**A PROJECT REPORT**

**ON**

**SOCIAL DISTANCE MONITORING USING**

**DEEP LEARNING ALGORITHM**

Submitted in partial fulfilment for the requirement of the award of

training and internship in

Data Analytics, Machine Learning and AI using Python



*Submitted By*

**Sidharth Kumar Mohanty (**C.V Raman Global University, Bhubaneswar**)**

*Under the guidance of*

**Mr. Bipul Shahi**

**DECLARATION**

This is to declare that the work in the Major Project entitled “**Social distance monitoring using Deep Learning algorithm”** submitted by Sidharth Kumar Mohanty in partial fulfillment of the requirements for the Internship is a bona fide work carried out by us under the supervision and guidance of **Mr. Bipul Sahai.** It does not contain materials copied from other published work, and the work is not published anywhere.

**Name & Signature of students:**

Sidharth Kumar Mohanty

**ACKNOWLEDGEMENTS**

We are pleased to acknowledge Mr. Bipul Sahai for their invaluable guidance during the course of this project work. We extend our sincere thanks for his immeasurable support during our project work.

We are also grateful to other members of the Diginique TechLab who co-operated with us regarding some issues.

**ABSTRACT**

To ensure social distancing protocol in public places and workplace, We have developed social distancing detection tool that can monitor if people are keeping a safe distance from each other and using face mask or not by analyzing real time video streams from the camera. Social distancing is a recommended solution by the World Health Organisation (WHO) to minimise the spread of COVID-19 in public places. The majority of governments and national health authorities have set the 2-meter physical distancing as a mandatory safety measure in shopping centres, schools and other covered areas. In this research, we develop a generic Deep Neural Network-Based model for automated people detection, tracking, and inter-people distances estimation in the crowd, using common CCTV security cameras. The proposed model includes a YOLOv4-based framework and inverse perspective mapping for accurate people detection and social distancing monitoring in challenging conditions, including people occlusion, partial visibility, and lighting variations. We also provide an online risk assessment scheme by statistical analysis of the Spatio-temporal data from the moving trajectories and the rate of social distancing violations. We identify high-risk zones with the highest possibility of virus spread and infection. This may help authorities to redesign the layout of a public place or to take precaution actions to mitigate high-risk zones. Face Detection has evolved as a very popular problem in Image processing and Computer Vision. Many new algorithms are being devised using convolutional architectures to make the algorithm as accurate as possible. These convolutional architectures have made it possible to extract even the pixel details. We aim to design a binary face classifier which can detect any face present in the frame irrespective of its alignment. We present a method to generate accurate face segmentation masks from any arbitrary size input image. Beginning from the RGB image of any size, the method uses Predefined Training Weights of VGG – 16 Architecture for feature extraction. Training is performed through Fully Convolutional Networks to semantically segment out the faces present in that image. Gradient Descent is used for training while Binomial Cross Entropy is used as a loss function. Further the output image from the FCN is processed to remove the unwanted noise and avoid the false predictions if any and make bounding box around the faces. Furthermore, proposed model has also shown great results in recognizing non-frontal faces. Along with this it is also able to detect multiple facial masks in a single frame. Experiments were performed on Multi Parsing Human Dataset obtaining mean pixel level accuracy of 93.884 % for the segmented face masks

**CONTENTS**

**DECLARATION …………………………………………………………………………………....2**

**CERTIFICATE………………………………………………………………………………………3**

**ACKNOWLEDGEMENTS…………………………………………………….……………………4**

**ABSTRACT ………………………………………………………………….………………………5**

**CONTENTS ………………………………………………………………………………………….6**

**INTRODUCTION……………………………………………………………………………………7**

**TECHNOLOGY & CONCEPT ……………………………………………….…………………9-13**

1. **CNN**
2. **OBJECT DETECTION**
3. **YOLO V3**

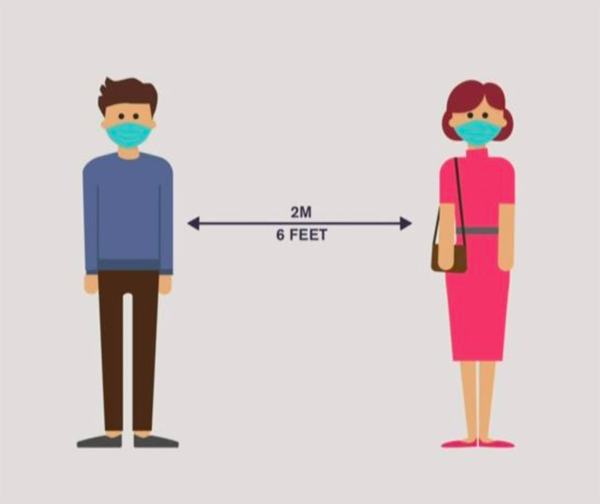
**RESULT / OUTCOME ………………………………………………………….…………………17**

**CONCLUSION ……………………………………………………………….……………………18**

**REFERENCES….………………………………………………………….……………………….19**

**INTRODUCTION**

Social Distancing – the term that has taken the world by storm and is transforming the way we live.In the world, transcending languages and cultures now Social distancing has become a mantra .

