PILLAI COLLEGE OF ENGINEERING,

New Panvel

Department of Information technology

Academic year : 2023-24

ADAPT LEARN HUB

An Inclusive Visual Learning Platform using Visual and Textual Feature extraction technique.

PROJECT CARRIED UNDER GUIDANCE OF : PROF. Jinesh Melvin

MEMBERS :

PANDEY PIYUSH YOGESH

PATIL HARSH SANDEEP

PATIL SAMARTH VILAS

TODKAR SMRUTI MANGESH

Abstract

The "Adapt Learn Hub" is a groundbreaking project leveraging artificial intelligence, specifically object detection and semantic segmentation, to revolutionize education for students with learning difficulties. By employing advanced technology, the platform identifies and comprehends diverse objects within educational materials, creating a tailored learning experience. Integrated Text-to-Speech capabilities cater to auditory learners, complemented by adaptive algorithms enabling educators to shape Customizable Learning Paths. Interactive Learning Modules and a multi-modal approach ensure engagement across various learning styles. The platform emphasizes mobile accessibility, progress tracking, and reporting functionalities, enhancing its adaptability.Designed with a user-friendly interface, the Adapt Learn Hub aims to break educational barriers, fostering an inclusive space where every student can thrive. Continuous collaboration with educators and specialists, coupled with iterative testing, drives the platform's evolution, promising to make education more accessible and engaging for all.

Introduction:

The recent advancements in information technology (IT) have spurred extensive research aimed at addressing everyday inconveniences. Consequently, numerous conveniences have been provided for people. However, students continue to face many challenges. One solution to these challenges is the Adapt Learn Hub, which utilizes object detection techniques to identify objects, convert them into text, search for details about the objects, and then convert that information into speech. This innovative approach integrates various technologies to enhance the learning experience and alleviate the difficulties faced by students.

Literature Survey

Sr no	RESEARCH PAPER LINK	Description	KEY FEATURES	year
1	Object Detection with Voice Feedback	Object Detection with Voice Feedback Rajat Lilhare, Jitendra Meena, Nikhil More, Shubhangi Joshi	M**ethodologies Used** : Yolov3 , CNN ,deep learning **Advantages**: automated , real time, improve accuracy **Disadvantages**: supervised training and limited data availability	2023
2	https://www.comp.nus.edu.sg/~changec/publications/thinwire.pdf	Realtime Visualization of Large Images over a Thinwire E.C. Chang and C.K. Yap and T.-J. Yenz	**Methodology:**Real-time visualization Thinwire transmission **Technology used:**Integration of realtime, thinwire, and large-scale image capabilities Enhanced user-control for realtime image visualization **Advantages:**Variable resolution over time **Limitations:** fixed-size viewing window	2022
3	https://ieeexplore.ieee.org/abstract/document/8546756	Toward Personalized Adaptive Gamification: A Machine Learning Model for Predicting Performance Christian Lopez, Conrad Tucker	**Methodology :** Data Collection, Continuous Update **Technology used :** Human-Computer Interaction (HCI): HCI principles and frameworks **Advantage :** Training Model: A machine learning model is trained using the extracted features to predict the individual's performance. **Disadvantage :** Poor Data Quality	2021

Problem definition

The problem entails leveraging AI, including object detection and semantic segmentation, to cater to students with learning difficulties. The goal is to create a customizable learning platform with Text-to-Speech, adaptive algorithms, and interactive modules for diverse learning styles. Mobile accessibility, progress tracking, and collaboration with educators are paramount for inclusivity and effectiveness.

Project Scope

The project encompasses the creation of an image analysis system, beginning with dataset collection and model training for object detection and semantic segmentation. The integrated system includes pattern analysis, gamification elements, and information-to-speech conversion. The overarching objective is to deploy a user-friendly application that achieves high accuracy in object identification, interactivity through gamification, and improved accessibility through information conversion to speech.

