March - April 2020
ISSN: 0193-4120 Page No. 5111 - 5119

Deep Belief CNN Based Artificial Vision

Tathagat Banerjee[1], KarthikeyanS[2] ,RohitBhargav Peesa[3], Priyanka Nair[4]
[2]Assistant Prof,
[1,2,3,4]School of Computer Science and Engineering, VIT-AP University, Amaravati, India
[1]banerjee.tathagat@vitap.ac.in, [2]karthikeyan.s@vitap.ac.in, [3]rohitbhargav.p@vitap.ac.in, [4]priyanka.nair@vitap.ac.in, ,

Article Info
Volume 83
Page Number: 5111 - 5119
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 27 March 2020*

Abstract

Data Analytics and Deep Learning have always tried to establish its significance to the medical scientific community. A decade ago, due to the lack of computational resources, its numerous efforts have not been able to embrace its name significantly. Today the medic community is largely astonished by the supreme pattern understanding and predictive analytics that data science under the names of deep learning and machine learning. Here in this paper we are presenting an architecture named vision. It has been used to establish a model design that can help a visually disabled people to be directed to their destination, by predicting Forward, Left and Right direction at each road step. We have maintained class division, non-overfitting and regularization at each step of our deep belief convolutional neural network which we get to know by high values for different recall and precision classes and on the other hand, the usage of Fully connected convolutional layers poses feature extraction techniques from three-dimensional images. The luxurious medical treatments are often out of reach of common public including some lethal risk factors for life, this algorithm not only solves the problem of navigation for the disabled but also path breaks this new arena of research and development. Transfer learning algorithms VGG and Inception which could attain an accuracy of 52 percent for real-time data even the results for precision and recall for low and with our architecture, we could gain accuracy of 91.41 percent along with precision and recall at an average of 89 percent.

Keywords; *Neural networks, Machine Learning, ANN, CNN, F-CNN, Transfer Learning, Vision*

I. INTRODUCTION

Human eye in the presence of light and functional rod and cone cells allows vision and it is able to differentiate numerous different colours which cherishes a scenic beauty, however there are some people who are unfortunate to witness the visual pleasure. A responsible human must try to assist them with the combination of knowledge and technology (by using AI) for the welfare of visually impaired people. The condition of poor visual perception or lack of complete vision is termed as blindness in humans. They cannot detect light source. Blindness is defined as visual defectiveness of less than 20/400, or a visual field loss to less than 10 degrees, in the better eye with best possible correction in human beings.

As per the survey of 2012, the report stated that the number of visually impaired people has increased to 285 million in the world, of which 246 million have low or partial vision and 39 million are completely blind. It was noted that the majority of people with low or poor vision are elder citizens over the age of 50 years and which precisely reside in developing nations.

Loss of vision can cease an individual to experience color and it cannot halt someone's dreams and aspirations. There are some blessed and talented famous personalities who were visually impaired but their will and knowledge made a humongous appreciation all around the world to name a few: Louis Braille - Known for Braille , Helen Keller - American lady who was deaf-blind writer, lecturer,

Published by: The Mattingley Publishing Co., Inc.

and even leading activist , Tiffany Brar -A great Social activist, who founded the Jyothirgamaya Foundation, HabenGirma - She changed history by becoming Disability rights advocate, first ever deaf blind graduate of Harvard Law School , Jacob Bolotin - the world's first complete blind physician fully licensed to practice medicine , GustafDalén - A great Swedish inventor and Nobel Prize winner , Dr. SatishAmarnath -first Indian Medical Microbiologist who became blind after an acid attack on his person but didn't give up on his passion.

Optical treatment is considered as complex and costly surgery in Medical Science. It can fulfil an individual's dream to see but, at the cost of their health. This paper proposes a better alternative which will be their 'vision' by navigating them to destination. Dataset used for this training this model is a dataset from Kaggle[7].

Transfer Learning is a technique where a model developed for one application or task makes use of the knowledge gained and applies it for another task. Model is based on transfer learning which was able to navigate visually impaired people in forward, left and right directions by detection [2] of direction signs considering them to be completely blind.

