INTRODUCTION TO INTERESTED DOMAIN

Artificial Intelligence and technology have been growing rapidly and has become the integral part of our life. Worldwide usage of internet, smart phones, information and communication devices, social media platforms are the influential reasons for the arise of large volume of data. These large volumes of data are growing in exponential way due to which artificial intelligence have been developing parallelly with the available resources for the utilization of that data for AI development and progression. Training various AI models for image classification, tokenizer for textual data, object classification and localization, image captioning, question answering and extremely growing generative AI with the available data have become very fruitful.

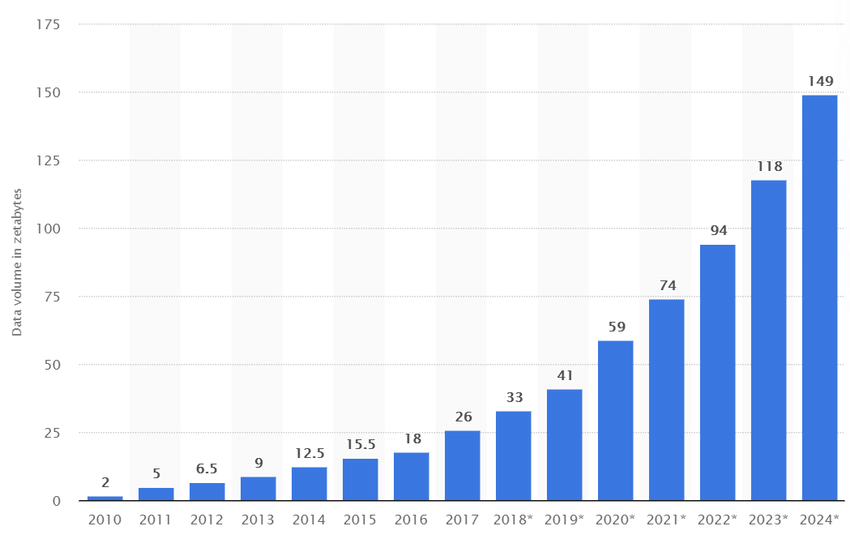


Figure :Data produced worldwide in 15 years

Object detection is one of the mostly used computer vision application in various AI and robotics applications, localization and classification of the objects from the images are seamlessly done by the state-of-art models such as YOLO, Efficient Net, F-RCNN, RCNN but the context which is another very important form of general intelligence of an AI model. The context generation part is done by the large language models such as GPT2, GEMMA, LLaMA. Models like LLaVA with multimodality can easily interpret the text and image and can be used for context-based object detection with the help of some prompt engineering on the LLM. But there have been very few efforts towards integrating these two different paradigms which opens up the huge potential in research areas such as contextual object detection in real time which opens up the opportunity of this efforts in sectors such as assistive devices for visually impaired individuals, locomotion and navigation of robots and robotic arms, augmented reality and so on.

Problem statement

Object detection has been used as the widely used computer vision tasks such as identifying and localizing the objects from the image or video along with detection in real time scenario as well. The object detection mainly revolves around the idea of first knowing whether the image/frame does or doesn’t contain the objects that have been trained to the object detection model and if the object is present then it moves towards localizing the object to find out which part of the image contain the classified object with the rectangular box. Transportation safety, autonomous vehicles, food manufacturing, biometrics and facial recognition, medical image analysis and video surveillance solutions are the application areas of the object detection. Although object detection performs well for these application areas, this project aims towards development of the AI models that have some form of human intelligence and ability which we call Artificial General Intelligence (AGI). This project will showcase the ability to provide the real time context in real world to have the human to see and think what is happening in the scene.

Project as a solution

This project proposes a solution to the traditional object detection algorithm with the limitation of context for the detected real time frames by integrating the language-based model with the vision-based object detection model with the finetuning of those models for the tailored output. This integration will also provide the research areas in the area of computer vision, natural language processing making possible the architectures like that of as Vision Transformer, Multi-modal models and so on.

BACKGROUND TO THE PROJECT

EXPECTED OUTCOMES AND DELIVERABLES

PROJECT RISKS, THREAT AND CONTINGENCY PLAN

AIM AND OBJECTIVES

METHODOLOGY FOR PROJECT DEVELOPMENT

SELECTED METHODOLOGY

GANTT CHART

WORK BREAKDOWN STRUCTURE

MILESTONES

CONCLUSION

REFERENCES

https://blog.roboflow.com/object-detection/