

A
Project Report
on
Sign Language detection using Deep learning

*Submitted in partial fulfillment of
the requirements for the award of the degree of*

Bachelor of Technology
in
Computer Science and Engineering

Submitted by
Sridhar Prasad Dash 1801106536
Srijeeth Panda 1801106537
Debashish Behera 1921106026

Under the guidance of

Mrs. Subhalaxmi Das
OUTR-B

Department of Computer Science and Engineering

Odisha University of Technology AND Research
Formerly College of Engineering & Technology
Bhubaneswar, Odisha - 751029

Department of Computer Science and Engineering

Odisha University Of Technology and Research, Bhubaneswar

Certificate

This is to certify that the thesis entitled **Sign language detection using Deep learning** is submitted by Sridhar Prasad Dash (1801106536), Srijeeth Panda (1801106536), Debashish Behera (1921106026) to the Department of Computer Science and Engineering , Odisha University of Technology and Research formerly College of Engineering and Technology, Bhubaneswar, is a record of bonafide research work under our supervision and we consider it worthy of consideration for partial fulfilment of the requirements for the award degree of Bachelor in Technology in Computer Science & Engineering under Odisha University of Technology and Research, Bhubaneswar.

Mrs. Subhalaxmi Das
(Project guide)

(Examiner)

Acknowledgement

It is our privilege and solemn duty to express our deepest sense of gratitude to Mrs Subhalaxmi Das, Lecturer, under whose guidance We carried out this work. We are indebted, for her valuable supervision, heart full cooperation and timely aid and advice till the completion of the thesis in spite of her pressing engagements. We would like to express our deep sense of gratitude to Mrs Jyotirmayee Routray, Coordinator of the Department for encouragement and inspiration throughout our project work. I wish to record my sincere gratitude to our respected Head of the Department, Mr Subhasish Mohapatra for his constant support and encouragement in preparation of this thesis.

We take this opportunity to express our hearty thanks to all those who helped us in the completion of our project work. We are very grateful to the author of various articles on the Internet, for helping us become aware of the research currently ongoing in this field.

We very thankful to our parents for their constant support and love. Last, but not the least, We would like to thank our classmate for their valuable comments, suggestions and unconditional support.

.

THANK YOU ALL.

1801106536 Sridhar Prasad Dash
1801106537 Srijeeth Panda
1921106026 Debasish Behera

Declaration

We certify that

- i The work contained in the thesis is original and has been done ourself under the general supervision of our supervisor.
- ii The work has not been submitted to any other Institute for any degree or diploma.
- iii We have followed the guidelines provided by the Institute in writing the thesis.
- iv Whenever We have used materials (data, theoretical analysis, figures, text) from the other sources, We have given due credit to them by citing them in the text of the thesis and giving their details in the references.
- v Whenever We have quoted written materials from other sources, We have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Date: 18th April 2022

1801106536 Sridhar Prasad Dash
1801106537 Srijeeth Panda
1921106026 Debasish Behera

Abstract

The goal of this project was to build a neural network able to classify which letter of the American Sign Language (ASL) alphabet is being signed, given an image of a signing hand. This project is a first step towards building a possible sign language translator, which can take communications in sign language and translate them into written and oral language. Such a translator would greatly lower the barrier for many deaf and mute individuals to be able to better communicate with others in day to day interactions. This goal is further motivated by the isolation that is felt within the deaf community. Loneliness and depression exists in higher rates among the deaf population, especially when they are immersed in a hearing world . Large barriers that profoundly affect life quality stem from the communication disconnect between the deaf and the hearing. Some examples are information deprivation, limitation of social connections, and difficulty integrating in society. Most research implementations for this task have used depth maps generated by depth camera and high resolution images. The objective of this project was to see if neural networks are able to classify signed ASL letters using simple images of hands taken with a personal device such as a laptop webcam. This is in alignment with the motivation as this would make a future implementation of a real time ASL-to-oral/written language translator practical in an everyday situation.

Keywords: Convolutional Neural Network, ReLU activation, Max Pooling , YoloV3 , Fast-RCNN

Content

CERTIFICATION	2
ACKNOWLEDGEMENT	3
DECLARATION	4
ABSTRACT	5
CHAPTER-1	
History of object detection.....	8
1.1.History of object detection	8
1.2.YOLO V3.....	8
1.3.Faster mask R-CNN.....	10
CHAPTER-2	
Literature survey.....	12
CHAPTER-3	
Workflow.....	13
CHAPTER-4	
Bounding Box detection.....	14
4.1.Bounding Box.....	14
4.2.Detection.....	14
CHAPTER-5	
Image segmentation.....	15
5.1.Image segmentation.....	15
5.2.Proposed Method.....	15
5.3.Edge based segmentation.....	15
CHAPTER-6	
Key point Detection.....	17
6.1.Key point.....	17
6.2.CNN Architecture.....	17
CHAPTER-7	
Sign language detection.....	20
7.1.Gesture/Action detection.....	21
CHAPTER-8	
8.1.Conclusion & Future scope.....	22
8.2.Reference.....	23

List Of Figures

Fig no	Figure name	Page no
1	History of object detection	8
2	Yolov3 architecture	8
3	Yolov3 feature extraction	9
4	Faster R-CNN Architecture	10
5	Bounding box output	14
6	Edge detection	16
7	Image segmenation output	16
8	Key point detection neural network	17
9	Enhanced Neural net	19
10	Key point output	19
11	Confusion matrix	21
12	Sign language detection	21

