

Using the Convolution Neural Network Attempts to Match Japanese Women's Kimono and Obi

Kumiko Komizo¹, Noriaki Kuwahara²
Graduate School of Engineering and Science
Kyoto Institute of Technology
Kyoto, Japan

Abstract—Currently, the decline in kimono usage in Japan is serious. This has become an important problem for the kimono industry and kimono culture. The reason behind this lack of usage is that Japanese clothing has many strict rules attached to it. One of those difficult rules is that kimonos have status, and one must consider the proper kimono to wear depending on the place and type of event. At the same time, the obi (sash) also has status, and the status of the kimono and obi must match. The matching of the kimono and obi is called “obiawase” in Japanese, and it is not just a matter of the person wearing the kimono selecting a pair that she likes. Instead, the first place you wear a kimono determines its status, and the obi must match that status and kimono. In other words, the color, material, meaning behind the pattern must be matched with obi. Kimono patterns may evoke the seasons or a celebratory event. All this must be considered. The kimono was originally everyday wear, and people were taught these things in their households, but with today's increasingly nuclear families, that person who could teach these things isn't nearby, adding to the lack of use of kimonos. Because of this, there has been interest in using CNN (Convolution Neural Network) from the digital fashion industry. We are attempting to use machine learning to tackle the difficult task of matching an obi to a kimono, using the CNN machines drawing the most attention today.

Keywords—Digital fashion; kimono; obi; convolution neural network

I. INTRODUCTION

In Japan, the obi and kimono have changed little by little, influenced by social position, lifestyle, the tastes of the times, and fashion. In 2019, Japan began a new era called Reiwa. Since the end of World War II in 1945, it has been said that people have been wearing kimonos less and opting more for Western clothes [1]. Today in Japan, kimonos are mostly worn to celebratory occasions such as weddings and graduations. Before 1945, people in Japan wore kimonos every day, and they were free to match colors as they liked. However, with the outbreak of World War II, the wearing of slacks-style monpei pants and a government decreed national uniform began. The reasoning was that during wartime, the kimono wasn't good for taking action, wasn't functional, and wasn't reasonable [1]. That's why the other type of clothing was accepted. Also, the fact that more households started to have sewing machines spurred a Western clothes boom, and more people began to wear Western clothes [2]. However, a certain number wore kimono to graduations and other celebratory events. According to the Kyoyuzen Komon Manufacturing

Survey and Report, at the peak of manufacturing in 1971, Kyoyuzen's total production was 16,524,684, from 1972, the numbers began to decline [3]. In 2018, production reached 388,902, which is just 24% of the peak of 1971 [4]. The same is true for the obi. If the shipment value index is set at 100 for the peak, which came in 1975, then the index has dropped to 33.4 by 2014 [5]. These numbers indicate how severe the crisis is in the kimono industry. Kimonos come with a lot of strict rules. When you wear a kimono, you first have to think about status. That status includes wearing the proper kimono for a special celebration or an official event (the first formal ranking), which would be a black tomesode, or a semi-formal kimono called a homongi for women invited to a wedding, etc. Komon kimonos and yukatas are commonly worn. Obis also have status, and one must match the status of the obi and the kimono. A double-woven obi (fukuro-obi) goes with the black tomesode or the semi-formal homongi, and the Nagoya obi and share-tai obi go with a komon kimono. Also, the type of material you should wear changes with the seasons. In the height of summer, for example, you should wear cool materials such as hemp. In winter, 100% silk, which is a relatively warm material, should be worn [6]. Kimonos have different designs and patterns depending on the season and the event. For example, stripes or regular hexagons connected at the top, bottom, left, and right in a turtle-shell pattern may be worn all year round. However, designs with the morning glory, which blooms in summer, should only be worn in summer [7]. One has to think about the color combinations of the kimono and obi as well as the matching material. There are many elements to consider. In the era when the kimono was everyday wear, the parents would teach their children the rules, and the children would teach the grandchildren. But in an age when 69.6% of the women of working-age population work (2019) [8], there is no time for families to teach the rules, and it is difficult to know if one's obi and kimono match. Today, when people rarely wear kimonos and they are used less and less, the opportunity to learn these rules becomes even rarer. This is one of the reasons that Japanese people have fewer and fewer chances to wear kimono. That's why this research has focused on something that has been garnering interest in recent years, the CNN (Convolutional Neural Network) to apply machine learning to see whether a kimono and obi match or not, automatizing the process and improving the rate of correct answers, and we want to provide an environment where Japanese can wear kimonos whenever they want. Therefore, the aims of our research are to create the training data necessary to develop a system that can match

kimono and obi by using CNN and deep learning technology, and to construct a network model which will learn by using that data. In order to consider the optimal model, we will compare a simple CNN model with the model using VGG16 which has high performance in image recognition.

