공 학 석 사 학 위 논 문

Deep Learning based on Meat Cooking Support for Color Vision Disorder Compensation

2024년 2월

부 경 대 학 교 대 학 원

인 공 지 능 융 합 학 과

Shota Chiba

공 학 석 사 학 위 논 문

Deep Learning based on Meat Cooking Support for Color Vision Disorder Compensation

지도교수 장 원 두

이 논문을 공학석사 학위논문으로 제출함.

2024년 2월

부 경 대 학 교 대 학 원

인 공 지 능 융 합 학 과

Shota Chiba

Shota Chiba의 공학석사 학위논문을 인준함.

2024년 2월 16일

위 원 장	공학박사	정 완 영	(인)
위 원	공학박사	김 종 남	(인)
위 원	공학박사	장 원 두	(인)

Table of Contents

Deep Learning based on Meat Cooking Support for Color Vision Disorder Compensation

Shota Chiba

Department of Artificial Intelligence Convergence, The Graduate School,

Pukyong National University

Abstract

People who suffer from color vision disorder (CVD) have difficulties recognizing colors and have limited capability to cook the meat, i.e., could not judge whether the meat is undercooked, or overcooked. Most CVD compensation approaches proposed in the existing studies focused on the problem of contrast loss experienced by CVD individuals, and these methods are not applicable to the meat-cooking task; while some studies proposed for this task may lose effect when the lighting condition changes. In this study, we propose a system for CVD support, which automatically determines the degree of cooking utilizing deep learning model and presents generated support information to CVD individuals using augmented-reality (AR) glasses. To enable the capability of the proposed system to perform the prediction, we create a meat-cooking dataset in this study. To confirm the effectiveness of the proposed method, quantitative and subjective evaluation experiments were conducted. The experimental results showed that the meat cooked with assistance of the proposed is better than those without the system.

색시력 장애 보상을 위한 딥러닝 기반 육류 요리 지원

Shota Chiba

부경대학교 대학원 인공지능융합학과

요약

색각 장애(CVD)를 앓고 있는 사람들은 색을 인식하는 데 어려움이 있고 고기를 요리할 수 있는 능력이 제한되어 있습니다. 즉, 고기가 덜 익었는지 또는 너무 익었는지 여부를 판단할 수 없습니다. 기존 연구에서 제안된 대부분의 CVD 보상 접근 방식은 CVD 개인이 경험하는 조영 손실 문제에 초점을 맞추었고, 이러한 방법은 육류 요리 작업에 적용할 수 없으며, 이 작업에 대해 제안된 일부 연구는 조명 조건이 변경되면 효과가 떨어질 수 있습니다. 본 연구에서는 딥러닝 모델을 활용하여 요리 정도를 자동으로 결정하고 증강 현실(AR) 안경을 사용하여 CVD 개인에게 생성된 지원 정보를 제공하는 CVD 지원 시스템을 제안합니다. 제안된 시스템의 예측 기능을 수행할 수 있도록 본 연구에서는 육류 요리 데이터 세트를 만듭니다. 제안된 방법의 효과를 확인하기 위해 정량적 및 주관적 평가 실험을 수행했습니다. 실험 결과는 제안된 방법의 도움을 받아 요리된 고기가 시스템이 없는 것보다 더 나은 것으로 나타났습니다.

I. Introduction

1. Research Background

The retina is one of the most important tissues in the eye. As shown in Figure 1, the retina contains two types of photoreceptors: rod cells and cone cells. Rods are sensitive to low-intensity light, while cones work in bright environment to distinguish colors and shapes. Cones can be further divided into three types: L-, M-, and S- cones, which are most sensitive to the wavelengths of light corresponding to red, green, and blue colors. The response of these three types of cones enables humans to distinguish different colors.