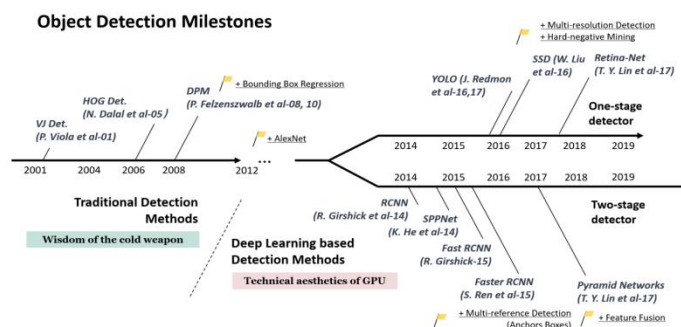
INTRODUCTION

1.1. HISTORY

Image grouping includes appointing a class mark to a picture, while object localization includes drawing a jumping box around at least one articles in a picture. Object identification is really difficult and joins these two errands and draws a jumping box around each object of interest in the picture and allots them a class mark. Together, these issues are alluded to as protest acknowledgment. Object acknowledgment is alludes to an assortment of related undertakings for recognizing objects in advanced photos.

District Based Convolutional Neural Networks, or R-CNNs, are a group of procedures for tending to protest restriction and acknowledgment assignments, intended for model execution.

You Only Look Once, or YOLO, is a second group of methods for object acknowledgment intended for speed and continuous use

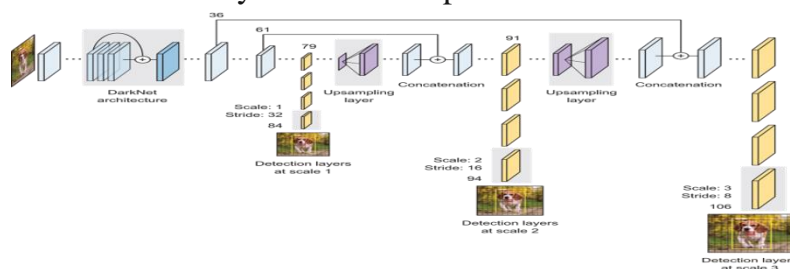


(Fig:1 History of Object detection)

The two state of the art object detection neural networks are YOLOV3 and mask Faster-RCNN. These two techniques are thoroughly discussed below.

1.2. YOLO V3

"You Only Look Once" or YOLO is a family of deep learning models designed for fast object Detection. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.

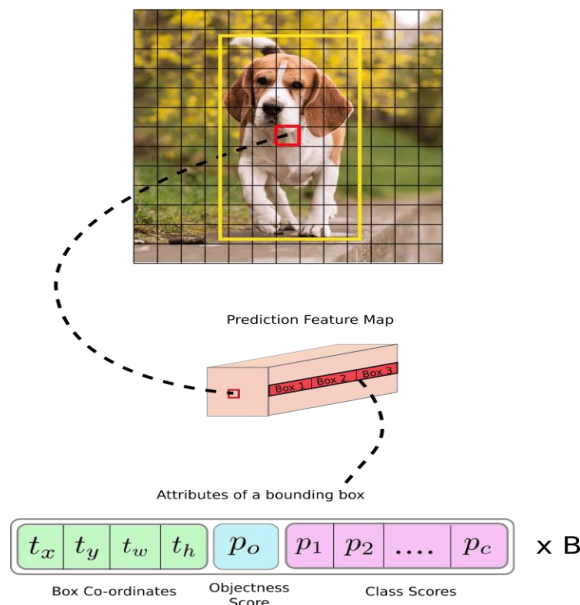


(fig:2 Architecture of Yolo v3)

Ventures for object Detection utilizing YOLO v3:

The data sources is a group of pictures of shape $(m, 416, 416, 3)$. YOLO v3 passes this image to a convolutional neural network (CNN). The last two elements of the above yield are straightened to get a result volume of $(19, 19, 425)$. Here, every cell of a 19×19 lattice returns 425 numbers. $425 = 5 * 85$, where 5 is the quantity of anchor boxes per lattice. $85 = 5 + 80$, where 5 is (pc, bx, by, bh, bw) and 80 is the quantity of classes we need to identify. The result is a rundown of bounding boxes alongside the perceived classes. Each jumping box is addressed by 6 numbers (pc, bx, by, bh, bw, c) . Assuming that we grow c into a 80-layered vector, each bouncing box is addressed by 85 numbers. At last, we do the IoU (Intersection over Union) and Non-Max Suppression to abstain from choosing covering boxes

Image Grid. The Red Grid is responsible for detecting the dog



(fig:3 Yolo v3 feature extraction)

Architecture details

YOLOv3 utilizes a variation of Darknet, which initially has 53 layer network prepared on Imagenet. For the undertaking of location, 53 additional layers are stacked onto it, giving us a 106 layer completely convolutional hidden engineering for YOLO v3. In YOLO v3, the identification is finished by applying 1×1 recognition parts on highlight guides of three distinct sizes at three better places in the organization. The state of detection kernel t is $1 \times 1 \times (B \times (5 + C))$. Here B is the quantity of bouncing boxes a cell on the component guide can foresee, '5' is for the 4 bounding box credits and one item certainty and C is the no. of classes. YOLO v3 uses binary cross-entropy for calculating the classification loss while object confidence and class predictions are predicted through softmax.