In this paper, we will first introduce the research trends in digital fashion with reference to some literature, and the novelty and usefulness of this research. Next, we will go on to describe the method for creating training data for judging the quality of arrangement of Japanese kimono and obi using a simple CNN or VGG16 model. Then, we will present the results of performing validation using a model trained by these data and comparing the correct answer rates, and finally our conclusion.

II. RELATED WORKS

Various fashion-related studies using artificial intelligence technology have been made [9]. For example, in [10], this study examined how semantic information extracted from clothes images using computer vision could be used to improve the user experience in online shopping. There are studies that can recommend an item set instead of one type of item. In order to recommend a set of fashion items, a tensor decomposition approach is utilized [11]. Heterogeneous graphs linking fashion items make up stylish outfits and link items to their attributes [12]. Another example of using CNN is a study that extracts styles from Amazon image sets so that they can be recommended by style rather than item category [13]. Some studies have shown that using CNNs to extract image and shape features of items provided better recommendations than text-based ones [14, 15]. On the other hand, there is a digital archiving system for Japanese kimonos [16], but there is no system that recommends Japanese kimono with appropriate obi. As mentioned above, the arrangement of Japanese kimono and obi involves the tacit knowledge of kimono experts, and there is no research that makes this possible by CNN or deep learning.

III. EXPERIMENT

A. Experiment using Images of Kimono and Obi to Determine Matches

We prepared images of 100 homongi kimonos (images contributed by Kyoto Yuzen Corporative) and 20 images of double-woven obis (images contributed by Company Kyoto Kimono Ichiba). We had three people participate by deciding whether the 20 obis matched with the 100 kimonos. The participants were one man (in his 20s) and two women (in their 40s and 50s). Examples of the kimonos are shown in Fig. 1 and examples of the obis are shown in Fig. 2. The size of the images was 391×324 pix for the kimonos, and 391×261 pix for the obis displayed on a 32-inch monitor (NEC Multi Sync Lcd V323). The participants viewed the images from 50cm away at a vertical angle of about 42° and a horizontal angle of about 70°. Furthermore, the participants viewed the images in a dark room so that lighting would not affect their view.

Fig.3 shows the screen shot of the program that the participants used for evaluation of matching kimonos and obis, and Fig. 4 shows the experimental scene.



Fig. 1. Example of the Kimono Images Shown to Participants.



Fig. 2. Example of the Double-Woven Obis Shown to Participants.



Fig. 3. Screen shot of the Program that the Participants used for Evaluation of Matching Kimonos and Obis.

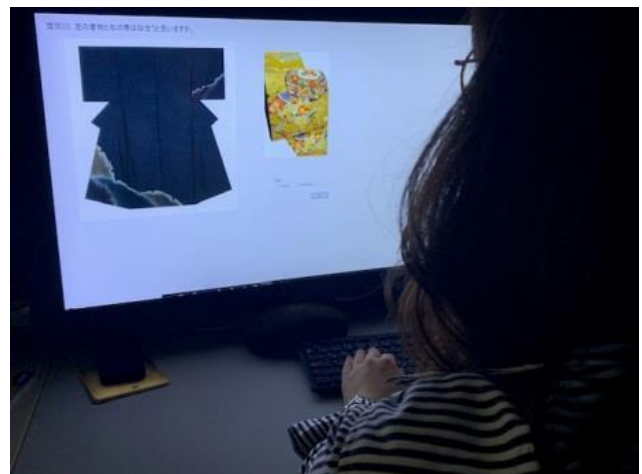


Fig. 4. Participants' view of Kimono and Obi on the Monitor.

