into dual phase that is fed with dual inputs, (1) images that are static form of input and (2) video's that are dynamic form of input. The AI as well as CV are competent and used in a broader range of applications such as "Independent Robot-Navigation, Independent path finding for Vehicles, and image recovery etc."

Based on the recent reports presented by the World health organization more than one hundred and eighty eight millions of people grieve with visual challenges, two hundred and eighteen suffer from reasonable to severe visual challenges whereas around thirty six thousands of people are almost blind. This indoor interpretation and the making out could be very useful for the visually challenged and the blind to sort out the directions easily in a messy indoor environment by identifying and eluding the hindrances and as well as examining the neighboring environments.

The paper has out forward the intellectual mechanism to aid the blind, visually challenged and the elderly in recognizing the surroundings and move around without any difficulties. According to the population records the count of the elderly would cross more than two billion by the end of 2060.There were many works devised to have a perfect making out and knowledge of the but it was quiet challenging because of the complexity that prevailed in the background of the inputs in static form, high sealing, confusing view points, heavy decorations, and as well as texture and scaling alterations that makes the indoor environment too messy.

As methods of making out he images and sorting out them lays as the primary stage in deducing the contents from the image, the examination becomes tedious. The automated mechanisms based on the AI are more prominent in conducting the highly complicated CV tasks. But the deducing of the indoor scenarios would be helpful for machines and as well as humans to especially the people with the visual challenges to make out the surroundings in the indoor scenario.

To enhance the accuracy in the making out the circumstance the, the paper utilizes the "deep neural learning network" to formulate a recognition aid that scope to clearly and more accurately present the scenarios in the indoor for the blind. This mechanism formulated is suggested to be used along with various embedded devices that are useful in assisting the people with visual challenges and blindness as well as the old people.

The mechanism developed was trained and tested with dataset that utilized the static form inputs collected under various circumstance and different indoor scenarios, to sort out the hindrance in the indoor and find out the path clearly for the blind.

The recognition aid with deep learning for the visually challenged is ordered with the related works in second part, the proposed design in the third, the performance analysis in fourth and conclusion in fifth.

2. Related works

Breuer, Thomas et al [1] related to the identification of the objects in the field of robotics designed a self-capable robot to service in the inland atmosphere. Geiger et al [2] utilized the "kitti vision benchmark suite to design the autonomous driving" Yu, et al [3] devised a "Learning to rank using user clicks and visual features for image retrieval." Rodríguez, et al [4] put forth the "Obstacle avoidance system for assisting visually impaired people."

Khan,et al [5] put forward a "Scene categorization with spectral features." Valanarasu, et al [6] developed a "Smart and Secure Iot and Ai Integration Framework for Hospital Environment." Raj, Jennifer S.et al [7] proposes "A Comprehensive Survey on the Computational Intelligence Techniques and Its Applications." Suma, V et al [8] present the. "Computer Vision for Human-Machine Interaction-Review." And the particulars WHO are presented in [9],

Koresh, M. H., et al [10]. "Computer vision based traffic sign sensing for smart transport." Manoharan, et al [11] presents the "Image Detection, Classification and Recognition for Leak Detection in Automobiles." Bashar, et al [12] presents the "Survey on Evolving Deep Learning Neural Network Architectures."

3. Proposed Work

Humans are capable of understanding and sorting out even highly complicated scenarios accurately, since the CNN are capable of performing equally the proposed work prefers the CNN with few modifications to improve its accuracy in making out. The CNN in the proffered work is "scaled up" to provide an enhanced and accurate performance. The scaling in the CNN is done as two phase, (i) structural-scaling and the network-scaling , where structural-scaling provides a proper balance in the resolution, width and depth and the network scaling offers the improvised functioning, in terms of accuracy. The dimensions of the network

13

are altered evenly utilizing the "Multifaceted Factor" this ensure the accuracy enhancement in recognizing the aid for the blind. Unlike the usual procedure followed in scaling including more layers as of ResNet, the Modified-CNN varies the dimensions of the network that the resolution, width and depth to attain a higher resolution, wider width and deeper depth. This mechanism is well suited for the mobile appliances as they utilize a "very effective methods to gather a minute number of parameters and the computation demands." "The multifaceted scaling represents highly logical components as the static form of input given to it are very large, and the network demands for more layers to heighten the accessible fields and deliver more channels to obtain a perfectly-tuned features, the agents utilized in scaling are not dependent."