****

This way of living has been forced upon us by the fastest growing pandemic the world has ever seen – COVID-19. Covid has so far infected almost 14 million people and claimed over 110K lives globally as per the World Health Organization. Around 213 countries have been affected so far by the deadly virus.

The biggest cause of concern is that COVID-19 spreads from person to person through contact or if you’re within close proximity of an infected person. Given how densely populated some areas are, this has been quite a challenge.

The novel generation of the coronavirus disease (COVID-19) were reported in late December 2019 in Wuhan, China. After only a few months, the virus was hit by the global outbreak in 2020. On May 2020 The World Health Organisation (WHO) announced the situation as pandemic1 , 2 . The statistics by WHO on 26 August 2020 confirms 23.8 million infected people in 200 countries. The mortality rate of the infectious virus also shows a scary number of 815,000 people. With the growing trend of patients, there is still no effective cure or available treatment for the virus. While scientists, healthcare organisations, and researchers are continuously working to produce appropriate medications or vaccines for the deadly virus, no definite success has been reported at the time of this research, and there is no certain treatments or recommendation to prevent or cure this new disease. Therefore, precautions are taken by the whole world to limit the spread of infection. These harsh conditions have forced the global communities to look for alternative ways to reduce the spread of the virus.

Face detection has emerged as a very interesting problem in image processing and computer vision. It has a range of applications from facial motion capture to face recognition which at the start needs the face to be detected with a very good accuracy. Face detection is more relevant today because it not only used on images but also in video applications like real time surveillance and face detection in videos. High accuracy image classification is possible now with the advancements of Convolutional networks. Pixel level information is often required after face detection which most face detection methods fail to provide. Obtaining pixel level details has been a challenging part in semantic segmentation. Semantic segmentation is the process of assigning a label to each pixel of the image.

In our case the labels are either face or non-face. Semantic segmentation is thus used to separate out the face by classifying each pixel of the image as face or background. Also most of the widely used face detection algorithms tend to focus on the detection of frontal faces. This paper proposes a model for face detection using semantic segmentation in an image by classifying each pixel as face and non-face i.e. effectively creating a binary classifier and then detecting that segmented area. The model works very well not only for images having frontal faces but also for non-frontal faces. The paper also focuses on removing the erroneous predictions which are bound to occur. Semantic segmentation of human face is performed with the help of a fully convolutional network. The next section discusses the related work done in the domain of face detection. In section III we describe the method followed for face segmentation and detection using semantic segmentation on any arbitrary RGB image. Finally, the generated facial masks are demonstrated in experimental results in section IV. Post processing on the predicted images has also been discussed at length which also entails the removal of erroneous predictions.

**Technology and Concepts**

**Deep Learining**

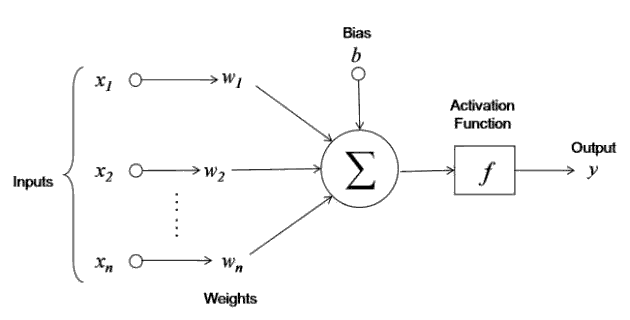
Deep learning is a part of machine learning which is used for detection of object like image ,audio etc .It includes statistics and predictive modelling and It also beneficial for data scientists who are tasked with collecting interpreting large amounts of data.

At its simplest, deep learning can be thought of as a way to automate [predictive analytics](https://searchbusinessanalytics.techtarget.com/definition/predictive-analytics). While traditional machine learning algorithms are linear, deep learning [algorithms](https://whatis.techtarget.com/definition/algorithm) are stacked in a [hierarchy](https://whatis.techtarget.com/definition/hierarchy) of increasing complexity and abstraction.

**Neural networks**

Neural networks are a popular type of machine learning model. A special case of a neural network called the convolutional neural network (CNN) is the primary focus of this thesis. Before discussing CNNs, we will discuss how regular neural networks work.

Neural networks were originally called arti\_cial neural networks, because they were developed to mimic the neural function of the human brain.



The neuron is trained by carefully selecting the weights to produce a desired output for each input.

**Convolutional Neural Network**

The convolutional neural network is a type of neural network model designed for working with 2D image data and they can be used with one-dimensional and 3D data.

