Using computer vision and machine learning with a view to building children's vocabulary

Kenneth Y T Lim[a], Kaushal Rajesh[b] and Alastair Pek[b]

[a]National Institute of Education, Singapore; [b]Independent authors

Abstract

This paper describes an independent research project carried out by a pair of junior college students under the mentorship of a Research Scientist at the National Institute of Education. Our project aimed to design a working device to aid kindergarten children in learning about common household equipment such as tables, chairs, and so on. We hoped that this device would help to solve the problem of kindergarten children not being able to learn during the Covid-19 pandemic due to the closure of kindergartens. We trained an Object Detection model and ran it on a Raspberry Pi with screen, speakers and camera connected to it. We took photos of different household equipment and labelled them using software in order to train these images for our Object Detection model. We made use of the Google Open Images and COCO image datasets to sift out the related images of the objects we wanted to train in the model. Overall, the use of the software such as TensorFlow helped us with the training of the different types of objects for the model and the use of Text-to-Speech software allowed us to incorporate the use of sound to project the pronunciation of the objects. Eventually, there are many more aspects we could explore with our prototype. We could make it so that it can say out the object names in different languages to help teach Mother Tongue to kindergarten children.

Introduction

With the recent ongoing Covid-19 pandemic, there are several restrictions and regulations imposed in Singapore and other countries. Due to the surge in the number of cases, the government had decided to close down Kindergartens, Primary, Secondary and Tertiary Institutions. Primary, Secondary and Tertiary Institution students are old enough to know how to use mobile phones or computers to attend online lessons for Home Based Learning; this leaves kindergarteners as one of the only groups who cannot adapt to the

CONTACT Kenneth Y T Lim, kenneth.lim@nie.edu.sg

new learning methods by themselves. This makes them more prone to learning losses, which would reduce their future productivity and lifetime earnings (Asian Development Bank, 2021). Also, it has to be acknowledged that kindergarteners are not really built for online learning. They are at the exploration stage of their lives, and hence it is important that they are provided with the means to explore; rather than sitting and staring at the screen (edutopia.org, 2020).

Hence, a user-friendly device was conceptualised which catered to kindergarteners and allows them to learn the names and pronunciations of different household items even while at home. This provides the kindergarteners an opportunity to learn object names and pronunciations.

Most importantly, it caters to their learning style which mainly consists of exploration. This is due to the fact that the device provides a hands-on form of learning for the kindergarteners, which will certainly interest them and provide enhanced learning (UNICEF, 2021).

Additionally, since this device is a form of technology, it helps these kindergarteners to understand more and become more familiar with it. This is useful as technology is pivotal in our society and in the future as well (Forbes, 2020).

This project was carried out in two separate segments - Object Detection and Raspberry Pi. The object detection model was first developed and trained. To do so, some pictures of household objects such as phones, scissors, pens and more were taken and manually labelled.

However, the vast majority of images were obtained from online databases such as the Open Images and COCO (Common Objects in Context) Image Databases. Thereafter, the images were labelled and used in training the Object Detection Model.

Subsequently, the code for Text-to-speech (TTS) was incorporated into the Object Detection Model. It was then transferred to the Raspberry Pi. After optimising the Raspberry Pi for the model, a screen, speakers and camera were attached to the Raspberry Pi module. The final device consists of the Raspberry Pi module and its attached components.

One of the main limitations of this project is the accuracy of the device. It is not foolproof and misidentifies objects from time to time. Methods to improve its accuracy such as increasing the size of our image dataset and training for more steps were tried out.

A few assumptions made about this project are that the device will be able to detect and pronounce the different objects accurately with no errors and that the kindergarteners are willing to accept this device as their learning material and make good use of it for their learning purposes.

Regarding the scope of the investigation, it has been narrowed down to 10 distinct objects such as fruits, vegetables and other accessories to ensure an optimum accuracy.

Objects that are very similar in shapes such as pencils and pens are eliminated as they hamper the accuracy of the device. On top of that, objects that are too large and too small - such as fridge and screw respectively are eliminated. The number of objects trained was also limited to ten because the higher the number of objects, the less accurate our current device will be. Last but not least, there is a need to make sure that the objects are in a very distinct shape so that the device can detect them accurately - such as tables, chairs, mobile phones, scissors, etc.

Literature review

Current learning methods

In Singapore, the most conventional way for kindergarteners to learn is through the use of a hands-on learning approach (the Apple Tree kindergarten, 2022). Kindergarteners usually interact with educational toys such as building blocks, toys and educational books such as storybooks. With the use of the device, they will get the chance to learn new objects through means of technology. It is simple and easy for them to use and can even be customised for their age group. It is a good idea to incorporate technology into kindergarteners' learning journey as it can make learning more engaging and fun. The mundane books and toys do not stretch kindergarteners' potential for learning and it is not as effective as the use of technology in learning (Raja and Nagasubramani, 2018). Also, research has shown that many younger children such as pre-schoolers have used technology before even touching books (OECD, 2022). Hence, there is a high chance that they will learn about the different objects faster than conventional methods such as reading books.