However, if one or more types of cones become abnormal, individuals cannot perceive colors like those with normal color vision. Such a symptom is so-called Color Vision Disorder (CVD). According to the report by Sharpe et al. [1], the incidences of CVD are 8% and 0.42% in males and females, respectively.

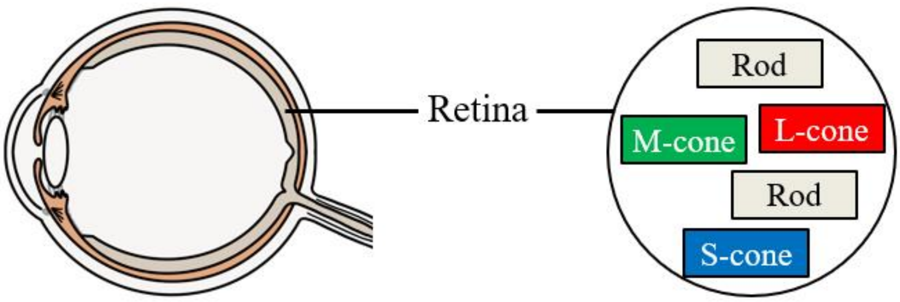


Fig. 1. Eye Structure

People with CVD have a reduced ability to distinguish colors and face the problem of color contrast loss. At the same time, they may get in trouble with the loss of semantic information. For example, people with CVD may be confused when they cook. People with normal color vision can infer the cooking degree of meat from its color appearance; however, it is not easy for people with CVD to complete this task. Fig. 2 shows an example of the meat-cooking. It can be found from Fig. 2(a) that it is easy for a person with normal color vision to distinguish the raw meat on the right side. Fig. 2(b) shows a simulated CVD's perception using a CVD simulation model [2]; in Fig. 2(b), it can be seen that there are almost no difference between well-done meat on the left side and the undercooked one on the right. Therefore, it is difficult for people with CVD to find an appropriate timing according to the color appearance of meat when they cook meat by themselves. In other words, if the meat is undercooked, it could be harmful to the health (e.g., upset stomach); on the other hand, if the cooking time is too long, the meat can be overcooked and does not taste good.

a) Normal Color Vision Perception b) Color Vision Disorder Perception

Fig. 2. Differences between normal vision and color vision disorder

2. Research Objectives

In this study, we aim to address the aforementioned issues by developing a system that allows individuals with CVD to engage in meat cooking. Actually, in the existing studies, a significant number of CVD compensation methods have been proposed; however, most of them compensate for the loss of color contrast. Zhu et al. [3] investigated image recoloring approaches for contrast compensation proposed in the past two decades. Although the assistive content generated by the image recoloring makes it easier to discriminate different colors, it fails to make CVD recognize the true colors. Although some assisting methods have been proposed for the meat-cooking, none of them can be adapted or used under different lighting conditions. On the other hand, object detection methods based on deep learning models have attracted much attention in the field of computer vision because of their high performance. In this study, we first divide meat with different cooking degrees into categories, such as raw, medium, and well-done; it's feasible for us to adopt object detection methods to automatically predict the cooking degree of meat. With a completed cooking degree prediction model, it is possible to utilize smartphone as a user interface for showing support information to CVD users. Moreover, Augmented-Reality (AR) glasses are wearable devices that can directly superimpose annotations (e.g., text, virtual objects, etc.) on the user's field of view. At the same time, user can utilize an AR glasses in a hand-free style.

3. Research Contents

In this study, we propose a meat-cooking assistance system based on deep learning model and AR glasses for CVD support. First, a deep learning model for object detection is adopted into the proposed system to predict the degree of cooking. To enable the model to complete the prediction task, a dataset for meat-cooking is constructed in this study. Furthermore, we propose a score formula to quantify the degree of meat based on the prediction results of the model. We also introduce AR glasses into the proposed system as an interface between the user and the deep learning model. Finally, quantitative and subjective evaluation experiments were conducted to demonstrate the effectiveness of the proposed system. The experimental results confirmed the effectiveness of the proposed system.