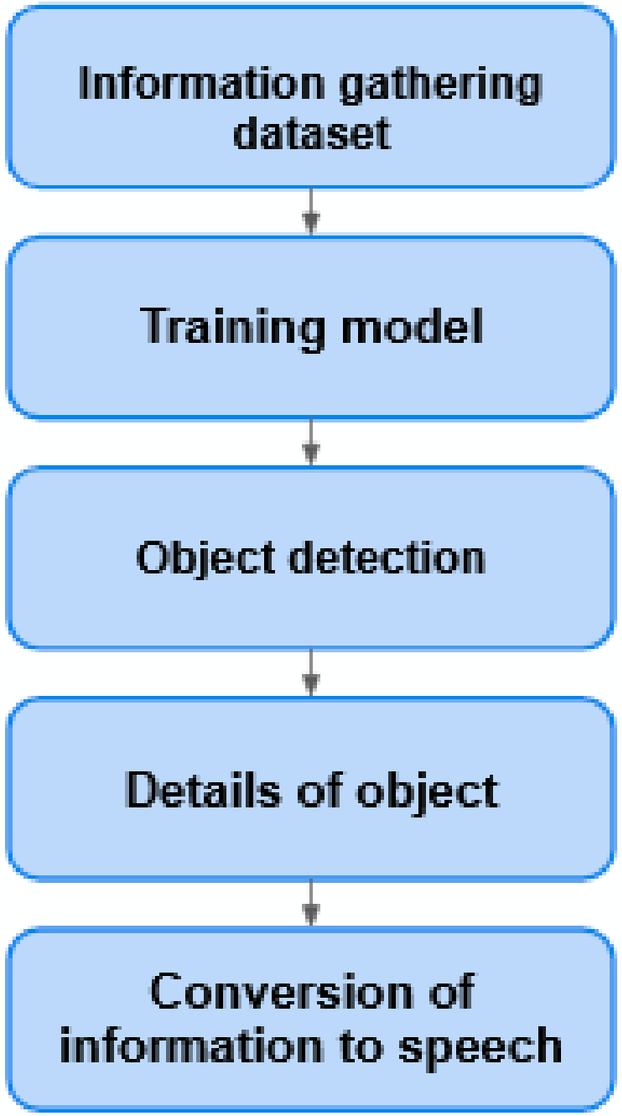
Identification of Tools ,Dataset or ideas to implement

Dataset :-

- MS COCO (Microsoft Common objects in context) dataset approx 3.30 lakh images and 80-90 classes
- Google images approx 9 million images and 600 classes
- Yolo v8 pre trained dataset

Tools and Algorithm:-

- Yolo v8 Model
- CNN - Fast RCNN
- FCN
- gTTS



(Existing System)

1 . Information Gathering Dataset:

- Collecting a diverse and representative dataset containing images and associated metadata relevant to the targeted object detection and classification task dataset used (Yolo v3 pre trained).

2. Training Model:

- Utilizing deep learning frameworks to train a neural network model on the gathered dataset, optimizing it for accurate object detection and classification.

3. Object Detection:

- Implementing the trained model to identify and locate objects within images, providing bounding box coordinates for each detected object algorithm used Yolov3.

4.Details of Object:

