ISSN 2278-3091
Volume 10, No.4, July - August 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse171042021.pdf
https://doi.org/10.30534/ijatcse/2021/171042021

Object Detection with Voice Sensor and Cartoonizing the Image

MD.Salar Mohammad[1], Bollepalli Pranitha[2], Shivani Goud Pandula[3] , Pulakanti Teja Sree[4]

[1]Sreyas Institute of Engineering and Technology , India , salarmohammad@sreyas.ac.in
[2]Sreyas Institute of Engineering and Technology , India , bollepallipranitha228@gmail.com
[3]Sreyas Institute of Engineering and Technology , India , shivanigoud1249@gmail.com
[4]Sreyas Institute of Engineering and Technology , India , tejasree.pulakanti@gmail.com

ABSTRACT

Object detection is a general term to describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs, identifying objects in live captured images. Object detection combines these two tasks and localizes and classifies one or more objects in an image. Object localization refers to identifying the location of one or more objects in an image and drawing abounding box around their extent. Image classification involves predicting the class of one object in an image. In this application SAPI.spVoice is used inorder to add voice. Voice sensor is used especially for the people who cannot see objects in a particular image.

We present YOLO, a new approach to object detection. YOLO, is a technique for object recognition designed for speed and real-time use. YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second.

Cartoonizing an image will transforms the image into a cartoon image. Today we can find countless numbers of photo editing applications on the internet that allow us to transforms images into cartoons on the internet. It's similar to BEAUTIFY or AI effect in cameras of modern mobile phones. It can be taken as smoothening of an image to an extent. It makes an image look vicious and like water paint, removing the roughness in colors.

So, this application will allow us to detect and identify the objects in an image along with voice sensor which converts annotated text to speech and transforms an image into a cartoon image without using any external tool.

Key words : Object Detection, YOLO-You Look Only Once, NMS- Non-Max Suppression, IoU-Intersection of Union,Cartoonizing,Voice Sensor-win32com.client.

PROBLEM STATEMENT

The main aim of this project is to recognize the objects in an image along with voice sensor which converts annotated text to speech and cartoonizing the image. When we look at images or videos, we can easily locate and identify the objects of our interest within moments. For Computers it is a very big task to detect the objects. Blind people cannot detect objects in an image. So, Our application uses voice sensor in order to help the blind people. This application mainly uses the new approach to Object detection i.e., YOLO(You Only Look Once).

Cartoonizing an image will transforms the image into a cartoon image. It's similar to beautify or AI effect in cameras of modern mobile phones. It can be taken as smoothening of an image to an extent. It makes an image look vicious and like water paint, removing the roughness in colors.

OBJECTIVE

Object detection is a key ability required by most computer and robot vision systems. The latest research on this area has been making great progress in many directions. The Objective of object detection with voice sensor is to detect the object within an image with corresponding class ids regardless of its position,scale,view within an image and converting the class ids to speech.

1. INTRODUCTION

1.1 Object Detection

Object recognition is to describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs. Image classification involves activities such as predicting the class of one object in an image. Object localization is refers to identifying the location of one or more objects in an image and drawing an abounding box around their extent. Object detection does the work of combines these two tasks and localizes and classifies one or more objects in an image. When a user or practitioner refers to the term "object recognition", they often mean "object detection". It may be challenging for beginners to distinguish between different related computer vision tasks.

So, we can distinguish between these three computer vision tasks with this example:

Image Classification: This is done by Predict the type or class of an object in an image.

Input: An image which consists of a single object, such as a photograph.

Output: A class label (e.g. one or more integers that are mapped to class labels).

Object Localization: This is done through, Locate the presence of objects in an image and indicate their location with a bounding box.

Input: An image which consists of one or more objects, such as a photograph.

Output: One or more bounding boxes (e.g. defined by a point, width, and height).

Object Detection: This is done through, Locate the presence of objects with a bounding box and types or classes of the located objects in an image.

Input: An image which consists of one or more objects, such as a photograph.

Output: One or more bounding boxes (e.g. defined by a point, width, and height), and a class label for each bounding box.

One of the further extension to this breakdown of computer vision tasks is object segmentation, also called "object instance segmentation" or "semantic segmentation," where instances of recognized objects are indicated by highlighting the specific pixels of the object instead of a coarse bounding box. From this breakdown, we can understand that object recognition refers to a suite of challenging computer vision tasks.

For example, image classification is simply straight forward, but the differences between object localization and object detection can be confusing, especially when all three tasks may be just as equally referred to as object recognition.