II. Related Works

1. CVD Support System

Based on the CVD simulation model [2, 4], image recoloring algorithms have been proposed to compensate for contrast loss. The survey paper [5] summarized the latest image recoloring algorithms for CVD compensation. The state-of-the-art image recoloring algorithm of Zhu et al. [3] considers the compensation for color contrast loss and the preservation of naturalness, and can be adapted to different degrees of CVD. However, these methods cannot provide effective information for discriminating whether meat is well-done or not.

Hishikawa et al. developed an application to judge the cooking degree of meat for people with CVD [6]. In this application, the LMS color space is first divided into four parts, each of which is classified as raw meat, undercooked, well-done, and overcooked, respectively. The application takes a picture of the meat with a camera and predicts the degree of cooking based on the average LM value of the meat within a specific frame. Before use this application, the lighting environment must be calibrated and fixed from then on. In addition, this application requires the meat to be located within a specific frame on the photo, and it cannot predict the cooking degree of multiple pieces of meat at once.

Tanuwidjaja et al. developed Chroma [7], a system that uses Google Glass to inform users about colors that are difficult to distinguish in scenes faced by people with CVD. When using Chroma to check whether meat is cooked or not, the user can set the red and pink colors to be highlighted, and if any part of the meat is highlighted, it means that the meat is not fully cooked. One problem with this system is that the set colors may not be highlighted correctly depending on the lighting environment.

2. Object Detection

YOLO [8], Faster R-CNN [9], and Mask R-CNN [10] are representative object detection methods based on deep neural networks. YOLO [8] and Faster R-CNN [9] predict the class names of objects and their locations in input images by bounding boxes. YOLO [8] is one-stage model that directly predicts class name and location of object from the input image; while Faster R-CNN [9] is two-stage model, which first extracts image features from the input image using Convolutional Neural Network (CNN) layers, and then predicts objects from the extracted image features. In addition to class name and bounding box, Mask R-CNN [10] further uses binary mask to segment object instance from the input image. And Mask R-CNN [10] achieves higher performance comparing to YOLO [8] and Faster-R-CNN [9].

III. Proposed Method

The framework of the proposed system is shown in Fig. 3. First, the user is asked to wear a HoloLens 2 on his head. The image captured by the camera on HoloLens 2 is sent over the network to the server. The server detects the meat and predict the cooking degree of meats within the captured image, and sends the results back to the HoloLens 2 and presents them to the user.

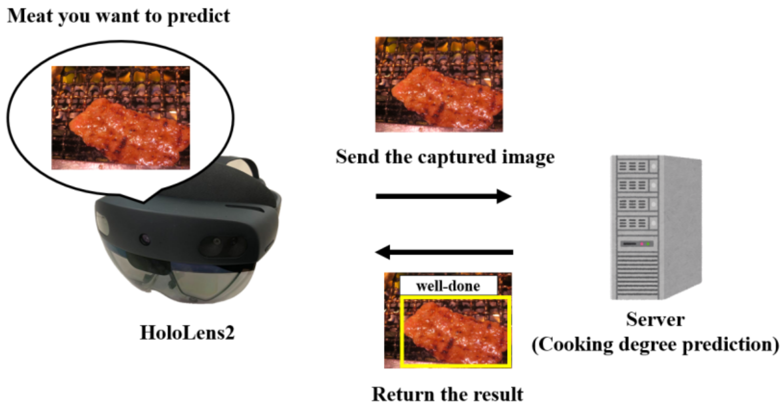


Fig. 3. Configuration of the proposed system and process

In the following section 1, we describe the Mask R-CNN [10] used for meat detection and cooking degree prediction. Section 2 describes the dataset prepared to enable the cooking degree prediction model. Section 3 describes the details of training the model using the dataset in Section 2. The final section, Section 4, describes the approach of presenting information to the user.