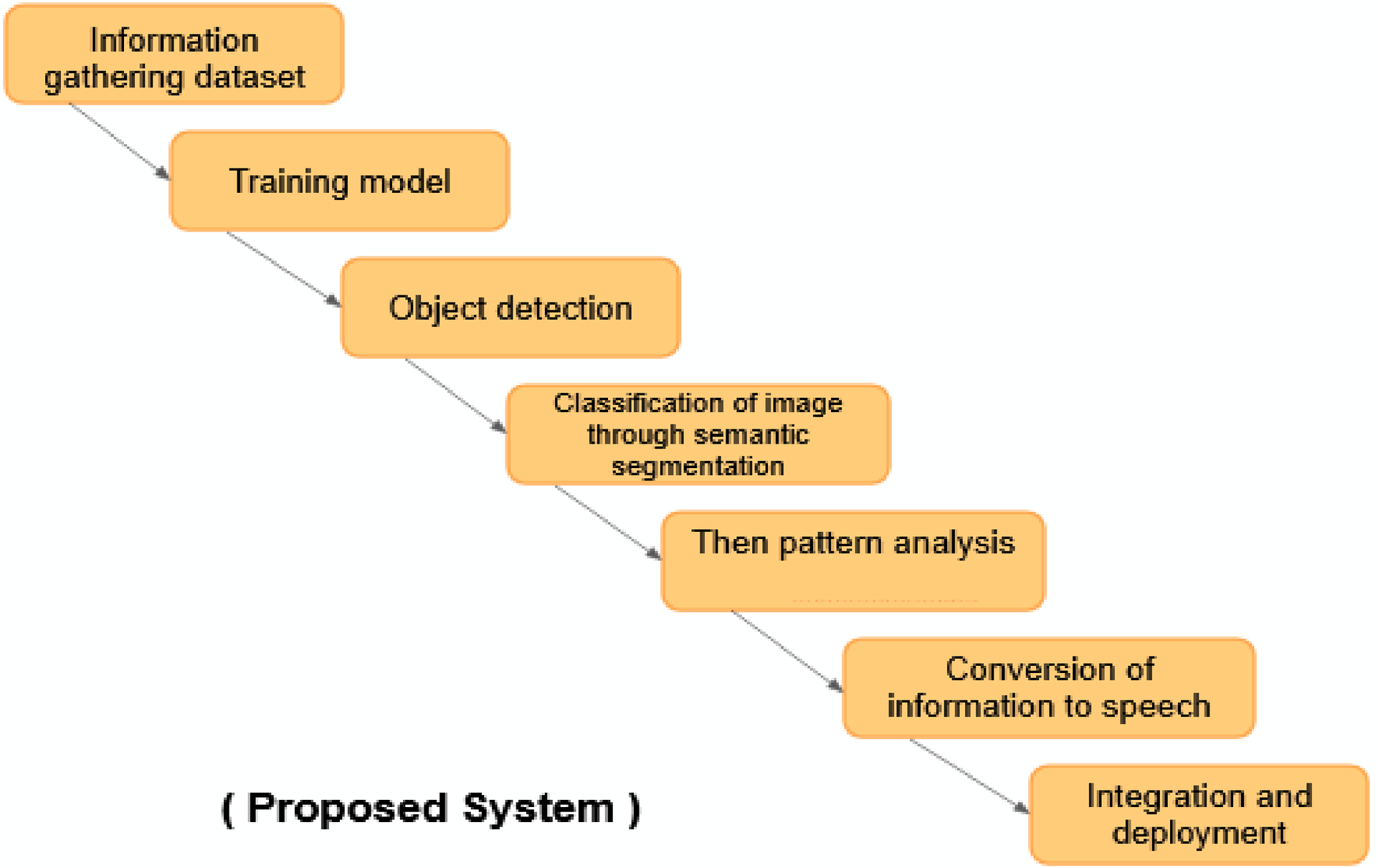
- Utilizing computer vision techniques, the model identifies and precisely locates objects within images, providing bounding box coordinates for each detected object, enabling a detailed understanding of the visual content.

5. Conversion of Information to Speech:

- Integrating a text-to-speech (TTS) with gTTS into the system to convert the analyzed information into spoken words, enhancing accessibility and user experience .

Gap Identification

- Semantic segmentation for classification of object
- Classification of image
- Pattern analysis and gamification
- Active Learning and recommendations



(Proposed System)

1. Information Gathering Dataset:

- Collecting a diverse and representative dataset containing images and associated metadata relevant to the targeted object detection and classification task Datasets used (COCO dataset and google images).

2. Training Model:

- Utilizing deep learning frameworks to train a neural network model on the gathered dataset, optimizing it for accurate object detection and classification.

3. Object Detection:

- Implementing the trained model to identify and locate objects within images, providing bounding box coordinates for each detected object algorithm required Yolo V8 and R-CNN.

