

Multiple Merged Structures Based on Image Recognition for Converting Application of Natural Language Artificial Intelligence Service

Hui-Yu Huang, Ming-Hsun Tsai, Ming-Shen Jian*

Dept. of CSIE, National Formosa University, Yunlin County, Taiwan 632

hyhuang@nfu.edu.tw, sean408031015@gmail.com, jianms@nfu.edu.tw*

Abstract— This research provide the Application of Natural Language Artificial Intelligence Service based on the Multiple Merged Structures through the Image Recognition. According to the object and category/class relationships, each object can be implemented corresponding to the on demand defined data structure or category. Based on the recursively object recognition, the information corresponding to the object data structure could be chained. By merging multiple structures from the image recognition, the image can be described by the natural language as the artificial intelligence service.

Keywords— Image Recognition, Data Structure, Artificial Intelligence, Natural Language, Application

I. INTRODUCTION

Recently, artificial intelligence (A.I.) services are generally used in industry, education, and even entertainment. To support the A.I. services, huge hardware resources for complex computing are needed. Although there are several popular application programming interfaces (APIs) provided by various companies through the network, some functions would be limited or the additional usage fee would be required depending on the computing missions of A.I.

Generally speaking, the small models of A.I. service could be managed and deployed more easily. However, due to the computing ability of the small models, to compute and deal with the complex scenario or mission would be difficult. However, small models are easily deployed and implemented. Therefore, to integrate or merge multiple models could be a possible solution.

With the limited computing resources, to parallel and distribute the A.I. computing missions into multiple computing cores is a possible method [1]. Although the processing unit cannot provide huge computing resources and performance, by dividing the original computing job into multiple tasks, the

small models could still deal with the A.I. computing works. Some systems would deploy the specific A.I. mission or computing service on the individual model. Hence, when the computing job is combined with various A.I. services required, each request could be delivered to the corresponding model.

With the development of A.I. services, to let the A.I. application draw an image even the melody with some text descriptions are popular [2-5]. Service users could provide the text description of some objects or emotions, the A.I. services could provide the possible and various images or music according to the data in the database and the past learning histories. Since the training data for the A.I. services are collected from various users and database, it means that one description would cause various feasible solutions or results.

On the other hand, based on the object recognition from the image processing, to transfer or translate from an image to the text description would also cause the various explanations or descriptions. With the importance, recognition accuracy, the size of the segmentation, etc., of the objects in the image, the description or explanation of the specific image would be more different. Therefore, how to let the A.I. service explains the image more similar to people with the order of importance and the topic of the story would be an issue.

In this research, we propose the recursive object segmentation and recognition procedure, called Multiple Merged Models, for the image description and explanation. In section 2, the related researches and works are introduced. Section 3 presents the proposed system. The verification results are given

in section 4. Finally, the section 5 provides the conclusion.

II. RELATED WORKS

Recently, the A.I. services could generate the image according to the given sentence or description which is divided into multiple keywords. One keyword would be used to generate the multiple possible images or objects. Finally, by merging these possible images various images corresponding to the description are presented.

To draw an object corresponding to the text description, the features of the specific object should be defined in advance. Traditionally, different object recognitions could be done depending on huge object data training. By labelling the various objects with huge image data, the object recognition and segmentation could be achieved [6].

However, to execute the A.I. services or image recognitions, especially for the data training, huge computing resources are needed. Except the centralized computing resource for the A.I. services, the parallel processing and distributed system based on the modular computing jobs was proposed [1]. Originally, the requested computing job of image recognition or A.I. service would contain multiple purposes or the recognition of various objects. For the parallel processing, the requested computing job could be divided into multiple tasks corresponding to multiple purposes or objects. Each task would be only for a single recognition target or computing process. Finally, the tasks are distributed to the multiple Computing nodes for further processing. Therefore, we can use the distributed system based on multiple computing nodes with less hardware cost instead of centralized A.I. computing.

In addition to the parallel processing and distributed system for computing, to properly describe the recognized objects, even to give the description corresponding to the original images are also the research issues. The first step is to recognize objects included in the images. By using YOLO algorithm and implemented service, users can give the images related to the target objects for training. After training, the YOLO service could identify and recognize the specific target object

included in the image. For example, according to the music spectrum or characteristics, the similar music instrument can be detected [3]. In addition, by collecting the image of the target object for YOLO algorithm training without anchor points, the target object can be recognized even under the water with high turbidity [4]. Even when the target object is less different in color from the background, the YOLO algorithm could still recognize the fig fruit [5].

After recognition, the found objects should be shown to the users. Based on the YOLO recognition results, the identified objects can be estimated and feedback to the users through audio [2]. In other words, the objects can be represented as the text plain content. It means that to give the description of the image is possible after object recognition.

III. PROPOSED SYSTEM

In this research, suppose that an object can be represented as a data structure. Various attributes included in the data structure are used to describe the corresponding object. Some attributes in the current object data structure can be pointed to another independent data structure as another new object with other new attributes. For example, a person can be an object with various attributes such as clothes, glasses, height, etc. An attribute, take clothes as an example, could be also an object with addition attributes which include color, material, sleeve length, etc. Therefore, similar to the tree structure, an object in the root node could be described and defined by more extending structures or data nodes. The following figure shows the possible data tree based on the data structure.