1. Deep Learning based Cooking Degree Prediction Model

In this study, Mask R-CNN [10] was used for meat detection and cooking degree prediction. The network structure of Mask R-CNN [10] is shown in Fig. 4. The input is an image, and the output are class label and its probability, and bounding box and mask for object location. The input image is first passed through a convolutional neural network (CNN) backbone to extract image features. Next, the extracted image features are passed through the Region Proposal Network to predict potential object regions. Finally, classes, probabilities, bounding boxes, and masks are predicted from the predicted candidate object regions using Fully Connected Layers and CNN, respectively. In this study, we use ResNeXt-101-FPN [11] as the CNN backbone. As loss functions, classification loss and mask loss use cross-entropy loss, and bounding box loss uses L1 loss. To give this deep learning model the ability to detect meat and predict the cooking degree, a meat cooking dataset is required.

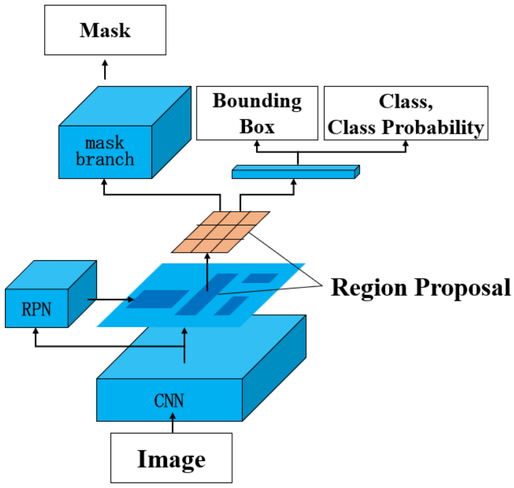


Fig. 4. Network structure of Mask R-CNN

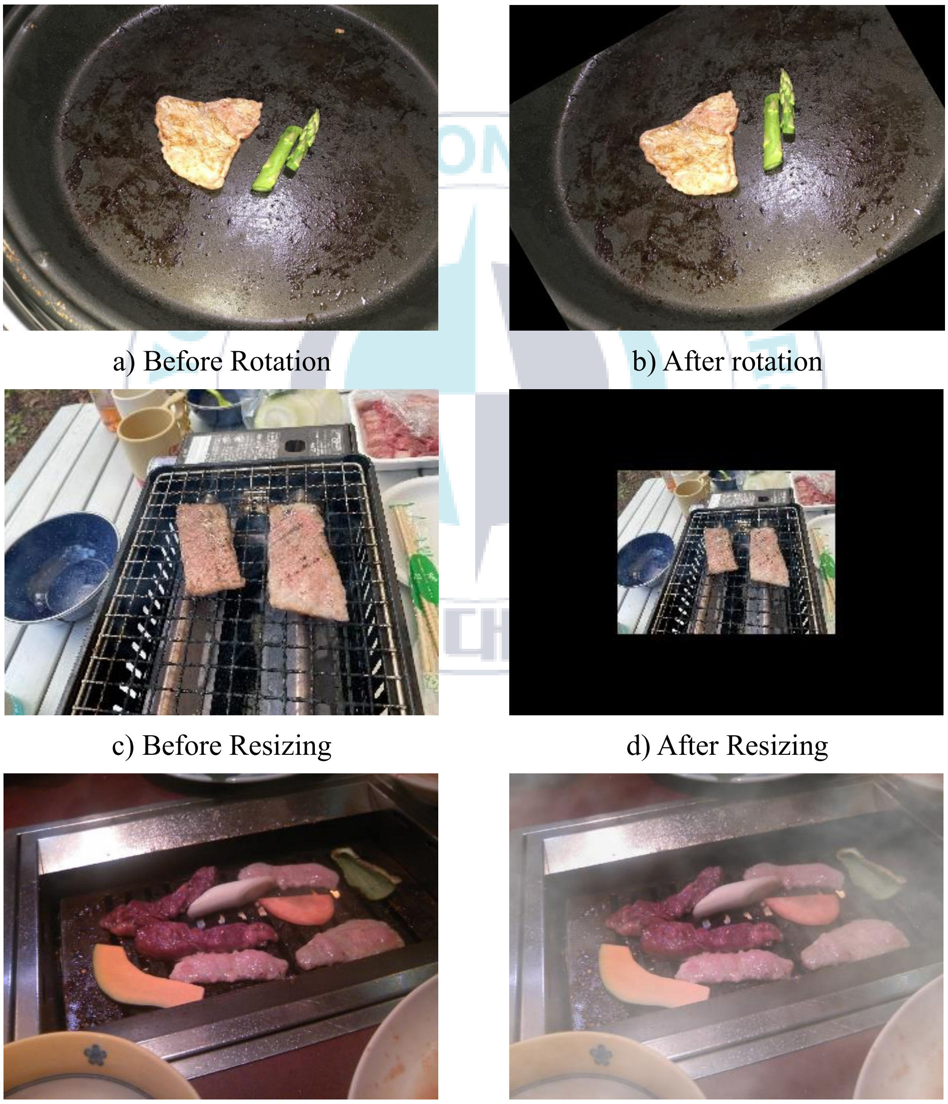
2. Meat Cooking Dataset

In order to enable a model to predict the cooking degree of meat, we collect meat image data as there are no dataset currently for this task. This study focused on beef, and we collected a total of 875 images of beef with varying cooking degrees of meat. The dataset created included data taken in a variety of environments, such as images taken indoors or outdoors while being cooking, and