The 100 kimono images and 20 obi images were combined into 2,000 combinations and input into the computer. The participants discussed whether the kimono and obi matched or not, and the majority opinion prevailed in this supervised learning exercise. Of the 2,000 combinations, 1,600 were chosen randomly as training data, and the remaining 400 were used as validation data. The 2,000 combinations divided into 1,600 used for training data and 400 used for validation data were randomly shuffled every time, and this process was repeated five times. The specs used this time are shown in Table 1. We used the Ubuntu 16.04 operating system and the Keras library. The GPU was GTX 1080 Ti. There were two different models for this research. The first used a simple CNN model (referred to below as CNN model). The other used the VGG16 model [17] (referred to below as VGG model). The VGG16 has a neural network that can finish learning about a convolution of more than 1 million images via the ImageNet database [18]. This network is 16 layers deep and is divided into objects by 1,000 (keyboard, mouse, pencil, types of animals, etc.).

B. Experiment with the Simple CNN Model

With the simple CNN model, we input 1,600 images of kimonos and 1,600 images of obis separately. The model structure that folded in the convolution of images in layer 1 (Conv2D) is shown in Fig. 5. The input was for both kimono images and obi images that were sorted into a convolution layer and a MaxPooling layer and applied to all bonded layers. They were bonded with the concatenate functions and sandwiched in with the other layers. Softmax was applied and a final output resulted. Furthermore, after each CNN layer and after the bonding of the 3rd layer with the other layers, the activation function ReLU was used in between the bonding of the third layer and the activation function and after the bonding of the 4th layer, Batch Normalization function was set. For the above model, Adam was used as an optimizer, and we used categorical cross-entropy for the loss function, setting the epoch number at 100 times. We repeated this series of process five times.

The format for each layer in Fig. 5 is shown in Table 2. Also, in Fig. 5, to be concise, we abbreviated the activation function to Batch Normalization.

TABLE I. THE CALCULATORS USED

Calculator	DeepStation DK-1000 Personal Edition
Operating System	Ubuntu 16.04
Library	Keras
GPU	GTX1080 Ti

TABLE II. FORMAT OF EACH LAYER

Input	Input([Vertical]×[Width]×[Channel])
Convolution Layer	Conv[kernel size]-[output]-[padding]
MaxPooling Layer	MaxPooling[pooling size]
Fully Bonded Layers	Dense[unit]
Dropout Layer	Dropout([Dropout ration])
Output	Output[No. of categories]

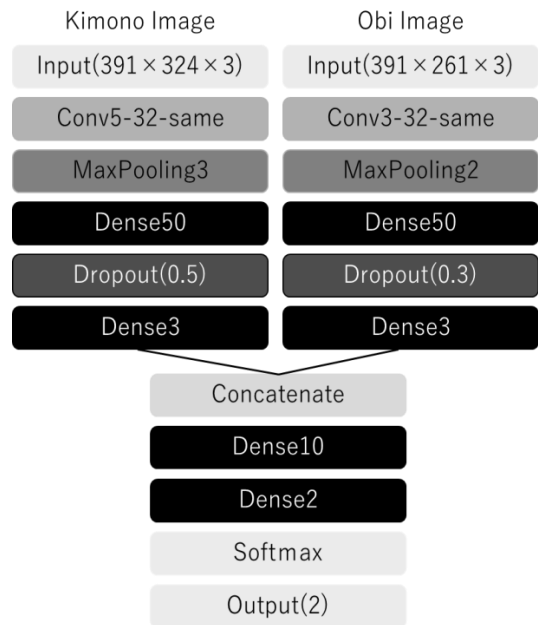


Fig. 5. Simple CNN Model Map.

C. Experiment using VGG16 Model

VGG is the model where the convolution layer and pooling layer are configured. VGG16 is when the convolution layer and other bonded layers are layered 16 deep. The unit model for VGG16 is shown in Fig. 6.