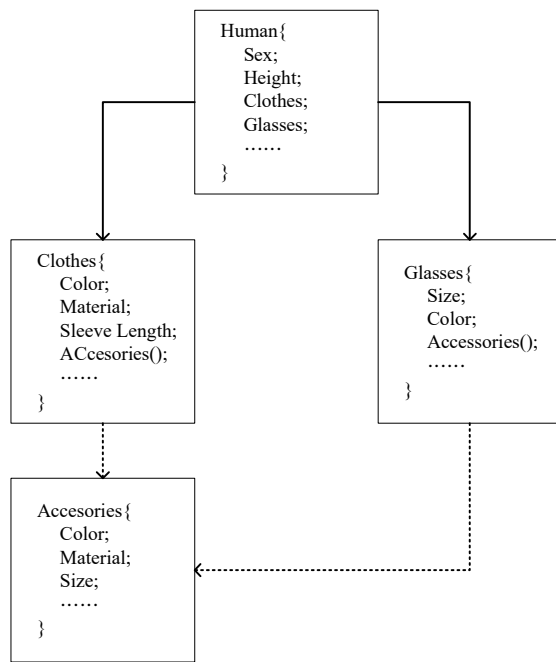


Figure 1. The UML diagram for the relationship between different classes of objects

In this research, some classes of data structure are defined as the root structures or root classes for the root object implementation. When a root object is recognized and segmented, the possible extending object according to the attributes in the parent class would be verified within the parent object segmentation area similar to the zoom in management of the image. Since the object would be only recognized and found within the limited area, the relationship between the parent object and the extending object can be stronger with higher accuracy.

Figure 2 presents the process flowchart. In the beginning, the image file would be reviewed and recognized according to the pre-defined root classes. If there are objects recognized that matching the definition of root classes, these objects are called root objects. In this research, an independent data structure corresponding to the recognized object is implemented. Therefore, one main object would be followed by multiple sub-objects. It also means that the data structure corresponding to an object would be pointed and chained extendedly by other structures when the attributes point to the structure of the sub-objects.

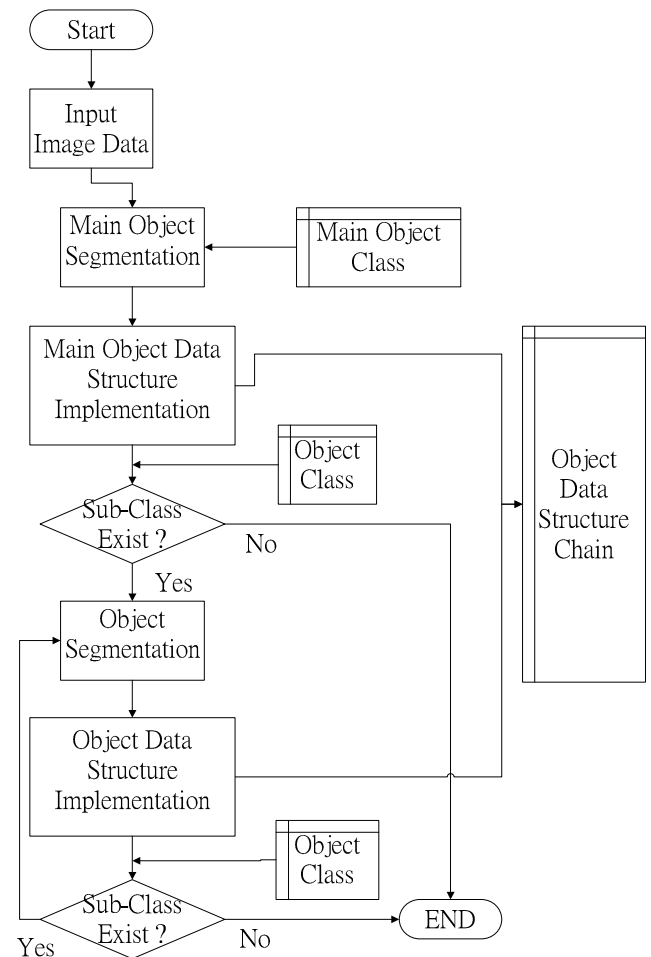


Figure 2. The flowchart of the object recognition and segmentation for the object data structure chain

In other words, according to the attributes included the data structure, the maximum amount of the objects recognized and segmented in the image can be decided.

Therefore, according to the Figure 2, all the found objects would be represented as multiple object data structure chains. Each main object would be extended as the data tree structure or multiple branch chains. Generally speaking, most recognized objects could be presented or expressed as a noun individually. In addition, the attributes of the structure corresponding to individual object are various adjectives such as color, size, etc. Considering the possible verbs related to the specific noun of the object, a suitable description of the current object can be developed.