images taken during the day or at night. After the image data were collected, labels of meat images were annotated by a volunteer with normal vision. One label corresponds to one piece of meat. For detail, the meat was classified into the following three classes according to the degree of cooking.

- raw beef (Number of labels : 1907)
- medium beef (Number of labels : 756)
- well-done beef (Number of labels : 1875)

In this study, the collected dataset was divided into training data and validation data, which contains 700 and 175 images, respectively.

Data augmentation was performed on the training data to improve the performance of the model. The data augmentation approaches included image rotation, resizing, and smoke effects rendering. As a result of the data augmentation, there are 1652 images in the training set. Fig. 5 shows examples of data augmentation (Fig. 5(a) before rotation, Fig. 5(b) after rotation, Fig. 5(c) before image resizing, Fig. 5(d) after image resizing, Fig. 5(e) before smoke effect, and Fig. 5(f) after smoke effect).



a) Before Rotation

b) After rotation

c) Before Resizing

d) After Resizing

e) Before Smoke Effect

f) After Smoke Effect

Fig. 5. Example of Data Augmentation

3. Training

To verify the effectiveness of data augmentation, the following two datasets were used to train the model, respectively.

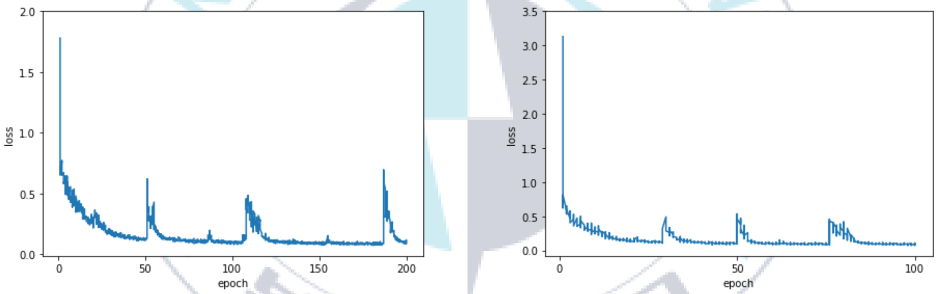
- Dataset before data augmentation

(train images : 700, validation images : 175)

- Dataset after data augmentation

(train images : 1652, validation images : 175)

