The application of 3D spatial object and gesture detection in children's education

[1,*] *Chuan-Shiuan Liang* (梁傳萱) , [1]*Zhi-Chi Wang* (王子綺) , [1]*Firdaus Golam* (劉冠標) ,
[1] *Chieh-Ming Yang* (楊哲旻) , and [1] *Jen-Yeu Chen* (陳震宇)

[1] Department of Electrical Engineering,
National Dong-Hwa University, Hualien, Taiwan
[*]E-mail: 410923037@gms.ndhu.edu.tw

ABSTRACT

Recently, there has been significant development in object detection across various fields. Building upon these developments, this thesis aims to explore the application of object detection technology in children's education. In this thesis, YOLOv7 and MediaPipe were chosen as the object detection models, along with hand landmarks detection, to develop a practical example—a game.

By integrating YOLOv7 and MediaPipe, we have created a game that employs real-time object detection through a camera, generates questions based on the detected objects, and requires players to use their fingers to point at the objects that match the given requirements. It offers both a leisure mode and a timer mode, each presenting different challenges and rewards.

This thesis showcases the potential application of object detection technology in game development. For further development, it can be applied to mobile applications or devices such as smart glasses, offering more exceptional gaming experiences and user interactions.

Keywords: *Children's education, MediaPipe, Object Detection, YOLOv7*

1. INTRODUCTION

With the rapid development of technology, the widespread application of object detection technology in various fields is becoming increasingly significant. Among them, applying object detection to children's education holds important motivation and potential value.

First of all, object detection technology can provide children with an interactive and visual learning environment. Traditional children's education relies mainly on textbooks and teachers' explanations, lacking practical experiences. However, through object detection technology, children can directly interact with objects, gaining a more immersive learning experience.

Secondly, object detection technology can offer personalized and autonomous learning experiences. The application of this technology contributes to promoting children's cognitive development, helping them construct their understanding of the world and fostering their learning abilities.

In summary, applying object detection technology to children's education holds important motivation. It can provide diversified learning experiences while stimulating children's learning abilities. Therefore, in this experiment, we chose to develop a game based on YOLOv7[1], with the focus on object recognition, providing children with a new learning option. The combination of technology and education helps break through the limitations of traditional teaching methods, offering children a more enriched, engaging, and effective learning experience. Looking ahead, we can expect the widespread application of object detection in children's education, bringing greater benefits to the learning and growth of the next generation.

2. METHODS

The methods of this experiment are as follows: we choose YOLOv7 to perform object detection in section 2.1 and MediaPipe[2] to perform gesture detection in section 2.2. In section 2.3, we create the Tkinter interface before opening the camera and the menu that can be called by gestures after opening the camera. And in section 2.4, we elaborate game modes and scoring, by creating two game modes: Leisure Mode and Timer Mode. In leisure mode, there is no time limit, and only the score accumulates. In timer mode, you can choose a time limit. In both modes, points are scored by clicking on objects in the environment with fingers after opening the camera.

2.1. YOLOv7

In this project, YOLOv7 was chosen as the object detection model. YOLOv7 is one of the best-performing object detection models currently available in the market. By using the pre-trained weights provided by YOLOv7, we are able to detect 80 common objects from the COCO dataset.

YOLOv7 is acknowledged for its speed and efficiency, allowing real-time object detection while maintaining accuracy. It effectively reduces the parameter count by 40% and the computational cost by 50%, through optimizations in model architecture and training process. The model architecture is optimized to achieve faster inference speed and higher detection accuracy and makes YOLOv7 an ideal choice for applications that require real-time detection.

2.2. MediaPipe

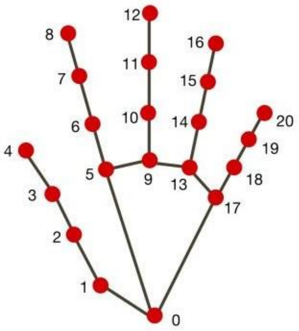


Fig. 1. Coordinate points of hand.

In this thesis, we used MediaPipe to implement hand tracking and gesture detection. MediaPipe is a multimedia machine learning framework developed by Google Research. By utilizing MediaPipe, we can easily implement functions such as hand tracking, face detection, and object detection. Considering the specific objective of detecting fingers pointing at objects, we focused on utilizing the hand landmarks detection feature of MediaPipe.

Once detecting a hand through the camera, the hand landmarks detection is immediately executed, and a loop is initiated to continuously retrieve the values of the 5 fingers and the 21 coordinate points on the palm, as depicted in Figure 1. In the game, the position of the fingertip of forefinger (the 8th coordinate point) serves as the reference for determining whether it falls within the four coordinates of the detected object. This forms the basis for making judgments in the game.