This paper consists of five more sections organized as follows: section II which describes the literature survey carried out related to the work specified in this paper; section III which provides insights regarding the implementation of model; section IV that describes the proposed architecture; section V consists of graphical plots related to accuracy of model; section VI the final conclusion on the model.

II. LITERATURE SURVEY

Computer vision is a field in today's technological world in which several scientists and researchers are striving to improve and invent great scientific inventions. Few methodologies and papers exist related to navigation for the visually impaired people based on appearance features and geometric properties. There are few navigation systems with cane, sunglasses, caps as guides which generally consist of GPS, cameras, speakers and ultrasonic sensors for obstacle avoidance as in [2-5] and few which provide robots assistants as alternative to dogs, these robots have the capacity to guide the ability to enhance GPS module to help , impaired reach the destination [6].

YingliTian et al. [2] is an approach which explains about door detection in environments of which person is unfamiliar with, using a single camera and a computer. Door detection is done by generic edge and corner detection instead of regular feature-based methods which detect based on the color, texture, etc since edges and corners in an image are the stable features when compared to appearance features like color, size. There are many existing technologies like GPS for navigation but they are hindered for indoor usage and only have been successful in navigating people in outdoor environment. The methodology followed here was visual information can be captured through camera mounted in sunglasses or in cap after which image processing part takes place in the computer and speech as an output would be provided in real time different neural architecture are implanted to serve this idea. We have made use of stable features like edges and corners along with appearance features like shapes, arrow sizes and majorly used colors for direction arrows.

Joao José et al.[3] have designed a small, easy to use prototype of smart stick which works as a navigation aid to the blind. This smart stick with the help of global navigation could guide its users to destination. It also has local navigation for sidewalks, corridors and it could prevent collision from obstacle which were static like a wall, door, etc as well as moving obstacles like animals, humans, etc. This device also consisted of a stereo camera that should be worn on the chest, a laptop in shoulder-strapped pocket slinging to shoulder belt rand speaker which would be the input to guide the user. Being similar to the regular white cane it was

March - April 2020
ISSN: 0193-4120 Page No. 5111 - 5119

neither an obstacle for others nor a heavy machine that could not be carried.

Vijay John et al. in paper [4] proposed a robust vision-based traffic light and arrow detection algorithm for vehicles [4]. In the paper, they detect all the three colors red, yellow and green traffic lights [5] along with the arrows which show directions. They have made use of pre-trained convolutional neural networks which was trained for the ImageNet classification, to localize traffic light's region of-interest (ROI) within the image captured by a monocular camera with the help of pre-generated saliency maps as constrain and then detect the color. Along with pre-generated saliency maps as constraint they also proposed to use optimal camera parameters consisting of gain, shutter speed, color balance and white balance for real-time image acquisition to enhance the TL detection accuracy. After detecting the color their algorithm is also capable of arrow by bounding box method. This model is capable to work in various illumination, environmental conditions and background noise

J. Park et al. in paper [5] proposed robot navigation using camera by identifying the presence of arrow signs. This paper was a feature-based approach to deal with image processing algorithms where navigation is done by identifying the arrows signs. The images are captured through the wireless camera. They have selected four different features namely Convex Area, Extent, Solidity, Eccentricity to isolate the arrow region in the image from the rest of the regions. The image is initially converted to binary image and then denoised, later it is passed into region segmentation and identification image is segmented to get the region of interest after which it is identified as respective arrow and after making the decision to either turn left or right based on the direction of the arrow signs finally the robot turns to respective directions.

III. IMPLEMENTATION

Preprocessing of image is the primary step in model. We have made use of Gaussian Blur to remove unwanted noise which deals with blurring image by a Gaussian function.

The implementation basically deals with basically 10 different layers, their idea, sight and benefits. As discussed earlier the image quality used for training purpose is of 128 x 128 x 3, so even before any layer training a max pooling layer of 2 x 2 is fitted. Max pooling is basically a simple mathematical operation used to sight fully reduce the dimensionality in such a way that no decrement to quality or importance is caused. Figure 1 below displays max pooling.