The balance in the scaling of the network components are attained using a multifaceted scaling, to obtain a uniform alteration among the parameters using the factor that is multifaceted. The following equations 1, 2 and 3 are framed in this regard.

Where, depth "D" is given by

$$D = A^\theta \text{ For } A \geq 1 \qquad\qquad (1)$$

Where width "W" is represented as

$$W = B^\theta \text{ For } B \geq 1 \qquad\qquad (2)$$

Where Resolution "R" is presented as

$$R = C^\theta \text{ For } C \geq 1 \qquad\qquad (3)$$

While the A, B, C are the values obtained by the "grid search" and the θ is the multifaceted factor. Along with this the alignments and the spots of the indoor environments are also observed. The method put forth is segregated as training process were the complete set is trained using the dataset and the testing process were the system become eligible to make out the environment. The following shows the flow diagram of the training and testing
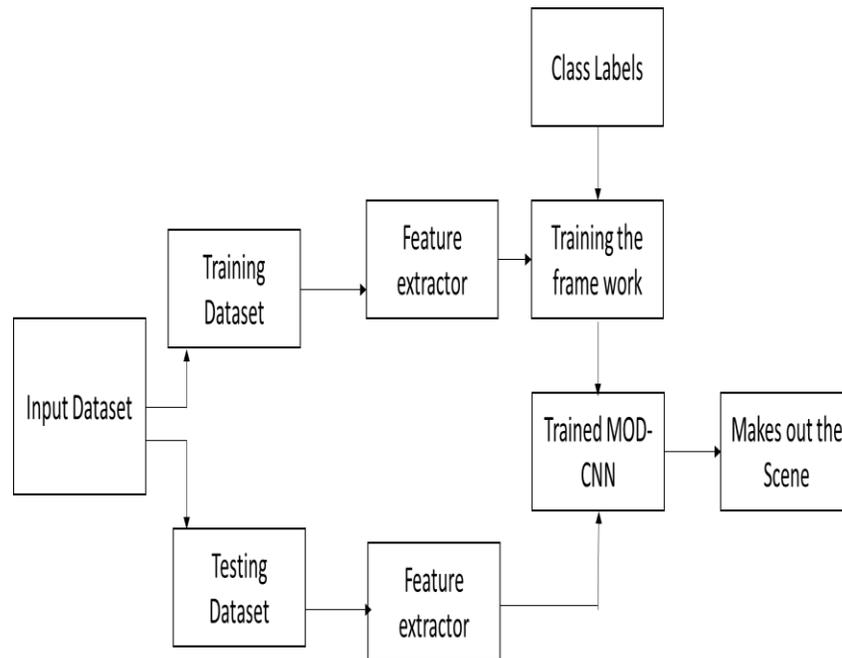
Figure.1 Training and Testing Phase

As shown in figure.1 The training phase incorporates the feature extraction process to make the system knowledgeable and the same feature extractor is used in the testing phase to provide the extracted features to note down how well it sort out the things going around in the environment based on the training done.