Using the device helps to increase the learning opportunities for kindergarteners as it can be used on a daily basis, be it at home or at their kindergartens. It also does not have a very steep learning curve as the kindergarteners just have to learn to point the camera towards an object and make use of the screen to learn the spelling of the word and listen to the pronunciation of the word using an earphone. This model makes learning more efficient and effective since kindergarteners will also be able to learn even when they are not attending kindergarten. They just have to carry this prototype and learn the names of the objects anywhere and everywhere they go, making learning convenient. Whenever kindergarteners are curious about the object's name and pronunciation, they can simply use the Object Detection model to identify the object. This makes learning on the go more interesting since they do not have to always be situated in a classroom where they will make use of toys and books to learn new knowledge.

Applications of object detection

The Object Detection model is not currently used to teach kindergarteners daily objects and they still follow the conventional way of learning. The use of Object Detection is very prevalent in today's technological era. There are examples such as autonomous driving systems in Tesla vehicles where they make use of this Object Detection technology to train the vehicle to recognise objects in the driver's view and prevent any collision when under auto-drive. Additionally, in the field of sports, the use of goal line technology in football also makes use of the Object Detection model to locate the position of the ball and it is able to decide whether the goal is allowed or disallowed (viso.ai, 2021).

With the prevalence of the use of the Object Detection model, it is also a good opportunity to incorporate this model into educational learning. This opportunity is able to stretch the potential of Object Detection in daily life usage. The use of Object Detection can make learning more enticing and meaningful for kindergarteners. It enhances the learning experiences for them so that it will be more memorable and they can better retain what they have previously learned.

Aims and objectives

The main aim of this project is to allow the kindergarteners to continue their learning even though their schooling has been interrupted due to the pandemic. Moreover, it provides a fun and interactive platform for them to play with, which encourages them to use it.
The objectives of this project are aided by questions such as:

1. How will the device seem appealing and attractive to the kindergarteners to ensure that they will use it as much as possible?
2. What if the device detects the objects wrongly and will they implicate the kindergarteners in terms of their learning and their understanding of different objects?
3. Will it be helpful and effective in terms of aiding their learning process? These questions are crucial in coming up with the aim for this project.

Methodology and materials

Methodology

The main programming language used to code the model and run it was Python 3. The tensorflow1&2 APIs were also heavily relied on to train the model.

The coding was done on notepad++. A few main functions that were defined included the Text-To-Speech (TTS) function and the real time prediction function. Another major component in this project was the training of the model.

The model refers to the object detection model, which was trained with sets of images. It helps to detect instances of objects in real-time feed obtained from the camera.

Text-to-speech function

The text-to-speech (TTS) function helps to convert text into audio. For example, when the model detects an object such as a Mobile Phone, the TTS function helps the device to say out 'Mobile Phone' through speakers connected to it. Depending on the object detected from the ten listed (ball, banana, broccoli, carrot, chopsticks, mango, mobile phone, paper towel, strawberry and watch), the TTS function will make the speakers say out the name of the object.

Training of the model

The training of the model is another pivotal part in bringing the device to fruition. Images were first gathered and labelled. Then they were trained to form the object detection model. Initially, software was to label images manually. However, that turned out to be infeasible due to the sheer volume of pictures needed to train the model to reach a higher level of accuracy. Thus, images from COCO dataset and Google Open Images Database were used.

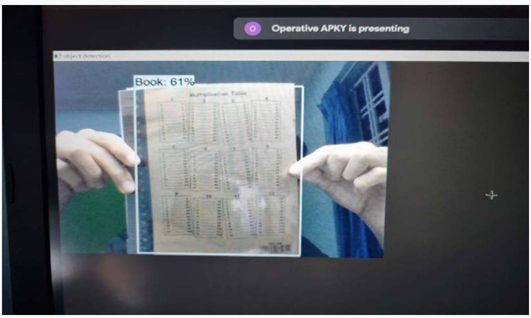
To prepare the image sets for training, 60% of the images and their corresponding labels to a folder were transferred to a folder called train while the rest of the images and their corresponding labels were transferred to a folder called test. By doing so, the model trains itself using the images in the 'train' folder and self-evaluates using the images in the 'test' folder. The training went on for 5000 steps. This means that the model went through 5000 iterations to improve its accuracy.

