Development of an Augmented Reality Application for Intuitive Infant English Learning*

Momoka Hagihara† and Tomoyuki Ishida
Fukuoka Institute of Technology, Higashi-ku, Fukuoka 8110295 Japan
mgm22103@bene.fit.ac.jp, t-ishida@fit.ac.jp

Received: June 21, 2022; Accepted: August 25, 2022; Published: September 10, 2022

Abstract

In this study, we developed an "English learning AR application for infants" that combines augmented reality (AR) and deep learning by linking visually familiar objects with English words. Infants, who are the target users of this study, can visually learn the English words of an object as AR information by holding a smartphone over an arbitrary object. We constructed 57 types of self-made datasets, prepared 200 images for each English word, and learned them 300 times. Additionally, we conducted a system evaluation questionnaire survey for 44 subjects to evaluate this AR application. After experiencing this application, the subjects evaluated the operability, relevance, readability, functionality, applicability, and learning effect. Because of the system evaluation questionnaire survey, although we could obtain high evaluations for many items, the result was that issues remained in functionality.

Keywords: Augmented Reality (AR), Markerless AR, Infant English Learning, Deep Learning

1 Introduction

Globalization is progressing remarkably in modern Japan. Therefore, it is essential to improve English ability, which is a prevalent international language. Consequently, the Ministry of Education, Culture, Sports, Science, and Technology of Japan revised the course of study and made English education in elementary schools compulsory from 2020. This new course of study acquires English proficiency that can actively use the four skills of "listening," "speaking," "reading," and "writing" throughout life. Especially in elementary school English education, from the middle grades to cultivate the basics of communication while becoming acquainted with the voice, and from the upper grades to develop the basics of communication ability of basic expressions "listening," "speaking," "reading," and "writing" [2, 3].

Alternatively, the rapid spread of mobile terminals, such as smartphones and tablets, with high processing power recently has made the Internet more familiar in our daily lives [4, 5]. This Internet access has made it easier for us to collect and disseminate various types of information. Therefore, the digitization of educational teaching materials has progressed, and different teaching materials using tablet terminals have been developed. Among them, AR teaching materials, which display information on top of each other in real space, are attracting significant attention. For example, various services such as teaching materials for learning the structure of three-dimensional models such as insects and planets by

Research Briefs on Information & Communication Technology Evolution (ReBICTE), Vol. 8, Article No. 8 (September 10, 2022)
DOI:*10.22667/ReBiCTE.2022.09.10.008*

*This paper is an extended version of the work [1] originally presented at the 25th International Conference on Network-Based Information Systems, Online, Sep. 7-9, 2022.

†Corresponding author: Graduate School of Engineering, Fukuoka Institute of Technology, 3-30-1 Wajiro-higashi, Higashi-ku, Fukuoka, 8110295, Japan

viewing them three-dimensionally by AR are provided. These services can be operated on mobile terminals and are applied in various fields such as tourism, education, welfare, and medical care [6, 7, 8, 9, 10]. AR is roughly divided into two types: location-based AR, which uses location information from GPS, and vision-based AR, which detects objects and displays information [11, 12, 13]. Furthermore, vision-based AR can be further divided into marker and markerless types. Most of the AR services used in the educational field are intended to be used in schools, so marker-type AR is widely used. Likewise, most of the AR services used in the academic field are intended to be used in schools, so marker-type AR is widely used. Therefore, in this study, we develop a markerless AR-based English learning application for infants by performing object recognition using image processing technology. Image processing techniques primarily use machine learning or deep learning. For example, in machine learning in image processing, humans must recognize color differences from machines to identify colors. Therefore, when it is necessary to recognize a huge amount of data, it is essential to recognize each data with human hands.

However, in deep learning, since the machine itself recognizes it, it is not necessary to recognize it manually. Therefore, it is possible to learn a huge amount of data easily, and it is possible to perform advanced and flexible processing. Deep learning is used for image recognition, speech recognition, and natural language processing. It is expected to progress in many fields, such as automated driving and medical equipment requiring high recognition accuracy. In this study, we have developed an "AR application for intuitive infant English learning" that combines AR and deep learning by linking visually familiar objects with English words.

The rest of the article is organized as follows. The related works are presented in Section 2. The objective of our study is described in Section 3. System architecture of our proposed AR application is explained in Sections 4. The prototype system is described in Section 5 and is analyzed in Section 6. The AR application is discussed in Section sect:Discuss, and the conclusion of our findings is given in Section 8.

2 Related Works

Konijn et al. [14] developed a system using social robots for language learning. This system allows primary school children to learn a second language with social robots as tutors. However, since primary school children learn a second language alongside social robots, it is not easy to use them outdoors. It also requires the cost of purchasing social robots.

Koca et al. [15] developed a system in which 3D objects and animation of animals are displayed when a camera captures a card with an animal. This system displays a 3D model with sound on the detected card when the captured frame matches the animal image in the 3D database. However, in this system, the user must have a card with an animal on it to experience the animal's 3D object and animation.

Eriksen et al. [16] have developed an augmented reality app to visualize chemical 3D models. This system allows users to freely view a 3D molecular structure model in 360 degrees by capturing the molecular structure printed on paper with the app's camera. However, in this system, the user must use a paper medium on which the molecular structure is drawn to view the 3D model of the molecular structure.

Haryanto et al. [17] developed an Android application to recognize animals using marker-based tracking methods. In this system, when the smartphone camera captures the marker, the 3D model of the aquatic animal drawn on the marker is superimposed and displayed. However, in this system, markers with drawn aquatic animals are required for users to view 3D models of aquatic creatures. Also, since it is implemented based on Android, only Android users can use this system.

Liu et al. [18] developed an English learning system that uses 2D barcodes and AR to assist English learning. The system uses 2D barcodes and an AR to provide students with a contextual English learning experience. The system uses 2D barcodes and an AR to provide students with a contextual English learning experience. However, this system requires a 2D barcode for users to learn English with AR.

Takeda et al. [19] developed a learning system that projects a molecular model onto a terminal by recognizing the structural formula of the molecule with a camera using a marker-type AR. This learning system was devised to deepen the understanding of molecules by associating markers with multiple learning methods, such as the quiz function of molecular composition. First, however, it must prepare AR markers for each molecular structure.