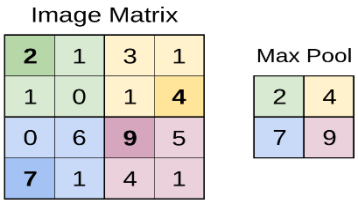


Fig-1. Max pooling on Image

The options available are minimum and average pooling, mathematically the requirement of max pooling is because on naturally taken images the gradients growth is simultaneous to important patterns and characteristics, thus maximum value is chosen via max pooling. After this the left-over image is 64 x 64 x 3.

Figure 2 below represents the flowchart of the algorithm used in our vision architecture. Architecture consists of pooling, flatten, dense, dropout layers along with Conv 2D and ANN.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 5111 - 5119

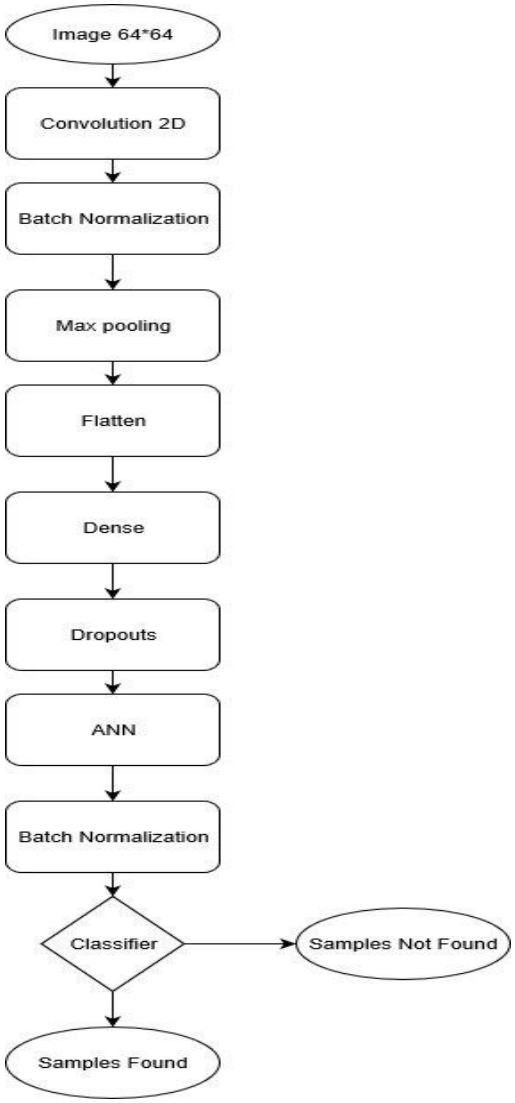


Fig-2. Flowchart of algorithm

Finally, the Vision architecture and the utility of each layer in the following is as follows:

1. Convolutional 2-D layer, helps to find insights from the images, which works in order range through the complete image matrix through the kernel convolute and create a new feature matrix.

2. Batch normalization, is applied after every layer since data is normalized data which ranges from scale 0 to 1 for a data which ranged from 0 to 255 earlier, irregularity in weights may cause on to dominate others which in turn affects the results.

3. Dropout is an important step to enhance models which do not overfit, it just randomly cancels out nodes in each iteration so that no node is of great importance to the models, neither single input features are heavily weighted, since under each backpropagation iteration the weights are decreased when dropped and failure of output occurs.

4. Flatten is used to convert the 3-dimensional matrix to one dimension. This is done in order to make the data eligible to be fitted in Artificial neural networks. However, it is important to note when training with more Convolutional layers flatten must be used only after the last Convolutional layer. Once flatten is used no more convolutional layers can be used.

5. Dense Directly refers to the Artificial neural network schema which is 64 nodes in one layer and 1 i.e. the output node in the last layer.

6. Hyper parameter Usage

All dropouts are 25% i.e. one fourth, so that overfitting can be removed.

a. For regularization, mostly Manhattan distance or L1 regularization is used.

b. The activation function are Leaking Rectilinear Units and Rectilinear units.

c. In last or the output node Sigmoid function is used.

d. Adam Optimizer function is used to find the global minima and not get stuck in local minima.

e. Binary Cross Entropy is used for loss function and metrics of evaluation is based on accuracy.

The model has been trained over 22,400 data images for 30 epochs, number of epochs are less as it is quite much as important to not to cause overfitting.

