Detection and Recognition of Text From Natural Camera Image using Deep Convolutional Network

Rashmi Kapoor, M. Sushama

Abstract—Amino acids are little bio-particles with different properties. The capacity to ascertain the physiochemical properties of proteins is pivotal in many research regions, for example, tranquilize plan, protein displaying and basic bioinformatics. The physiochemical properties of the protein decides its collaboration with different atoms and subsequently its capacity. Foreseeing the physiochemical properties of protein and translating its capacity is of extraordinary significance in the field of medication and life science. The point of this work is to create python based programming with graphical UI for anticipating the physiochemical and antigenic properties of protein. Thus the instrument was named as ASAP-Analysis of protein succession and antigenicity expectation. ASAP predicts the antigenicity of the protein succession from its amino corrosive arrangement, in light of Chou Fasman turns and antigenic file. ASAP computes different physiochemical properties that is required for invitro tests. ASAP utilizes standardization esteems that expansion the affectability of the apparatus.

Keywords: Amino acids, antigenicity, normalization and Protein modeling.

1. INTRODUCTION

The field of technology in the modern day has progressed to such an extent that every complex problem has a well defined solution. With the increase in the advancement, the adverse effect on human lifestyle has also increased. The repetitive usage of smart devices and the gradual increase in the radiation has an impact on the whole ecosystem.

Recent statistics show that a lot of people are facing health issues due to excessive usage of Digital Devices. The reasons that drive a person to use these devices are many. It also has to do with the reason that these smart devices are available in most of the places in the world. Technology is in a way helping us to improve the efficiency and execute things in a faster method but excessive usage of this is leading to several health consequences. One such example that can be derived from the excessive usage of Smart Devices is the impact on

human visual sensory system. People feel stressful from prolonged usage of mobile phones and computers. The blue light emitted from these devices cause gradual blindness in the long-run. Visual Impairment is the major health challenge that modern day people are facing due to the usage of smart devices. It would be of a great help to the person if a specific portable tool is available to carry along and understand thethings around. The current project considers a scenario where

a person suffering from a visual impairment needs a tool to carry around and correlate the information which cannot be seen. Some products are available in the international market like one shown below but they are very costly (between 1500$ to 2000$):

1. Assisted Vision Smart Glasses: They are constructed using transparent OLED displays, two small cameras, a gyroscope, a compass, a GPS unit, and a headphone. Most visually impaired people can distinguish light and dark, these glasses can make anything that's close to the wearer brighter, so they can discern people and obstacles. The main problem with theses glasses is they are very costly and cannot identify text from images.

2. A wearable device called Horus is using a combination of computer vision, machine learning and audio cues to improve the lives of visually impaired people. Developed by a Swiss startup called Eyra, Horus consists of a headband with stereo cameras on one end that can recognize text, faces and objects. Information from the cameras is fed via a 1m cable into a smartphone-sized box containing a battery and a NVIDIA Tegra K1 processor. This provides GPU-accelerated computer vision, deep learning and sensors that process, analyze and describe the images from the cameras.

Apart from this one more device, available in market is "figure reader". This MIT Media Labs project is a wearable device, a very chunky ring that sits on the finger and is capable of detecting and interpreting 12-point printed text as the user scans his or her finger across it. It reads aloud in real-time. Small vibrations alert the wearer to any deviation off the line. Seeing AI, an app developed by Microsoft AI & Research. It essentially narrates the world for blind and low-vision users, allowing them to use their smartphones to identify everything from an object or a color to a dollar bill.

But when the exact location of text is not known or the distance between the user and text is much more, these scanner based devices will not be much affective.

In India many researchers are working in the same field to utilize artificial intelligence and deep learning neural network to help blind people. Dr. Amit Ray is working in Compassionate AI lab, to utilize AI for the benefit of blinds. In 2017, BrailleMe, the award-winning product developed by Surabhi Srivastava, IIT Bombay, made it possible for the visually impaired to access any digital information instantaneously in their own tactile script. The price of the product was just 300$.

Revised Version Manuscript Received on August 19, 2019.

Dr. Rashmi Kapoor, Assistant Professor, DepartmentofEEE, Vnrvjiet, Hyderabad, Telangana, India.

Dr. M. Sushama, Professor, Department of Electrical Engineering, JNTUH College of Engineering, Kukatpally Hyderabad, Telangana, India

II. PROCEDURE DISCUSSION & RESULTS

A. Text detection and recognition system

Feature detection from an image, without giving the exact information about the feature, is what a deep convolutional neural network can do. This can be utilized to detect the text from any natural scene image with varying backgrounds. Deep convolutional neural network need to be trained properly with sufficient data set to achieve the goal.