Running of the model

To run the model, another function called Real Time Prediction was used. This was the main function and included all other required functions to successfully run the prototype such as the TTS function, detect function (where images of the different objects are recognised) and etc. This thus led to the end result, where the laptop webcams were able to detect objects (Fig 3) and subsequently say out the name of the object.

Figure 1

A book being recognised by the trained model running on a laptop webcam.



Since the device was for kindergarteners to use, the code had to be exported to a Raspberry Pi device.

Materials

List of equipment:

- Raspberry Pi
- HDMI Cable
- Speakers
- Screen
- Camera
- SD card and adaptor

Figure 2 shows the screen.

Figure 2

Screen

The function of the screen is to allow kindergarteners to see the image captured by the camera. They will be able to see if the camera is pointed to their desired object so that they can learn the spelling of the object when shown on the mini screen.
Figure 3 shows the camera module.

Figure 3

Camera Module

As seen from Figure 3, the camera module captures the image of the object for the Object Detection model to work. The prototype requires a small and compact camera due to the prototype being relatively small. Hence, this camera module is the most suitable for this prototype.

Figure 4 shows the Raspberry Pi.

Figure 4

Raspberry Pi

With reference to Figure 4, the Raspberry Pi serves as the CPU of a desktop. The Raspberry Pi is able to work as a mini version of a CPU and hence is able to keep the device at a small form factor in order to be user-friendly for kindergarteners.

Figure 5

Entire set-up

Figure 5 shows the overall set-up of our prototype.

Results and discussion

Figures 6 to 9 highlight the testing of the Object Detection model from the laptop, which had not been integrated into the Raspberry Pi yet. The model was trained with ten specific objects because based on the findings and conclusions over the course of the project, those objects with the same size and shape will lead to the inaccuracy of the model. Hence, the selected objects for the Object Detection model are mostly distinct and unique in shape and sizes. Additionally, another point to take note of is the number of objects trained in the model. This is because the larger the number of objects trained, the less consistent and accurate the model will detect the objects correctly. Hence, it is necessary to limit the number of objects so as to maintain the consistency and accuracy of the model. During the testing process of the Object Detection model, it was discovered that it detects any object as a Mobile Phone when the image is not focused. The most accurate and consistent detection is obtained when a mobile phone is shown directly in front of the camera and focused in image as shown in Figure 6.

Figure 6

'Phone' Object Detection

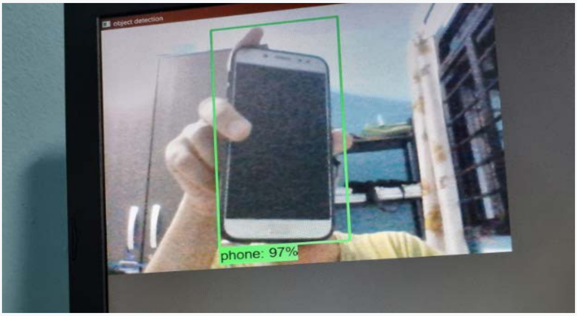
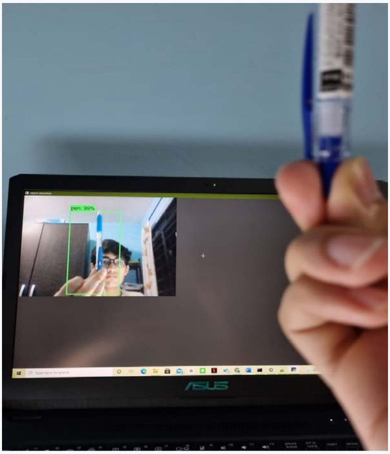


Figure 7

'Pen' Object Detection



As for objects such as the pen (Figure 7), the Object Detection model has its limits because it is not able to identify and distinguish a pen and a pencil clearly even after training sufficient images of these two objects. This is due to the resemblance of the two objects in terms of the size and shape. Hence, it is necessary to avoid training images of these types of objects so as to maintain accuracy and consistency of the model.

Figure 8

'Coin' Object Detection

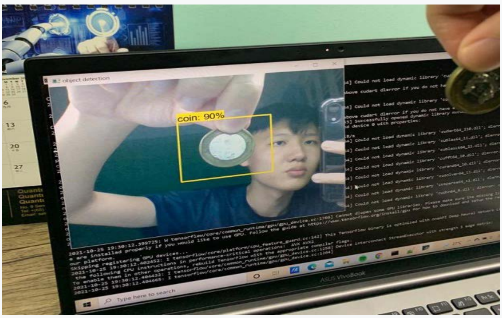
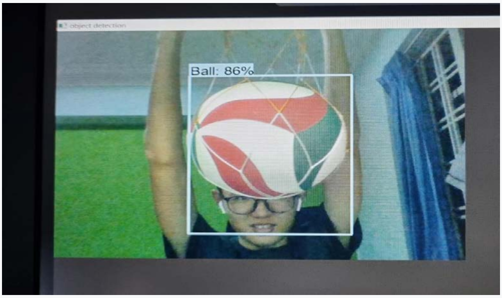