ReLu computes maximum value of (x,0) where x is the input of neuron. ReLu reduces computational cost, accelerates Stochastic Gradient Descent (SGD) [1]. However, it suffers a few drawbacks due to

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 5111 - 5119

large gradient sometimes weights are set too high for it to get activated. This issue is resolved by Leaky ReLu which instead of setting negative values as zeros replaces them by multiplying x with small values α (where α<1) and finds max(αx,x) [1] . Moreover combination of ReLu and Leaky ReLu avoids saturation

IV. PROPOSED ARCHITECTURE

The proposed architecture in figure 3 demonstrates various activation functions and training parameters used for analysis. The colored boxes showcase the dropout implementation.

The final classification uses SoftMax function and distinguish between three sets of data category.

An image of size 64x64 is sent to preprocessing unit so as to clean up the noise and apply basic preprocessing principles

and sliding window so as to find arrow in image. Later the image is sent a 2D convolutional neural network which has activation function tanh

Batch normalization and L1 regularization are applied after which is sent to max pooling whose process is shown in figure 1. Then, they are activated by ReLu activation function after which the image is flattened and sent it

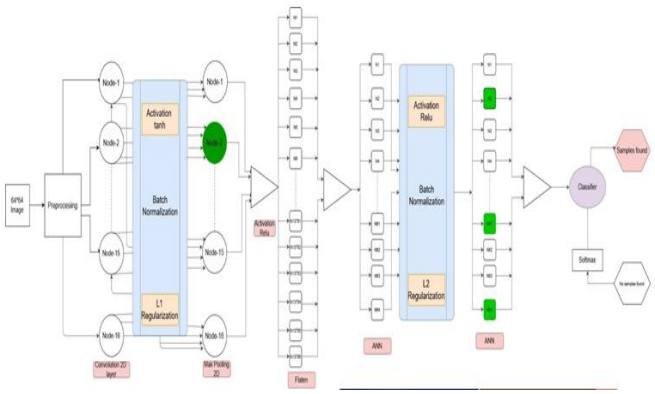


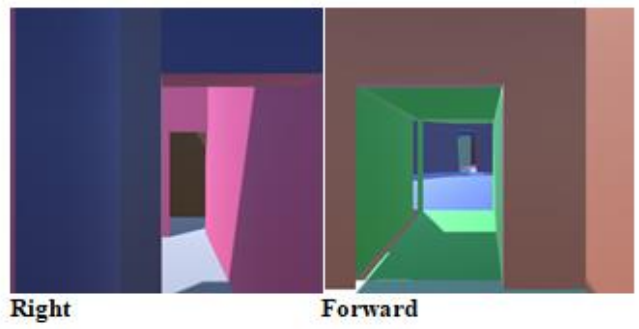
Fig 3: Architecture of the model

through ANN where again we apply batch normalization, L2 regularization and activate the outputs using ReLu and send it to ANN afterit to ANN after which classifier checks whether detected arrow is matching left, right or forward arrow using its previous experiences. If they are found then the output would be arrow detected and if they are not detected, output is sent to be activated using SoftMax activation function to give another try and still if they are not found then there would be no output.

Dataset Insights:

A human first recognize the view in front of him and then decides an action by taking Step in forwarding, sideways, left, right or more complex movement. The same is done during data gathering stage i.e. we captured image first and labeled as forward if the agent can move forward or right if he needs to turn right to go out of the door/room.

The dataset contains 3,245 direction images with unequal instances of different categories biased towards forward. Further, we have modified the data set with three different categories as Front, Left and Right. There we use the concept that is if we allow three independent categories to train through train generator we shall run into a problem called continuous labeling i.e. a long sequence of the single category might train the model to learn the single category only. Thus, we keep randomly available different images in such a way that all categories co-exist over a linear interval of time



Hence, by this, we not only dealt with the problem of class division, label series but also created a shuffled distributed and non-skewed data, for our predictive analysis.

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 5111 - 5119

The typical workflow consists of approach of Computer vision in the following way: The data for first person walking simulator is gathered in a game environment where room's maps are created and images are captured and labelled walk by Forward, Left and Right classes. Typically human walk is more complicated, but we in this stage only collected three type of action. In the aim to predict the direction to come out of the room.

