International Journal of Advanced Development in Science and Technology
Volume : 3 Issue : 03 April 2021 www.ijadst.com e-ISSN: 2582-1059

REAL-TIME OBJECT AND TEXT DETECTION SYSTEM USING DETR AND OCR ALGORITHM

[1] R Nithya, [2]S Sree Varshini, [3] K Nattar Kannan, [4] R K Kapilavani,
[1,2,]Student, [3,4] Supervisor,
PrinceShriVenkateshwaraPadmavathy Engineering College, Ponmar.

ABSTRACT

Object detection and object localization are the main implementations of deep learning. Object detection, as one of the most fundamental and challenging problems in Computer vision, has received great attention in recent years. Generally, Object detection includes detecting instances of visual objects of various specific classes. Training an enormous amount of data set to effectively detect various objects is impossible and is a major challenge for object detection. Text in natural scenes typically adds semantic information to the scene.

It can add identification to various detected objects like name, brand, or type of a product and recognize information from nameplates or written boards. The necessity of the implementation of object and text detection in real-time for intelligent transportation systems and automatic mobility systems plays a vital role. Our proposed project is to build a system that detects both text and object in real-time that can act as an aid in autonomous driving systems (ADS) and Smart mobility services.

We used Detection Transformer (DETR) algorithm which is a fresh design for object detection systems based on transformers and bipartite matching loss for direct set prediction, and Optical Character Recognition (OCR) for text detection.

Our proposed system can identify objects based on MS COCO datasets and can determine the distance and direction of the detected objects along with the object name. Thus, our proposed system can be an effective technological solution to identify objects and text in front of them and can be used in various computer vision applications.

Keywords-Objectdetection, Object localization, deep learning, OCR, Detection Transformation.

I. INTRODUCTION

Object detection is a process of finding all the possible instances of real-world objects, such as human faces, flowers, cars, etc. in images or videos, in real-time with utmost accuracy. The object detection technique uses derived features and learning algorithms to recognize all the occurrences of an object category. Object detection technique helps in

the recognition, detection, and localization of multiple visual instances of objects in an image or a video. It provides a much better understanding of the object as a whole, rather than just basic object classification.

In our project, for object detection (DETR) Detection transformer is used.The main ingredients of the new framework, called Detection Transformer or DETR, are a set-based global loss that forces unique predictions via bipartite matching, and a transformer encoder-decoder architecture.

Given a fixed small set of learned object queries, DETR reasons about the relations of the objects and the global image context to directly output the final set ofpredictions in parallel.The new model is conceptually simple and does not require a specialized library, unlike many other modern detectors. DETR demonstrates accuracy and run-time performance on par with the well-established and highly-optimized Faster RCNN baseline on the challenging COCO object detection dataset. Moreover, DETR can be easily generalized to produce panoptic segmentation in a unified manner. We show that it significantly outperforms competitive baselines.

In our project, OCR is used for text detection. OCR (Optical Character Recognition) is an extensive technology to recognize text inside images, such as scanned documents, photos, and name cards. OCR technology is used to convert implicitly all kinds of text in images and also text such as typed, handwritten, or printed into machine-readable text data.The process of OCR is most commonly used to turn hard copy legal or historic documents into PDFs. Once placed in this soft copy, users can edit, format, and search the document as if it was created with a word processor.Text recognition involves two steps: first, detecting and identifying a bounding box for text areas in the image, and within each text area, individual text characters. Second, identifying the characters. Using this, Detection of text is performed.

II. RELATED WORK

"The Survey of the Four Pillars for small objects"[2] The project focuses on improving the performance of object detection performance comparing to generic object detection architectures.

The project states and explains the ways to improve the performance of object detection for real-world applications, such as self-driving cars, unmanned aerial vehicles, and robotics. The State-of-art datasets for small objects are collected and the performance of different methods on these datasets are reported. The state-of-art small object detection networks are investigated along with the differences and modifications to improve the detection

performance comparing to generic object detection architectures.

The work is done by Zhu et al. [7]. The authors use information from text regions in natural scenes to improve object/scene classification accuracy. The authors combine visual features extracted from the full image with features extracted only from detected text regions. The project does not recognize the text from the images.