2.3. User Interface

In this experiment, the Tkinter library was utilized for the development of the user interface, as illustrated in Figure 2. This interface grants users the freedom to select the game mode and the objects to be recognized, empowering them to explore, choose items that resonate with their personal interests and prevent detection errors that could result in questions about non-existent objects in the user's environment, as illustrated in Figure 3. The game offers two distinct modes: Leisure Mode and Timer Mode. In Timer Mode, users have the option to set a time limit of 1 minute, 3 minutes, or 5 minutes, providing an exhilarating experience as they strive to achieve high scores and surpass records within the confines of the designated time frame. Conversely, Leisure Mode does not involve time constraints and focuses solely on score accumulation, allowing players to accumulate points at their own pace, fostering a relaxed and unhurried game experience.

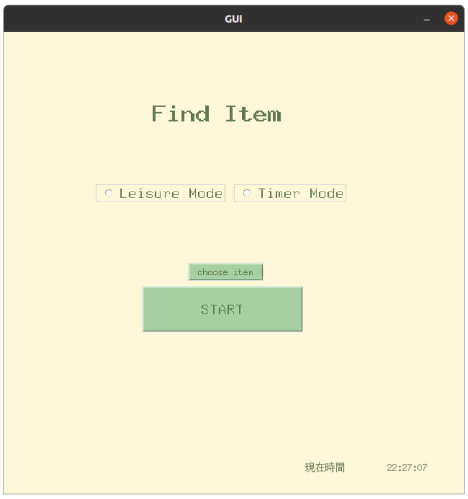
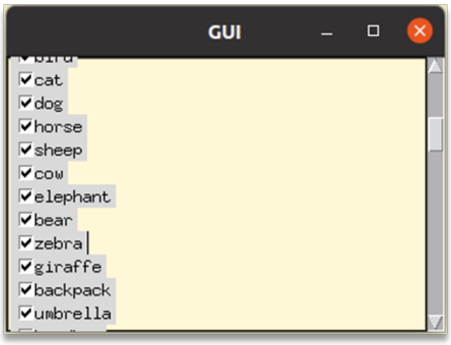


Fig. 2. Mode selection interface



☑ cat
☑ dog
☑ horse
☑ sheep
☑ cow
☑ elephant
☑ bear
☑ zebra
☑ giraffe
☑ backpack
☑ umbrella

Fig. 3. Item selection interface

Upon opening the camera, we have implemented a gesture-based method using finger angle calculations to access the menu by performing a specific hand gesture, allowing the player to choose a new game mode or adjust the game time, as illustrated in Figure 4. Once the menu is accessed, blocks representing the Leisure Mode and Timer Mode will be displayed on the screen. When the player moves their finger onto one of these blocks, a button labeled "Confirm" will appear. By keeping the finger on the Confirm button, the player can proceed to the corresponding page. If the Timer Mode is selected, the player will first enter a page to select the time limit. Once the time limit in the Timer Mode expires, blocks will appear on the screen, giving the player the option to restart the game or return to the mode selection page.

This design provides usability and user experience, making mode switching and restarting the game more convenient and intuitive, as illustrated in the flowchart depicted in Figure 5.

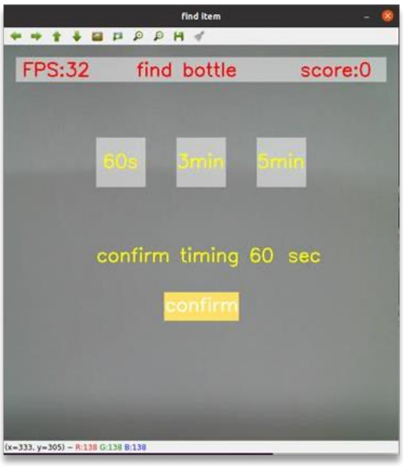


Fig. 4. Mode selection interface after opened the camera.

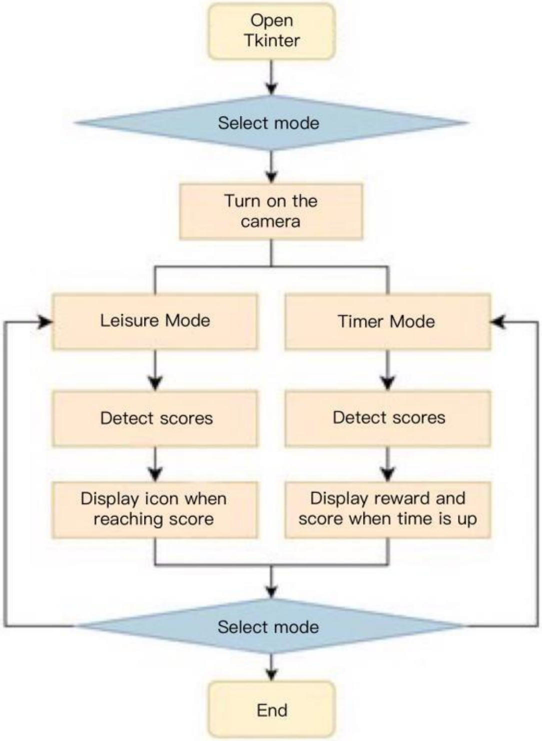


Fig. 5. Flowchart of game modes

2.4. Game modes and scoring

Once the game starts and the camera is activated, object detection begins. The detected objects are stored in an array, from which the questions are generated to ensure that the requirements of the questions align with the actual situation. This prevents situations where the question's requirements are inconsistent with the user's surrounding environment, thus ensuring that the questions can be completed. Additionally, the coordinates of the detected objects, including the top right, bottom right, top left, and bottom left coordinates, are extracted and passed into the function. Furthermore, hand landmark detection using MediaPipe is performed to facilitate the evaluation process.

