IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Ai Based Picture Translation

Mr. Annappa Swamy D R [1], Nishitha[2], Pradeep L K[3], Rajatesh M D[4], Sahithi Rai[5]

Associate Professor, Dept. of Computer Science & Engineering, Mangalore Institute of Technology & Engineering, Moodabidri, India[1]

Student, Dept. of Computer Science & Engineering, Mangalore Institute of Technology & Engineering, Moodabidri, India[2,3,4,5]

Abstract: This work presents the development of an AI-based image translation web application designed to recognize objects within an image and provide translations in Hindi, English, and Kannada. The system integrates advanced computer vision and natural language processing techniques to deliver accurate and real-time translations. Object detection is carried out by YOLOv5s (You Only Look Once version 5 Small) which is an extremely effective deep learning model developed for real-time object identification. OpenCV applies its power for operations such as image preprocessing- image resizing, normalization, and noise reduction, ensuring high-quality inputs for the model. The system utilizes TensorFlow for processing in neural networks and PyTorch for training and fine-tuning the deep learning models involved in object recognition. To enhance its functionality, the application also employs Tesseract OCR (Optical Character Recognition) to get the text from images that shall be translated into the selected languages. With a user-friendly interface, users can upload images and receive immediate translations of recognized objects and text in Hindi, English, or Kannada. This AI-powered image translation tool offers a practical solution for overcoming language barriers, benefiting diverse user groups in fields such as education, travel, and everyday communication.

Index Terms **- AI-based Translation, Object Detection (YOLOv5s), Deep Learning, Optical Character Recognition (OCR), Multilingual Translation.**

I. INTRODUCTION

Increasing demand for uninterrupted communication among different languages around the world has highlighted the problems individuals face and different communities in transcending a linguistic gap. One of the primary obstacles is the inability to quickly translate written or labelled content in images, which can hinder understanding and accessibility in various contexts such as education, travel, and daily interactions. Traditional translation methods often rely on text input, leaving a gap when users need to translate objects or text present in images. To address this challenge, the proposed project aims to develop a web-based application that leverages advanced AI techniques for real-time object recognition and multilingual translation. The system utilizes state of-the-art deep learning models, including YOLOv5s for object detection, OpenCV for image preprocessing, and Tesseract OCR for extracting text from images. Once objects are detected, the system translates them into three major languages—Hindi, English, and Kannada—using a combination of TensorFlow and PyTorch for model training and inference. This is an AI-powered translation tool, which has an easy interface that allows users to upload images; the system later identifies and translates the objects and text detected into the specified language. The application is intended to help in achieving more inclusive communication through overcoming the language barriers for a vast number of users, such as tourists and students, and thus enhancing everyday communication between people from different linguistic backgrounds. With such sophisticated computer vision and natural language processing technologies, this system offers a workable and accessible solution to meet multilingual communication needs.

II. LITERATURE SURVEY

[1] This study focuses on the development of an AI-based language translation system for real-time classroom use, aiming to assist non-native Arabic-speaking teachers. The system integrates with classroom devices like digital podiums and projectors, providing instant translation through Google Cloud Translation API. It uses machine learning models to translate English text into Arabic, helping students improve vocabulary and comprehension. The evaluation showed that the system effectively enhanced student engagement and learning outcomes by providing immediate translations during lectures.

[2] This study focused on evaluating real-time language translation systems using advancements in artificial intelligence and machine learning. It assessed their accuracy, efficiency, and application in diverse language contexts, emphasizing their potential to break language barriers and promote global communication. The research addressed challenges in translation, such as linguistic nuances .

[3] This study focused on applying AI techniques, particularly neural machine translation (NMT) and natural language processing (NLP), to improve language translation systems. The research explored advancements in deep learning and fuzzy logic for enhancing accuracy and contextual understanding. It highlighted the potential of these technologies in addressing linguistic nuances, facilitating global communication, and expanding multilingual applications while overcoming challenges in translation precision and cultural adaptability.

III. SCOPE AND METHODOLOGY

Scope

The project aims to develop an easy-to-use web application for real-time image translation. Using AI technologies, the system will recognize objects and text in images and translate them into Hindi, English, and Kannada. It will use contain learning models like YOLOv5s for object detection, OpenCV for image processing, and Tesseract OCR for text extraction. The objective is to provide an interface as simple as possible through which users can upload images and receive quick translations. Also, the project aims at increasing language support and enhancing the accuracy of object recognition so that it will be a very valuable tool for language barrier-breaking across education, travel and routine communication domains

Methodology

The project begins by collecting a large and varied dataset of images with objects and text. These images come from public datasets or user-generated content, and include things like animals, vehicles, signs, and text labels.