Humans can detect and identify objects present in an image. The human visual system is fast and accurate and can also perform complex tasks like identifying multiple objects and detect obstacles with little conscious thought. The availability of large sets of data, faster GPUs, and better algorithms, we can now easily train computers to detect and classify multiple objects within an image with high accuracy. We need to understand terms such as object detection, object localization, loss function for object detection and localization, and finally explore an object detection algorithm known as "You only look once" (YOLO).

Image classification also involves assigning a class label to an image, whereas object localization involves drawing a bounding box around one or more objects in an image. Object detection is always more challenging and combines these two tasks and draws a bounding box around each object of interest in the image and assigns them a class label. Together, all these problems are referred to as object recognition.

Object recognition refers to a collection of related tasks for identifying objects in digital photographs. Region-based Convolutional Neural Networks, or R-CNNs, is a family of techniques for addressing object localization and recognition tasks, designed for model performance. You Only Look Once, or YOLO is known as the second family of techniques for object recognition designed for speed and real-time use.

1.2 Cartoonizing an Image

Image Processing – In the field of the research processing of an image consisting of identifying an object in an image, identify the dimensions, no of objects, changing the images to blur effect and such effects are highly appreciated in this modern era of media and communication. There are multiple properties in the Image Processing. Each of the property estimates the image to be produced more with essence and sharper image. Each Image is examined to various grid. Each picture element together is viewed as a 2-D Matrix. With each of the cell store different pixel values corresponding to each of the picture element.

2. LITERATURE SURVEY

[1] Joseph Redmon, Santosh Divvala, Ali Farhadi - Unified, Real-Time Object Detection : A unified model for object detection which is easy to build and is trained straight on full images. The model was built to detect images accurately, fast and to differentiate between art and real images.[2] Chengji Liu, Yufan Tao - Degenerative model: A degenerative model built for detecting degraded images like blurred and noisy images .This model performed better in terms of detecting degraded images and coped better with complex scenes. [3]Wenbo Lan, Song Wang - YOLO Network Model : The number of detection frames can reach 25 frames/s, which meets the demands of real-time performance.[4]Rumin Zhang, Yifeng Yang - The images of the common obstacles were labeled and used for training YOLO. The object filter is applied to remove the unconcern obstacle. Different types of scene, including pedestrian, chairs, books and so on, are demonstrated to prove the effectiveness of this obstacle detection algorithm.[5]Zhimin Mo1, Liding Chen1, Wen-jing - Identification and detection

automotive door panel solder joints based on YOLO. The YOLO algorithm, proposed identifies the position of the solder joints accurately in real time. This is helpful to increase the efficiency of the production line and it has a great significance for the flexibility and real-time of the welding of automobile door panels.[6]Gatys first proposed a neural style transfer (NST) method based on CNNs that transfers the style from the style image to the content image. They use the feature maps of a pre-trained VGG network to represent the content and optimize the result image.The results for cartoon style transfer are more problematic, as they often fail to reproduce clear edges or smooth shading.[7]Li and Wand obtained style transfer by local matching of CNN feature maps and using a Markov Random Field for fusion (CNNMRF). However, local matching can make mistakes, resulting in semantically incorrect output.[8]Chen proposed a method to improve comic style 9466 transfer by training a dedicated CNN to classify comic/noncomic images.[9] Liao proposed a Deep Analogy method which keeps semantically meaningful dense correspondences between the content and style images while transferring the style. They also compare and blend patches in the VGG feature space.

3. METHODOLOGY

Object detection is done using YOLO algorithm. YOLO is a single stage detector.win32com.client is used to convert the annotated text to speech. To achieve the basic cartoon effect, a bilateral filter and edge detection is used. The bilateral filter will reduce the color palette, or the numbers of colors that are used in the image. It reduce noise in an image.

3.1 Technique of detection

3.1.1 YOLO

All the previous object detection algorithms have used regions to localize the object within the image. The network does not look at the complete image. Instead, parts of the image which has high probabilities of containing the object. YOLO or You Only Look Once is an object detection algorithm much is different from the region based algorithms which seen above. In YOLO a single convolutional network predicts the bounding boxes and the class probabilities for these boxes. To help increase the speed of deep learning-based object detectors, YOLO uses a one-stage detector strategy.

YOLO works by taking an image and split it into an SxS grid, within each of the grid we take m bounding boxes. For each of the bounding box, the network gives an output a class probability and offset values for the bounding box. The bounding boxes have the class probability above a threshold value is selected and used to locate the object within the image.

Image classification and localization are applied on each grid. YOLO then predicts the bounding boxes and their corresponding class probabilities for objects.