V. RESULTS

1.Accuracy

The model has been trained the model with VGG, Mobile net and Inception but none of the model good give us expected results when applied on real time data highest was inception which could only attain a maximum accuracy of 52.3 percent on real time data it. The architecture has been modified as per the requirements by stacking up a few more layers of deep neural networks which was of type deep belief nets named as vision. Vision architecture could attain an accuracy of 93.37% percent and 92.34% percent on validation data and test data respectively. However, it was 73.6 percent applied on real time data better in comparison to other models.

2.Classification Report

As per the report classification score i.e figure 5 for predicting majorly all fields are above 85% which is of great advantage to fast growing medical field.

The F1 score is pretty much balanced at 91% over all. We could also see a low of percentage of 84% at different recall and precision classes. This suggests that class division problem, overfitting and underfittinghas been reduced to a major extent.

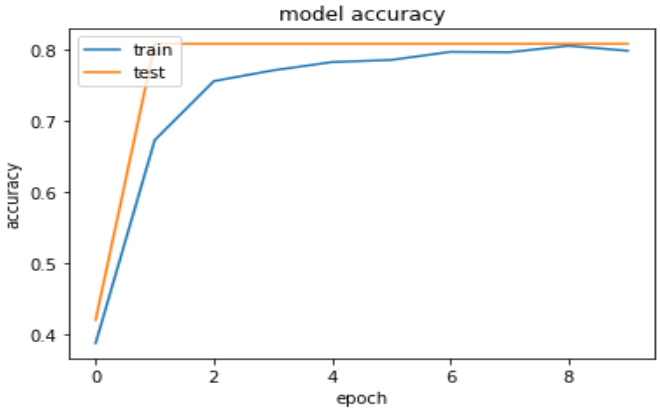


Fig 4: Accuracy vs Epoch

	precision	recall	f1-score	support
Forward	0.98	0.74	0.84	238
Left	0.84	0.99	0.91	177
Right	0.86	0.97	0.91	218
accuracy			0.89	633
macro avg	0.89	0.90	0.89	633
weighted avg	0.90	0.89	0.89	633

Fig 5: Classification Report of the model

N=3245	Predicted No	Predicted Yes	
Actual No	TN=758	FP=218	976
Actual Yes	FN=135	TP=2134	2269
	893	2352	

N=3245	**Predicted No**	**Predicted Yes**	
Actual No	**TN=758**	**FP=218**	**976**
Actual Yes	**FN=135**	**TP=2134**	**2269**
	893	**2352**	

Table 1: Confusion Matrix

True positive (TP) is an outcome where the model correctly predicts the positive class. [16]

True negative is an outcome where the model correctly predicts the negative class. [16]

False positive is an outcome where the model incorrectly predicts the positive class. [16]

False negative is an outcome where the model

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 5111 - 5119

incorrectly predicts the negative class. [16]

F1 Score=2TP÷(2TP+FP+FN)

Accuracy=(TP+TN)÷(TP+TN+FP+FN)

Precision=TP÷(TP+FP)

Recall=TP÷(TP+FN)

VI. Loss and Accuracy Curves

3.1 Accuracy Plot

From the accuracy figure it could be definitely be concluded that the growth of the training accuracy is rather linear than the test accuracy, hence it shows how variant and multidisciplinary the test set is being used. However, the steady and aggressive growth chain of both train and test demonstrate normal traits of a neural network. Long and multi-layer deep belief network training the model has been able to generalize the traits shown in not only on the train-set but has gathered a good insight of test-set even.

The Final training shows the rise of test and training accuracies, the high kurtosis point is still a fair enough statement to quote as peaks and falls over duration of about 300 epochs suggest the randomness, variance and skewness of the data. We are also able to vividly demonstrate the de-normalized distribution of the data.