1.3. Faster R-CNN

After the improvement in engineering of item detection network in R-CNN to Fast R-CNN. The training and detection time of the network decrease decline extensively, yet the network isn't quick to the point of being utilized as a continuous framework since it takes around (2 seconds) to produce yield on an information picture. The bottleneck of design is a particular hunt calculation. Accordingly K He et al. proposed another design called Faster R-CNN. It doesn't utilize particular inquiry rather they propose another locale proposition age calculation called Region Proposal Network. We should examine the Faster R-CNN engineering.

Faster R-CNN architecture contains 2 networks:

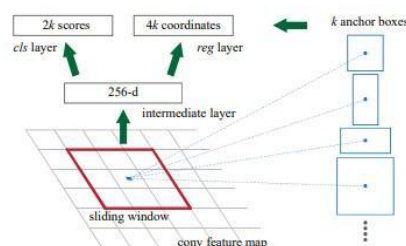
1. Region Proposal Network (RPN)
2. Object Detection Network

Region Proposal Network (RPN)

This region proposition network takes convolution highlight map that is produced by the spine layer as information and results the anchors created by sliding window convolution applied on the information include map.

Anchors

For each sliding window, the network produces the most extreme number of k -anchor boxes. By the default the worth of $k=9$ (3 sizes of $(128*128, 256*256$ and $512*512)$ and 3 angle proportion of $(1:1, 1:2$ and $2:1)$) for every one of various sliding situation in picture. In this way, for a convolution include guide of $W * H$, we get $N = W * H * k$ anchor boxes. These locale recommendations then passed into a middle layer of $3*3$ convolution and 1 cushioning and 256 (for ZF) or 512 (for VGG-16) yield channels. The result created from this layer is passed into two layers of $1*1$ convolution, the arrangement layer, and the relapse layer. the relapse layer has $4*N$ ($W * H * (4*k)$) yield boundaries (signifying the directions of bouncing boxes) and the arrangement layer has $2*N$ ($W * H * (2*k)$) yield boundaries (indicating the likelihood of item or not object).



(fig:4 faster R-CNN Architecture)

Training and Loss Function (RPN)

As a matter of first importance, we eliminate all the cross-limit secures, thus, that they don't expand the misfortune work. For an average 1000×600 picture, there are generally 20000 (~ $60 \times 40 \times 9$) secures. In the event that we eliminate the cross-limit secures, there are about 6000 anchors left for each picture. The paper likewise utilizes Non-Maximum Suppression in view of their characterization and IoU. Here they utilize a decent IoU of 0.7. This additionally lessens the quantity of anchors to 2000. The benefit of utilizing Non-Maximum concealment that it doesn't hurt exactness also. RPN can be prepared start to finish by utilizing backpropagation and stochastic angle plunge. It produces every scaled down group from the anchors of a solitary picture. It doesn't prepare misfortune work on each anchor rather it chooses 256 irregular anchors with positive and negative examples in the proportion of 1:1. On the off chance that a picture contains <128 up-sides, it utilizes more regrettable examples. For preparing RPNs, First, we really want to appoint paired class mark (climate the concerned anchor contains an article or foundation). In the quicker R-CNN paper, the creator utilizes two circumstances to allocate a positive mark to an anchor.

These are :

- those anchors which have the highest Intersection-over-Union (IoU) with a ground-truth box, or
- an anchor that has an IoU overlap higher than 0.7 with any ground-truth box.

Object Detection Network

The object detection network utilized in Faster R-CNN is a lot of like that utilized in Fast R-CNN. It is additionally viable with VGG-16 as a spine organization. It likewise utilizes the RoI pooling layer for making locale proposition of fixed size and twin layers of softmax classifier and the jumping enclose regressor is additionally utilized the expectation of the item and its bouncing box.

Since the bottleneck of Fast R-CNN design is locale proposition age with the particular hunt. Quicker R-CNN supplanted it with its own Region Proposal Network. This Region proposition network is quicker when contrasted with specific and it likewise further develops locale proposition age model while preparing. This likewise assists us with decreasing the general recognition time when contrasted with quick R-CNN (0.2 seconds with Faster R-CNN (VGG-16 organization) when contrasted with 2.3 in Fast R-CNN).

Faster R-CNN (with RPN and VGG shared) when prepared with COCO, VOC 2007 and VOC 2012 dataset creates mAP of 78.8% against 70% in Fast R-CNN on VOC 2007 test dataset). District Proposal Network (RPN) when contrasted with selective search , likewise contributed imperceptibly to the improvement of mAP.

LITERATURE SURVEY

Tanuj Bohra et al. proposed a constant two-way gesture based communication correspondence framework assembled utilizing picture handling, profound learning and PC vision. Strategies, for example, hand discovery, skin variety division, middle haze and form identification are performed on pictures in the dataset for improved outcomes. CNN model prepared with an enormous dataset for 40 classes and had the option to anticipate 17600 test pictures in 14 seconds with a precision of close to 100%.