We need to pass the labelled data to the model in order to train it. Suppose we have divided the image into a grid of size 3 X 3 and there are a total of 3 classes which we want the objects to be classified into. Let's say the classes are Pedestrian, Car, and Motorcycle respectively. So, for each grid cell, the label y will be an eight dimensional vector:

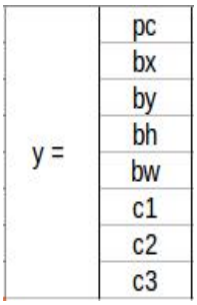

Figure3.1-Y Vector

In Figure 3.1 Y-Vector

- p_c defines whether an object is present in the grid or not (it is the probability)
- b_x, b_y, b_h, b_w specify the bounding box if there is an object
- c_1, c_2, c_3 represent the classes. So, if the object is a car, c_2 will be 1 and c_1 & c_3 will be 0, and so on

3.1.2 Non-Max Suppression

One of the most common problems with object detection algorithms is that rather than detecting an object just once, they might detect it multiple times. The Non-Max Suppression technique cleans up this up so that we get only a single detection per object. Taking the boxes with maximum probability and suppressing the close-by boxes with non-max probabilities.Discard all the boxes having probabilities less than or equal to a pre-defined threshold (say, 0.5).

3.1.3 win32com.client

win32com.client module is used to add voice that converts the annotated tect to speech. Specifically, SAPI.SPvoice is used.

3.1.4 Cartoonizing an image

The process to create a cartoon effect image can be initially branched into 2 divisions –To detect, blur and bold the edges of the actual RGB color image. To smooth, quantize and the conversion of the RGB image to grayscale. The results involved in combining the image and help achieve the desired result.

4. IMPLEMENTATION

During the implementation phase, code is generated from the deliverables of the design phase, and is the longest phase of the software development life cycle. For a developer, this is the most vital stage of the life cycle because it is where the code is created. The implementation phase may overlap with the design and testing phases. There are numerous tools (CASE tools) available to automate the production of code based on information gathered and produced during the design phase.

5. SYSTEM ARCHITECTURE

The design phase's goal is to start organizing a solution to the problem, such as a necessity document. This section describes how the opening moves from the matter domain to the answer domain. The design phase meets the system's requirements. The design of a system is most likely the most important factor in determining the quality of the software package. It has a significant impact on the later stages, particularly testing and maintenance.

The style of the document is the result of this section. This document works similar to a blueprint of solution and is used later in implementation, testing, and maintenance. The design process is typically divided into two phases: System Design and Detailed Design.

System design, also known as top-ranking design, seeks to identify the modules that should be included in the system, the specifications of those modules, and how they interact with one another to provide the desired results.

All of the main knowledge structures, file formats, output formats, as well as the major modules within the system and their specifications square measure set at the top of the system style. System design is the method or art of creating the design, components, modules, interfaces, and knowledge for a system in order to meet such requirements. It will be read by users because it applies systems theory to development.

The inner logic of each of the modules laid out in system design is determined in Detailed Design.Throughout this section, the fine print of a module square measure is sometimes laid out in a high-level style description language that is independent of the target language within which the software package will eventually be enforced.

The main goal of system design is to distinguish the modules, whereas the main goal of careful style is to plan the logic for each of the modules.

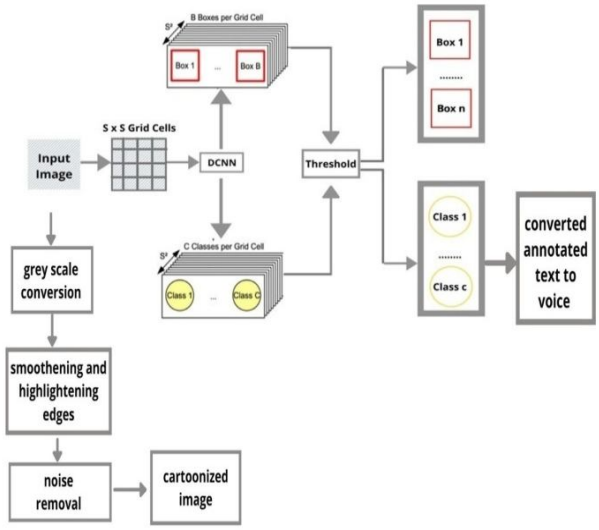

Figure 5.1 Architecture diagram

Figure 5.1 shows that , the User will upload or capture the image.It is the preference given to user whether to upload or capture the image. The given image is then divided into S X S grid cells by YOLO algorithm which is then given as a forward pass to DCNN .Then prediction of Bounding boxes will takes place and its corresponding class ids are taken into consideration. There is a possibility that more than one bounding box may be predicted for single object. So, Non-Max Suppression along with IoU have to be done.NMS will only keeps the highest score boxes.