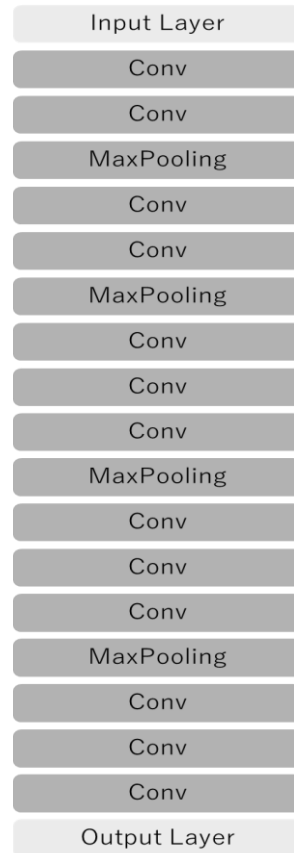


Fig. 6. Structure of VGG16.

As with the CNN model, a total of 2,000 combinations were used from the input data of 100 images of kimonos and the 20 images of obis. The participants discussed whether the kimono and obi matched or not, and the majority opinion prevailed in this supervised learning exercise. Of the 2,000 combinations, 1,600 were chosen randomly as training data, and the remaining 400 were used as validation data. The 2,000 combinations divided into 1,600 used for training data and 400 used for validation data were randomly shuffled every time, and this process was repeated five times.

Then, we input 1,600 images of kimonos and 1,600 images of obis separately and sent them to the VGG model. At this point, we used the learning convolution parameters of ImageNet. The learning went to the 15th and later layers, a process known as transfer learning. After that, all the layers were bonded with concatenate functions, Softmax was applied and a final output resulted. Furthermore, after the 3rd and 4th bonded layers were equipped with Batch Normalization, we used the activation function ReLU, which was set after the bonding of the 3rd layer and after Batch Normalization was set. SGD was used as an optimizer on the above model and the learning rate was set to 0.0001. We used categorical cross-entropy for the loss function. After a number of trials, about 50 epochs gave us the peak training results, so we set the epoch number to 50 times. We repeated this series of process five times. The formation of the different layers shown in Fig. 7 is the same as Table 2.

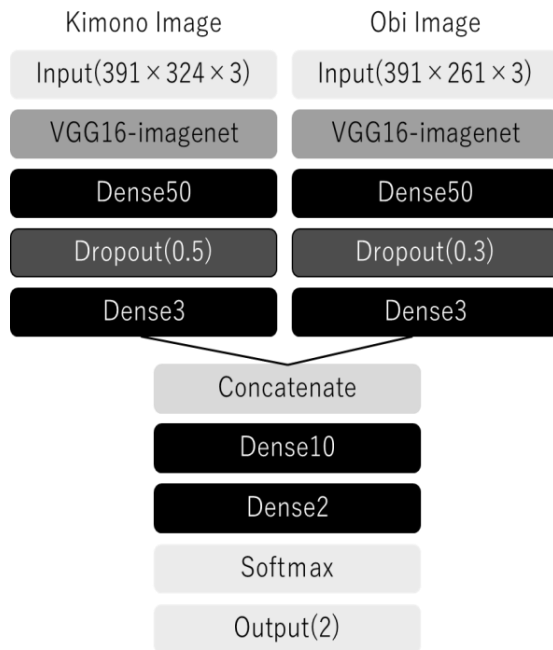


Fig. 7. Model using VGG16.

IV. RESULT

A. Results with the Simple CNN Model

Fig. 8 is a graph showing the transition of the accuracy rate when training data and validation data were used once. The results for the CNN model using just one layer unit is shown in Table 3. Using 100 epochs repeated five times, the average was 75.4%.

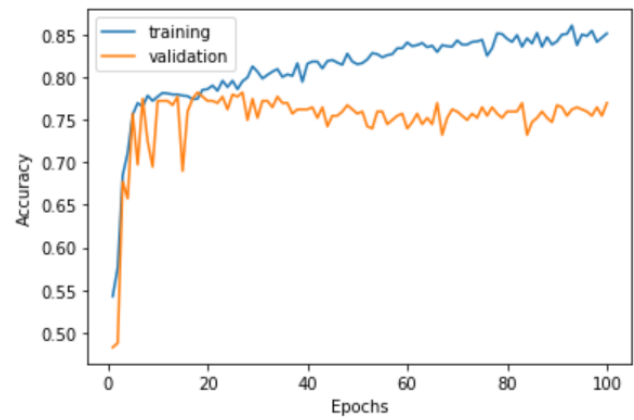


Fig. 8. Transition of Accuracy Rate for One Time of Training and Validation.