4. Classification through Semantic Segmentation:

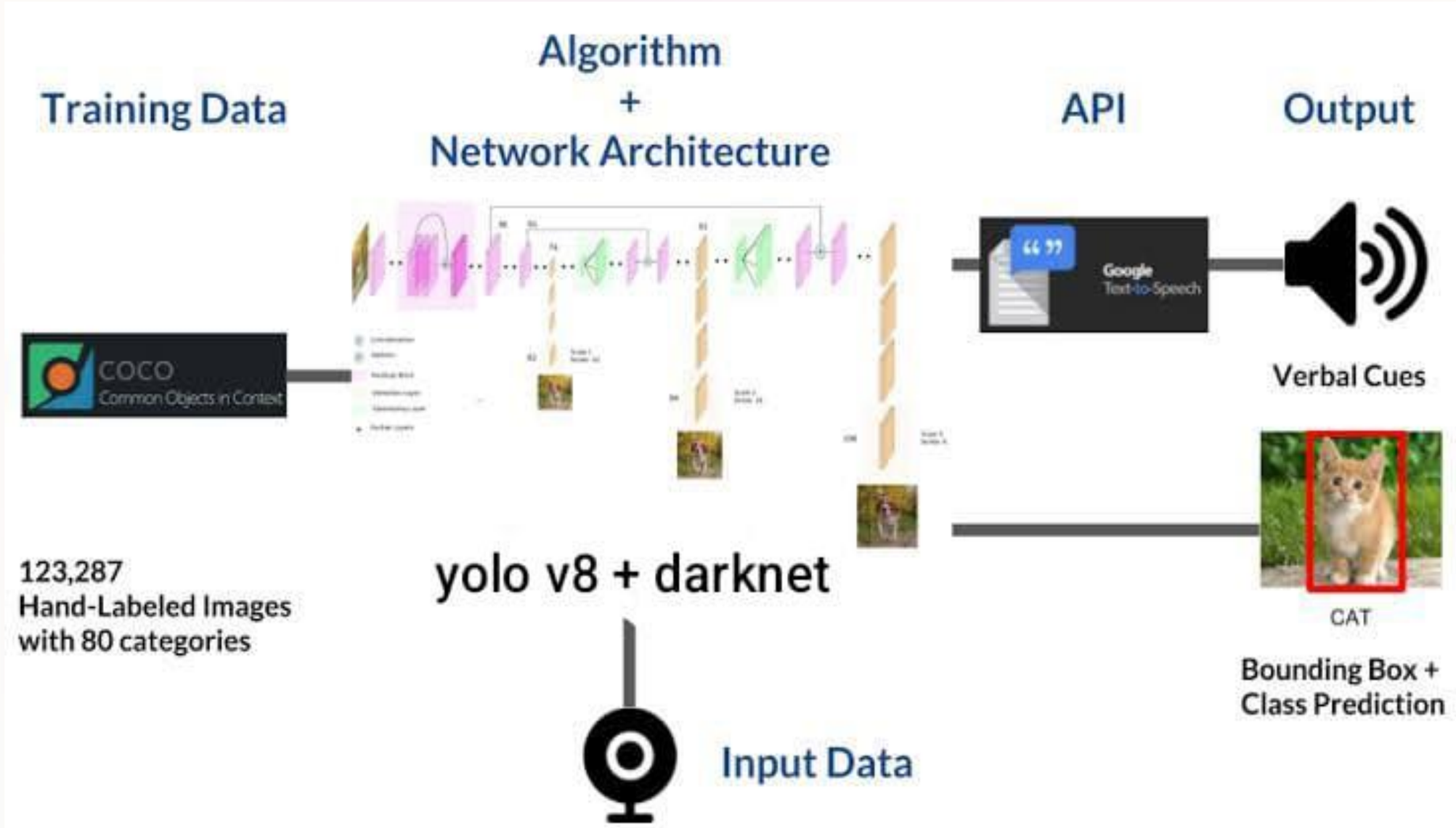
- Extending the model capabilities to perform semantic segmentation, allowing for pixel-level classification of objects(with FCN) within images for more detailed analysis .

5. Conversion of Information to Speech:

- Integrating a text-to-speech (TTS) into the system by using tool gTTS to convert the analyzed information into spoken words, enhancing accessibility and user experience.