The proposed system enables a visually impaired person to understand text on sign boards, banners, hoardings. This system captures the image from its surroundings using a camera and the image will be internally processed and the speech output is given through the speaker or earpiece connected to it.

There are two main blocks: image processing block and voice processing block, In the image processing block the image captured using camera is converted to text. In voice processing block the output of the previous block i.e., text extracted from the captured image is converted to speech. For convenience of the user the voice may be altered to masculine or feminine voice.

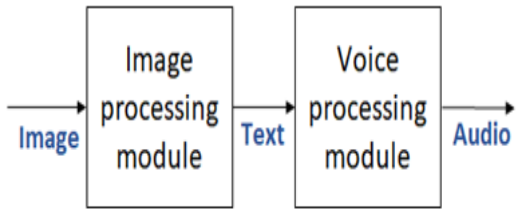

Fig.1: Block diagram Image processing block

Image processing block

- Image is acquired using a camera.
- Text is separated from the image after processing in a pre- trained deep network.
- Here we get .txt file from a .png file or .jpg file Voice processing block

Voice processing block

- Now the text file is further converted into speech using a text to speech synthesizer.
- There are two ways to do this: one is text to phoneme conversion where text is compared with the words present in dictionary and giving output, other one is learning based speech output approach.Final Stage

B. Implementation

OpenCV is installed in the Raspberry Pi to perform the image processing. OpenCV is Open Source Computer Vision which is a set of libraries including all the programs that support real-time computer vision. This OpenCV installed in the Raspberry Pi supports in executing the image processing captured with the camera.

The camera used in the project is the Pi Camera which is interfaced with the Raspberry Pi using specific commands. The Speaker can be of a standard audio output device such as Headphones / Earphones which helps the user to listen to the voice output.

The capturing of image and the reduction of text using the microcontroller is done using the OCR technology. Optical Character Recognition abbreviated as OCR is a technique used to capture the text from an image and display it as an understandable language.

The reduction of text from the image and conversion of text to speech is achieved through the TTS technique. The Text-to-Speech technique is a process in which the normal language text is converted into speech. This technique is performed through a speech synthesizer which in our case is the microcontroller. The assembly of all the three components along with the OpenCV software makes the prototype fully functional and helps the user to understand the image in front.

C. Convolutional neural network

Artificial intelligence is bridging gap between humans and machine constantly. It helps the machine to see the world as a human. This can be done by deep learning where output is predicted for a given input. Here we are using Convolutional Neural Network(CNN) , a deep learning algorithm which takes input and allots confidence to different characteristics in the input image which helps in differentiating one from another.

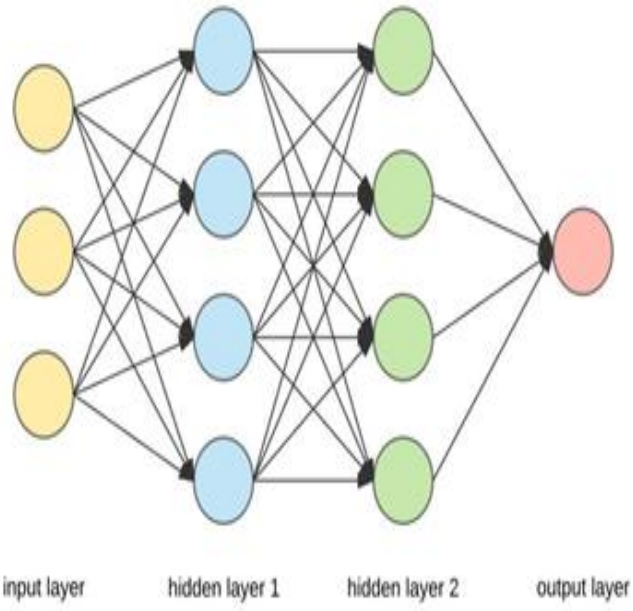

fig.1 Convolutional Neural Network

A series of convolution and pooling operations are performed followed by fully connected layers to get the output for a given input.

Convolution layer is the main building block CNN. Convolution means merging two sets of data. Here, the filter, also called as kernel is convolved with input image to get a featured map.