CNN is the convolutional which gives the network its name. That layer performs an operation called as “convolution“.

In the context of a CNN, the convolution is a linear operation that involves in the multiplication of a set of weights with input, much like a traditional neural network. Given that technique was designed for 2D input, that multiplication is performed between array of input data and a 2D array of weights, called as filter or a kernel.

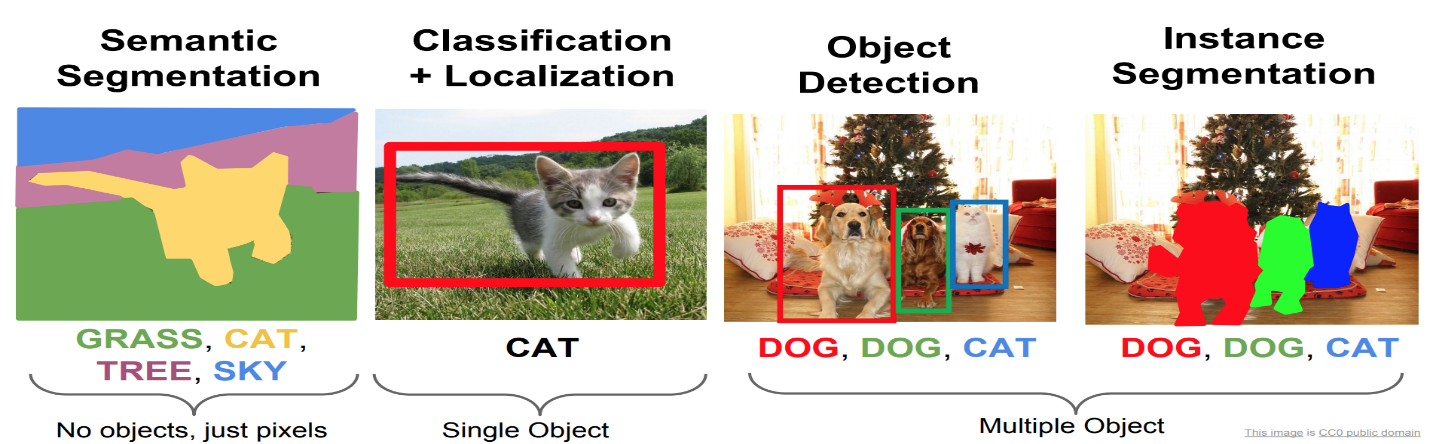
Using a filter smaller than the input is intentional as it allows the same filter which is set of weights that to be multiplied by the input array multiple times at different points on the input. Specifically, the filter is applied systematically to each overlapping part or filter-sized patch of the input data, left to right, top to bottom.

A multi-layer network typically includes three types of layers: an input layer, one or more hidden layers and an output layer. The input layer usually merely passes data along without modifying it. Most of the computation happens in the hidden layers. The output layer converts the hidden layer activations to an output, such as a classification. A multilayer feed-forward network with at least one hidden layer can function as a universal approximator.

**Object detection**

Object detection is the classical problems of computer vision (CV). It is similar to other computer vision tasks, because it involves creating a solution that is invariant to deformation and changes in lighting and viewpoint. What makes object detection a distinct problem is that it involves both locating and classifying regions of an image . The locating part is not needed in, for example,

whole image classification.



To detect an object, we need to have some idea where the object might be and how the image is segmented. This creates a type of chicken-and-egg problem, where, to recognize the

class and shape of an object, we need to know its location that to recognize the location of an object, we need to know its class and shape. Some visually dissimilar features, such as the clothes and face of a human being, may be parts of the same object, but it is difficult to know this without recognizing the object first.Iterating over the problem of localization plus classification we end up with the need for detecting and classifying multiple objects at the same time. Object detection is the problem of finding and classifying a variable number of objects on an image. The important difference is the “variable” part. In contrast with problems like classification, the output of object detection is variable in length, since the number of objects detected may change from image to image.

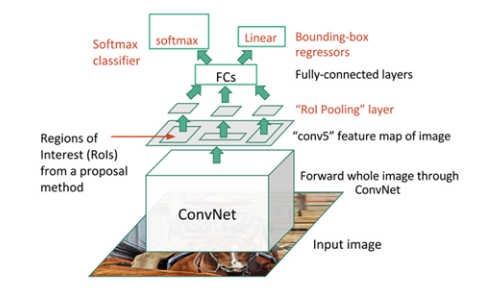
Since the mid-2000s some point and shoot cameras started to come with the feature of detecting faces for a more efficient auto-focus. While it’s a narrower type of object detection, the methods used apply to other types of objects as we’ll describe later.