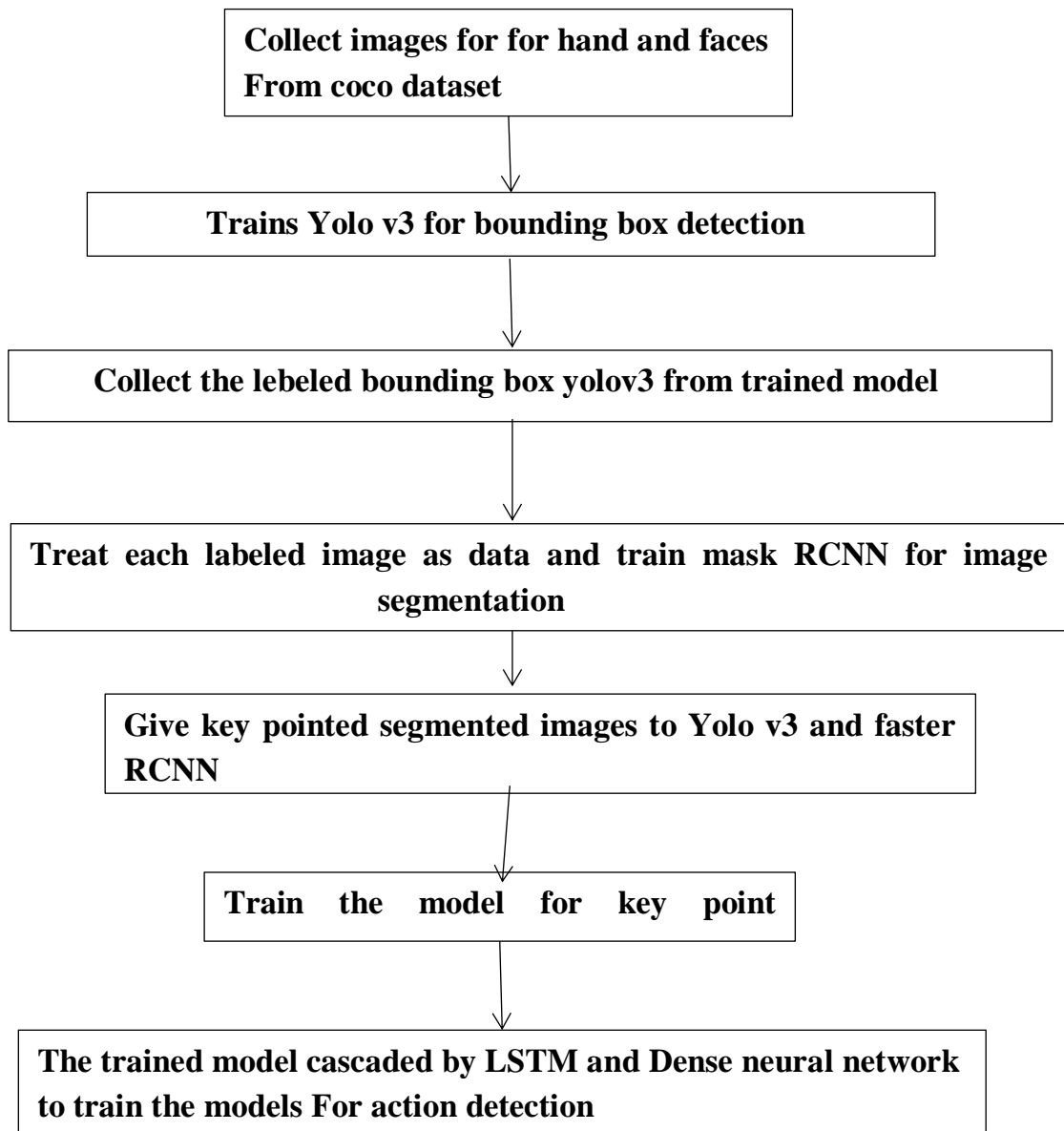
Joyeeta Singha and Karen Das proposed a framework for indian communication via gestures acknowledgment from a live video. The framework includes three phases. Preprocessing stage incorporates skin sifting and histogram coordinating. Eigen values and eigen vectors are being considered for include extraction stage and Eigen esteem weighted euclidean distance for order. Dataset comprised 480 pictures of 24 indications of ISL endorsed by 20 individuals. Framework was tried on 20 recordings and accomplished an exactness of 96.25%.

Muthu Mariappan H. and Dr. Gomathi V. planned an ongoing communication through signing acknowledgment framework as a convenient unit utilizing form discovery and fluffy c-implies calculation. Shapes are utilized for recognizing face, left and right hand. While fluffy c-implies calculation is utilized to segment the info information into indicated number of groups. Framework was carried out on a dataset that contained recordings recorded from 10 endorsers for a long time and sentences. It had the option to accomplish precision of 75%..

Salma Hayani et al. proposed a Bedouin communication through signing acknowledgment framework in light of CNN, enlivened from LeNet-5 [13] . Dataset contained 7869 pictures of Bedouin indications of numbers and letters. Different analyses were led by fluctuating the quantity of preparing sets from half to 80%. 90% precision was gotten with 80% preparation dataset. The creator has likewise contrasted the outcomes acquired and AI calculations like KNN (k-closest neighbor) and SVM (support vector machine) to show execution of the framework. This model was simply picture based and it tends to be stretched out to video based acknowledgment..

Kshitij Bantupalli and Ying Xie worked on american communication through signing acknowledgment framework which works on video groupings in view of CNN, LSTM . A CNN model named Inception was utilized to remove spatial elements from outlines, LSTM for longer time conditions. Different investigations were directed with fluctuating example sizes and dataset comprises of 100 unique signs performed by 5 underwriters and greatest exactness

WORK FLOW AND METHODOLOGY



BOUNDING BOX DETECTION

5.1. Bounding box

Image annotation has helped computer vision to develop on a scale that it has never done over the years. It includes different techniques over different use cases. The process of annotating images can be as simple as drawing rectangles over objects in the images. But this process serves a greater purpose in later computer vision tasks. Bounding box is one such technique of image annotation.. A bounding box is a rectangular construction superimposed over a picture including exceptionally significant elements of a specific item living in it. It is one of the easiest and low time taking strategies of picture comment. The annotator frames the objects of the pictures in a case according to the venture prerequisites.