TABLE. III. RESULTS FROM THE SIMPLE CNN MODEL

	Accuracy Rate
1 time	77.0%
2 times	74.3%
3 times	78.3%
4 times	70.0%
5 times	77.3%
Average	75.4%

B. Results from using the VGG Model

Fig. 9 shows the transition of the accuracy rate on the fourth attempt of five samples using training data and test data. The results from the VGG model are shown in Table 4. When 50 epochs are repeated 5 times, the average is 76%.

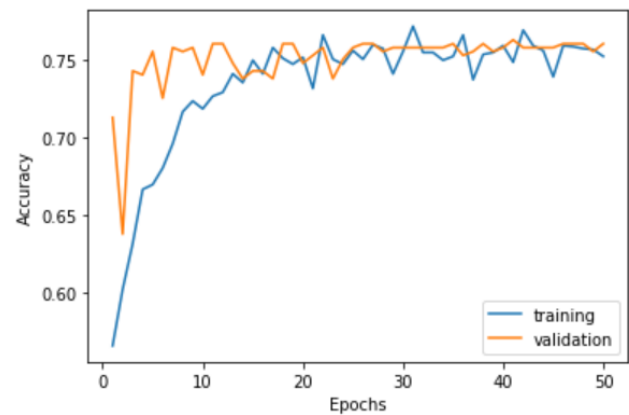


Fig. 9. Transition of the Accuracy Rate on the 4th Time using Training and Validation Data.

TABLE. IV. RESULTS FROM THE VGG16 MODEL

	Accuracy Rate
1 time	74.3%
2 times	78.0%
3 times	77.5%
4 times	76.0%
5 times	74.0%
Average	76.0%

V. DISCUSSION

This time, we aimed to automate the arrangement of Japanese kimono and obi by using CNN and Deep Learning technologies. Then, we compared the average accuracy rates of the simple CNN model that has only one CNN layer and the model utilizing VGG model with each model repeating the epoch numbers five times using cross validation. The results showed that the simple CNN model averaged 75.4%, and the VGG16 model averaged 76.0%. The difference is 0.6%.

We hypothesized that the VGG model would produce far greater results, but that did not happen. This time, with a simple task of deciding whether a kimono and obi were a good match or not, the CNN model using just one layer averaged more than 75%, a perfectly acceptable result. It can be seen that there is no reason to use the many CNN layers available to VGG16. However, as shown in Fig. 7, there is a distinct gap in the CNN model between the accuracy rate of the training data and the validation data. This indicates the possibility of overtraining.

Consequently, we were able to research the accuracy rate when deciding whether obis matched certain kimonos, but deciding on the right kimono attire is not as simple as just whether the obi and kimono match. Actually, there are many rules for deciding which obis match which kimonos. The most important is the status of the kimono. In the long history of the kimono, this is something that started being said in 1976 [19]. Today, even though there are various styles for wearing a kimono and people express themselves freely with their clothing, conventional kimono rules cannot be ignored. Especially in public places and events, some people will continue to honor these rules. When we asked kimono experts about what to consider when matching kimonos and obis, they always said "status" first [20], but for average people who don't have deep knowledge of the kimono, it's clear that they focus only on color.

In other words, it became clear that they only focused on the compatibility of colors. In this research, when we created the supervised learning exercise, of our participants, the one man had no knowledge of the kimono, and he said in an interview that he only considered colors when deciding whether a kimono and obi go together. However, the two female participants said that they rejected his decision which had ignored the rules of matching kimono and obi. That said, this research showed that both the CNN model and the VGG model produced accuracy rates of higher than 75%, and these results might reflect not only the color combination but also other rules such as the status, texture, the seasons, and the meaning of the patterns of these items. This means that the tacit knowledge of the kimono experts can be implement by using these technologies.

VI. CONCLUSION

We attempted to build a system that automates the arrangement of Japanese kimono and obi using CNN and deep learning techniques. To this end, we created the original training data. In preparing the training data, considering the rules of matching kimono and obi, and training the model with the data, it was possible to determine whether or not the arrangement was appropriate with an accuracy of about 75% using both techniques.