IoU =areaof Intersection of Bounding boxes/area of Union of Bounding boxes

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. In order to convert the annotated text to speech Win32.com module from OpenCV library specifically SAPI.sp voice is used.

In order to convert the image given by the user to cartoon ,first it is converted to grayscale. cvtColor(image, flag) is a method in cv2 which is used to transform an image into the colour-space mentioned as ‘flag’. Here, our first step is to convert the image into grayscale. Thus, we use the BGR2GRAY flag. This returns the image in grayscale. A grayscale image is stored as grayScaleImage.

To smoothen an image, we simply apply a blur effect. This is done using medianBlur() function. We use bilateralFilter which removes the noise. It can be taken as smoothening of an image to an extent.Here, we will try to retrieve the edges and highlight them. This is attained by the adaptive thresholding technique.We perform bitwise_and on two images to mask them. This finally CARTOONIFY our image!

6. RESULTS

Figure6.1 command for uploading the image

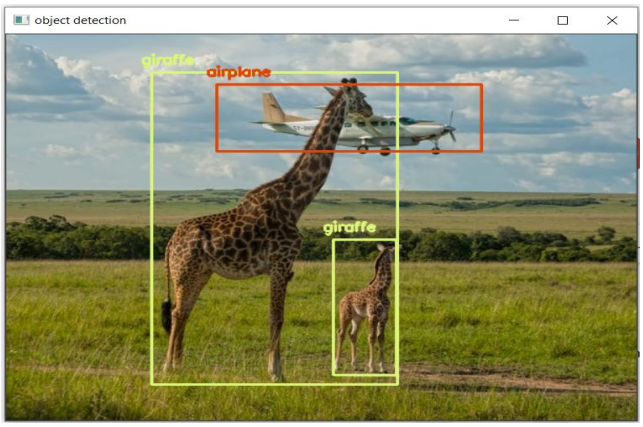

Figure 6.2 image window with the detected objects

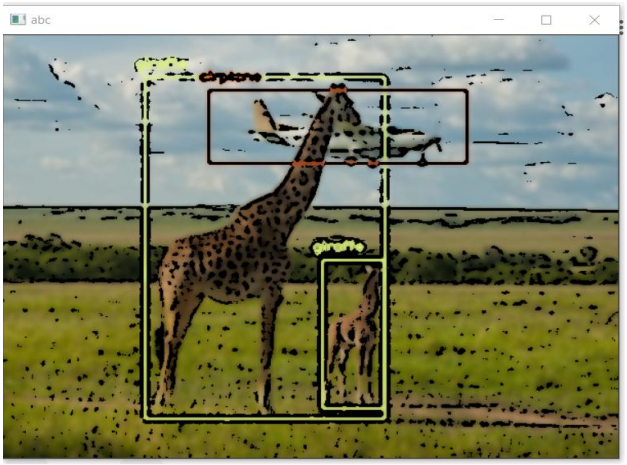

Figure 6.3 image window with cartoonized image

Figure 6.1 shows the command for uploading the image. Figure 6.2 refers to an Image Window that will be displayed to the user with detected objects . Figure 6.2 refers to an Image Window that will be displayed to the user as a separate window with Cartoonized Image.