5.2. Detection

The annotators frames the items in boxes according to the undertaking necessities. While searching for a vehicle the calculation just pursuits in the jumping boxes marked vehicles as opposed to searching for it in the entire picture. The bounding box contains organizes which has data about where precisely the article lives in the picture.

The picture shows the directions of the bouncing box explanation. To observe the vehicle from this picture, the calculation tends the framework to glimpse just inside these directions as opposed to checking out at the entire picture for the vehicle. Along these lines facilitating the recognition occupation of the model.Be that as it may, only a solitary jumping box can't empower a 100 percent expectation rate in the model. For this reason, we want to take care of the machines with a bigger number of bouncing boxes or just "preparing information" for improved discovery of articles in the picture.



(fig:5 bounding box output)

IMAGE SEGMENTATION USING Mask R-CNN

6.1. Image segmenation

Image segmentation is a strategy wherein an advanced picture is separated into different subgroups called Image fragments which helps in diminishing the intricacy of the picture to make further handling or investigation of the picture less complex. Division in simple words is allocating marks to pixels. All image components or pixels having a place with a similar class have a typical mark relegated to them. For instance: Let's take an issue where the image must be given as contribution to protest discovery. Instead of handling the entire picture, the finder can be inputted with a district chosen by a division calculation. This will keep the identifier from handling the entire picture subsequently diminishing induction time.



6.2. Approach for Image segmeantion

Similitude approach: This approach depends on identifying closeness between picture pixels to frame a section, in view of an edge. ML calculations like bunching depend on this sort of way to deal with portion a picture.

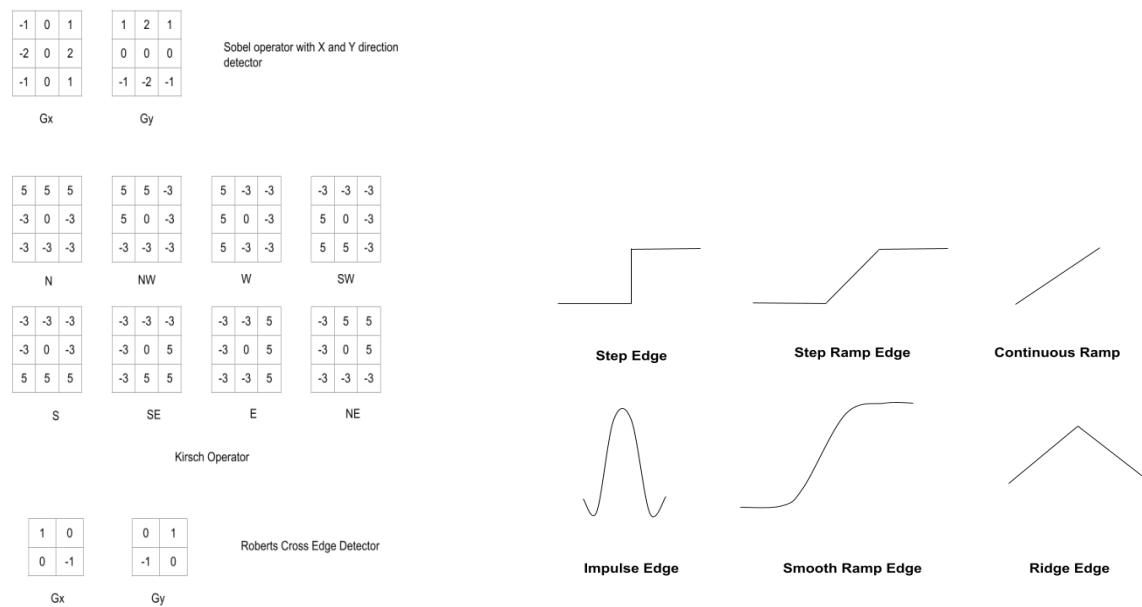
Irregularity approach: This approach depends on the intermittence of pixel force upsides of the picture. Line, Point, and Edge Detection procedures utilize this kind of approach for acquiring middle of the road division results which can be subsequently handled to get the last sectioned picture.

6.3. Edge based Image segmentaion

Edge-put together division depends with respect to edges found in a picture utilizing different edge location administrators. These edges mark picture areas of brokenness in dim levels, variety, surface, and so forth. At the point when we move starting with one area then onto the next, the dark level might change. So on the off chance that we can observe that brokenness, we can track down that edge. An assortment of edge discovery administrators are accessible however the subsequent picture is a middle of the road division result and ought not be mistaken for the last fragmented picture. We need to perform further handling on the picture to the portion it. Extra advances remember joining edges portions acquired into one section for request to diminish the quantity of fragments instead of lumps of little lines which could

frustrate the course of area filling. This is done to acquire a consistent boundary of the item. The objective of edge division is to get a middle division result to which we can apply locale based or some other kind of division to get the last fragmented picture.

Edges are generally connected with "Extent" and "Bearing". Some edge identifiers give the two bearings and extent. We can utilize different edge locators like Sobel edge administrator, watchful edge finder, Kirsch edge administrator, Prewitt edge administrator, Robert's edge administrator, and so on.



(fig:6 Edge detection)



(fig:7 Image segmenation output)

KEY POINT DETECTION

7.1. KEY POINT DETECTION

In this segment, we'll first officially form the keypoints discovery issue and present execution measurements. To take care of this issue, we construct two models as baselines, which are acknowledged in light of straightforward brain organization furthermore, convolutional brain network individually. Then, at that point, as the significant piece of this part, the Inception Model will be presented and examined exhaustively. Particularly, we'll exhibit its great benefits over conventional cnn models.