Fujitsuka et al. [20] developed a penmanship learning support system. Here, the user wears a head-mounted display (HMD) and recognizes the AR marker for learning penmanship while watching the movement of the teacher's brush displayed at hand. However, since this system uses HMD as a platform, it is not easy to use.

Iwashita et al. [21] developed teaching materials to understand radiation's characteristics using AR technology and OpenCV. However, since this system uses OpenCV for image processing, manually recognizing each feature point is necessary, and enormous processing must provide several educational materials.

Wakahara et al. [22] developed an English learning system using marker-type AR. In this system, the user randomly rearranges the markers of the alphabet group selected by the system and learns the meaning of the completed word from the voice and image. However, the recognition rate decreases when the number of AR markers is five or more.

Ando et al. [23] developed an AR-based relevant information system for supporting book purchases by using the spine image of a book as an AR marker. Simultaneously, by browsing the book information related to itself, the user can acquire the position information by object recognition and browse the related information. However, since ARKit, which is a framework for iOS development, is used, the use of this system is limited to iOS users.

Yamazaki et al. [24] have developed an AR system that projects annotation information on documents to support and navigate document entry work by using projection-type AR. Since this system uses a projector, multiple people can share information without wearing a device such as a tablet or HMD. However, this system requires devices such as a projector and a detection camera, and its usage is limited.

Motonobu et al. [25] developed a clean-up support system, technology, and image recognition using AR and OpenCV, respectively. This system extracts feature points using OpenCV and identifies books. However, the human cost is an issue because many images must be processed to increase the recognition rate.

3 Research Objective

Many systems use AR markers in related studies on the learning theme. However, these systems have problems in that the installation cost of the AR marker is high, the number of users is limited depending on the terminal used, and the place where the system is used is limited. Additionally, in image processing, using machine learning, it is necessary to extract the feature points of things one by one, and this is a problem that processing much data requires human cost. Therefore, to solve the problems mentioned in the related study, we developed an English word-learning application for infants that combines markerless AR and deep learning. This English word-learning application superimposes English words in real space when the terminal's camera catches a familiar object. This application stimulates infants' motivation and interest in learning English and deepens their interest in it. In particular, this study raised interest by feeling closer to English to learning English in elementary school. In realizing this application, YOLO, an algorithm of the object detection method, is used. Also, in image processing, we can support various familiar objects such as animals and fruits by creating a self-made data set and learning by deep learning. The development of an English word-learning AR application for intuitive infant English education is, therefore, realizable.

4 System Architecture

The system architecture is shown in Figure 1. This system consists of a mobile agent, a content server, and a learning dataset.

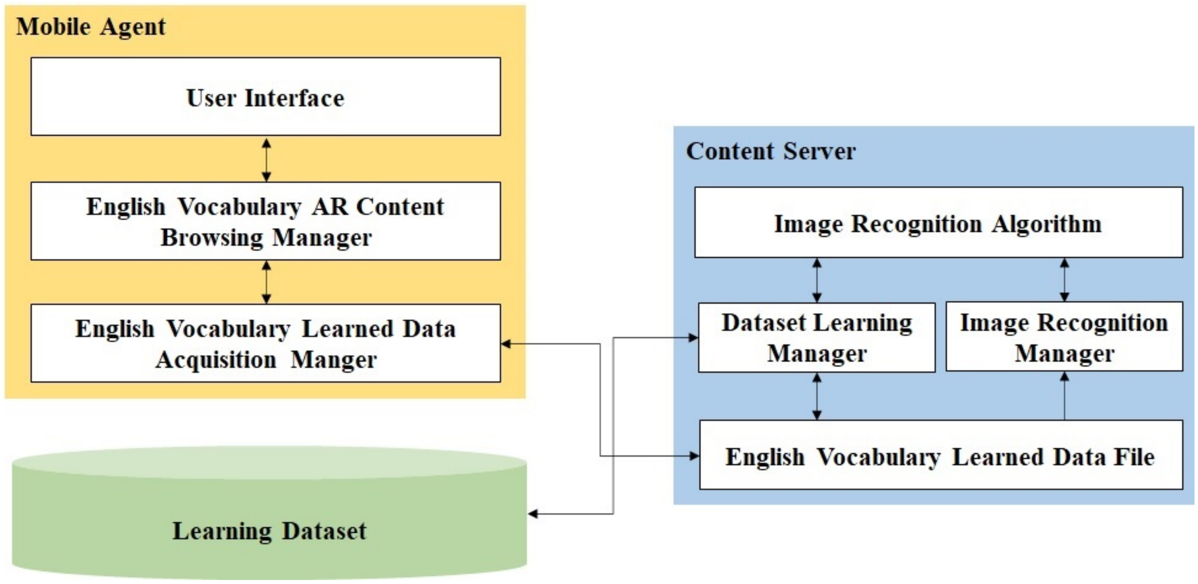

Figure 1: System architecture

- Mobile Agent
  The mobile agent consists of a user interface, an English vocabulary AR content-browsing manager, and an English vocabulary learned data acquisition manager. The user interface is an interface part for users to use the contents of this system and provides an application for mobile agents. The English vocabulary AR content-browsing manager requests object recognition from the content server's English vocabulary learned data file via the English vocabulary learned data acquisition manager and superimposes the AR content on the user interface. Likewise, the English vocabulary learned data acquisition manager requests object recognition from the content server's English vocabulary learned data file in response to the request of the English vocabulary AR content-browsing manager.

- Content Server
  The content server consists of an image recognition algorithm, a dataset learning manager, an English vocabulary learned data file, and an image recognition manager. The image recognition algorithm is an algorithm that can perform object detection and image recognition at the same time. Object recognition is performed using the learned data in this algorithm. The dataset learning manager learns using the data that the learning dataset provides. The English vocabulary learned data file stores the learned data by the dataset learning manager. The learned data generated by the dataset learning manager is provided to the English vocabulary learned data acquisition manager

and the image recognition manager of the mobile agent. The image recognition manager uses the learned data in the English vocabulary learned data file to perform object recognition.

- Learning Dataset
  The learning dataset stores the learning data provided to the mobile agent. Additionally, the dataset contains images of recognizable objects and their corresponding labels.