IV. VERIFICATION RESULTS

In this research, the image recognition is implemented based on YOLO v8. When a new

image comes, the central of the image will be decided first. Generally speaking, the object in the central location would be the main object or motif of the image. After first round of YOLO, the main object with the biggest size or the nearest position of the image centre can be recognized. Hence, the procedure in Figure 2 can be used for the data information chain development. For example, in the Figure 3, there are 3 persons are recognized with the bigger segment size. Since these 3 persons can be implemented as 3 individual objects according to the same data structure/class called ‘person’, it means that these three objects would have the same importance. Then, according to the data structure of person, there are multiple attributes which may also point to another structure. In this research, shoes, pants, emotions, etc., are the possible attributes which would need the on demand defined data structure for description.



Figure 3. The example of the object recognition according to the proposed system

Then, according to the on demand defined structure corresponding to the results of image recognition, the various recognized objects related to the main object can be found and chained. Table 1 shows the data structure chain corresponding to the individual main object.

TABLE 1. DATA STRUCTURE CHAIN CORRESPONDING TO INDIVIDUAL FOUND MAIN OBJECT

No.	Data Structure Chain	
	Main Object	Sub-Class / Object
1	person	[footwear, white] [footwear, white] [romper, black] [Happy]
2	person	
3	person	[trouser, pink] [Sad]

Hence, by merging multiple structures in the individual data structure chain corresponding to the main object, the suitable description can be given. For example, the object number 1 can be described as “in this image, a person who wears white shoes and black romper is happy”.

V. CONCLUSIONS

Based on the proposed merging data structure chain, multiple objects corresponding to the main object can be represented as the normal natural language. In addition, the structures of recognized objects can be correctly connected to the main object structure. In other words, based on the data chain or even the tree structure, more suitable text plain content can be given.

REFERENCES

- [1] M. -S. Jian, J. -T. Gui, X. -M. Wu, “Uncentralized Artificial Intelligence Computing Agent with the Distributed Training and Computing Tasks Based on Open Source Cloud Proxy,” CSCE 2023 - ICAI, Las Vegas, USA, p.93. (ISBN:1-60132-518-5)
- [2] M. Mahendru and S. K. Dubey, “Real Time Object Detection with Audio Feedback using Yolo vs. Yolo v3,” 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 734-740, doi: 10.1109/Confluence51648.2021.9377064.
- [3] C. Dewi, R. -C. Chen, Hendry and Y. -T. Liu, “Similar Music Instrument Detection via Deep Convolution YOLO-Generative Adversarial Network,” 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), Morioka, Japan, 2019, pp. 1-6, doi: 10.1109/ICAwST.2019.8923404.
- [4] X. Wang, X. Jiang, Z. Xia and X. Feng, “Underwater Object Detection Based on Enhanced YOLO,” 2022 International Conference on Image Processing and Media Computing (ICIPMC), Xi'an, China, 2022, pp. 17-21, doi: 10.1109/ICIPMC55686.2022.00012.
- [5] W. Yijing, Y. Yi, W. Xue-fen, C. Jian and L. Xinyun, “Fig Fruit Recognition Method Based on YOLO v4 Deep Learning,” 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Mai, Thailand, 2021, pp. 303-306, doi: 10.1109/ECTI-CON51831.2021.9454904.
- [6] A. Torralba, B. C. Russell and J. Yuen, “LabelMe: Online Image Annotation and Applications,” in Proceedings of the IEEE, vol. 98, no. 8, pp. 1467-1484, Aug. 2010, doi: 10.1109/JPROC.2010.2050290.

Hui-Yu Huang received the B.S. degree in Electronic Engineering from Feng Chia University, Taiwan, in 1992, and the M. S. degree in Electrical and Computer Engineering from Yuan Ze Institute of Technology, and the Ph.D degree in Electrical Engineering from National Tsing Hua University, Taiwan, in 1994 and 2002, respectively. Since 2006, she has been with the Department of Computer Science and Information Engineering at National Formosa University in Taiwan, where she is now a professor. Her research interests include computer vision, multimedia security, fuzzy and neural network applications, and content-based image retrieval system. Dr. Huang is a member of the Chinese Association of Image Processing and Pattern Recognition



Ming-Hsun Tsai currently is a master degree student of Dept. Computer Science and Information Engineering at National Formosa University. His current research interests are in the area related to object recognition and segmentation, and image processing. He joins the Cloud Computing and Intelligent System Lab. (CCIS Lab.) from 2023.



Ming-Shen Jian was born in Kaohsiung City, Taiwan in 1978. He received the B.S. from the National Chiao Tung University, HsinChu, and Ph.D degrees in Computer Science and Engineering from the National Sun Yat-sen University, Kaohsiung, Taiwan in 2007. From 2018, he was an Associate Professor and director with the National Formosa University Cloud Computing and Intelligent System Laboratory. Currently he is also an IEEE Senior Member. Since 2009, he has been an Assistant Professor with the Computer Science and Information Engineering Department, National Formosa University. He is the author of four books, more than 50 articles, and at least 15 invention patents. His research interests include IOT development and application, Big Data, Optimal Solution, Intelligent System, and Cloud Computing. He was a Secretary of the Taiwan Association of Cloud Computing. Dr. Jian was a recipient of the IEEE sponsored international conference Paper Award in 2016, 2017, and 2018.