7.2. NEURAL NETWORK ARCHITECTURE

In the first place, we utilize one secret layer brain network model as a benchmark. We reshape the 96×96 picture to a 9216×1 vector as the contribution of our organization. The neuron number of the secret layer is 100. What's more, the result layer yields 15 sets of directions of the 15 keypoints, giving 30 numbers altogether. The misfortune work is characterized as the Euclidean distance between the ground truth and result keypoints vectors. To join quicker, we use Nesterov energy as the update rule for slope plummet. The clump size is 1 and we repeat for 400 ages during the preparation stage. To accomplish better keypoints identification precision (lower misfortune), we construct a convolution brain network model as an progressed standard. The design outline of our CNN is displayed in Fig. 1. The number displayed over each layer shows the size of its comparing enactment, and the portrayal underneath each layer depicts its capacity. The hyperparameters of those convolutional layers in the organization are outlined in Table. 1. Both secret layers have 500 neurons and the result layers create a 30-component vector, same as the previous one secret layer brain organization, addressing the directions of the keypoints.

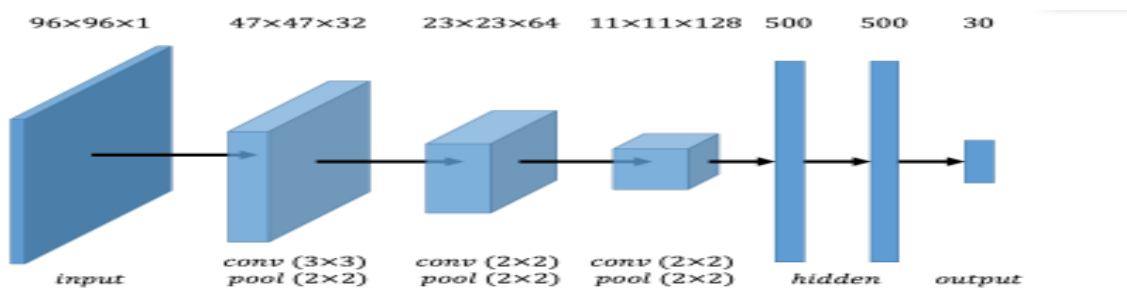


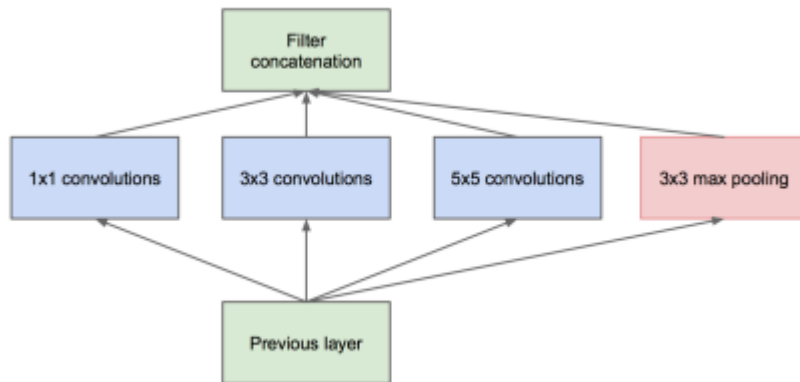
Figure 1. Convolution Neural Network Architecture

Table 1. Hyperparameters of CNN

	conv layer 1		conv layer 2		conv layer 3	
	conv	pool	conv	pool	conv	pool
filter	3×3	2×2	2×2	2×2	2×2	2×2
stride	1	2	1	2	1	2
pad	0	0	0	0	0	0

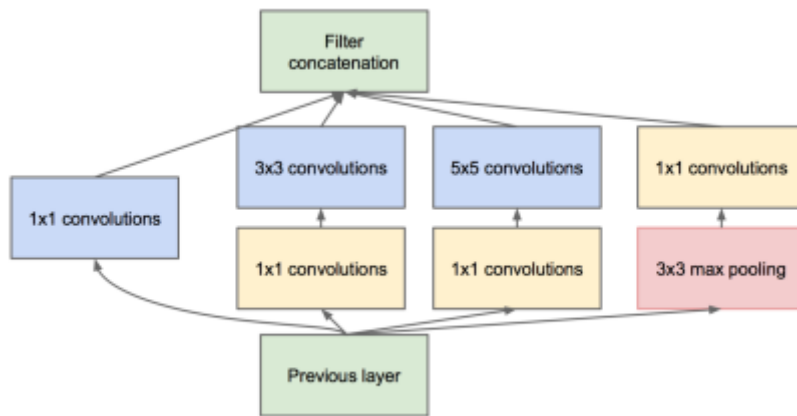
(fig:8 Architecture of key point detection CNN)