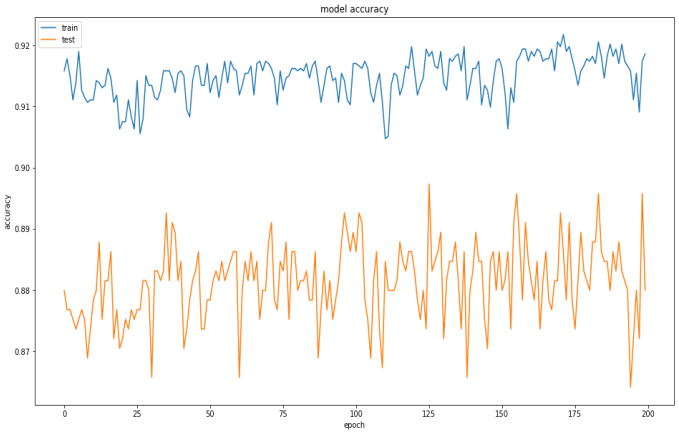


Fig 6: Final Accuracy vs epochs

3.2. Loss Plot

The curves suggest the reach of lowest possible point for test set; however, train set seems to have raised up by just a fraction. This tradeoff is irreplaceable and has to be encountered with potential skewness and kurtosis is evenly dispersed because of linear pulse kind of modulation. The plots demonstrate that Global minima is achieved and lowest point loss is targeted.

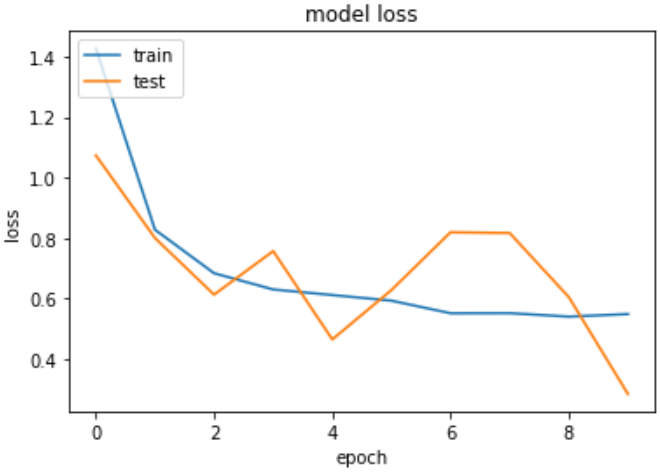


Fig 7 : Loss vs Accuracy Final

3.3. Training and validation Loss

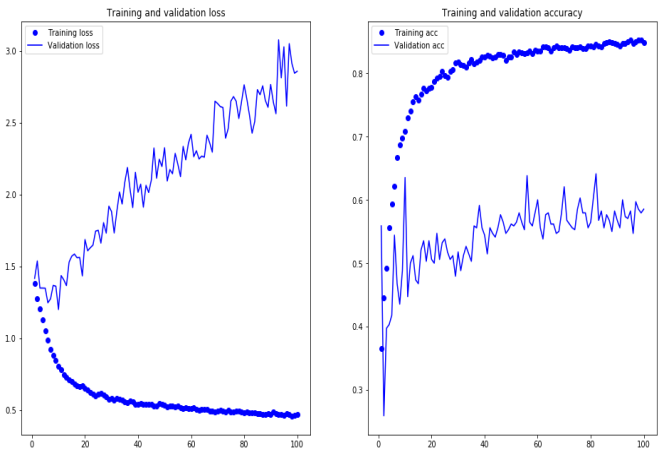


Fig 8 : Training and validation Loss

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 5111 - 5119

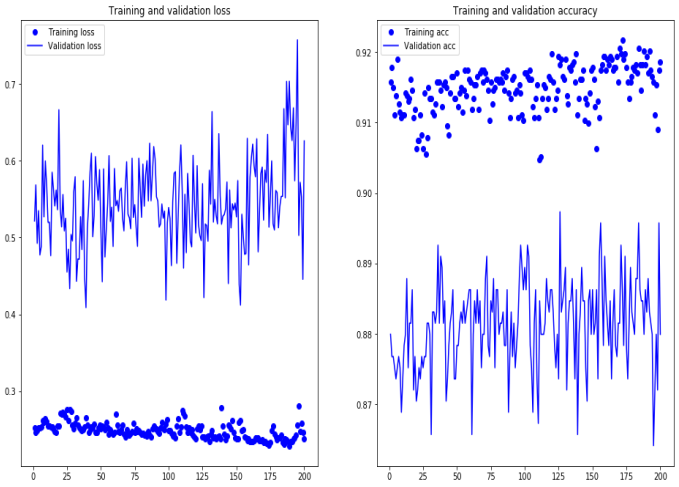


Fig 9: The graphical final representation of validation to training loss graph

The peak finder path has been taken by the algorithm and loss has been minimized even for the validation set of results. It can be said that the kurtosis and skewness observed until 100 epoch paths is still less than that observed in the test scenario.