5 Prototype System

The home screen of the mobile agent application is shown in Figure 2. When the user launches the application, the home screen is displayed. The home screen consists of a "START !!" button for using the AR camera function and a "How to use" button for viewing how to use the application.

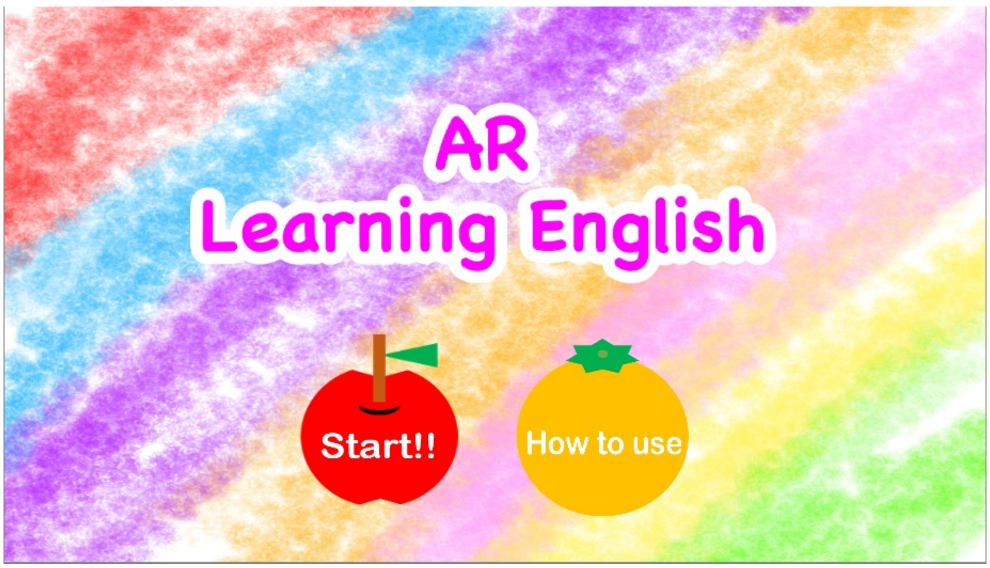

Figure 2: Home screen of the mobile agent application

The AR camera function recognizes an object captured by the camera and displays the information on the recognized object in the AR. When the user presses the "START !!" button, a caution screen for using the AR function, as shown in Figure 3, is displayed. After that, the camera started when the user confirmed the caution screen and pressed the "OK" button.

Figure 4 and 5 show the execution screens in which an object is recognized using the AR camera function, and the information about the recognized object is displayed as AR.

To realize object recognition, we used YOLOv5 [26]. YOLO is an object detection algorithm that can perform object detection and classification at the same time. In this study, we constructed a self-made dataset. We used Kaggle [27] for the image dataset, a website that handles datasets and prepares 200 images for each object. Table 1 shows the 57 types of dataset items.

Table 1: Self-made dataset items

Recognizable objects			
Dolphin	Fish	Jellyfish	Octopus
Shark	Shell	Shrimp	Squid
Turtle	Whale	Ant	Bee
Beetle	Butterfly	Cicada	Dragonfly
Grasshopper	Ladybug	Mantis	Spider
Cabbage	Corn	Cucumber	Eggplant
Green pepper	Lettuce	Mushroom	Onion
Potato	Pumpkin	Cherry	Grapes
Kiwi fruit	Lemon	Peach	Pineapple
Strawberry	Watermelon	Lion	Pig
Snake	Tiger	Alligator	Koala
Monkey	Panda	Quail	Rabbit
Duck	Frog	Goat	Gorilla
Hippopotamus	Kangaroo	Penguin	Polar bear
Rhinoceros			