To prepare the data, the images are resized to a standard format, and pixel values are normalized. Data augmentation is the use of techniques such as rotating, flipping or scaling images to increase dataset diversity and therefore improve generalization while avoiding overfitting.

The core part of the system is the YOLOv5s model, as it has been devised to identify and classify objects in the images. Still, if an image contains text embedded, Tesseract OCR is applied to capture the same text. The system further proceeds to translate the detected objects and texts in the identified languages, Hindi, English, or Kannada. The model trains from preprocessed images. Optimizing performance is given by hyperparameter tuning, including learning rate and batch size; the number of layers of the model enables accurate object detection and efficient text extraction by the system.

Users can upload images to the application, which will process the images in real time. First, the system detects and classifies objects using the YOLOv5s model. If text is present, Tesseract OCR extracts it, and then the text is translated into the selected languages. The translation is powered by a machine learning model that ensures accurate translations. The system delivers fast and accurate results, making it user-friendly and effective.

Metrics, such as accuracy, precision, and recall, have to be used to evaluate the system's performance. These helps analyze whether the system is effective at detection, text extraction, or in translation. The system is tested with different types of images to make sure it works well in real-world scenarios, and adjustments are made to improve its performance where needed.

The application is designed to be simple and easy for everyone to use. Users can upload images quickly and get immediate translations of objects and text. The image processing system has been changed to accommodate images from specific locations, such as uploading from smartphones for immediate response. Application supports multiple languages since it enables communication and proper understanding of various visual contents in other languages and thus offers access to a wider audience. The final system integrates object detection, text extraction, and translation into a single, smooth workflow. This makes the AI-powered application a practical tool for recognizing objects, extracting text, and translating it into multiple languages. Therefore, the language barrier within everyday communication, education, travel, and other fields would not be an issue for its users to easily understand and even interact with images in a specific language.

IV. System Architecture

This project follows a systematic process flow to detect objects in an image and translate their names into Kannada, ensuring a smooth transition from image input to final display. The process begins with the user uploading an image containing one or more objects, which serves as the raw input for the system. The image, in any format such as JPEG or PNG, is stored for further processing. Next, the uploaded image undergoes preprocessing to enhance its quality, involving steps like resizing, noise reduction, and contrast adjustments. This ensures the image is standardized for accurate object detection. The preprocessed image is then analyzed using OpenCV, where pre-trained models detect and classify objects by providing bounding boxes and labels. The identified object is translated in English as names for the objects identified in the image, while later, as the names of the identified objects are translated, the contextually appropriate translation is returned from the Google Cloud Translation API using natural language processing. Finally, the system displays the object names in both English and Kannada on a user-friendly interface, ensuring clear and accessible functionality. Each step builds on the previous one, transforming the input image into a meaningful multilingual output through a streamlined and efficient workflow.

