Food plating referencer: A Visual Plateware Selection System Based on Image Masking and Similarity Search

Risa Takahashi[1] [0009−0009−2615−0502] and
Mitsunori Matsushita[2] [0000−0003−0174−7993]

[1] Kansai University, Osaka, Japan
k246655@kansai-u.ac.jp
[2] Kansai University, Osaka, Japan
m_mat@kansai-u.ac.jp

Abstract. Visually appealing food presentation enriches the dining experience. The impression a dish (in this paper "dish" is used exclusively to denote plated food) creates is dramatically influenced by the food, its plating, the tableware, and other furnishings. This research aims to realize visually enjoyable dining by diversely enhancing dishes through various plating methods on a selection of crockery. The proposed system takes user-owned or store-sold plates as "plates accessible to users" queries and displays food images utilizing similar plates. Conversely, specific foods can be used as queries to display images of dishes used the plate similar to those accessible to users. From this, users would gain inspiration and plating suggestions for combining menu items with plates accessible to them. The system includes pairs of "dish images" and "images with only the extracted plate portions of those dishes" as its dataset. Based on color information, the query plate image is used for a similarity search among the plate images in the dataset. This identifies images of dishes with plates similar to the query plate. The dataset's images have the plates' centers removed, whereas the query plate images do not. We conducted preliminary experiments on whether this difference affects similarity calculations. Based on those results, this paper verified whether computational similarity evaluation matches human similarity evaluation as a first step toward measuring the system's usefulness.

Keywords: Dish · Plate · Plating · Dining · Quality of Life · Similarity Calculations · Plate Similarity · Image Masking · Oneformer · Augnet.

1 Introduction

Dining is one of the vital activities that enriches our daily lives, in addition to being a source of nutrition. In particular, Japan offers a unique environment where people can enjoy a diverse range of dishes from around the world, including traditional Japanese cuisine, Western, Chinese, and various ethnic foods[3]. The

[3] https://www.statista.com/statistics/1358528/cities-with-most-michelin-starred-restaurants-worldwide/(confirmed on 26 June, 2025).

elements that contribute to dining pleasure include how the dish(in this paper "dish" is used exclusively to denote plated food) tastes and its visual quality, both contributing to the richness of the experience. No matter how delicious a dish may be, people would probably decline it if it looked unappetizing [5]. Enhancing the visual appeal of dishes increases diners' expectations and the perceived deliciousness of the food [17]. Furthermore, the visual appeal is provided by the food and the complex interplay of elements throughout the entire dining space. For example, even with the same menu of "a chicken steak, cherry tomatoes, broccoli, and rice," completely different impressions can be constructed depending on how the elements are combined. When a chicken steak, cherry tomatoes, broccoli, and rice are plated together, and served with a knife and fork, it becomes a "Western-style chicken steak." Conversely, if the chicken steak is plated on a flat plate, the cherry tomatoes and broccoli are each placed in small bowls, and rice is served in a rice bowl accompanied by chopsticks, it becomes a "Japanese-style chicken steak." In Japan, dishes are often styled to emphasize both taste and visual appeal through the strategic selection of ingredients. For instance, chopped parsley is sprinkled on the Western-style chicken steak, while chopped green onions are sprinkled on the Japanese-style chicken steak. As this example shows, even with the same menu, the visual impression can change dramatically depending on how they are plated, the choice of cutlery, and the presentation through elements such as interior design and lighting effects. In this research, we focused on *plateware choice* and *plating*, which influence the impression of the dishes as the primary subject of this study. The plateware choice involves selecting a suitable option for serving the prepared foods by considering color, shape, and material. Plating encompasses various serving aspects, such as whether to garnish [15] the dish, offer flat, or high-piled portions [18] and how to arrange the components .

Restaurants often adopt a consistent styling approach, with tableware playing a key role in contributing distinctive characteristics. In this context, "style" encompasses not only Japanese versus Western distinctions but a wide range of visual impressions. For example, elegant plates are used in upscale restaurants with chic interiors, while relaxed cafes use plates with casual designs [21]. Thus, many restaurants enhance the attractiveness of their dishes by considering how well their concept matches the plates they use.

Regular householders often enjoy cooking with attention to visual appeal and plating dishes beautifully on their favorite plates. Plating that matches the dining context [7] can enhance the attractiveness of the dishes. For example, when serving homemade dishes to celebrate a partner's birthday, plating food elegantly on sophisticated crockery can create a special atmosphere that emphasizes aesthetic maturity. On sweltering summer days, presenting the foods on glass plates may afford diners a sense of coolness. When friends with children visit, using playful plates such as yellow star-shaped or red flower-shaped ones and garnishing with colorful vegetables can create a dining table that children can enjoy visually [8]. As these examples suggest, cooks can express their culinary intentions [10] and create a dining space where guests feel more comfortable by staging the dining table through plate selection and plating that matches the dining context.