Object Reading: Text recognition for object recognition [8]. This uses text recognition to aid in visual object class recognition. To this end, we first propose a new algorithm for text detection in natural images. The proposed text detection is based on saliency cues and a context fusion step.

III. PROPOSED FRAMEWORK

Our proposed project is to build a system that can detect both text and object in real-time, that can act as an aid in ADS and Smart mobility services. When an object is detected in front of the camera while enabling live stream, bounding boxes are applied mentioning the name (label) of the objects. The distance and direction of the detected objects during live stream are calculated and displayed. When a text is detected in front of the camera while enabling live stream, a bounding box is applied to the text and it is displayed. Thus, our proposed system can be an effective technological solution to identify objects and text in front of them and can be used in various computer vision application.

A.OBJECT DETECTION

Object detection is done using DETR with a pre-trained dataset. The dataset consisting of 91 object classes is collected from MS COCO.End to End modelling is done applying the DETR technique.In real-time, input is taken from live streaming, and detection of object classes are executedin a prompted dialogue box with bounding boxes locating the objects with their label.

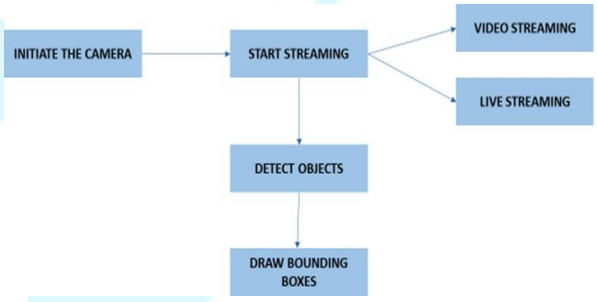


Figure 1:Object detection

B. TEXT DETECTION

Firstly, detecting and identifying a bounding box for text areas in the image.Second, the detected text area is pre-processed to remove noise and the result is passed to segmentation part to segment the individual characters. The segmented characters are normalized and passed to OCR where it will be converted into encoded text. The characters are recognized using Template matching.

International Journal of Advanced Development in Science and Technology
Volume : 3 Issue : 03 April 2021 www.ijadst.com e-ISSN: 2582-1059

When a text is placed in front of the camera the text is effectively identified. In real-time, the text area is identified and detected using a bounding box in live streaming. Text is detected and read in the prompted dialogue box and the result is shown.

C. DISTANCE AND DIRECTION

The distance from the system is calculated when the object mode is chosen.The bounding box is constructed on an object and the distance from the object is calculated based on the bounding box measurement taken along the x and y-direction.Bounding boxes are one of the most popular and recognized tools when it comes to image processing for image and video annotation projects.On successful object detection, the direction and distance of the object are effectively calculated and displayed.For instance, when a mouse object is detected, the resulting output tells that the mouseobject is on the right and with approximate distance.

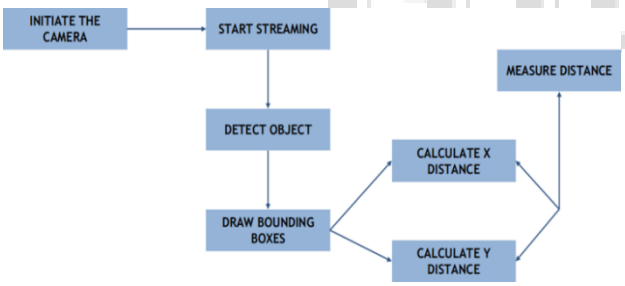
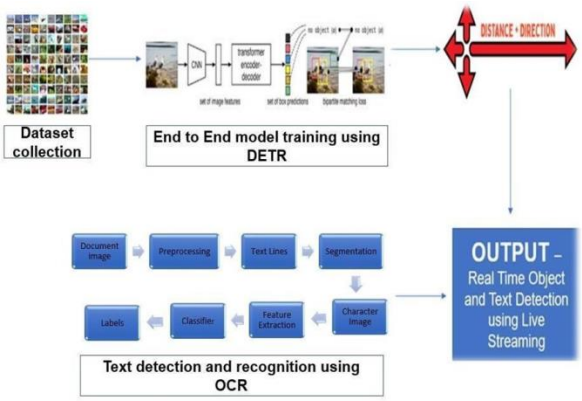


Figure 2: Distance Measurement

IV. ARCHITECTURE DIAGRAM



V. RESULTS AND DISCUSSION

FINAL OUTPUTS OBTAINED:

On initiation of the camera, the object infront is effectively detected using the DETR architecture. The below images (Figure 3 and 4) shows the object detection of the object infront of the camera.