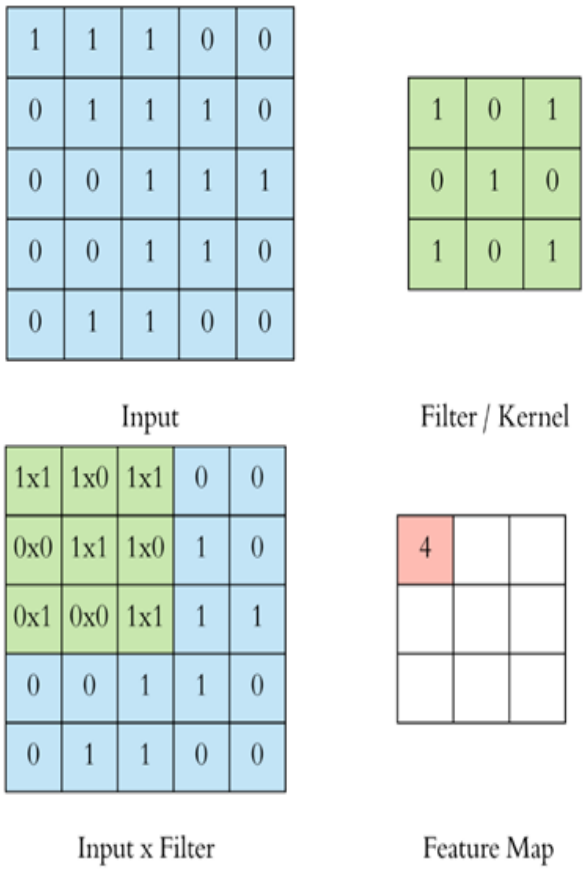

Fig.2 An example showing input and filter and Filter sliding over the input

Consider the fig.2 Where blue area is the input and green area is the filter. This filter slides over the input, where, matrix multiplication is done element wise and the sum gives feature map as shown in fig 2. In the above mentioned example we have considered in 2D but in general the image is taken as a 3D matrix. Usually we use many filters which slide on input and result in different feature maps which are combined together to get a single output from the convolution layer. Stride tells us by how much value the filter slides over the input. Generally, the stride value is 1. If we want less overlap we can take higher stride value. In fig.2 we can observe that feature map size is not same as the input. So, we use padding by bordering the input with zeros. Now, the dimensions of both input image and feature map will match. By doing this the possibility of image to shrink is eliminated.

Pooling is used to reduce the dimensions. The height and width of the feature map is reduced but the depth is maintained the same. There are two types of pooling

- max pooling
- average pooling

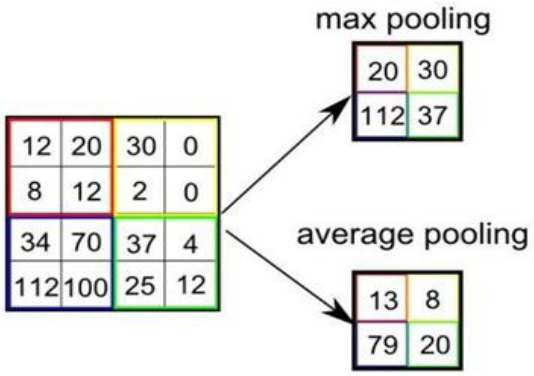

Fig.3 Types of pooling

Max pooling which is commonly used type of pooling considers the maximum value from the pooling window. Average pooling takes all the values from a pooling window and computes the average value as shown in fig.3.

Fully connected layers are added after the convolution and pooling layers to complete the CNN architecture. The output from the convolution and pooling layers is 3D but the output from the fully connected layer is 1D. So, the output from the last pooling layer is flattened to make it 1D.

In the fig1 the hidden layers are considered to be convolution layer and pooling layer whereas output layer is considered as a fully connected layer.

Text detection is a tough task because

- The light may be so bright causing saturation of the image or the light may not be sufficient enough
- The surface of the text may be reflective which make it tough to capture the image due to reflection and refraction phenomena's.
- The resolution of the camera may below standard value
- When compared to a scanner the sensor noise is high of a camera
- The images may be blurred at times
- The text may be at an angle which makes it hard to detect the text

III. HARDWARE REQUIREMENT& RESULTS

The components required for the hardware implementation are as follows

- Raspberry Pi
- Power supply
- Camera
- Speaker
- Mouse/ Push button
- HDMI cable

Raspberry pi is a small computer whose size is of a credit card. A mouse and keyboard can be used to operate it when connected to a display. We have chosen raspberry pi as it supports python, the language in which the code is written. And also, the cost of the pi is low and it is portable. Here, we are using Raspberry pi 3 B+ model.

Specifications :

Raspberry PI 3B+:

- SOC - Broadcom BCM2837B0
- CPU - 1.4 GHz
- Memory - 1GB
- Networking - Ethernet, 2.4/5 GHz wireless
- Storage - MicroSD slot
- 40 pin GPIO
- Power Source - 5V
- Ports: HDMI, audio-video 3.5mm jack, 4xUSB, Camera Interface, Display Interface, Ethernet.