As is referenced over, one unavoidable shortcoming of multi-facet cnn design is its superfluously enormous measure of boundaries. Because of expanded layer and boundary size, the organization is leaned to overfit and neglects to sum up. In addition, countless boundaries causes emphatically enormous computational asset utilization also. To beat such downsides showed over, one principal approach to tackling this issue is to make a few extreme changes from the engineering level. Szegedy et al proposed such a model named Inception Model , which inclines toward scantily associated designs over completely associated ones, even inside the convolutions. This technique addresses the huge boundary issue in a general sense, however likewise is more steady according to an organic viewpoint. Notwithstanding, this thought may not fill in as well as we expected by and by due to the enhancement approach in equipment level. With regards to mathematical calculation on non-uniform meager information structures, customary computer it are not extremely wasteful to ing frameworks. This is additionally the motivation behind why we want to invest such countless amounts of energy in the examination of effective procedure on scanty information structures, for example scanty grid augmentation. In such a case, to keep the equilibrium between computational asset utilization and computational productivity, an ideal construction is by all accounts like an internationally meager convolutional brain organization, which is made out of thick designs locally. Worldwide sparsity ensures the quantity of boundaries is restricted while nearby thickness gives productive calculation. A run of the mill square of the Inception Model .In each layer, we utilize various channels with various sizes 1×1 , 3×3 and 5×5 rather than one channel in customary cnn layers. This approach accomplishes purported worldwide sparsity since we split an immense measure of boundaries from one-size channel up into a few gatherings relating to various channel sizes. In the interim, it ensures nearby thickness as well on the grounds that every convolution activity can be acknowledged in a customary way.



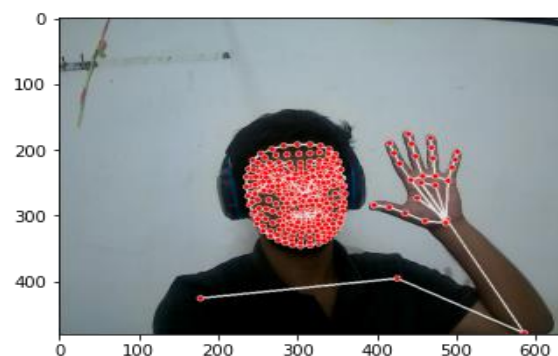
In any case, the above credulous form of Inception Model square is as yet wasteful in calculation. Expect to be the input size is $S \times S \times C$ where $S \times S$ is the picture size and C addresses the channel number, then for each channel 1×1 , 3×3 and 5×5 , the quantity of channels are overall a similar C . An enormous C will prompt countless calculations due to 5×5 channel; while in the event that C is decreased to a moderate worth, then there will be less point in utilizing 1×1 channel since it may not create very much summed up highlights. More or less, this square requirements further acclimations to change the quantity of channels of various size channels. To acknowledge such a design outlined over, the creators raise a high level square. which is utilized in the last Inception Model. The principle distinction from the construction in Fig. 2 is that 1×1 convolutional channels are embedded before 3×3 and 5×5 channels (and later max pooling). In such a manner, we first utilize 1×1 convolutions to figure

decreases before costly convolutions like 3×3 and 5×5 . These 1×1 convolutions can likewise be utilized as corrected direct enactments. Thus, we bear more channels while processing 3×3 and less for 5×5 . Table shows some particular quantities of profundities in each channel. The 3×3 and 5×5 decreases address the two 1×1 convolutions embedded before them. The input profundity is 128, which is diminished to 96 and 16 preceding 3×3 and 5×5 channels. The numbers in intense address the profundities of convolutional yields, which amount to 256 ($= 64 + 128 + 32 + 32$) altogether.



(fig:9 Enhanced neural net)

One more noteworthy benefit of Inception Model is its adaptability to be adjusted to various layers. In lower layers, which we center more around neighborhood districts of the info picture, like edge data, what makes a difference more is really more modest size channels like 1×1 and 3×3 . Then again, in any case, in higher layers, where highlights like surface are more leaned to be caught, we most likely need to utilize bigger size channels like 5×5 . Utilizing such a model square, we can change the quantity of profundities among various channels as indicated by our requests at various layer levels. In the accompanying trials, we didn't assemble an organization utilizing Inception Model squares. All things considered, we removed highlights from a pretrained Inception Model for picture arrangement, and information those elements to three unique organizations, 1) one secret layer brain organization and 2) cnn whose structures are very like the previous baselines, and 3) cnn with dropout. As a rule, we think about execution (misfortune and train/test time) among those 5 organizations to assess the adequacy of those extricated highlights contrasted with crude picture.



(fig:10 key point detection)

Sign language and Action Detection

8.1. GESTURE CLASSIFICATION

1. The approach which we used for this project is : Our approach uses two layers of algorithm to predict the final symbol of the user.

Algorithm Layer 1:

1. Apply gaussian blur filter and threshold to the frame taken with opencv to get the processed image after feature extraction.
2. This processed image is passed to the CNN model for prediction and if a letter is detected for more than 50 frames then the letter is printed and taken into consideration for forming the word.
3. Space between the words are considered using the blank symbol.

Algorithm Layer 2:

1. We detect various sets of symbols which show similar results on getting detected.
2. We then classify between those sets using classifiers made for those sets only.

CNN Model

1st Convolution Layer : The input picture has resolution of 128x128 pixels. It is first processed in the first convolutional layer using 32 filter weights (3x3 pixels each). This will result in a 126x126 pixel image, one for each Filter-weights.

1st Pooling Layer : The pictures are downsampled using max pooling of 2x2 i.e we keep the highest value in the 2x2 square of array. Therefore, our picture is downsampled to 63x63 pixels.

2nd Convolution Layer : Now, these 63 x 63 from the output of the first pooling layer is served as an input to the second convolutional layer. It is processed in the second convolutional layer using 32 filter weights (3x3 pixels each). This will result in a 60 x 60 pixel image.