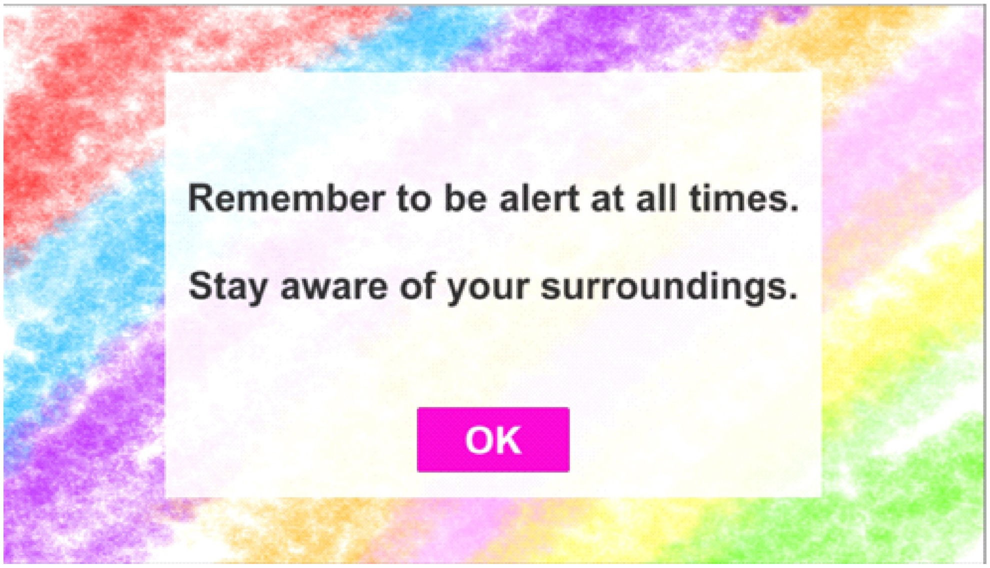

Figure 3: Caution screen for using the AR function

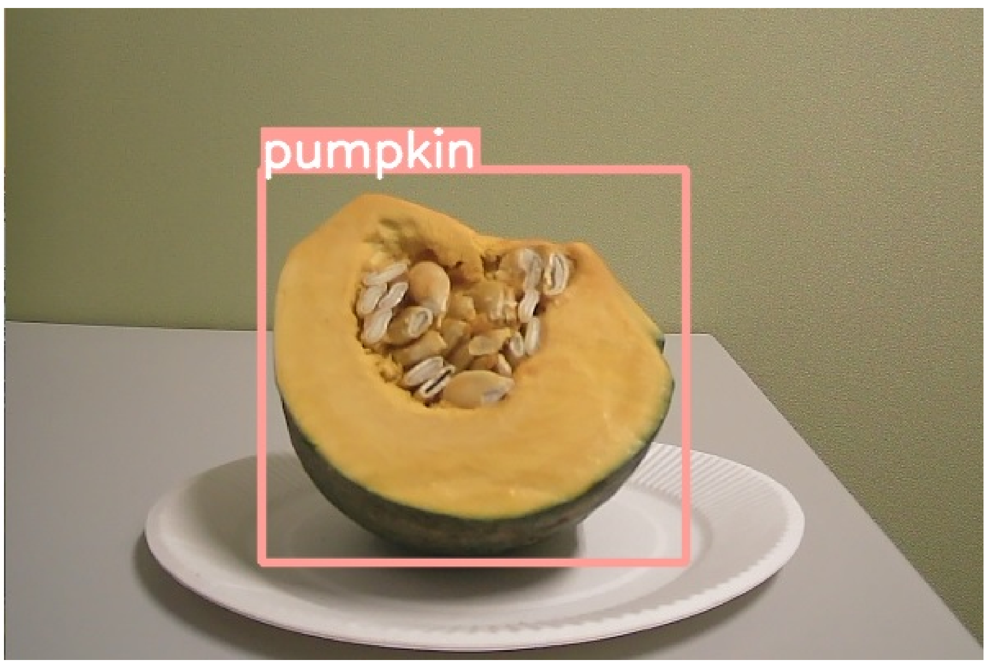

Figure 4: AR display screen when recognizing a pumpkin

Fig. 6 shows the sequence of this application.

- Mobile agent (User)
  When the user, a mobile agent, selects the "Start!" button on the home screen, the AR camera starts. First, the AR camera requests information about the captured object from the content server. When the content server recognizes the object successfully, the mobile agent acquires the object information from the learned data and browses the information as an AR.

- Content server

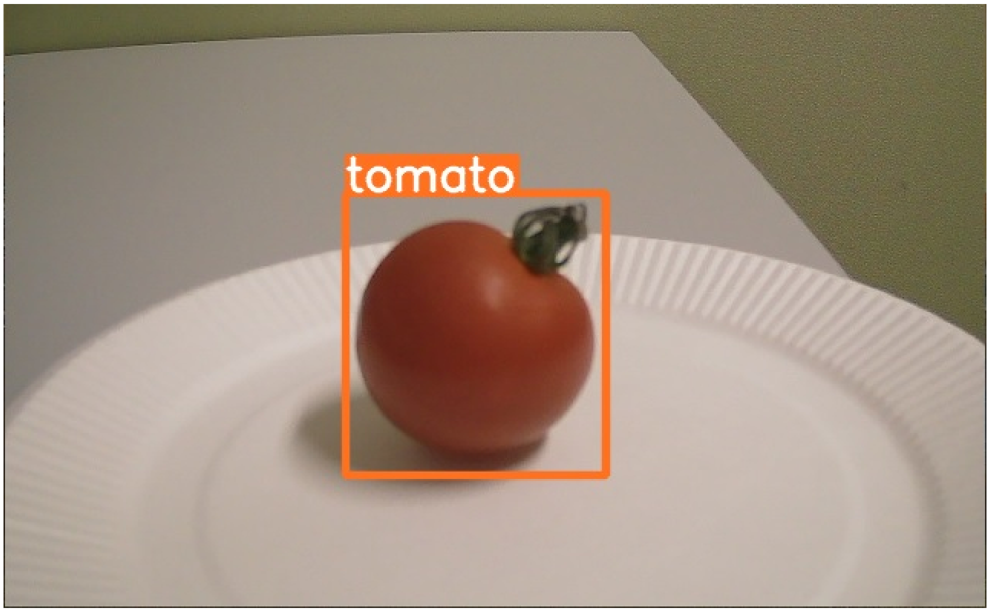

Figure 5: AR display screen when recognizing a tomato

Upon request from the mobile agent, the content server provides the learned data to the mobile agent when it recognizes an object. On the one hand, this server acquires images and labels from the learning dataset and learns the data.

- Learning dataset
  On the content server, the learning dataset stores the images and their corresponding labels necessary for learning.

6 System Evaluation

We conducted a questionnaire survey of 44 subjects to evaluate the mobile agent application's operability, relevance, readability, functionality, applicability and learning effect. The subjects were three teens, thirty-seven 20s, two 40s, and two 50s (Figure 7).

Figure 8 shows the operability results of the mobile agent application. Regarding the operability of the mobile agent application, all subjects answered "easy" or "somewhat easy," and we could confirm the mobile agent application's high operability.

Figure 9 shows the relevant results of the mobile agent application. Regarding the relevance of the mobile agent application, 43 subjects answered "relevant" or "somewhat relevant," and we could confirm the high relevance of the mobile agent application.

Figure 10 shows the readability results of the mobile agent application. Regarding the readability of the mobile agent application, 43 subjects answered "easy to understand" or "somewhat easy to understand." Therefore, we could confirm the high readability of the mobile agent application. A subject commented, "I think this application is good because the English word of AR is superimposed in real-time in an easy-to-understand manner."

Figure 11 shows the functionality results of the mobile agent application. Regarding the functionality of the mobile agent application, 33 subjects answered "satisfied" or "somewhat satisfied." However, nine