Model learning was performed until the loss converged. Finally, the loss converged at 150 epochs before the data augmentation and at 60 epochs after the data augmentation. Fig. 6 shows plots of the loss values per epoch before data augmentation (Fig. 6(a)) and after data augmentation (Fig. 6(b)).



a) Before Data Augmentation b) After Data Augmentation

Fig. 6. Loss trends by epoch

Examples of detections for the validation data before and after data augmentation are shown in Fig. 7. Meat surrounded by a red bounding box indicates that the prediction is correct, while those meat surrounded by blue bounding boxes indicates that the prediction result is incorrect. Fig. 7(a) and 7(b) show that before image enhancement, the prediction result for one of the seven meats was wrong, but after image enhancement, the correct prediction result was obtained for the meat for which the prediction result was wrong. Fig. 7(c) and 6(d) show that before image enhancement, the prediction result was wrong for one of the four meats, but after image enhancement, the correct prediction result was obtained for the meat for which the prediction result was wrong. Fig. 7(e) and 7(f) show that before image enhancement, the prediction results were wrong for all of the two pieces of meat, but after image enhancement, the correct prediction results were obtained for all of the pieces of meat.

a) Before Data Augmentation 1　　b) After Data Augmentation 1

c) Before Data Augmentation 2　　d) After Data Augmentation 2

e) Before Data Augmentation 3　　f) After Data Augmentation 3

Fig. 7. Example of detection for validation images

4. Information Presentation

Although it is possible to present meat labels such as "raw" and "medium" directly to the user, the information is difficult to read. In this study, we quantize the cooking degree of meat using a score s, which is calculated using the following equations:

$$s = \frac{(1 \times P_{raw}) + (2.5 \times P_{med}) + (5 \times P_{well})}{P_{raw} + P_{med} + P_{well}} \tag{1}$$

$$score = 2s - 5 \qquad (2)$$

where P_{raw}, P_{med} and P_{well} are the class probabilities that the meat is predicted to be raw, medium, and well done, respectively, and a higher score value indicates that the meat is well done. Finally, the predicted location of the meat and the score values are returned to HoloLens 2. To simplify the information presented on HoloLens 2, those bounding boxes for meats whose scores are less than 3 were hidden, which is considered as undercooked meat. When both sides of meat are predicted with scores near 5, then user can pick the piece of meat to eat.

IV. Evaluation

To confirm the effectiveness of the proposed system, quantitative and subjective evaluation experiments were conducted.

1. Quantitative Evaluation

The evaluation is performed by comparing the value of each evaluation metric for the validation data. Precision, recall, and average precision (AP, area under Precision-Recall curve) were used as the evaluation metrics. Each evaluation metric is expressed by the following equations (3) and (4). Predictions were considered correct if the IoU (equation (5)) between the correct label and the predicted bounding box was greater than 0.5.

$$Precision = TP / (TP+FP) \tag{3}$$

$$Recall = TP / (TP+FN) \tag{4}$$

$$IoU = Area\ of\ Intersection / Area\ of\ Union \tag{5}$$

The results of the quantitative assessment are shown in Table I. Table I shows that the bbox AP_{50} increased for all classes after the data augmentation. However, the values of Precision and Recall of the classification increased in some classes and decreased in others. We also see that the MEDIUM class is lower than the other two classes in all evaluation scales. One possible reason is the small number of data collected: the MEDIUM class has less than half as many labels as the other two classes, and there is a bias in the number of data between the classes. Another reason is that the medium class is located between the raw and well-done classes, making it difficult to classify.