The layers of the CNN are devised as shown i n the figure. 2 the modified CNN is composed initially with one normal convolutional layer of 3*3 followed by the altered convolutional layers (M-CONV) that are capable of mapping the encoded features to a lower dimension space for the purpose of minimizing the complications in the computations with size of kernel equal to 6 and the filter size varying from 3*3 to 5*5. Further the pooling layer is employed followed by the fully connected layer.
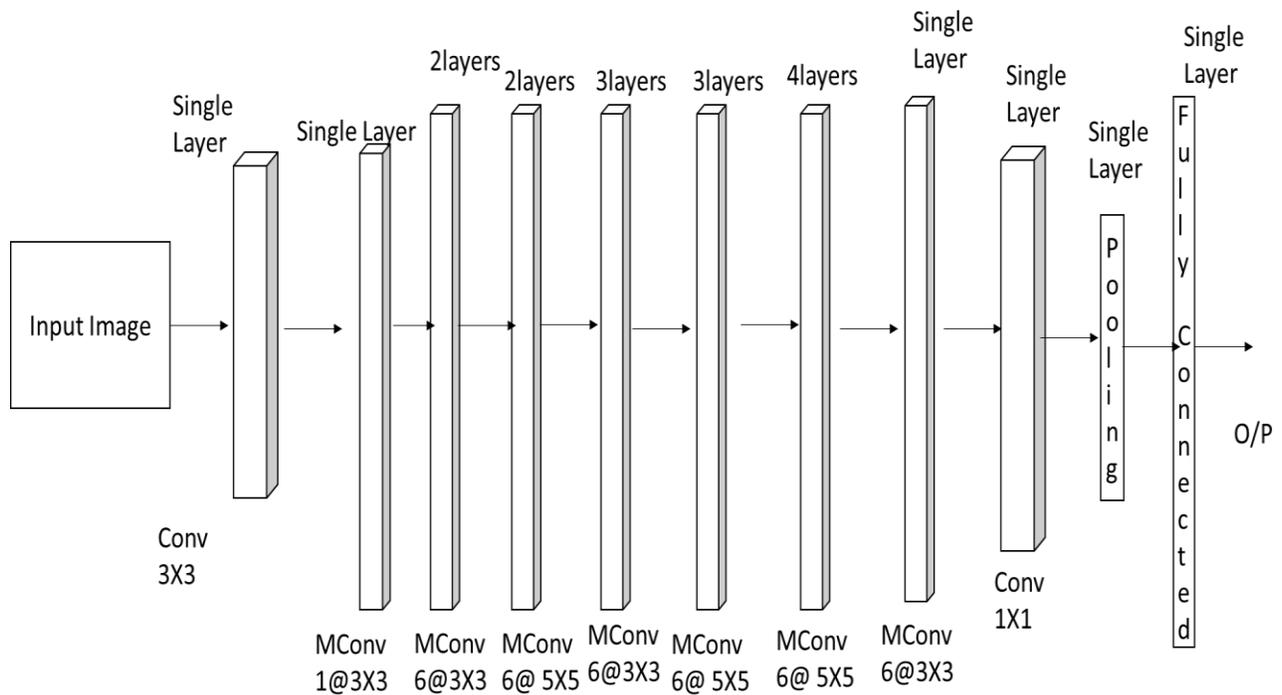
15

Figure.2 Modified-CNN Structure.

The proffered procedure is a highly robust and efficient for making out the scenes in the indoor environment, but still there are few issues faced in the methodology due to the over fitting problem, so to elude the problem the future work of the paper is carried out utilizing he Caps Net to attain an even higher accuracy than the proffered.

4. Performance Analyses

The proffered was evaluated on the own dataset created by gathering images of the indoor environments under different scenarios such as , reception, sit out, living area, utility area, rest room kitchen and the bed room. The procedure to devise a perfect indoor recognizer for the visually challenged and the elderly were designed using the "python as well as the tensor-flow", further the analyses was carried out in the laboratory with encompassed with "E5-2683 V4 processor, and tesla graphic processor, with the memory of 12 GB". The modified convolution layers used, allowed to utilize the loss function (cross entropy) in the training

stage to determine the variations in the probability of the happenings taking place in the indoor environment and its destination and minimizes the variations.

The deep learning process that relies on the transfer learning is utilized to frame a robust recognition, the accuracy of the system in the training and the testing was evaluated based on the images sets that were provided as input. The parameters used in the analyses is shown in the table.1 below

Parameters	Values
Iterations for Training	12,000
Rate of Learning	.011
Training Dataset	70%
Testing Dataset	30%

Table.1 Experiment Parameters

The results observed in the figure.3 shows the accuracy in the testing and training achieved by the model put forth based on the images gathered under various scenarios in the indoor environment for testing.
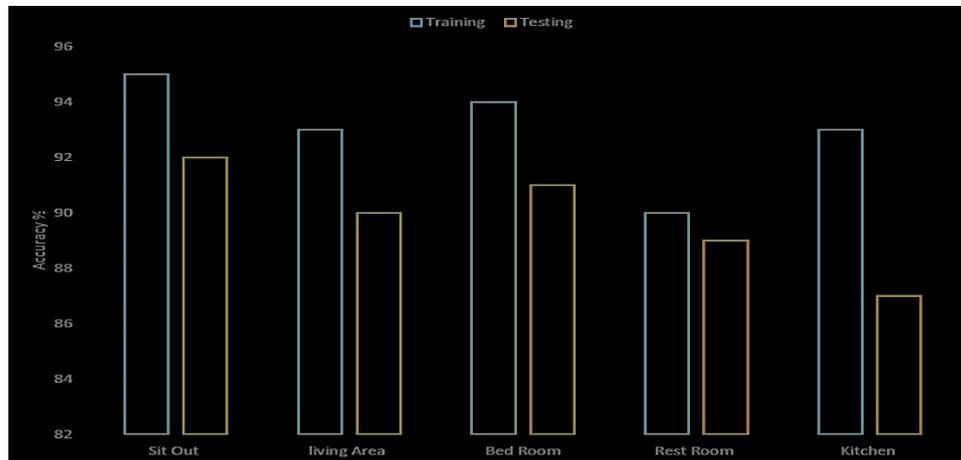