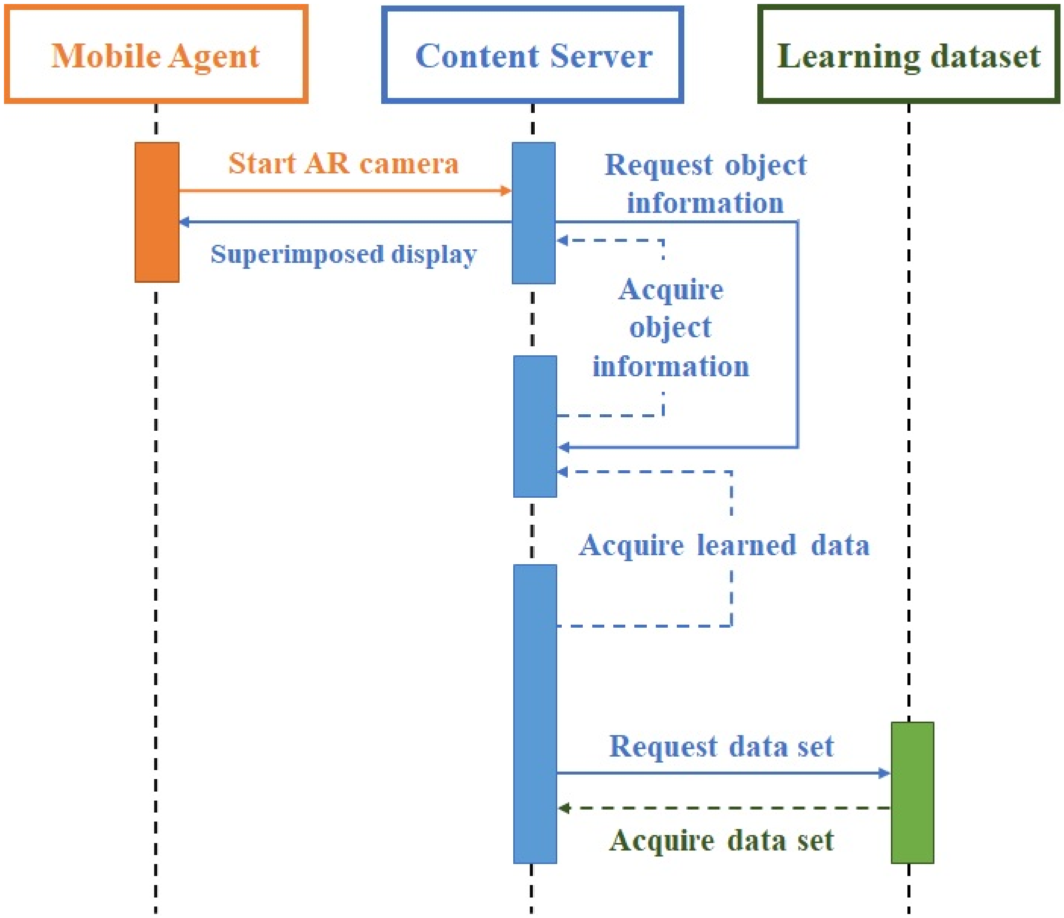

Figure 6: Sequence of this application

Figure 7: Subject's age ($n = 44$)

subjects answered "no opinion," and one commented, "I think it would be a better application if there were a voice reading function."

Figure 12 shows the applicability results of the mobile agent application. Regarding the applicability of the mobile agent application, 41 subjects answered "possible" or "somewhat possible." Therefore, we could confirm the high applicability of the mobile agent application. Furthermore, subjects commented, "I think it can be applied to tablet lessons in elementary school," and "I think it can be applied to the scientific field as well."

Figure 13 shows the learning effect results of the mobile agent application. Regarding the learning effect of the mobile agent application, 42 subjects answered "effective" or "somewhat effective." Therefore, we could confirm the high learning effect of the mobile agent application. Furthermore, a subject commented, "I thought that the learning effect would be even higher if not only the AR display of English words but also the AR display of Japanese."

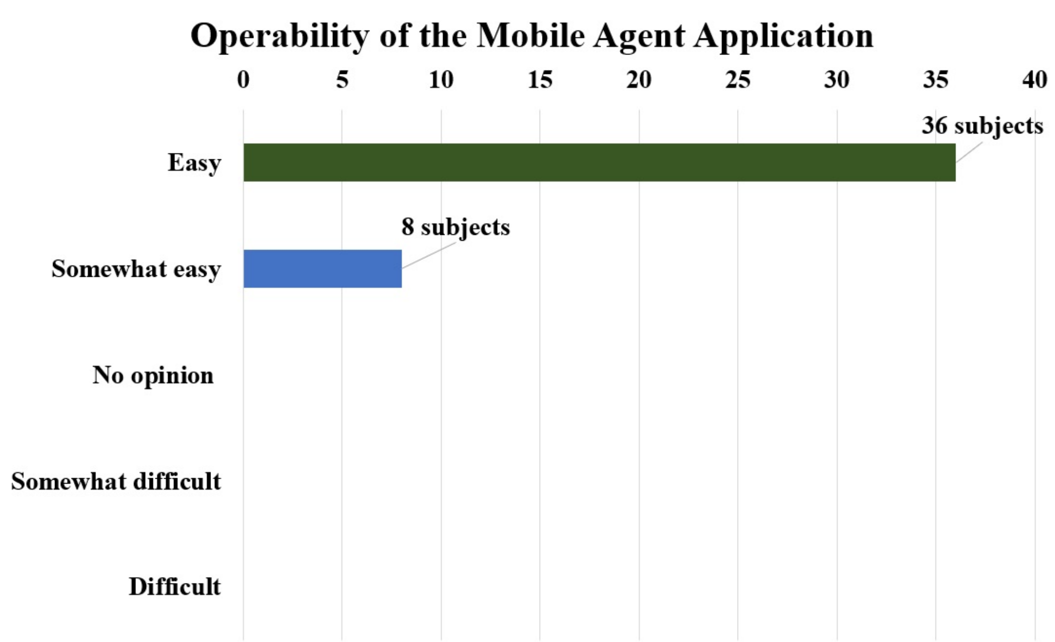

Figure 8: Operability of the mobile agent application (n = 44)

7 Discussion

The self-made dataset created in this study used 200 images for each English word and was learned 300 times. However, because the number of images trained by YOLOv5 and the amount of training was small, there was the problem that it sometimes reacted to unexpected objects. Therefore, it is necessary to improve the accuracy by increasing the number of learning images and the number of learnings per English word. Additionally, the dataset trained with YOLOv5 is twelve types of English words. However, with the twelve types of English words, the range of English words that infants can learn is limited. Therefore, in the future, it will be necessary to increase the number of datasets so infants can learn various familiar objects. We have developed an application that allows infants to learn English words visually by superimposing the recognized object's English words in real space. However, learning the actual pronunciation only by superimposing English words is impossible. Additionally, in the questionnaire survey, we received a comment from the subject that "it would be a better application if there