In the game, players are required to find the objects that meet the requirements of the questions and point at them using their index finger to earn five points. When the system determines that the requirement is met, a success message will be displayed, and the name of the object will be spoken out loud, as flowchart shown in Figure 6. In Leisure Mode, as the player's score increases, different rewarding icon will appear on the screen, adding fun and challenges to the game. In Timer Mode, after the time limit expires, icons representing different scores will be displayed, as shown in Figure 7.

This game design offers an engaging way to perform object detection and challenge players with various questions, while motivating them to continuously improve their scores. Players will strive to achieve the game objective by searching for and identifying the correct objects, which helps enhance their observational skills and responsiveness.

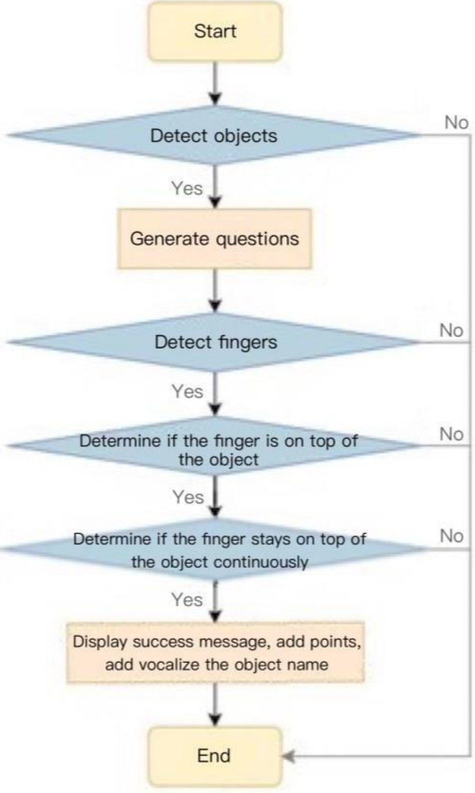
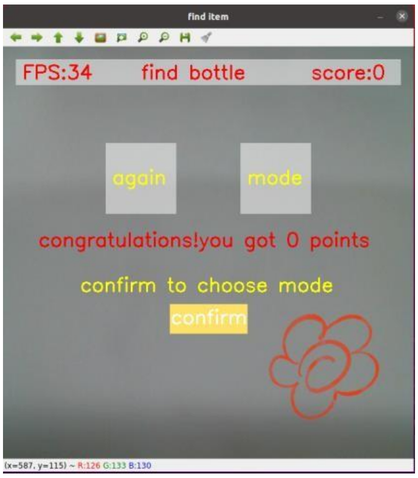

Start
Detect objects
No
Yes
Generate questions
Detect fingers
No
Yes
Determine if the finger is on top of the object
No
Yes
Determine if the finger stays on top of the object continuously
No
Yes
Display success message, add points, add vocalize the object name
End

Fig. 6. Flowchart of detection



FPS:34 find bottle score:0

congratulations!you got 0 points

confirm to choose mode

(x=587, y=115) ~ R:126 G:133 B:130

Fig. 7. Example of time-up.

3. CONCLUSION

In this experiment, we used Python programming language to write the entire program to explore the concepts of object detection and game challenges. However, Python, despite its convenience, is not an efficient programming language due to its interpreted nature. Initially, during the experiment, the program had a very low frame rate, and the screen experienced lagging issues. To improve the screen speed, we decided to use GPU for acceleration. However, MediaPipe cannot run on the GPU in the Windows system, so we had to run Python on the Ubuntu system. However, NVIDIA's support for the Linux system is not very friendly, and we encountered many version incompatibility issues during the installation process, requiring us to recompile and find other solutions. However, in the end, we succeeded in increasing the frame rate from the original 21-23 frames to about 30 frames.

In the future, we hope to develop a mobile application or integrate the game into smart glasses to provide a better gaming experience. However, given our current time constraints, we have not had the opportunity to refactor the program into a mobile application or adapt it for smart glasses.

These developments necessitate additional research and development, with a focus on platform compatibility, interface design, and user experience. If an opportunity arises in the future to migrate the game to mobile devices or smart glasses, it will offer a wider range of game options and a more convenient user experience. We remain vigilant in tracking advancements in this domain and eagerly anticipate the chance to materialize this concept. Doing so would enable us to bring the game to a broader user base, delivering increased value and enjoyment to society.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 111-2622-8-259-001-TE2.

REFERENCES

[1] C.Y. Wang, A. Bochkovskiy, and H.Y.M. Liao, “YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 2023.

[2] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.L. Chang, and M. Grundmann, “MediaPipe Hands: On-device Real-time Hand Tracking,” CVPR Workshop on Computer Vision for Augmented and Virtual Reality, 2020.