Table I. Model Evaluation Results

		raw	medium	well-done
Before Data Augmentation (150 epoch)	bbox AP_{50}	0.879	0.648	0.846
	Classification Precision	**0.967**	0.605	**0.918**
	Classification Recall	0.865	**0.649**	0.863
After Data Augmentation (60 epoch)	bbox AP_{50}	**0.897**	**0.672**	**0.849**
	Classification Precision	0.940	**0.712**	0.916
	Classification Recall	**0.892**	0.604	**0.905**

2. Subjective Evaluation

Next, we describe the details of the subjective experiment. Four volunteers (males in their 20s to 50s) with CVD participated in our experiment. Before the experiment, the volunteers were given a color vision test to determine the degree and type of symptoms. The color vision tests were the Ishihara Color Vision Test Chart, Farns-Munsell Panel D-15, and the Farns-Munsell 100 Hue Test. The Ishihara Color Vision Test Chart was a test in which participants were asked to answer the numbers in a chart. Panel D-15 and the 100 Hue Test consisted of a test in which participants were asked to line up the caps on a color chart. The test results indicated that two subjects are with protan disorder and two are with deutan disorder.

The HoloLens 2 used in the proposed system displays a score and a bounding box for meat with a score equal to or higher than 3, indicating the degree of cooking of the meat. Fig. 8 shows an example of cooking degree prediction using the proposed system.

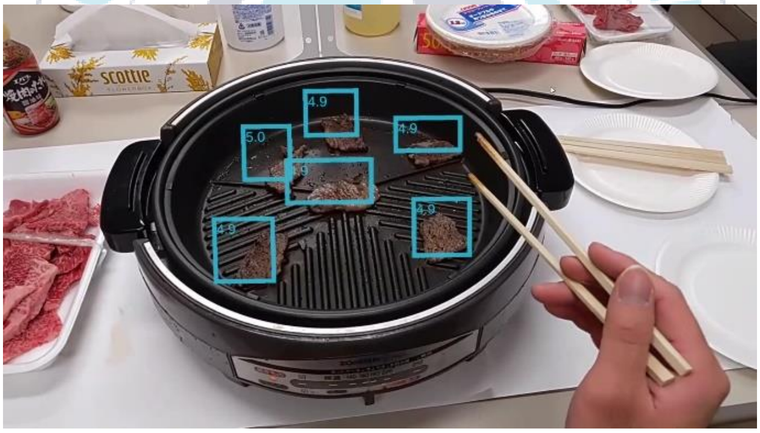


Fig. 8. Example of Proposed System Operation

The procedure for the evaluation experiment is as follows.

Step 1:

- Without the proposed system, a participate cook two pieces of meat by himself and pick them up when he thinks the meat is well-done.

Step 2:

- With the proposed system, a participant cook two pieces of meat and pick them up when the score reaches about 5.

- Repeat steps 1 and 2 for three times.

We also asked the participants to rate the following items on a 5-point scale.

- Deliciousness of each piece of meat
- Reliability for cooked meats with the proposed system

Delicious was rated as 1 for not delicious, 3 for neither, and 5 for delicious. For reliability, 1 was unreliable, 3 was neither, and 5 was reliable. The delicious ratings were conducted after the participants had eaten each piece of meat. For the reliability rating, participants were asked to rate the reliability of the cooked meat without and with the proposed system at the end of the experiment.

The results of the subjective evaluation experiment are shown in Fig. 9. The evaluation without the proposed system is shown by blue bars, while the evaluation with the proposed system is shown by red bars. As can be seen from Fig. 9, both the average deliciousness and the average reliability increased with the proposed system. However, when we look at the evaluation values of taste and reliability for Volunteer 1 and the evaluation value of taste for Volunteer 3, the results show that the evaluation with the proposed system was lower than that without the proposed system. When we asked the reason for the lower evaluation, we received the opinion that the meat cooked with the proposed system was still insufficient in terms of the degree of cooked meat. As for the degree of reliability, the participants said that they usually take longer time when cooking meat by themselves, so when they took the meat as instructed by the system, the cooking time was shorter than their usual experience, and the degree of cooked meat seemed a little insufficient, so the degree of reliability decreased. These results indicate that although the proposed system can suggest whether the meat is well-done, it does not necessarily mean that the user can cook the meat, which is consistent with their personal preferences.

