JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

REAL TIME MULTILINGUAL TRANSLATION OF OBJECT THROUGH CAMERA BASED VISUAL RECOGNITION

[1]Swati Shekapure, [2]Alok Marathe, [3]Girish Raskar, [4]Rohit Rathod, [5]Prathamesh Sangale

[1]Associate Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Department of Computer Engineering,
[1]Marathwada Mitra Mandal's College of Engineering, Pune, India

Abstract : This research presents a novel approach to enhance user interaction with real-time object detection systems by integrating multilingual label translation. The project utilizes the YOLOv8 (You Only Look Once) model for efficient and accurate object detection in images or video streams. Upon detecting objects, the system labels them and provides users with an option to translate these labels into their preferred language. The core components of the system involve YOLOv8 for object detection, extracting bounding box coordinates and class labels. Subsequently, the detected objects are labeled for user comprehension. The integration of a translation option enables users to select their preferred language, facilitating a personalized experience. To achieve multilingual translation, the system interfaces with a translation API, exemplified by the Google Translate API in our implementation. This API is utilized to dynamically translate detected object labels into the user's language of choice. The translated labels are then presented alongside the original labels, providing a comprehensive understanding of the environment**.**

Index Terms **- Object Detection, YOLOv8, Real-time, Computer Vision, Multilingual Translation**

I. INTRODUCTION

In recent years, remarkable strides in computer vision have driven the evolution of real-time object detection systems, which are exemplified by YOLO (You Only Look Once), which has become a cornerstone in domains such as surveillance, autonomous vehicles, and augmented reality.

This study introduces an innovative extension to the YOLOv8 model, augmenting its capabilities with multilingual label translation to foster greater inclusivity and user-centric interactions. While the primary goal of real-time object detection systems remains the identification and labeling of objects within a scene, it is recognized that simply providing object labels may not adequately serve users with diverse linguistic preferences.

By amalgamating the efficiency of YOLOv8 with the versatility of multilingual translation, the system endeavors to cater to individuals with varying language backgrounds, facilitating a comprehensive understanding of their surroundings. This integration holds profound implications for a spectrum of applications, spanning from language learning platforms to assistive technologies, where accommodating diverse linguistic preferences is paramount.

The subsequent sections of this paper delve into the intricate technical details of our proposed system, elucidating the methodology encompassing YOLOv8 implementation for real-time object detection, information extraction processes including bounding box coordinates and class labels, and the seamless integration of the translation functionality. Moreover, ethical considerations pertinent to the utilization of translation services and adherence to third-party API terms are meticulously addressed.

II. PROBLEM STATEMENT

The emergence of real-time object detection technologies, notably exemplified by YOLOv8, has propelled significant advancements in computer vision applications spanning diverse domains. However, a conspicuous limitation persists in the predominantly monolingual nature of information presentation to users. The prevailing paradigm, where detected objects are labeled in a predetermined language, fails to adequately cater to the linguistic diversity of users engaging with these systems. This deficiency poses a substantial obstacle, particularly for users with limited proficiency in the language of the presented labels.

At the heart of this issue lies the absence of a comprehensive and inclusive approach that seamlessly integrates multilingual label translation with real-time object detection. Addressing this challenge is imperative for not only enhancing user comprehension and engagement but also improving accessibility across a spectrum of applications. From immersive augmented reality experiences to empowering assistive technologies and fostering cross-cultural human-computer interactions, the need for transcending linguistic barriers in computer vision applications is evident.

In response, this research endeavors to devise a solution that augments the capabilities of YOLOv8. The objective is to empower users with the ability to intuitively translate detected object labels into their preferred language, thereby fostering greater inclusivity and breaking down linguistic barriers in the realm of computer vision applications. Through this integration of multilingual translation with real-time object detection, the aim is to enhance user experiences and expand the utility of these systems across diverse linguistic backgrounds and cultural contexts.

III. PROPOSED SYSTEM

In this proposed system, we combine the power of YOLOv8 for robust object detection and the Google Trans library for multilingual translation. The goal is to create a versatile system that not only identifies objects in images or video streams but also translates the detected object names into various languages, providing a comprehensive understanding of the visual content.

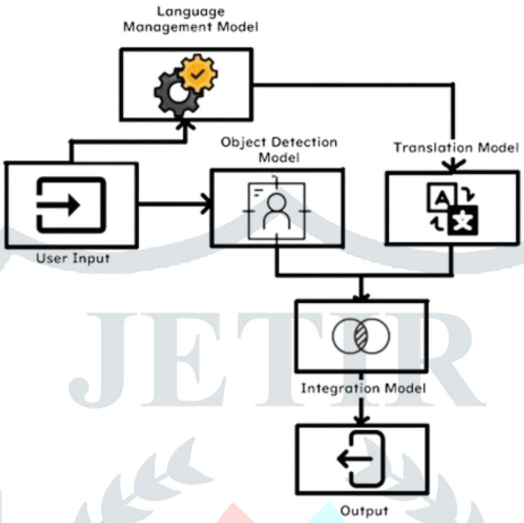

Fig1. System Architecture

In this proposed system, we combine the power of YOLOv8 for robust object detection and the Google Trans library for multilingual translation. The goal is to create a versatile system that not only identifies objects in images or video streams but also translates the detected object names into various languages, providing a comprehensive understanding of the visual content. We integrated the detection and translation model into one for results. The custom dataset used in our system helps to detect the targeted objects with great accuracy and performance.