**Steps For Object Detection**

1. First we take an image as input
2. Then we use region-based CNN to generate multiple regions from this image as sub-segmentations
3. The technique then combines the similar regions to form a larger region based on colour, texture, size, and shape compatibility and similarity
4. Finally these regions produces the object locations (Region of Interest).
5. Pass all these regions to the CNN and classify them into various classes
6. Once we have classified each region into its corresponding class, combine all these regions to get the original image with the detected objects

input image sub-segmentations region of interest

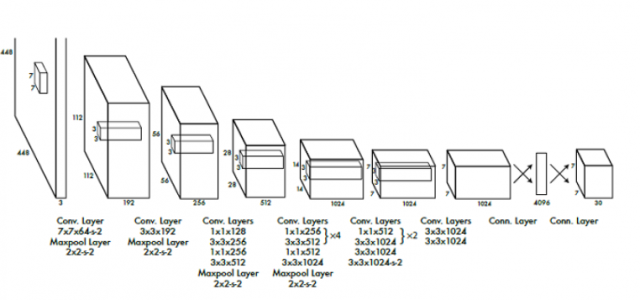


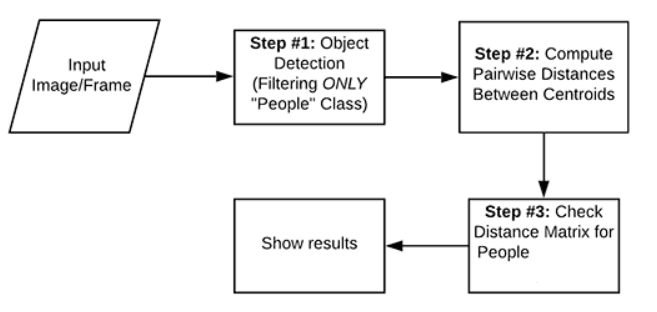
**YOLO V3**

YOLO is a advance convolutional neural network for doing image detection in live video stream. These bounding boxes are weighted by the predicted probabilities.YOLO is popular because it achieves high accuracy while also being able to run in real-time. The algorithm “only looks once” at the image in the sense that it requires only one forward propagation pass through the neural network to make predictions. YOLO V3 is an improvement over previous YOLO detection networks. Compared to prior versions, it features multi-scale detection, stronger feature extractor network, and some changes in the loss function. As a result, this network can now detect many more targets from big to small. And, of course, just like other single-shot detectors, YOLO V3 also runs quite fast and makes real-time inference possible on GPU devices. Well, as a beginner to object detection, you might not have a clear image of what do they mean here. But you will gradually understand them later in my post. For now, just remember that YOLO V3 is one of the best models in terms of real-time object detection as of 2019.

First of all, let’s talk about how this network look like at a high-level diagram (Although, the network architecture is the least time-consuming part of implementation). The whole system can be divided into two major components: Feature Extractor and Detector; both are multi-scale. When a new image comes in, it goes through the feature extractor first so that we can obtain feature embeddings at three (or more) different scales. Then, these features are feed into three (or more) branches of the detector to get bounding boxes and class information.

The feature extractor YOLO V3 uses is called Darknet-53. You might be familiar with the previous Darknet version from YOLO V1, where there’re only 19 layers. But that was like a few years ago, and the image classification network has progressed a lot from merely deep stacks of layers. ResNet brought the idea of skip connections to help the activations to propagate through deeper layers without gradient diminishing.

**How It Works:**

****

**RESULT & OUTCOMES:**



**CONCLUSION**

So here we conclude that Social distance monitoring and face mask detection system is very important for out society.Our project also has some disadvantages but we should always move forward by taking the advantages and our future aspects is that we will try to put this in applications with some exciting graphics and to add some new creativity in this project. We were able to generate accurate face masks for human objects from RGB channel images containing localized objects. We demonstrated our results on Multi Human Parsing Dataset with mean pixel level accuracy. Also the problem of erroneous predictions has been solved and a proper bounding box has been drawn around the segmented region. Proposed network can detect non frontal faces and multiple faces from single image. The method can find applications in advanced tasks such as facial part detection.

**REFERENCES**

1. T.-H. Kim, D.-C. Park, D.-M. Woo, T. Jeong, and S.-Y. Min, “Multi-class classifier-based adaboost algorithm,” in Proceedings of the Second Sinoforeign-interchange Conference on Intelligent Science and Intelligent Data Engineering, ser. IScIDE’11. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 122–127.
2. P. Viola and M. J. Jones, “Robust real-time face detection,” Int. J. Comput. Vision, vol. 57, no. 2, pp. 137–154, May 2004.
3. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, Dec 2001, pp. I–I.
4. J. Li, J. Zhao, Y. Wei, C. Lang, Y. Li, and J. Feng, “Towards real world human parsing: Multiple-human parsing in the wild,” CoRR, vol. abs/1705.07206.
5. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
6. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014.
7. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,”