6. Integration and Deployment:

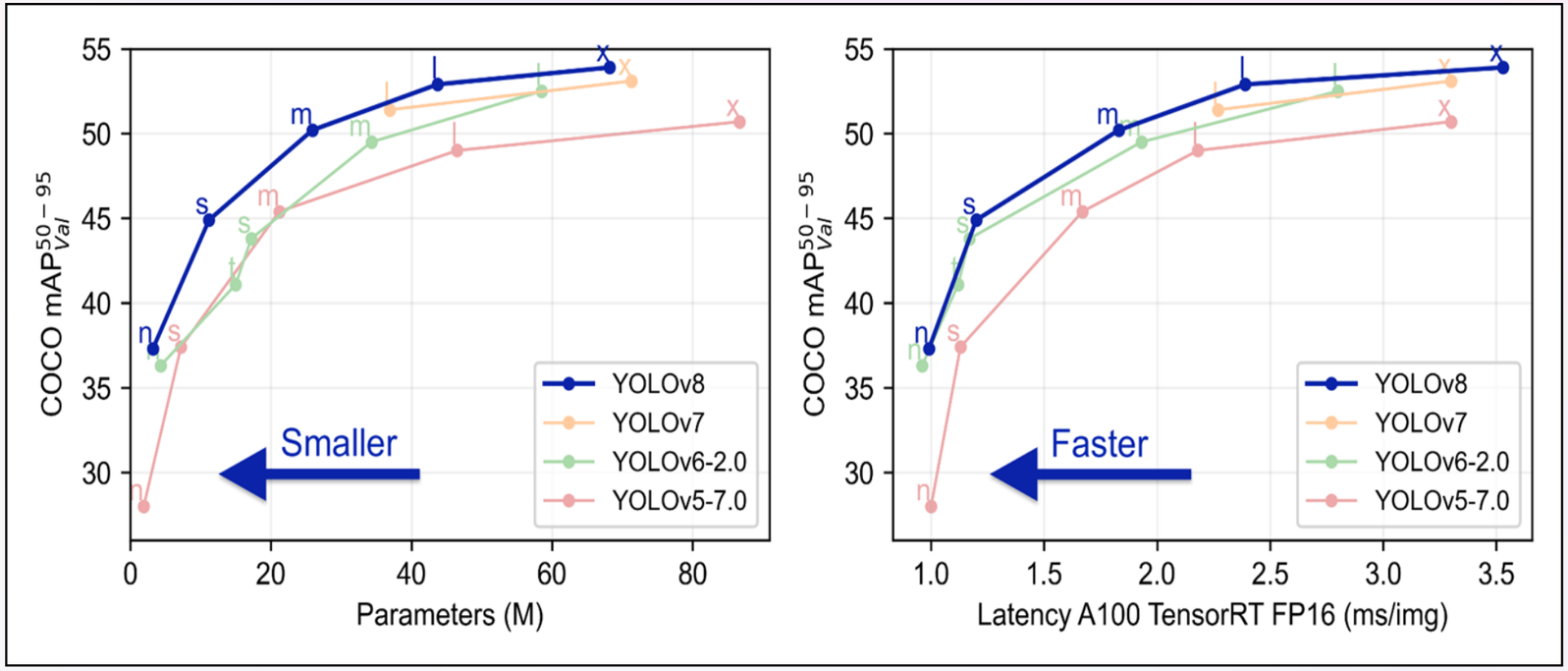
- Integrating the entire system into a cohesive application and deploying it on the desired platform, ensuring seamless functionality and accessibility for end-users.


Training Data
Algorithm
+
Network Architecture
API
Output
COCO
Common Objects in Context
Google
Text-to-Speech
Verbal Cues
123,287
Hand-Labeled Images
with 80 categories
yolo v8 + darknet
CAT
Bounding Box +
Class Prediction
Input Data

Methodology And Algorithms

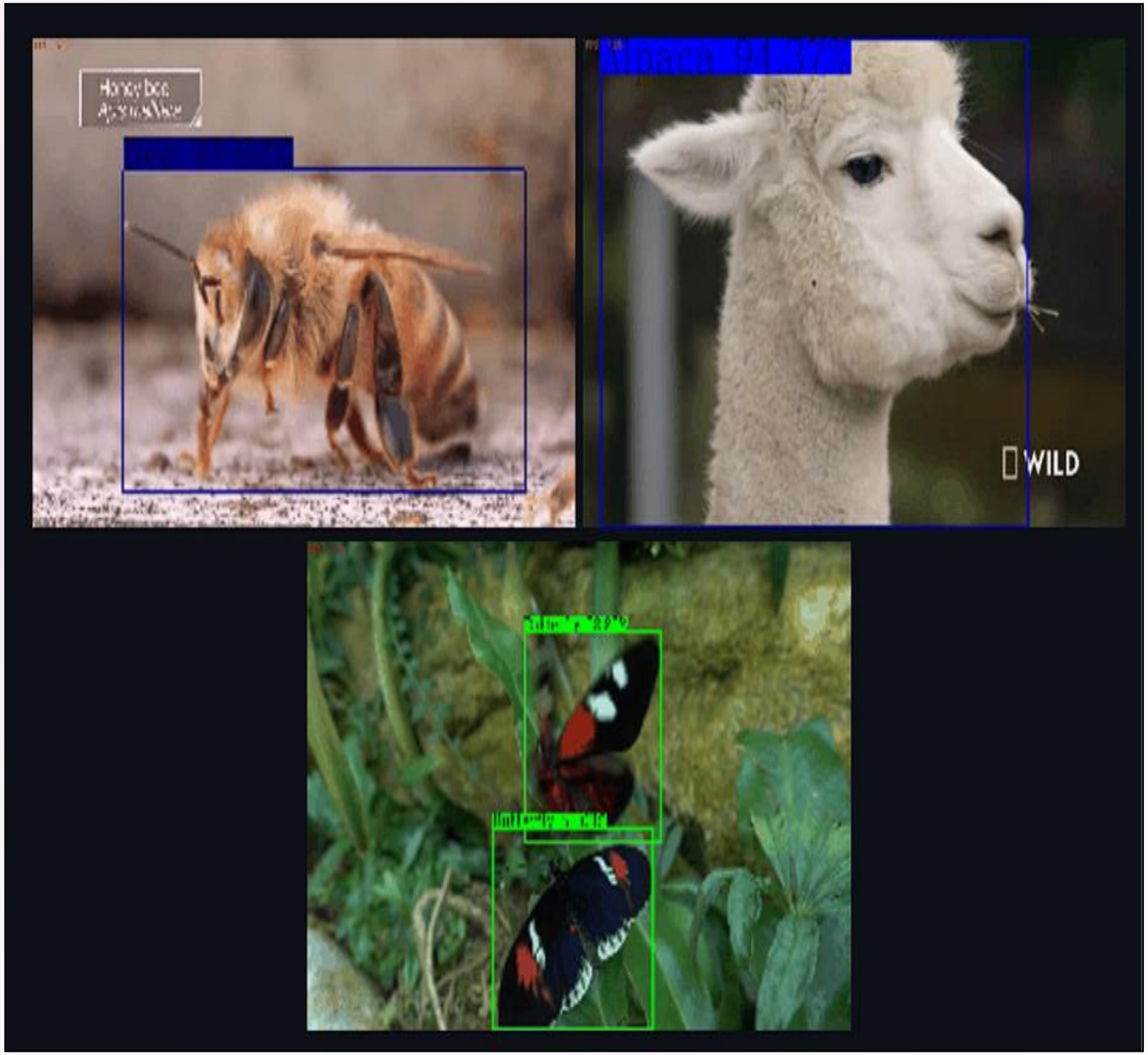
1 . YOLO V8

YOLOv8 advances object detection technology by delivering faster, more accurate, and resource-efficient performance across a wide range of applications.



YOLOv8: Object Detection

- Fast and accurate (Ultralytics).
- Enhanced architecture for small/large objects.
- Real-time use: surveillance, driving, medical.
- Supports detection, segmentation, classification.
- Optimized for edge devices.



2. COCO dataset:

The COCO (Common Objects in Context) dataset is a large-scale, high-quality dataset widely used for object detection, segmentation, and image captioning tasks. Developed by Microsoft, it contains over 330,000 images with more than 1.5 million object instances spanning 80 object categories like people, animals, and everyday objects.

The below image represents a complete list of 80 classes that COCO has to offer.

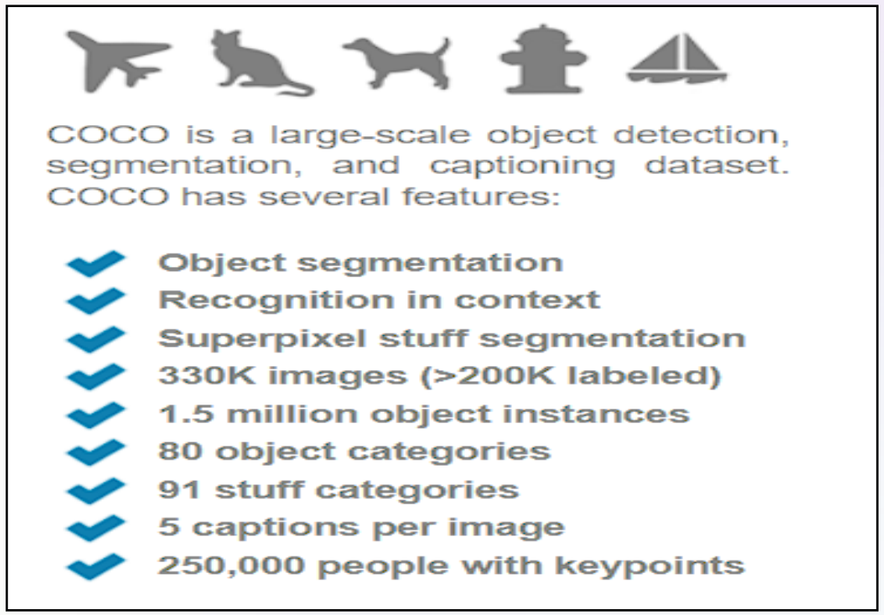
person	fire hydrant	elephant	skis	wine glass	broccoli	dining table	toaster
bicycle	stop sign	bear	snowboard	cup	carrot	toilet	sink
car	parking meter	zebra	sports ball	fork	hot dog	tv	refrigerator
motorcycle	bench	giraffe	kite	knife	pizza	laptop	book
airplane	bird	backpack	baseball bat	spoon	donut	mouse	clock
bus	cat	umbrella	baseball glove	bowl	cake	remote	vase
train	dog	handbag	skateboard	banana	chair	keyboard	scissors
truck	horse	tie	surfboard	apple	couch	cell phone	teddy bear
boat	sheep	suitcase	tennis racket	sandwich	potted plant	microwave	hair drier
traffic light	cow	frisbee	bottle	orange	bed	oven	toothbrush

COCO Dataset

- Complex, real-world environments with varied object scales and occlusions.
- Annotations: bounding boxes, segmentation masks, keypoints for tasks like pose estimation.
- Contextual labeling adds realism (e.g., person on bike).

YOLOv8 Performance on COCO

- mAP (IoU = 0.5): 70-80% accuracy.

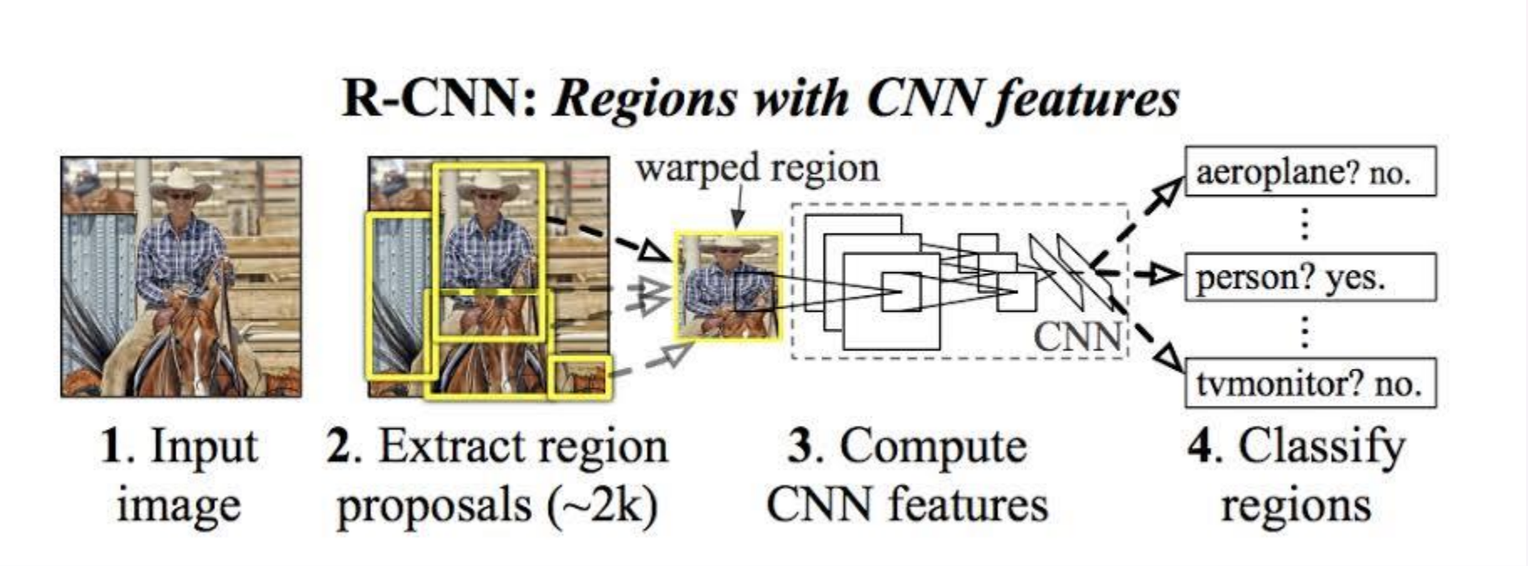


3. CNN (Convolutional Neural Networks)

- Deep learning model for image recognition and classification.
- Key feature: convolutional layers for automatic feature extraction.
- Handles spatial hierarchies in images.
- Widely used in image recognition, object detection, and facial recognition.

Performance

- High accuracy in image classification tasks (e.g., MNIST, ImageNet datasets).



4. GTTS (Google Text-to-Speech)

- Converts text into natural-sounding speech using Google’s TTS engine.
- Supports multiple languages and accents.
- Simple API for integration into various applications.
- Commonly used in voice assistants, audiobooks, and accessibility tools.

Performance

- High-quality, real-time speech synthesis with support for multiple voices.

Conclusion:

- Tailored Learning: Identifies objects in educational materials.
- Text-to-Speech: Adaptive algorithms for auditory learners.
- Customizable Paths: Educators shape learning using visuals and text.
- Interactive Modules: Engages various learning styles.
- Mobile Learning: Anytime, anywhere.

References

[1] Yang R and Yu Y (2021) Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis. Front. Oncol. 11:638182. doi: 10.3389/fonc.2021.638182

[2] Sharma, M.; Lim, J.; Lee, H. The Amalgamation of the Object Detection and Semantic Segmentation for Steel Surface Defect Detection. Appl. Sci. 2022, 12, 6004. https://doi.org/10.3390/app12126004

[3] Chang, EC., Yap, C., & Yen, TJ. (2021). Real Time visualization of large images over a thinwire. In IEEE Visualization '97 (Late Breaking Hot Topics) Also: CD proceedings of conference

[4] Yang, Owen & Cuccia, David & Choi, Bernard. (2021). Real-time blood flow visualization using the graphics processing unit. Journal of biomedical optics. 16. 016009. 10.1117/1.3528610.

[5] arXiv:2402.09378v1 [eess.AS]

[6] Kumar, Y., Koul, A. & Singh, C. A deep learning approaches in text-to-speech system: a systematic review and recent research perspective. Multimed Tools Appl 82, 15171–15197 (2023). https://doi.org/10.1007/s11042-022-13943-4

References

[7] Sharma M, Lim J, Lee H. The Amalgamation of the Object Detection and Semantic Segmentation for Steel Surface Defect Detection. Applied Sciences. 2022; 12(12):6004. https://doi.org/10.3390/app12126004

[8] Stephen, Okeke & Mishra, Deepanjali & Sain, Mangal. (2020). Real Time object detection and multilingual speech synthesis. 1-3. 10.1109/ICCCNT45670.2019.8944591.

[9] Owen Yang, David J. Cuccia, Bernard Choi, "Real-time blood flow visualization using the graphics processing unit," J. Biomed. Opt. 16(1) 016009 (2021) https://doi.org/10.1117/1.3528610

[10] K. Manikanta Vamsi et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1042 012019