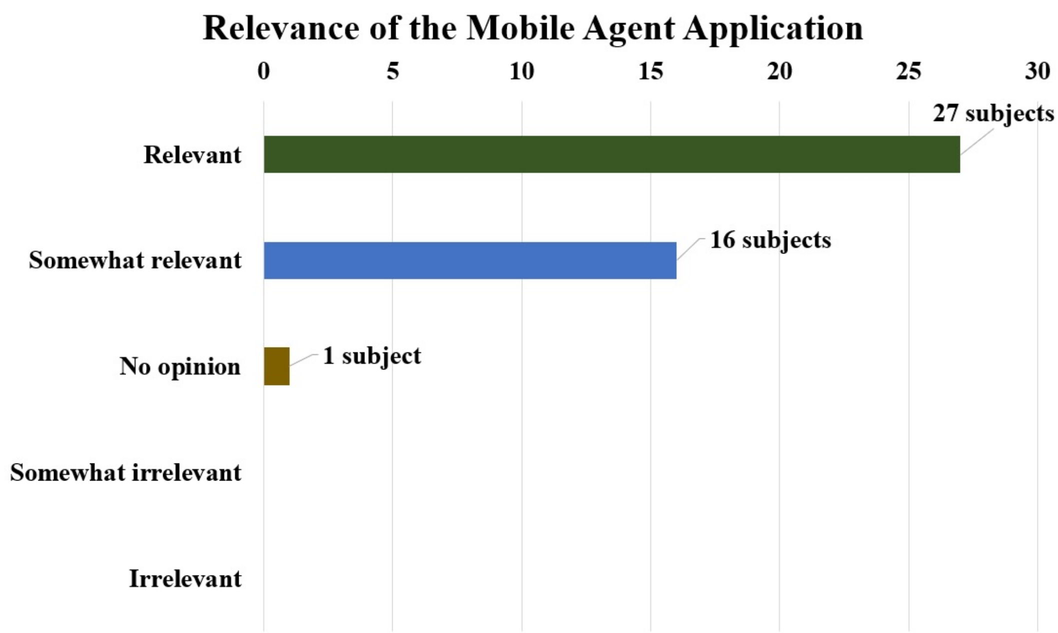

Figure 9: Relevance of the mobile agent application (*n* = 44)

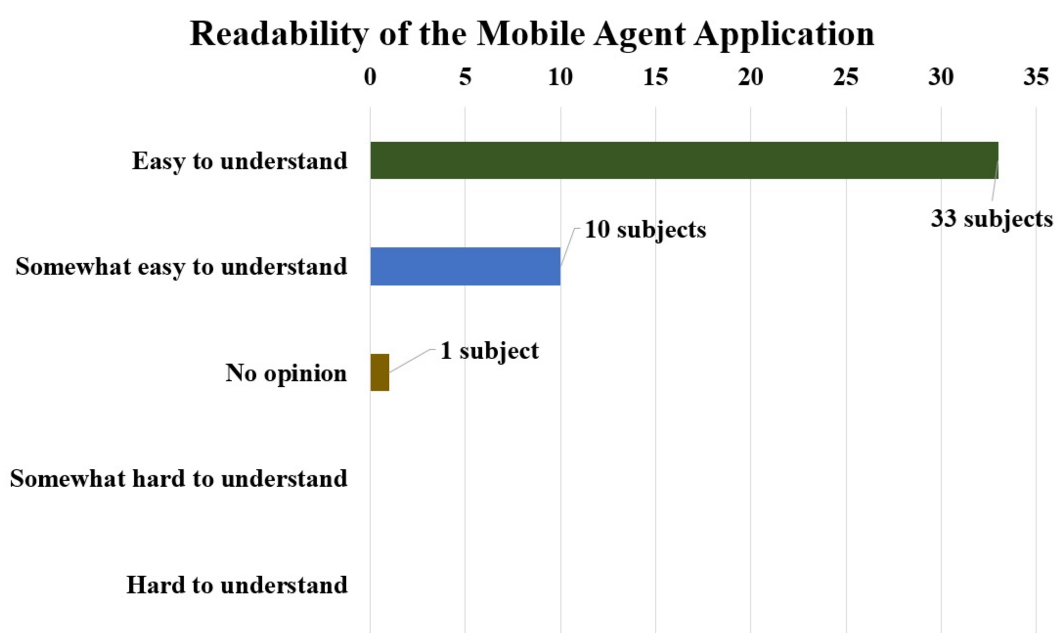

Figure 10: Readability of the mobile agent application (*n* = 44)

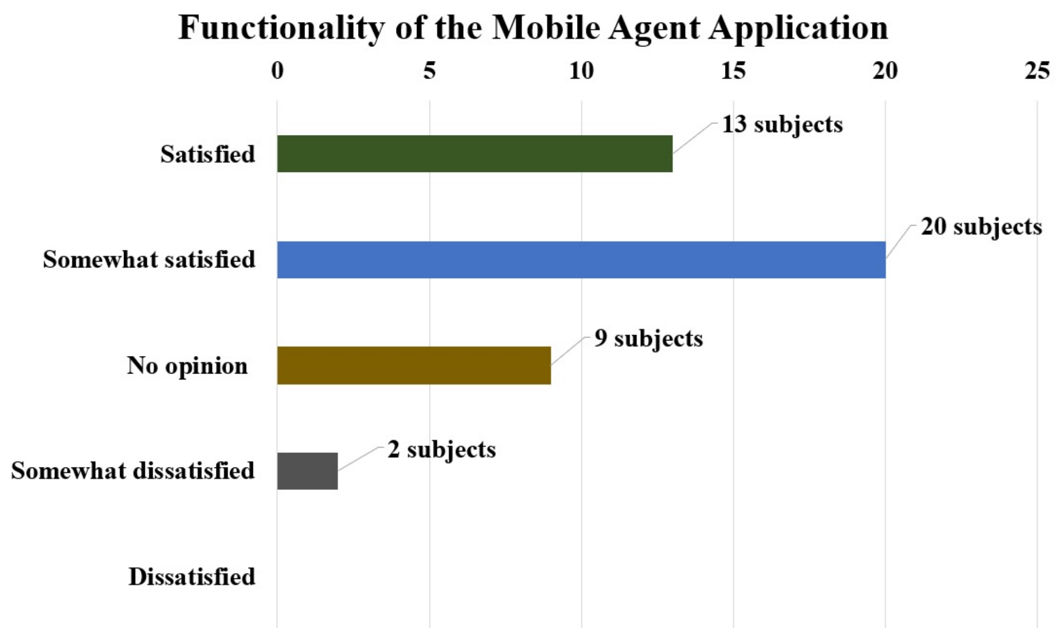

Figure 11: Functionality of the mobile agent application (n = 44)