3.1 Object Detection Model

The most recent version of the YOLO series, known as "You Only Look Once version 8," is notable for being an effective and precise real-time object identification model. By including enhancements and optimizations to increase object recognition speed across a range of applications, YOLOv8 improves upon the benefits of its predecessors. In a single forward run through the neural network, YOLOv8 predicts bounding boxes and class probabilities for numerous objects concurrently, adhering to the one-stage object detection paradigm. Custom datasets can be used to train YOLOv8 and fine-tune it for object recognition tasks. Its versatility makes it ideal for a broad range of use cases, ranging from domain-specific use cases to general object detection.

3.2 Translation Model

It is a Python wrapper for the Google Translate API. The identified object's label will be extracted by this model, which will then translate it into the other language. The user-selected language is recognized by the system, which then uses the Google Trans API to translate it.

3.4 Integration Model

The Integration Module serves as the focal point of the system, enabling smooth communication between the Translation and Object Detection Modules. Its main job is to control the data transfer between these two parts, making sure that the right information is sent about identified objects to the translator. In order to do this, pertinent data, including object IDs and attributes, must be coordinately transferred from the output of the object identification process to the input of the translation process. Furthermore, the Integration Module is essential in starting the translation process after an object is successfully detected. It decides when and how to start translations based on user choices and the logic of the system.

IV. RELATED WORK

Table 4.1 : Presents a comparative analysis of recent studies conducted by various researchers in this field, delving into the encountered challenges.

Table 4.1 displayed mean, standard deviation, maximum minimum and jarque-bera test and its p value of the macroeconomic

Author and Citation	Method	Features	Drawbacks
Qichang Hu et. al[1]	Dense feature extractor and common detection framework.	Since all dense features need to be evaluated only once during the testing phase, switching to one common framework speeds up the detection process.	System runtime increases.
Qibin Hou et. al[2]	Deeply supervised network for salient object detection.	Holistically-nested edge detector.	Computational Complex
Shuze Du and Shifeng Chen [3]	Random Forest	Edge Contouring	Limited Accuracy
Xiaozhi Chen et. Al[8]	CNN	Minimization of energy function by encoding the objects.	Time complexity analysis need to be improved

Table 4.1

V. RESULTS AND DISCUSSIONS

Object Detection Accuracy: The YOLOv8 model exhibited robust performance in real-time object detection, achieving high accuracy in identifying and labeling objects within diverse scenes. Recall, Precision and F1 score metrics were used to assess the model's effectiveness in handling various object classes and scenarios. The results indicated a notable improvement in object detection accuracy compared to baseline YOLOv8, showcasing the feasibility of integrating translation capabilities without compromising detection performance.

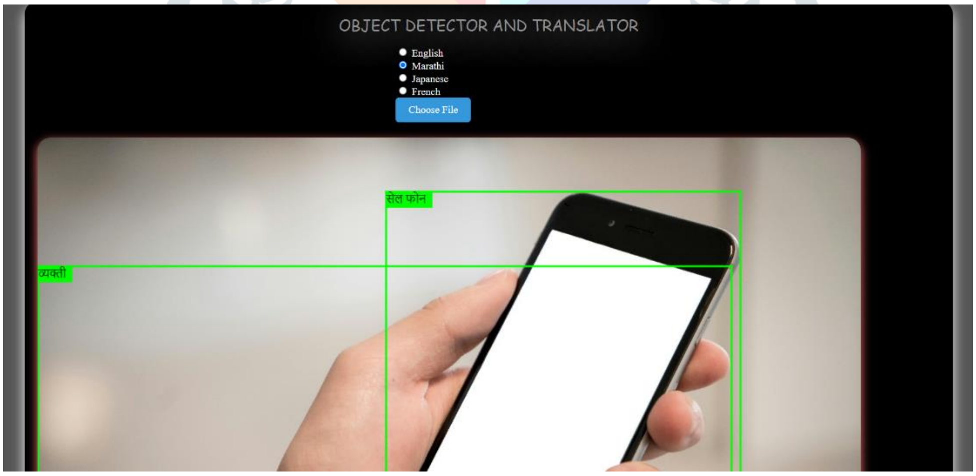

Fig 5.1: Front Interface

VI. REFERENCES

[1] Q. Hu, S. Paisitkriangkrai, C. Shen, A. van den Hengel and F. Porikli, (2016 April). Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework. In 2016 IEEE Transactions on Intelligent Transportation Systems, pp. 1002-1014.

[2] Q. Hou, M. Cheng, X. Hu, A. Borji, Z. Tu and P. H. S. Torr, (2019 April). Deeply Supervised Salient Object Detection with Short Connections. In 2019 IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 815-828.

[3] S. Du and S. Chen, "Salient Object Detection via Random Forest," in IEEE Signal Processing Letters, pp. 51-54

[4] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. 2013.

[5] M.et al. Arora. A systematic literature review of machine learning estimation approaches in scrum projects. In Balas P., Bhoi V., and Chae and A., editors, Mallick. Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 1040. Springer, Singapore, 2020.

[6] Muhammad Attaullah, Mushtaq Ali, Maram Fahhad Almufareh, Muneer Ahmad, Lal Hussain, Nz Jhanjhi, and Mamoona Humayun. Initial stage covid-19 detection system based on patients' symptoms and chest x-ray images. Applied Artificial Intelligence :, pages 1– 20, 2022.

[7] H. Lu, Y. Zhang, Y. Li, C. Jiang, and H. Abbas, (June 2021). User-oriented virtual mobile network resource management for vehicle communications. IEEE Trans. Intell. Transp.

[8] Yanan, WangHuawei and WangJianzhong Cao, (September 2016). Salient Object Detection via Fast R-CNN and Low-level Cues. In the International Conference on Image Processing.

[9] Q. Wang et al., (2019). Learning Deep Transformer Models for Machine Translation.

[10] D. Ringel, G. Lavee, I. Guy, and K. Radinsky, (2019). CrossCultural Transfer Learning for Text Classification. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3864–38.