As we move further in this research, we proceeded with our research only using the learning convolution parameters on the 15th and later layers based on the learning convolution parameters of the VGG16. Using VGG16 ImageNet as a foundation, we will look at whether the accuracy rate can be improved through learning when using kimono and obi images on all layers. There is also potential for the simple CNN model to over-train, so we would also like to expand our data and see if we can validate an improved accuracy rate.

The results of this research as related to Internet business shows that it would be possible to help customers with no kimono knowledge by recommending the right combinations. Also, there could be an application that examines a photo of your kimono and obi and tells whether they match or not, giving the kimono wearer more confidence. There are various ways to wear kimono, and people express themselves freely through their clothes. But even today, we can't ignore the rules. Especially at public places and celebratory events, things will change a little as we go on, but traditions will be handed down. We hope that this research will help preserve the kimono culture.

REFERENCE

- [1] Kazuko Koizumi, Kimono of the showa era, KAWADE SHOBO SHINSHA, 2006 (in Japanese).
- [2] Masuda Yoshiko, Japanese clothing history, Yoshikawa Kobunkan, 2010 (in Japanese).
- [3] Kyoyuzen Komon Manufacturing Survey and Report, Kyoto Yuzen Cooperative, 2016 (in Japanese).
- [4] Kyoyuzen Komon Manufacturing Survey and Report, Kyoto Yuzen Cooperative, 2019 (in Japanese).
- [5] Outline of Nishijin machinery industry survey, The 21st Nishijin Machinery Industry Investigation Committee, 2016 (in Japanese).
- [6] Analysis of the thought process for choosing suitable Kimono for Customer by the expert, Kumiko Komizo, Takashi Furukawa Noriaki Kuwahara, Kazunari Morimoto, LNCS, Vol. 9745, pp. 559-567, 2016.
- [7] Kenzo Fujii, Kimono pattern, SEKAIBUNKA PUBLISHING INC., 2009 (in Japanese).
- [8] White Paper on Gender Equality 2019, Gender Equality Bureau, Cabinet Office, 2019, ISBN-10: 4906955959 (in Japanese).
- [9] Haosha Wang and Khaled Rasheed, Artificial intelligence in clothing fashion. In Proc. on the International Conference on Artificial Intelligence (ICAI), 2014.

- [10] Zequn Li, Honglei Li, and Ling Shao, Improving Online Customer Shopping Experience with Computer Vision and Machine Learning Methods. In Proc. of International Conference on HCI in Business, Government and Organizations. pp.427-436, 2016.
- [11] Yang Hu, Xi Yi, and Larry S Davis. Collaborative fashion recommendation: A functional tensor factorization approach. In Proc. of the 23rd ACM international conference on Multimedia, pp.129-138, 2015.
- [12] Hanbit Lee and Sang-goo Lee, Style Recommendation for Fashion Items using Heterogeneous Information Network, In Proc. of Recommender Systems 2015.
- [13] Qiang Liu, Shu Wu, and Liang Wang, DeepStyle: Learning User Preferences for Visual Recommendation. In Proc. of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.841-844, 2017.
- [14] Satya Keerthi Gorripati and Anupama Angadi, Visual Based Fashion Clothes Recommendation with Convolutional Neural Networks, International Journal of Information Systems & Management Science, Vol. 1, No. 1, 2018.
- [15] Ryosuke Saga and Yufeng Duan, Apparel Goods Recommender System Based on Image Shape Features Extracted by a CNN, Proc. of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), DOI: 10.1109/SMC.2018.00406.
- [16] Haruko Takahashi, An Image Digital Archive for Substantiating the Acculturation of Clothing Culture in Japan, Proc. of 2015 International Conference on Culture and Computing (Culture Computing), DOI: 10.1109/ Culture.and.Computing.2015.13.
- [17] Simonyan, Karen and Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv: 1409.1556, 2014.
- [18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In Proc. CVPR, pp. 2-9, 2009.
- [19] Motoshige Ito and Yajima Takatoshi, Kimono culture and Japan, Nikkei Publishing Inc., 2016 (in Japanese).
- [20] Choosing an Obi Suitable for Kimonos Eliciting a Sense of Hannari - From a color-based perspective -, Kumiko Komizo, Noriaki Kuwahara, The Journal of Ergonomic Technology, Vol.19, No.1, pp.45-56, 2019.