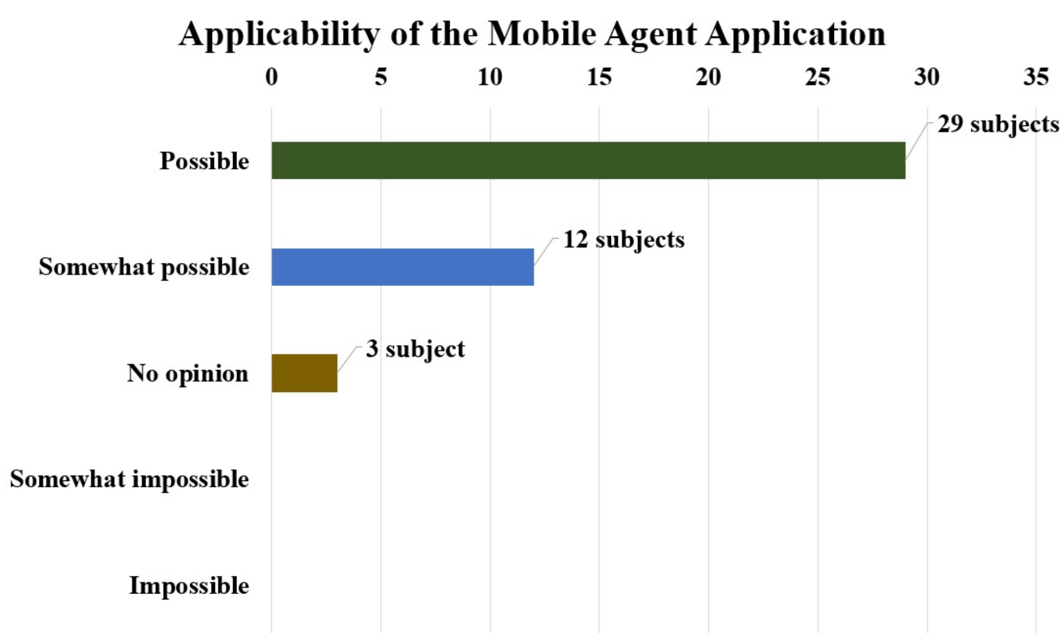

Figure 12: Applicability of the mobile agent application (n = 44)

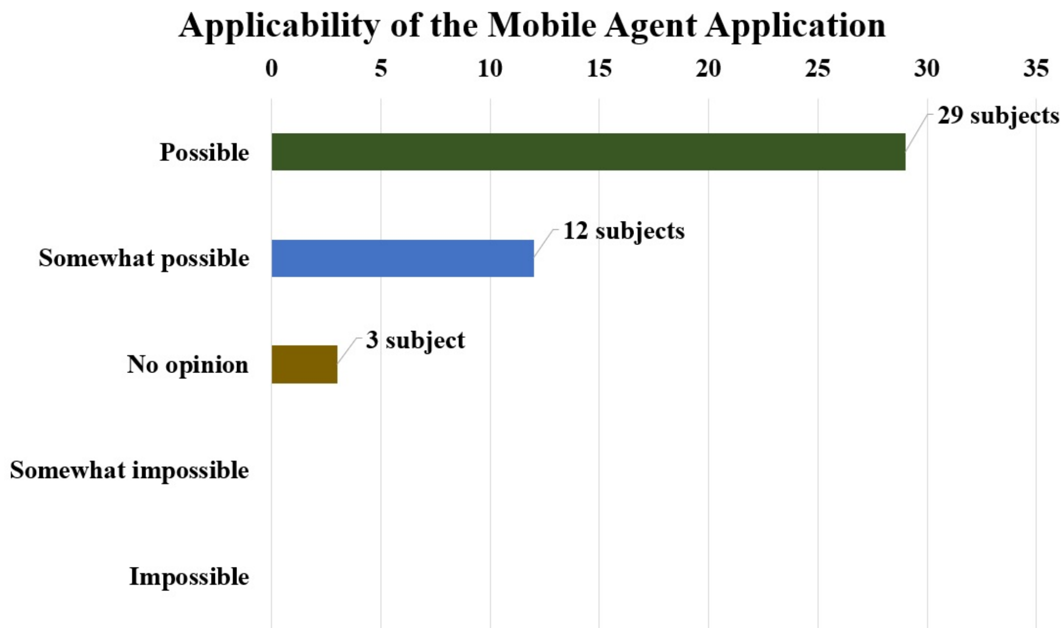

Figure 13: Learning effect of the mobile agent application (n = 44)

were a voice reading function." Therefore, Additional to the visual function, it is necessary to consider the addition of a voice function that can be learned aurally.

8 Conclusion and Future Works

In this study, we developed and evaluated an English word-learning application for infants that combined markerless AR and deep learning. In the developed learning application, YOLOv5 was used as an object detection algorithm, and an infant's unique data set related to English words was constructed in image processing. This application superimposes familiar English words in a real space with the camera of a tablet terminal. Using this application, the infants can stimulate their motivation and interest in learning English and deepen their interest in it. Additionally, a system evaluation questionnaire survey was conducted on 44 subjects to evaluate the operability, relevance, readability, functionality, applicability, and learning effect of the English word-learning AR application. As a result, we obtained high evaluations on many items. However, from the evaluation results of functionality, future issues, such as the addition of a voice reading function, were clarified.

References

[1] M. Hagihara and T. Ishida. Implementation of an english word learning ar application for infant english education. In *Proc. of the 25th International Conference on Network-Based Information Systems, Online*, volume 526 of *Lecture Notes in Networks and Systems*, pages 328–334. Springer, Cham, September 2022.

[2] MEXT. Report on the future improvement and enhancement of english education: Five recommendations on the english education reform plan responding to the rapid globalization. `https://www.mext.go.jp/b_menu/shingi/chousa/shotou/102/houkoku/attach/1352464.htm`, [Online; Accessed on September 1, 2022].

[3] MEXT. English education reform plan corresponding to globalization. `https://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2014/01/31/1343704_01.pdf`, [Online; Accessed on September 1, 2022].