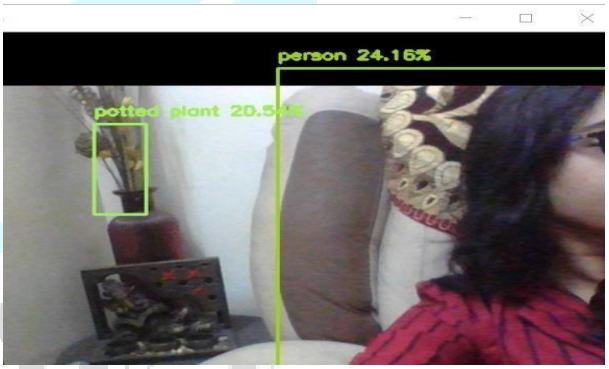


Figure 3

Figure 4

International Journal of Advanced Development in Science and Technology
IJADST Volume : 3 Issue : 03 April 2021 www.ijadst.com e-ISSN: 2582-1059

On successful object detection the direction and distance of the object are effectively calculated and displayed:

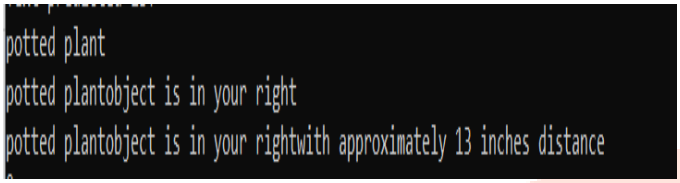


Figure 5

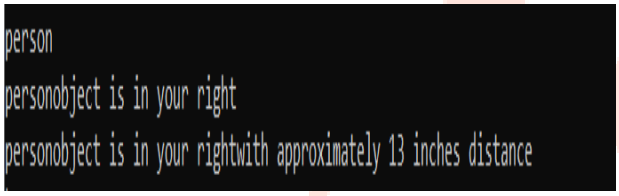


Figure 6

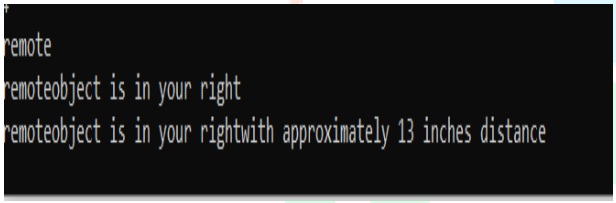


Figure 7

(Figure 5,6,7 are the outputs for the detected objects in Figure 3 and 4)

When a text is placed in front of the camera the text is effectively identified. The below images (Figure 8 and 10) shows the text being effectively displayed:

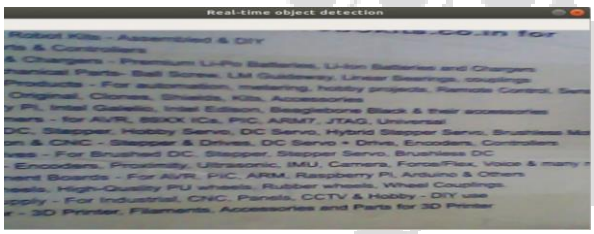


Figure 8- Detecting text during live stream

On text being successfully detected in front of the camera, the text in front of the camera is translated and read in the terminal. The below images(Figure 9 and 11) showsthe detected text from the images (Figure 8 and 10) in the output terminal.

Figure 9- Display of the detected text in termina.