Figure 9

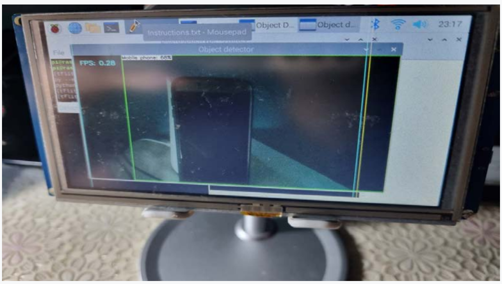
'Ball' Object Detection



As for the coin (Figure 8), the Object Detection model's accuracy and consistency started to decline as the camera focused on objects that are circular or spherical. The model will persist in detecting the coin as other types of round objects such as a ball as seen in Figure 9 and vice versa. It is quite seldom that the model will detect the coin correctly at high accuracy and consistency as shown in Figure 8.

Figure 10

Object Detection on our Prototype



Unlike the previous figures under this segment where the model was running on our laptop webcams, Figure 10 shows the result when the Object Detection model is transferred from the laptop to the Raspberry Pi. The transfer of the model was successful and tested well. However, the model took a longer time to detect an image as compared to running the model on a laptop. The CPU difference resulted in the model to only be able to accommodate for fewer frames per second when compared to running it on a laptop. This could be due to the lower processing power of the Raspberry Pi, making it slower at detecting objects as compared to laptops. Hence, there are many rooms for improvements to patch up the flaws. With a longer duration for the project, training a larger dataset of images or changing our prototype altogether will be possible and doable.

Conclusion

The Raspberry Pi device seeks to help kindergarteners learn the names of various objects through the use of visual and auditory stimuli. Overall, the use of software such as Notepad++ and TensorFlow allowed for the training of the model with different types of objects and the use of Text-to-Speech software allowed the incorporation of sound to project the pronunciation of the names of the objects. On top of that, the dataset of images are collected from Google Open Images and COCO images. They have a huge variety of images ranging from different objects, which makes it convenient to sieve out the related images of the objects needed to train in the model.

However, it is also very evident that improvements need to be made to the model and that it needs to reach a higher level of optimisation in the Raspberry Pi in order to gain a higher level of accuracy.

Eventually, there are many more aspects that can be considered and explored with the device. It is possible to train the Object Detection model to say out the names of the different objects in different languages to also help teach Mother Tongue to kindergarteners. If given better equipment, it is possible to train a larger variety of object types so that the kindergarteners can learn and reap more benefits from the device.

References

Asian Development Bank. (2021, April). *Learning and Earning Losses from COVID-19 School Closures*. Asian Development Bank. Retrieved January 5, 2022, from https://www.adb.org/sites/default/files/publication/692111/ado2021-special-topic.pdf

Edutopia.org. (2020, October 21). *Why Teaching Kindergarten Online Is So Very, Very Hard*. Edutopia.org. Retrieved January 5, 2022, from https://www.edutopia.org/article/why-teaching-kindergarten-online-so-very-very-hard

Forbes. (2020, July 8) *Technology For A Better Future*. Forbes. Retrieved January 8, 2022, from https://www.forbes.com/sites/alainapercival/2020/07/07/technology-for-a-better-future/?sh=4073ffa82783

OCED. *What do we know about children and technology?* - OECD. Retrieved January 10, 2022, from https://www.oecd.org/education/ceri/Booklet-21st-century-children.pdf

Raja, R. & Nagasubramani, P.. (2018). Impact of modern technology in education. *Journal of Applied and Advanced Research. 3*. 33. 10.21839/jaar.2018.v3iS1.165. Retrieved January 8, 2022, from https://www.researchgate.net/publication/325086709_Impact_of_modern_technology_in_education

The Apple Tree kindergarten. *The benefits of hands-on learning for children.* Retrieved January 15, 2022, from https://appletreekindergarten.com/blog/benefits-hands-learning-children/

UNICEF. (2021, December 10). *UNICEF Lego Foundation Learning through Play.* Retrieved January 8, 2022, from https://www.unicef.org/sites/default/files/2018-12/UNICEF-Lego-Foundation-Learning-through-Play.pdf

Viso.ai. (2021, August 9) *Object Detection in 2021: The Definitive Guide*. Retrieved January 10, 2022, from https://viso.ai/deep-learning/object-detection/