[4] MIC. White paper information and communications in japan 2020. `https://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2020/2020-index.html`, [Online; Accessed on September 1, 2022].

[5] MIC. White paper information and communications in japan 2017. `https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h29/pdf/index.html`, [Online; Accessed on September 1, 2022].

[6] E.Z. Barsom, M. Graafland, and M.P. Schijven. Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical Endoscopy*, 30(10):4174—-4183, 2016.

[7] G. Yamamoto, A. I. W. Lübke, T. Taketomi, and H. Kato. A see-through vision with handheld augmented reality for sightseeing. In *Proc. of the 16th International Conference on Human-Computer Interaction (DAPI'14), Crete, Greece*, volume 8530 of *Lecture Notes in Computer Science*, pages 392–399. Springer, Cham, June 2014.

[8] H. Furata, K. Takahashi, K. Nakatsu, K. Ishibashi, and M. Aira. A mobile application system for sightseeing guidance using augmented reality. In *Proc. of the 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligence Systems, Kobe, Japan*, pages 1903–1906. IEEE, November 2012.

[9] D. D. Sumadio and D. R. A. Rambli. Preliminary evaluation on user acceptance of the augmented reality use for education. In *Proc. of the 2nd International Conference on Computer Engineering and Applications (ICCEA'10), Bali, Indonesia*, pages 461–465. IEEE, March 2010.

[10] D. Wagner, D. Schmalstieg, and M. Billinghurst. Handheld ar for collaborative edutainment. In *Proc. of the 16th International Conference on Artificial Reality and Telexistence (ICAT'06), Hangzhou, China*, volume 4282 of *Lecture Notes in Computer Science*, pages 85—-96. Springer, Berlin, Heidelberg, November 2006.

[11] Junchul Chun and Seonho Lee. A vision-based 3d hand interaction for marker-based ar. *International Journal of Multimedia and Ubiquitous Engineering*, 7(3):51–59, July 2012.

[12] H. Miyajima, N. Yata, and Y. Manabe. Ar sightseeing system with proximity detection and markerless ar. In *Proc. of the 2018 International Workshop on Advanced Image Technology, Chiang Mai, Thailand*, pages 1–3. IEEE, January 2018.

[13] K. C. Brata and D. Liang. An effective approach to develop location-based augmented reality information support. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(4):3060–3068, August 2019.

[14] E. A. Konijn, B. Jansen, V. M. Bustos, V. L. N. F. Hobbelink, and D. P. Vanegas. Social robots for (second) language learning in (migrant) primary school children. *International Journal of Social Robotics*, 14:827—-843, 2022.

[15] B. A. Koca, B. Çubukçu, and U. Yüzgeç. Augmented reality application for preschool children with unity 3d platform. In *Proc. of the 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT'19), Ankara, Turkey*, pages 1–4. IEEE, October 2019.

[16] K. Eriksen, B. E. Nielsen, and M. Pittelkow. Visualizing 3d molecular structures using an augmented reality app. *Journal of Chemical Education*, 97(5):1487—-1490, April 2020.

[17] E. V. Haryanto, E. L. Lubis, A. Saleh, Fujiati, and N. I. Lubis. Implementation of augmented reality of android based animal recognition using marker based tracking methods. *Journal of Physics: Conference Series*, 1361(1):1–5, 2019.

[18] T.-Y. Liu, T.-H. Tan, and Y.-L. Chu. 2d barcode and augmented reality supported english learning system. In *Proc. of the 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS'07), Melbourne, Australia*, pages 5–10. IEEE, July 2007.

[19] I. Takeda and Y. Kenji. Development of molecular structure indicate system using ar technology. In *Proc. of the 82nd National Convention of IPSJ, Ishikawa, Japan*, pages 543–544. IPSJ, March 2020.

[20] T. Fujiitsuka, J. Iwakura, S. Yamashita, and H. Arai. A penmanship learning support system using augmented reality. In *Proc. of the 2014 IEICE General Conference, Nitaga, Japan*, page 163. IEICE, March 2014.

[21] S. Iwashita and A. Oda. Development of radiation teaching materials using image processing technology.

Research reports of Kumamoto National College of Technology, 9(2018):55–60, 2018.

[22] M. Wakahara, Y. Matsuda, Y. Simosato, and R. Hamakawa. The proposal of the next generation type english learning system using augmented reality. In *Proc. of the 74th National Convention of IPSJ, Aichi, Japan*, pages 317–318. IPSJ, March 2012.

[23] R. Ando, T. Ozono, and T. Shintani. Developing an ar-based relevant information system for supporting book purchase. In *Proc. of the Forum on Information Technology 2019, Okayama, Japan*, pages 73–76. IPSJ, September 2019.

[24] K. Yamazaki and H. Abe. A projection-based ar system by annotating view management based on force-directed graph. *IPSJ SIG Technical Report*, 2018-DCC-18(2018):1–7, 2018.

[25] T. Motonobu and T. Yoshino. Proposal of a clean-up support system by using color presentation method. In *Proc. of the 80th National Convention of IPSJ, Tokyo, Japan*, pages 325–326. IPSJ, March 2018.

[26] Inc. GitHub. ultralytics. `https://github.com/ultralytics/yolov5`, [Online; Accessed on September 1, 2022].

[27] Kaggl. Find open datasets and machine learning projects. `https://www.kaggle.com/datasets`, [Online; Accessed on September 1, 2022].

Author Biography

Momoka Hagihara received the B.S. degree in Fukuoka Institute of Technology in 2022. Currently she is taking a master's course at Graduate School of Engineering, Fukuoka Institute of Technology. Her research interests include Augmented Reality, Virtual Reality, Deep Learning, Image Processing, and Educational Technology.

Tomoyuki Ishida received the B.S. and M.S. degrees in Software and Information science from Iwate Prefectural University in 2004 and 2006, and Ph.D. degrees in the same University in 2010. Currently he is a professor in Department of Information and Communication Engineering, Fukuoka Institute of Technology, Japan. His research interests include Web Geographic Information System for local governments, Disaster Management System, Safety Confirmation System, Regional Disaster Prevention Planning, Virtual Reality and Tele-Immersion. He is a member of Virtual Reality Society of Japan (VRSJ), Information Processing Society of Japan (IPSJ) and Visualization Society of Japan (VSJ).