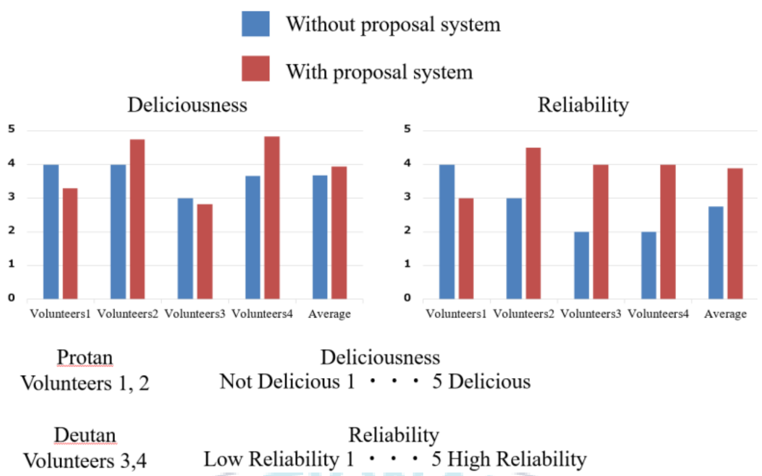


Fig. 9. Example of Proposed System Operation

V. Conclusion

This study proposed a meat cooking support system for people with CVD. A meat-cooking dataset was also collected in this study. Quantitative and subjective evaluation experiments were conducted to confirm the effectiveness of the proposed system.

For the future work, detecting a certain cooking degree of meat, which satisfies the preference of individual user, is important. At the same time, extending the proposed system to predict those meats other than beef is also important.

References

[1] L. T. Sharpe, A. Stockman, H. Jägle and J. Nathans, "Opsin genes cone photopigments color vision and color blindness," in Color Vision: From Genes to Perception, pp. 3-51, 1999.

[2] H. Brettel, F. Viénot and J. D. Mollon, "Computerized simulation of color appearance for dichromats,". Journal of the Optical Society of America A, vol. 14, no.10, p.2647-2655, 1997.

[3] Z. Zhu, M. Toyoura, K. Go, K. Kashiwagi, I. Fujishiro, T-T. Wong and X. Mao, "Personalized Image Recoloring for Color Vision Deficiency Compensation," IEEE Transactions on Multimedia, vol. 24, pp. 1721-1734, 2021. https://doi.org/10.1109/TMM.2021.3070108

[4] G. M. Machado, M. M. Oliveira, and L. A. F. Fernandes, "A Physiologically-based Model for Simulation of Color Vision Deficiency," IEEE Transactions on Visualization and Computer Graphics 15, no. 6, pp. 1291-1298, 2009.

[5] Z. Zhu and X. Mao, "Image recoloring for color vision deficiency compensation: a survey," The Visual Computer 37, pp. 2999–3018, 2021.

[6] Y. Hishikawa, S. Katsura and S. Sunaga, "Development of an application for dichromats judging the doneness of meat based on color changes of Yakiniku," Journal of the color science association of japan, vol. 41, no. 3, pp. 141-144, 2017.

[7] E. Tanuwidjaja, D. Huynh, K. Koa, C. Nguyen, C. Shao, P. Torbett, C. Emmenegger and N. Weibel, "Chroma: A Wearable Augmented-Reality Solution for Color-Blindness," UbiComp '14, pp. 799-810, 2014.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," In CVPR, pp. 779-788, 2016.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," In NIPS, 2015.

[10] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," In proceedings of the IEEE International Conference on Computer Vision, pp. 2961-2969, 2017.

[11] S. Xie, R. Girshick, P. Dollár, Z. Tu and K. He, "Aggregated Residual Transformations for Deep Neural Networks," In CVPR, pp. 5987-5995, 2017.