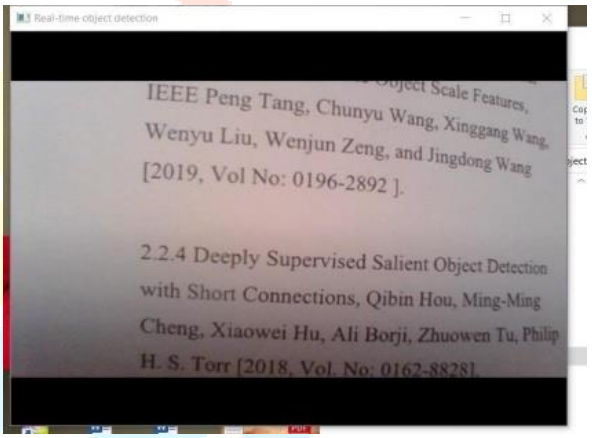


Figure 10 – Detecting text during live stream

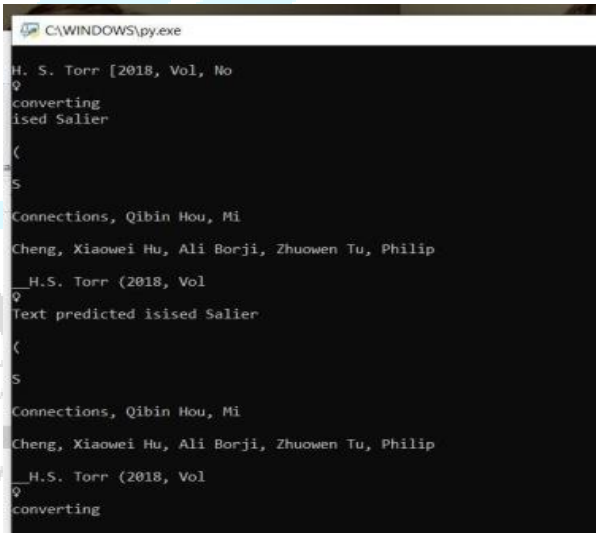


Figure 11- Display of the detected text in terminal

VI.CONCLUSION

Though text detection and object detection are of different implementations of the same domain, the necessity of the

implementation of object and text detection in real-time for intelligent transportation systems and other automatic mobility systems plays a vital role.Our proposed project can detect both text and objects in real-time mentioning their distance and direction.Thus, our proposed system can be an effective technological solution to identify objects and text in front of them and can be used in various computer vision application.It can act as an aid in ADS and Smart mobility services.

VII. REFERENCES

[1] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillo, and Sergey Zagoruyko: "End-to-End Object Detection with Transformers" Facebook AI Research paper, published-May 27 2020

[2] GuangChen ,Member, IEEE, Haitao Wang , Kai Chen, Zhijun Li , Senior Member, IEEE, Zida Song, Yinlong Liu , Wenkai Chen, and Alois Knoll, Senior Member, IEEE," A Survey of the Four Pillars for Small Object Detection: Multiscale Representation, Contextual Information, Super-Resolution and Region Proposal[2020]"

[3] Object Detection in High Resolution Remote Sensing Imagery Based on Convolutional Neural Networks With Suitable Object Scale Features, IEEE Peng Tang, Chunyu Wang, Xinggang Wang, Wenyu Liu, Wenjun Zeng, and Jingdong Wang [2019, Vol No: 0196-2892]

[4] Number Plate Recognition Using OCR Technique Er. Kavneet Kaur1 , Vijay Kumar Banga2

[5] Deeply Supervised Salient Object Detection with Short Connections, Qibin Hou, Ming-Ming Cheng, Xiaowei Hu, Ali Borji, Zhuowen Tu, Philip H. S. Torr [2018, Vol. No: 0162-8828].

[6] Weakly Supervised Object Detection via Object-Specific Pixel Gradient, Yunhang Shen, RongrongJi , Senior Member, IEEE, Changhu Wang, Senior Member, IEEE,Xi Li, and Xuelong Li , Fellow, IEEE [IEEE 2018, Vol .No.: 2162-237X].

[7] Zhu, Q., Yeh, M.C., Cheng, K.T.: Multimodal fusion using learned text concepts for image categorization. In: ACM MM. (2006)

[8]SezerKaraoglu, Jan C. van Gemert, and Theo Gevers Intelligent Systems Lab Amsterdam (ISLA), University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands