Simple Action Model: Enabling LLM to Sequential Function Calling Tool Chain

Rajat Sandeep Sen
Department of AI&DS
SJCET Palai
Kottayam, Kerala
rajatsandeepsen2025@ai.sjcetpalai.ac.in

Neena Joseph
Department of AI&DS
SJCET Palai
Kottayam, Kerala
neenajoseph@sjcetpalai.ac.in

***Abstract*—Today LLMs are everywhere, it is making human internet life a lot easier than ever. Everyday new sophisticated models are releasing. But these models are not good enough to become the personal assistant like the Jarvis from Sci-fi movie IronMan. This paper proposes a way to enable any LLM to execute complex requirements in real world applications. By leveraging state-of-the-art Large Language models, we can create a simple action model that can understand environment around them. Eventually these models can help or assist humans in real time applications. The Sequential Function Calling Tool Chain System aims to bridge the gap between human language understandings and computer programming.**

***Index Terms*—Large language model, Action model, OpenAPI format and Function Calling Tools**

I. Introduction

Now on a world without AI is impossible to imagine. New models will comes everyday, and new architecture will improve its performance. Multi Modality or Mixture of Models are not enough to become the personal assistant. Graph based models takes up lot of computation in the case of complex function calling. Simple Question, ”what if the one good model with one request is enough to execute multiple functions sequentially?”.

The integration of openAPI schema with LLM can make a decent action model that can understand the API specs of that specific applications. The large language model will take prompts and decides what to do. From programming side, we'll parse the information and execute. Simple action model is done.

By implementing a custom JSON parser alongside with an OpenAPI schema type system it is possible to utilize one model as main decision making inside an application. This technology works faster compared to other solutions, but requires much more intelligent language model.

This helps to provide stable and intelligent AI assistant in any applications uphold the integrity and standards API formats. This research paper also leads to the exploration of Multiple and Sequential Function calling tool chain that utilizes State-of-the-Art Language model to enhance the security, time efficiency,flexibility for building AI applications.

Purpose of this paper is to explore and improve existing solutions of Function calling tools available in open source software community. This paper also aims to provide a new approach to prompting the LLMs without fine tuning.

II. Literature Survey

These literature reviews provides knowledge about the existing researches done by various scholars and open source developers.

III. Methodology

The approach of Sequential Function calling tool is using the popular and newest OpenAPI schema 3.0. By fetching the schema json from the backend url is converted to type parser like Zod (typescript) at build time. The prompt captured from the user is fed to LLM with proper type annotation of OpenAPI schema. Any popular open source model can be used to understand and extract object/json data from users prompts. The extracted data are then loaded into a custom JSON parser that supports extra keywords from standard JSON format. The parser moves the data to OpenAPI clients libraries like OpenAPI-Fetch (typescript) which fetch result from the backend. The result is then passed to the next function in the chain. The process is repeated until the end of the chain. But LLM is used once just to write ”what to do in this environment” according to users prompts and type definition.

A. Schema Preparation

The preparation of the dataset for an offline exam proctoring system that utilizes artificial intelligence is a crucial step in ensuring the accuracy and efficiency of the system. This process involves collecting relevant data, cleaning and preprocessing, and then structuring it in a way that is conducive to training machine learning model. The recorded footage of exams were obtained and a YOLOv8 model trained by Ultralytics on COCO dataset was used to detect the students on the frame, the detected images of students were isolated from the input frames and saved into a folder. Later these obtained images were resized to 224*224px and labeled using the roboflow platform into two classes, ”cheating ” and ”non-cheating”.The dataset contained a total of 2813 images which were splitted into training, validation, and testing at a ratio 70:20:10

graphicx

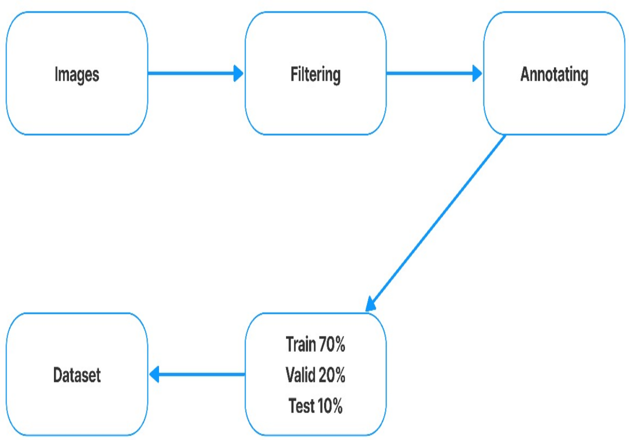

Fig. 1. Block Diagram Dataset Creation.

B. Model Architecture

The AI-based offline exam proctoring system utilizes a specific model architecture.At the forefront of the YOLO series, YOLOv8 is distinguished by its real-time accuracy and intricate design, which blends spatial attention modules with convolutional neural networks (CNNs) for improved feature identification [13]. The 53 convolutional layers that make up the CSPDarknet53 backbone are key to its design; they effectively maximize feature extraction, which is important for identifying the students in the examination hall . YOLOv8 is suited to different processing requirements and comes in nano to extra- large configurations. The medium to extra-large variants are beneficial for detecting the students more accurately. graphicx

For the classification of data, the VGG-16 Classifier has been employed. This classifier is known for its effectiveness in accurately categorizing various types of data. By utilizing this classifier, the system can efficiently determine whether a given frame falls under the ”cheating” or ”not cheating” category.

To ensure consistency and compatibility, the input frames are standardized to a size of 224 pixels by 224 pixels. This standardization allows for seamless processing and analysis of the frames, enabling the system to effectively detect any instances of cheating during the exam.

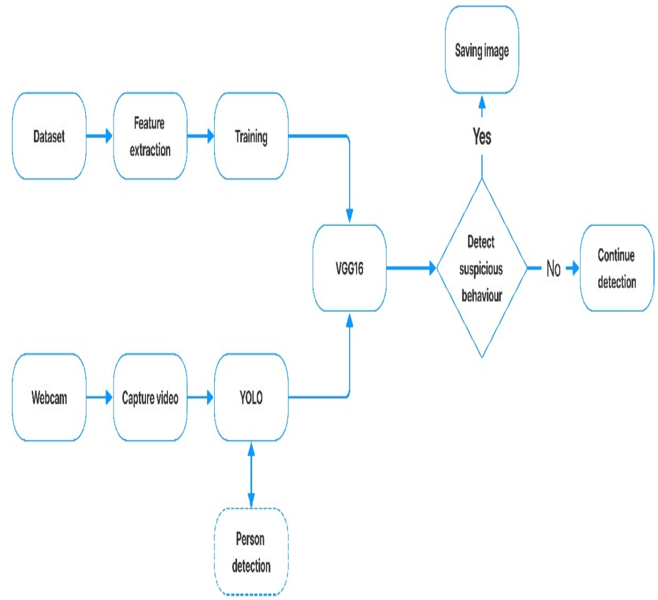

Fig. 2. Block Diagram of Proposed System.

C. Model Training

The design of the offline exam proctoring system utilizing artificial intelligence has been structured to ensure efficient monitoring and supervision during examinations. This model architecture incorporates advanced AI technology to detect and prevent any form of cheating or misconduct during offline exams.

YOLOv8 pre-trained on COCO dataset, enabling it to learn and generalize patterns by exposing it to real-time environment. The Convolutional Neural Network (CNN) backbone serves as a feature extractor,learning essential patterns from input images.

A VGG16 Image classifier trained on a custom dataset to identify the ”cheating” and ”non-cheating” behaviour. The transfer learning technique is used in the classifiers training to reduce the time and effort to train the classifier.

TABLE I
VGG16 CLASSIFIER

VGG16 parameters	**parameter values**
Batch size	32
Epochs	50
Optimizer	adam
Class mode	Categorical
loss	categorical_crossentropy

D. Object Detection

The video feed of the exam environment is segmented into a grid of cells, and employing a YOLOv8-like methodology, the object detection model detects students within each cell.

This grid-based methodology enables the system to perform simultaneous monitoring of the entire exam setting, facilitating comprehensive detection coverage throughout the entirety of the video feed.

Fig. 3. Predicted Output From the Model

E. Classification

The classification of offline exam proctoring systems has undergone a significant transformation with the integration of artificial intelligence (AI). By leveraging a dataset that has been meticulously annotated in roboflow, the system can effectively discern between the various behaviors displayed by students during exams using VGG16 classifier, specifically focusing on distinguishing between cheating and non-cheating actions.

F. Output and Utilization

The utilization of offline exam proctoring system powered by AI not only promotes academic integrity but also improves the overall efficiency of the examination process. With AI technology, institutions can automate the monitoring and evaluation of exams, saving time for both educators and students. This streamlined approach enhances the productivity of exam administration while maintaining the quality and accuracy of assessments.

Along with real-time detection, images of student who attempted to commit malpractices during the exam are also saved to a folder for later cross references with their confidence score.

IV. Implementation and Result

A. Implementation

The study used an advanced object identification model called YOLOv8 (You Only Look Once version 8) to achieve real- time and accurate detection of students in the class. YOLOv8 is widely recognized for its remarkable precision and effectiveness in object recognition, rendering it a perfect match for the specific object detection task. The VGG16 classifies the images into required classes.

A centralized system with a good graphical processor act as the central hub of the entire system which houses an object detection model(YOLOv8)and an image classification model(VGG16)

B. Performance Evaluation

The result in Table II demonstrates that in the study of YOLOv8 for the detecting the students in exam hall, the model showed outstanding performance across key metrics. YOLOv8 continuously maintained high precision, as seen by its high Mean Average Precision (mAP) , guaranteeing the accuracy of 95% in detecting students. It minimized false positives with an exceptional precision score of 0.94. With closely no misses, the model's recall of 0.96 demonstrated its ability to accurately identify all the students . All of these findings demonstrate how reliable and accurate YOLOv8 is an appropriate model to detect students in the classroom, which makes it an important tool for offline exam proctoring.YOLOv8 is the most trending and accurate model which has accuracy close to perfect predictions. combining the state-of-the-art YOLOv8 and VGG16 provided the project with a level of perfection that is unmatchable with humans.Humans has blind spots and other limitations but the EX-GUARD does not have such limitations.

The VGG16 classifier which was trained in house using our custom data set showed some of the best result.the trained model could adapt to the varying demands and situations of the examination hall.The VGG16 Provided an accuracy of 92% in its evaluation phase.

While comparing with other models like AlexNet, the VGG16 provided better output regarding the ability to classify the student's behaviors. Combining the YOLOv8 and VGG16 the system performed better than the humans.

TABLE II
Performance Evaluation:YOLOv8

Metric	YOLOv8
Accuracy	95%
Precision	0.94
Recall	0.96
F1-score	0.95

C. Findings

From Fig. 6, which illustrates the usefulness of the YOLOv8 model for detecting students and VGG16 for classification, the study derives a number of noteworthy findings. The following are the main findings:

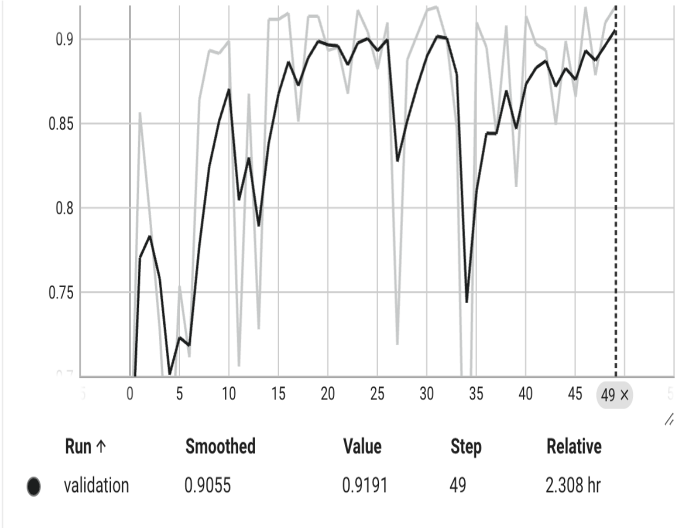

Fig. 4. Accuracy of Vgg16

1) High detection Accuracy: The high Mean Average Precision (mAP) score indicates that the YOLOv8 model typically attained a high level of accuracy. This high accuracy ensures correct detection of exam takers.
2) High classification accuracy: The high accuracy of VGG16 classifier(92%) plays a major role in correctly classifying between cheating and non cheating.
3) Precision and Recall: The model's 0.94 precision and 0.96 recall scores highlights its ability to reduce false positives, a crucial aspect of detecting only students from each frame.
4) Balanced F1-Score: At 0.95, the F1-Score, a measure of accuracy that is balanced, was observed. This measure shows that the YOLOv8 model successfully reduces false positives and false negatives while maintaining good detection accuracy.
5) Practical Deployment: We have developed a practical and easily understandable automated proctoring system by integrating pre-trained YOLOv8 model with our custom-trained VGG16 classifier.

By providing a user-friendly interface for monitoring road conditions, this system makes real-time analysis and reporting easier.

These indicate the resilience and effectiveness of the YOLOv8 model in detecting students and VGG16 model in classifying the students into ”cheating” and ”non cheating”. The foundation for improved road maintenance and safety is laid by this research, which may find use in damage prevention and real-time monitoring.

D. Comparison with State-of-the-Art methods

In the realm of offline exam proctoring systems, leveraging deep learning models for real-time detection and analysis has become increasingly prevalent. This script integrates several state-of-the-art components, notably utilizing YOLO (You Only Look Once) for object detection, particularly focusing on identifying individuals, typically students, within an exam setting. YOLO offers real-time object detection with impressive accuracy, efficiently bounding boxes around detected objects. YOLOv8 is the best-performing object detection algorithm[17] in the field with the highest accuracy and performance measures.

the system also uses VGG16 classifiers for classifying images, we trained the AlexNet with the same dataset as that of the VGG16. Referring to Table II, the following outcomes of the comparative study between VGG16 (the proposed model) and AlexNet on the same dataset are shown:

TABLE III
COMPARISON OF THE PERFORMANCE METRICS FOR VGG16 AND ALEXNET

Metric	VGG16	AlexNet
Accuracy	92%	64%
Loss	68	76

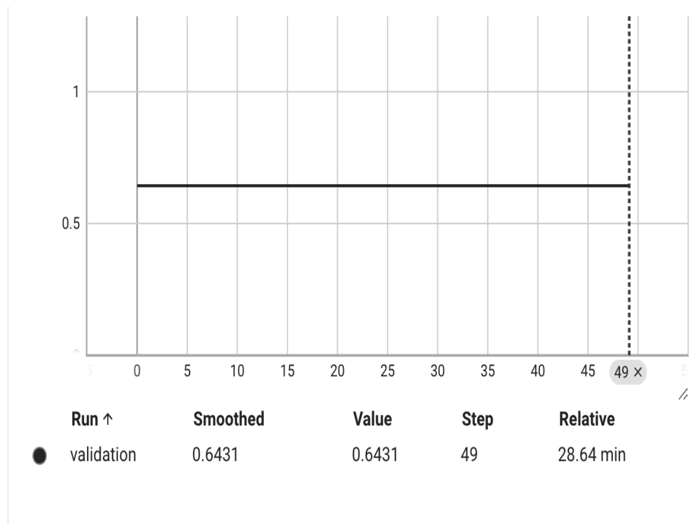

Fig. 5. Accuracy of AlexNet

V. CONCLUSION

The offline exam proctoring system using AI is an utilization of a combination of YOLO object detection for detecting people in a video stream and a pre-trained VGG16 model for determining whether detected individuals are engaged in cheating behavior during an exam.

In essence, the system processes each frame of the video stream, identifying people using YOLO, and then analyzing each detected person's behavior using the VGG16 model. If the VGG16 model predicts that the person is potentially cheating based on certain features extracted from their behavior or surroundings, such as looking at unauthorized materials, it saves an image of the person for further review. Conversely, if the model determines that the person is not engaged in cheating behavior, it continues processing the video stream.

This process is repeated for each frame of the video, allowing the system to continuously monitor and flag potential instances of cheating during an exam.

.

REFERENCES

[1] Tong Liu ,AI proctoring for offline examinations with 2-Longitudinal-Stream Convolutional Neural Networks 10 December 2022

[2] Musa Dima Genemo ,Suspicious activity recognition for monitoring cheating in exams 24 February 2022

[3] Fairouz Hussein , Ayat Al-Ahmad , Subhieh El-Salhi , Esra'a Alshdaifat , Mo'taz Al-Hami ,Advances in Contextual Action Recognition: Automatic Cheating Detection Using Machine Learning Techniques ,31 August 2022

[4] Rhitvik Pasricha ,Prathamesh Churi ,A Systematic Review on AI-based Proctoring Systems: Past, Present and Future September 2021

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi You Only Look Once: Unified, Real-Time Object Detection ,9 May 2016

[6] Tanzila Saba, Amjad Rehman, Nor Shahida Jamail1, SouadLarabi Marie-Sainte, Mudassar Raza, and Muhammad Sharif Categorizing the Students' Activities for Automated Exam Proctoring using Proposed DeepL2-GraftNet CNN Network and ASO Based Feature Selection Approach, March 2021

[7] M. Ghizlane, B. Hicham, and F. H. Reda, "A New Modelof Automatic and Continuous Online Exam Monitoring,"in 2019 International Conference on Systems of Collaboration Big Data, Internet of Things Security (SysCoBIoTS), 2019, pp. 1-5.

[8] Mabrouk, A.B., Zagrouba, E.: Abnormal behavior recognition for intel-ligent video surveillance systems: a review. Expert Syst. Appl. 91,480–491 (2018)

[9] He, H., Zheng, Q., Li, R., Dong, B.: Using face recognition to detect "Ghost Writer" cheating in examination. In: International Confer-ence on E-Learning and Games, pp. 389–397 (2018)

[10] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable object detection using deep neural networks. In ComputerVision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 2155–2162. IEEE, 2014.

[11] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik. Simultaneous detection and segmentation. In Computer Vision–ECCV 2014, pages 297–312. Springer, 2014.

[12] Saba, T., Rehman, A., Jamail, N. S. M., Marie-Sainte, S. L., Raza, M., Sharif, M. (2021).Categorizing the students' activities for automated exam proctoring using proposed deep L2-GraftNet CNN network and ASO based feature selection approach. IEEE Access, 9, 47639–47656. https://doi.org/10.1109/ACCESS.2021.3068223 [13]Ultralytics Accessed: 04.08.2023 [Online] Available: https://www.ultralytics.com/