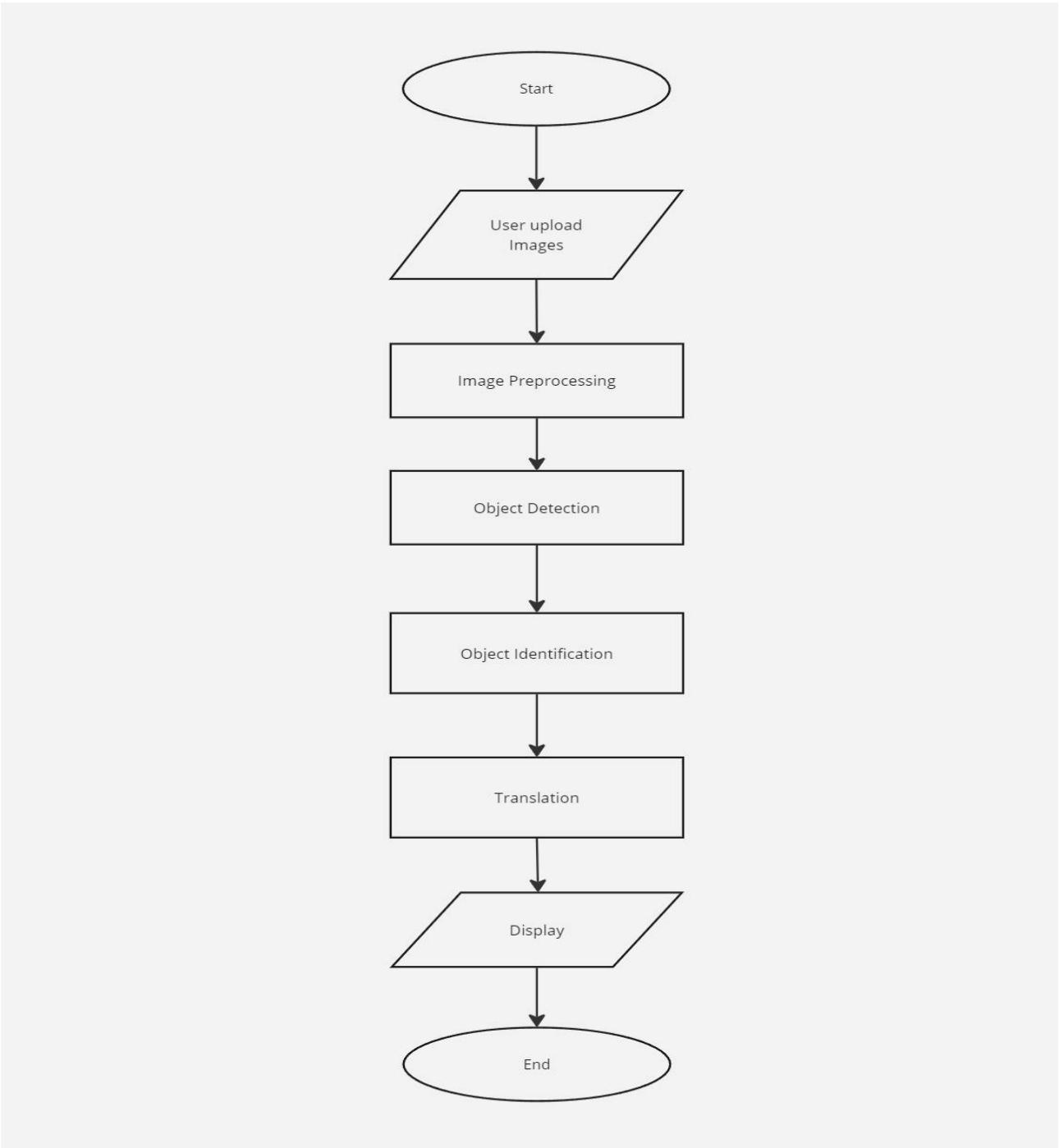

Figure 4.1: System Architecture

V. CONCLUSION

This paper presents the AI-based Picture Translation Web Application effectively integrates advanced technologies like YOLOv5s for object detection, Tesseract OCR for text extraction, and a machine translation model for real-time multilingual translation. The user can upload images for analysis to detect objects within the images and extract all readable text, which would then be translated into Hindi, English, or Kannada. By combining object recognition and text translation in one seamless platform, the application provides a practical solution for overcoming language barriers in a variety of contexts, such as education, travel, and everyday communication. It allows for an interactive user interface that can accommodate anyone regardless of technical capability. Its real-time processing and results make it beneficial to the user in translating rapidly, which also allows it to recognize objects. As the system continues to improve in terms of model accuracy and additional language support, it holds significant potential for bridging communication gaps and facilitating multilingual interactions, contributing to a more connected and inclusive world.

REFERENCES

[1] "Artificial Intelligence Based Language Translation Platform", Prince Sattam Bin Abdulaziz University, Wadi Ad Dawaser, 11990, Saudi Arabia *Corresponding Author: Manjur Kolhar. Published 23 January 2021.

[2] "The Impact of Artificial Intelligence on Language Translation: A Review", Yasir Abdelgadir Mohamed, (Member, IEEE), Akbar Khanan, Mohamed Bashir, Abdul Hakim H. M. Mohamed, (Senior Member, IEEE), Mousab A. E. Adiel, and Muawia A. Elsadig .23 February 2024.

[3] "Real Time Language Translation Using AI and ML", Prudhvi Sandeep Varma Pamudhurthy, Rajesh Komali, Abdul Basha Panyam, Ashok Katarukonda and Hiren Raithatha.1 March 2024.