Figure.3 Training and attesting Accuracy

17

The Table.2 below provides the comparison of the proposed model with the "deep filter bank [5], places VGGnet +Spectral Feature [5], the feature level FOSNET-CCG and Mixed CGM-CCG [8] on the dataset framed based on the images gathered from the utility area of the house.

Dataset	Proposed		Deep Filter Bank		VGGnet+Spectral features		FOSNET-CCG		FOSNET-CCM-CCG	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Kitchen	95	92	80	72	82	75	89	80	90	88
Bed Room	93	90	82	74	82.45	76	88	80.12	90.1	85
Rest Room	94	91	80.1	72.5	82.78	76.45	87.5	80.35	90.13	87
Sit Out	90	89	80.23	72.56	82.88	76.49	86	80.78	90.29	88
Living Area	93	87	80.34	72.89	83	77	87	83	90.3	89

Table.2 Comparison of Results on the Dataset Used

5. Conclusion

Since the making out of the indoor environment are still complex in the CV the proposed frame work utilizes the Modified-CNN with the deep learning capability to recognize the happenings and hindrances that affect the smooth movement of the people who are visually challenged and the elderly in the indoor environment. The proposed method was evaluated based on the accuracy attained in the training and the testing. The results observed were compared with the existing system, and proved to perform better, but still certain issue prevailed due to the problem of over fitting, to overcome the issue the future work of the paper is to develop a recognizer using the Caps Net to further improve the accuracy of recognition.

References

[1]   Breuer, Thomas, Geovanny R. Giorgana Macedo, Ronny Hartanto, Nico Hochgeschwender, Dirk Holz, Frederik Hegger, Zha Jin et al. "Johnny: An autonomous service robot for domestic environments." *Journal of intelligent & robotic systems* 66, no. 1-2 (2012): 245-272.

[2]   Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354-3361. IEEE, 2012.

[3]   Yu, Jun, Dacheng Tao, Meng Wang, and Yong Rui. "Learning to rank using user clicks and visual features for image retrieval." *IEEE transactions on cybernetics* 45, no. 4 (2014): 767-779.

[4]   Rodríguez, Alberto, Luis M. Bergasa, Pablo F. Alcantarilla, Javier Yebes, and Andrés Cela. "Obstacle avoidance system for assisting visually impaired people." In *Proceedings of the IEEE Intelligent Vehicles Symposium Workshops, Madrid, Spain*, vol. 35, p. 16. 2012.

[5]   Khan, Salman H., Munawar Hayat, and Fatih Porikli. "Scene categorization with spectral features." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5638-5648. 2017.

[6]   Valanarasu, Mr R. "Smart and Secure Iot and Ai Integration Framework for Hospital Environment." *Journal of ISMAC* 1, no. 03 (2019): 172-179.

[7]   Raj, Jennifer S. "A Comprehensive Survey on the Computational Intelligence Techniques and Its Applications." *Journal of ISMAC* 1, no. 03 (2019): 147-159.

[8]   Suma, V. "Computer Vision For Human-Machine Interaction-Review." *Journal of trends in Computer Science and Smart technology (TCSST)* 1, no. 02 (2019): 131-139.

[9]   https://www.who.int/world health orgaization

[10]   Koresh, M. H., and J. Deva. "Computer vision based traffic sign sensing for smart transport." *J. Innovative Image Process.(JIIP)* 1, no. 01 (2019): 11-19.

[11]   Manoharan, Samuel. "Image Detection, Classification and Recognition for Leak Detection in Automobiles." *Journal of Innovative Image Processing (JIIP)* 1, no. 02 (2019): 61-70.

[12]   Bashar, Abul. "Survey on Evolving Deep Learning Neural Network Architectures." *Journal of Artificial Intelligence* 1, no. 02 (2019): 73-82.