Figure 6.4 command for capturing the image

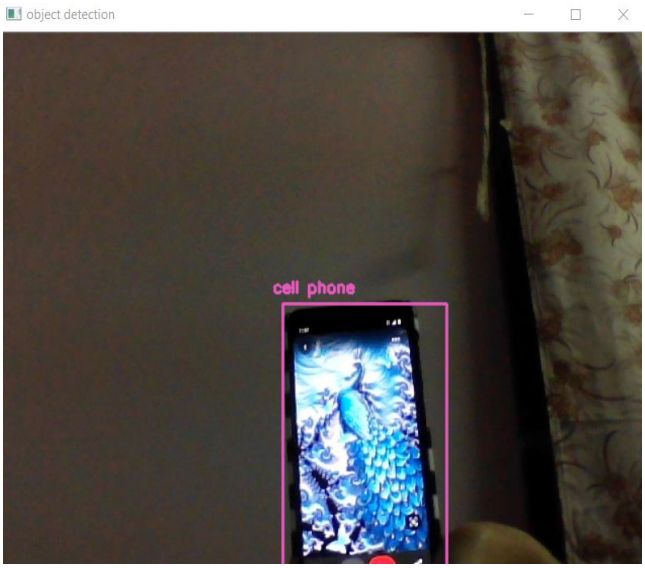

Figure 6.5 image window with the detected objects

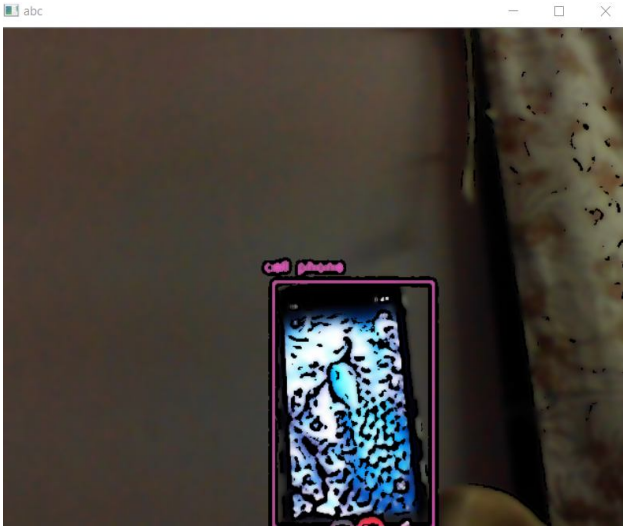

Figure 6.6 image window with cartoonized image

Figure 6.4 shows the command for capturing the image. Figure 6.5 refers to an Image Window that will be displayed to the user with detected objects . Figure 6.6 refers to an Image Window that will be displayed to the user as a separate window with Cartoonized Image.

7. CONCLUSION

Object detection with Voice Sensor and Cartoonizing an Image can be used widely to provide the blind with privacy and convenience in everyday life. Also, it is expected to be applied to industrial areas where diminished visibility occurs, such as coal mines and sea beds, to greatly help production and industrial development in extreme environments.

This application aims to enable people with visual impairment to live more independently. People with visual impairment will be able to overcome some threats that they may come across in their day to day life that may be either while reading a book or traveling through the city by making efficient use of the application and its associative voice feedback. Thus, helping visually impaired people to **'See Through the Ears'**.

Cartoonizing an image will transforms the image into a cartoon image. Today we can find countless numbers of photo editing applications on the internet that allow us to transforms images into cartoons on the internet. It's similar to BEAUTIFY or AI effect in cameras of modern mobile phones. It can be taken as smoothening of an image to an extent. It makes an image look vicious and like water paint, removing the roughness in colors.

8. FUTURE SCOPE

- Object detection is a key ability for most computer and robot vision system. Although great progress has been observed in the last years, and some existing techniques are now part of many consumer electronics (e.g., face detection for auto-focus in smartphones) or have been integrated in assistant driving technologies.
- In the fields of healthcare and security systems. In the domain of healthcare, medical image analysis can be performed using image extraction or object detection systems for computer vision predictive analytics and therapy. Identification of cancer cells in tissue biopsy may serve as an example for the above technique.
- It is impossible for humans to reach the depth parts of sea as they cannot handle pressure. So, Object detection systems for nano-robots or for robots is used to explore areas that have not been seen by humans.

9. ACKNOWLEDGEMENT

We have tried our best to present Paper on the "Object Detection with Voice Sensor and Cartoonizing the Image" as clearly as possible. We are also thankful to our Guide Prof.Md Salar Mohammad for providing the technical guidance and suggestions regarding the completion of this work. It's our duty to acknowledge their constant encouragement, support and, guidance throughout the development of the project and its timely completion. We are also thankful to Prof. Abdul Nabi Shaik(HOD, Computer Science and Engineering), without his support and advice our project would not have shaped up as it has.

REFERENCES

1. Aditya Raj, Manish Kannaujiya, Ajeet Bharti, Rahul Prasad, Namrata Singh, Ishan Bhardwaj " Model for Object Detection using Computer Vision and Machine Learning for Decision Making " International Journal of Computer Applications (0975 – 8887) .
2. Cartoonizing an image,https://data-flair.training/blogs/cartoonify-image-opencv-python/
3. Global data on visual impairment, World Health Organization.
4. https://www.who.int/blindness/publications/globaldata/en/
5. Google cloud Text to Speech, https://cloud.google.com/text-to-speech
6. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi "You Only Look Once: Unified, Real-Time Object Detection".
7. OpenCV, https://opencv.org/
8. Python programming language, https://www.python.org/
9. Rafael C. Gonzalez and Richard E. Woods "Digital Image Processing". Pearson 2018.
10. Selman TOSUN, Enis KARAARSLAN "Real-Time Object Detection Application for Visually Impaired People: Third Eye".
11. win32com.client , https://pbpython.com/windows-com.html