2nd Pooling Layer : The resulting images are downsampled again using max pool of 2x2 and is reduced to 30 x 30 resolution of images. 5. 1st Densely Connected Layer : Now these images are used as an input to a fully connected layer with 128 neurons and the output from the second convolutional layer is reshaped to an array of $30 \times 30 \times 32 = 28800$ values. The input to this layer is an array of 28800 values. The output of these layer is fed to the 2nd Densely Connected Layer. We are using a dropout layer of value 0.5 to avoid overfitting. 2nd Densely Connected Layer , Now the output from the 1st Densely Connected Layer are used as an input to a fully connected layer with 96 neurons.











Final layer: The output of the 2nd Densely Connected Layer serves as an input for the final layer which will have the number of neurons as the number of classes we are classifying (alphabets + blank symbol).

Activation Function : We have used ReLu (Rectified Linear Unit) in each of the layers(convolutional as well as fully connected neurons). ReLu calculates $\max(x,0)$ for each input pixel. This adds nonlinearity to the formula and helps to learn more complicated features.It helps in removing the vanishing gradient problem and speeding up the training by reducing the computation time.

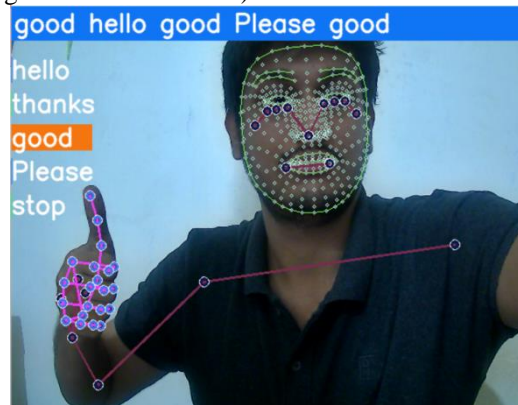
Pooling Layer : We apply Max pooling to the input image with a pool size of (2, 2) with relu activation function.This reduces the amount of parameters thus lessening the computation cost and reduces overfitting.

Dropout Layers: The problem of overfitting, where after training, the weights of the network are so tuned to the training examples they are given that the network doesn't perform well when given new examples.This layer "drops out" a random set of activations in that layer by setting them to zero.The network should be able to provide the right classification or output for a specific example even if some of the activations are dropped out. **Optimizer :** We have used Adam optimizer for updating the model in response to the output of the loss function. Adam combines the advantages of two extensions of two stochastic gradient descent algorithms namely adaptive gradient algorithm(ADA GRAD) .

Table VII: Confusion matrix for distinguishing between the 5 hand shapes using shape context.

Hand Shape	Classified As				
					
	95	4	0	1	0
	0	98	0	1	1
	6	0	93	1	0
	2	7	0	91	0
	1	1	0	1	97

(fig:11 confusion matrix)



(fig:12 sign language detection)

CONCLUSION

- In this report, a practical ongoing vision based american communication through signing acknowledgment for D&M individuals have been created for asl alphabets. We accomplished last exactness of 95.0% on our dataset. We can work on our expectation subsequent to executing two layers of calculations in which we confirm and anticipate images which are more like one another. This way we can identify practically every one of the images given that they are shown appropriately, there is no clamor behind the scenes and lighting is sufficient.
- We are wanting to accomplish higher precision even if there should arise an occurrence of complex foundations by evaluating different foundation deduction calculations. We are additionally considering working on the preprocessing to anticipate signals in low light circumstances with a higher precision

REFERENCES

1. Wang H, Chai X, Zhou Y, Chen X. Fast sign language recognition benefited from low rank approximation. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2015. 2015.
2. Singha J, Das K. Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique. arXiv preprint arXiv:1303.0634. 2013; 4(2): p. 188-195.
3. Kalsh EA, Garewal NS. Sign Language Recognition System. International Journal of Computational Engineering Research. 2013; 03(6): p. 15-21
4. Tewari D, Srivastava S. A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self-Organizing Map Algorithm. International Journal of Engineering and Advanced Technology (IJEAT). 2012; 2(2): p. 165-170.
5. Raheja JL, Mishra A, Chaudary A. Indian Sign Language Recognition Using SVM 1. Pattern Recognition and Image Analysis. 2016 September; 26(2).
6. Kishore PVV, Prasad MVD, Prasad CR, Rahul R. 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. In International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Association with IEEE; 2015. p. 34-38.
7. Goyal, Sakshi & Sharma, Ishita & Sharma, Shanu. Sign Language Recognition System For Deaf And Dumb People. International Journal of Engineering Research & Technology (IJERT). 2013 April; 2(4).
8. Huang J, Zhou W, Li H, Li W. Sign language recognition using 3D convolutional neural networks. In Multimedia and Expo (ICME), 2015 IEEE International Conference on; 2015: IEEE. p. 1-6.
9. Chai X, Li G, Lin Y, Xu Z, Tang Y, Chen X. Sign Language Recognition and Translation with Kinect. The 10th IEEE International Conference on Automatic Face and Gesture Recognition. 2013;: p. 22-26. 12. Pigou L, Dieleman S, Kindermans PJ, Schrauwen B. Sign Language Recognition using Convolutional Neural Networks. In Workshop at the European Conference on Computer Vision; 2014; Belgium. p. 572-578.
10. Greg C. Lee & Fu-Hao Yeh & Yi-Han Hsiao. Kinect-based Taiwanese sign-language recognition system. Multimed Tools Appl. 2014 October.
11. Zhang LG, Chen Y, Fang G, Chen X, Gao W. A Vision-Based Sign Language Recognition System. In Proceedings of the 6th International Conference on Multimodal Interfaces; 2004; Pennsylvania: ACM. p. 198-204