Pi camera v2 :

- Sony IMX219 Sensor.
- 8 MP camera capable of taking picture of 3280 x 2464 pixels.
- Capturing video at 1080p 30fps , 720p 60fps and 640 x 480p 90fps resolutions.
- Supports the latest version of Raspbian OS.
- Supports Raspberry Pi 1, Pi 2 and Pi 3 and Models A, B and B+.
- Applications of Pi Camera: CCTV security, auto motion detection, time lapse photography.

Power supply:

- PSU Current Capacity : 2.5 A
- Total Peripheral Current Draw from USB : 1.2 A
- Active Current Consumption from Bare-Board : 500 mA

Speaker:

- Earphones can also be used as audio output.
- Standard Speaker can be used with audio amplifier.
- Bluetooth speaker can also be used for wireless audio output.

Display Output:

- 1.3/1.4a HDMI cable
- Transfer speed of up to 10.2 Gbps

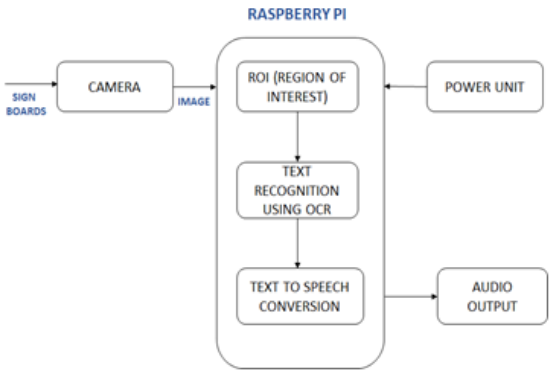

Fig.4 Hardware block diagram

Before running the code in the command window there are necessary software for efficient running of the EAST Text Detector. They are as follows:

1. Python 2.7
2. OPENCV 3.4.x to 4.0.x
3. Tesseract OCR

As these are open sourced softwares they are free of cost and can be used over various platforms such as Windows, MAC, Linux etc.

Fig.5 output of text detection block from a picture taken at main gate of VNRVJIET

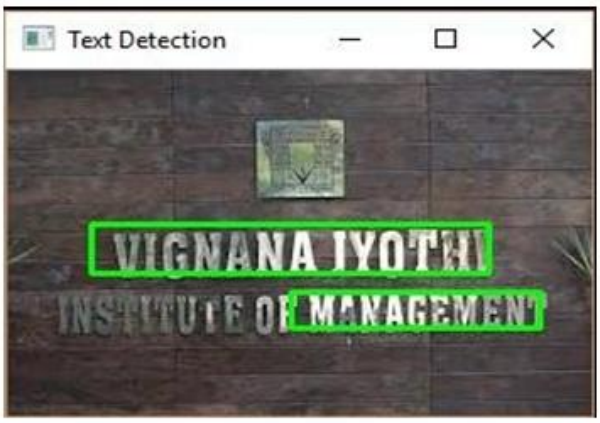

Fig.6 output of text detection block from another picture taken at VNRVJIET

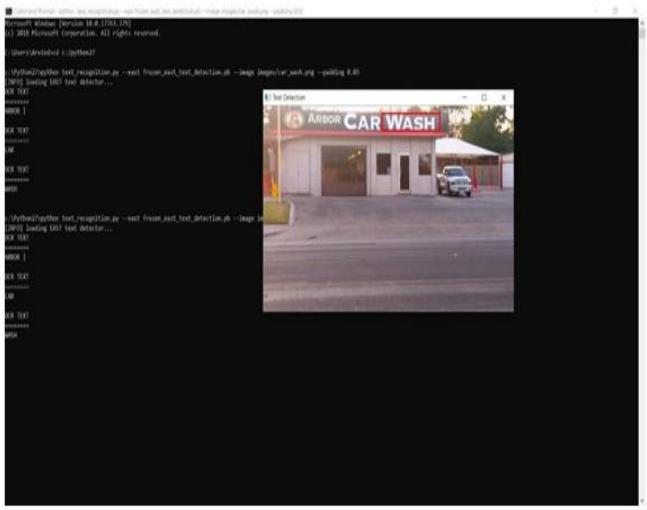

Fig.7 output of text detection and recognition block for a sample picture

REFERENCES

1. Kwang In Kim, Keechul Jung "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm" - Pattern Analysis and Machine Intelligence, IEEE

2. Chucai Li et.al "Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons", IEEE Transactions on Mechatronics, June 2014
3. R. Lienhart and A. Wernicke entitled "Localizing and segmenting text in images and videos," ,IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 4, pp. 256 –268, 2002.,
4. Chucai Li and Ying Li Tian entitled "Text string detection from natural scenes by structure based partition and grouping," IEEE Trans. Image Process., vol. 20, no. 9,pp. 2594– 2605,Sep. 2011