VII. CONCLUSION

The understanding of vision to the disabled field where Data science and Deep learning is still developing.Over the already present models which have been discussed about tackling images of direction by artificial neural networks and transfer learning have shown that accuracy scores have been limited to at max 67% of dependable accuracy. However, proposed model has been able to formulate and achieve accuracy of 88.6% and instance of test score of 89.3%. This model set is open sourced under the name by Vision model. The Adam optimizer and loss function Sparse categorical cross entropy can be further optimized or modified for hyper parameter tuning, however even in its present state in the need of the hour our model is able to accurately determine prediction to different direction, as the project Vision. The social and economic welfare targeted about is of heavy importance over to which further enhancement and study can bring many modifications to architecture

REFERENCES

[1]. Lecture notes,CS231N, Stanford University, 2015. http://cs231n.github.io/

[2]. YingliTian,Xiaodong Yang, Aries Arditi,K. Miesenberger et al.,Computer Vision-Based Door Detection for Accessibility of Unfamiliar Environments to Blind Persons, ICCHP 2010, Part II, LNCS 6180, pp. 263–270, 2010.

[3]. José, Miguel Farrajota, Joao M.F. Rodrigues (2013), “A Smart Infrared Microcontroller Based Blind Guidance System”, Hindawi Transactions on Active and Passive Electronic Components,Vol.3, No.2, pp.1-7, June 2013.

[4]. Vijay John, L. Zheming and S. Mita, "Robust traffic light and arrow detection using optimal camera parameters and GPS-based priors," 2016 Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), Tokyo, 2016, pp. 204-208.

[5]. J. Park, W. Rasheed and J. Beak, "Robot Navigation Using Camera by Identifying Arrow Signs," 2008 The 3rd International Conference on Grid and Pervasive Computing - Workshops, Kunming, 2008, pp. 382-386.

[6]. Megalingam R.K., Vishnu S., Sasikumar V., Sreekumar S. (2019) Autonomous Path Guiding Robot for Visually Impaired People. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore

[7]. https://www.kaggle.com/hamzafar/look4me

[8]. M. F. Haque, H. Lim and D. Kang, "Object Detection Based on VGG with ResNet Network," *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, Auckland, New Zealand, 2019, pp. 1-3.

[9]. Nan Wang, Wei Liu, Chunmin Zhang, Huai Yuan and Jiren Liu, "The detection and recognition of arrow markings recognition based on monocular vision," *2009 Chinese*

Published by: The Mattingley Publishing Co., Inc.

Control and Decision Conference, Guilin, 2009, pp. 4380-4386.

[10]. D. Sil, A. Dutta and A. Chandra, "Convolutional Neural Networks for Noise Classification and Denoising of Images," *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, Kochi, India, 2019, pp. 447-451.

[11]. L. Wendling and S. Tabbone, "A new way to detect arrows in line drawings," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 935-941, July 2004.

[12]. V. John, L. Zheming and S. Mita, "Robust traffic light and arrow detection using optimal camera parameters and GPS-based priors," *2016 Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, Tokyo, 2016, pp. 204-208.

[13]. G. Maier, S. Pangerl and A. Schindler, "Real-time detection and classification of arrow markings using curve-based prototype fitting," *2011 IEEE Intelligent Vehicles Symposium (IV)*, Baden-Baden, 2011, pp. 442-447.

[14]. R. Kulkarni, S. Dhavalikar and S. Bangar, "Traffic Light Detection and Recognition for Self Driving Cars Using Deep Learning," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-4.

[15]. S. Prasad and S. Sinha, "Real-time object detection and tracking in an unknown environment," *2011 World Congress on Information and Communication Technologies*, Mumbai, 2011, pp. 1056-1061.

[16]. Classifications notes, Google's Machine learning crash course