This research targets "people who care about the visual appearance of dishes," regardless of whether they are in restaurants or at home. We aim to expand their repertoire of plate choices and plating, providing support for more diverse and contextuallyappropriate dining presentations. For this, reference to previous plating examples is appropriate. When using food types as the starting point, a user preparing a pasta recipe enters "pasta" as a search query which returns pasta-related images in the search results, enabling reference to the types of plates used and the plating methods employed. However, the plates used in those images may not necessarily be like any the user owns. Conversely, when using plates as the starting point, even when conducting searches using plate images, the results are limited to product images of similar plates, with the plated dishes seldom appearing. A more intuitive and practical method is to reference plating examples that use "plates accessible to users," such as those they already own or those available in stores. In this research, we developed a system that pseudo-references plating examples using dish images using plates similar to those users can access. This system expands the opportunities to utilize self-owned plates and introduces novel food presentation approaches, targeting quality of life (QoL) improvement through more diverse dining experiences.

2 Related Work

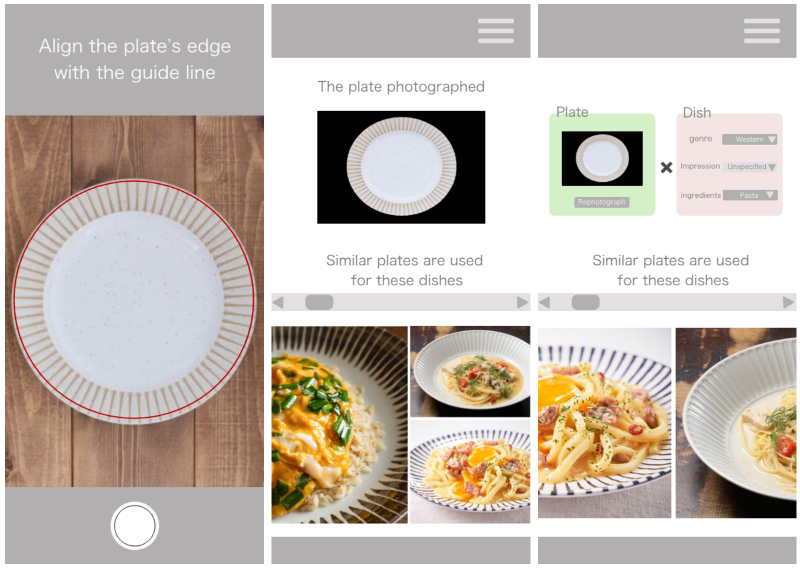
2.1 Idea reference using similar images

Research on referencing and recommending usage examples or ideas using similar images has been conducted not only for plates. Particularly in the fashion domain, there are various approaches to support coordination and purchase decision-making[6, 13]. The mix-and-match compatibility framework [12] provides recommendations not only based on visual similarity but also on how an item is used in a coherent context. This aligns with our goal of supporting idea formation through usage-oriented example retrieval.

Regarding interior and lifestyle scenes, similar problems to this research have been conducted, such as searching for "examples of similar styles" using interior images. The MMIS dataset [11] provides interior scene images paired with detailed textual descriptions, including style attributes and object layouts. This multimodal structure[23] enables usage-oriented example retrieval, aligning with our goal of presenting reference examples based on contextual arrangement rather than purely visual similarity.

2.2 Assistance with plateware selection

Research has been conducted to assist users in selecting plateware. The "meals–plates exploration cycle framework" [20] searches for food-plate combinations while reflecting users' plate preferences [14], based on the assumption that there is no



(a)The plate is photographed (b)The search result is presented (c)Additional Search Conditions

Fig. 1. Prototype of the system interface

single correct plate for a specific dish. Based on this assumption, this study aims to develop a system that enables users to more visually and intuitively assess the compatibility between their plates and foods.

"Plates impression estimate system" [21] assigns its impressions to plates based on the assumption that the interior of restaurants and the style of plates used there are consistent. In this research, they created a plate dataset by assigning impression words such as "modern" and "retro," collected from restaurant review information on gourmet sites [3] regarding the plates used in those restaurants. One approach to utilizing this dataset in this research is presenting dish images using plates owned by users, based on the impression the user intends to make.

3 Proposed Method

We developed a system that presents dish images using plates similar to those accessible to users.

3.1 Interface

The system interface functions as follows. First, a photo is taken of the plate to which the user will refer as a plating example, such that it fits within the presented frame(Fig. 1-(a)). Then, multiple dish photos using plates similar to the photographed plate are presented(Fig. 1-(b)). In this process, plate photos, dish categories, and desired dining table ambiance can be made selectable as search queries(Fig. 1-(c)). This enables us to provide contextually suitable dish examples when users are considering specific dishes or situations. Users can reference the photos and gain inspiration regarding the types of food to plate, the garnishing methods, the plating techniques, and other aspects.

3.2 Dataset

The dataset of dish images was prepared to meet the following requirements.

1. Images of food that is beautifully plated and photographed.
   Since the purpose is to showcase "visual appeal" users must be able to confirm the content clearly.
2. The dataset comprises images of various flat plates.
   Referencing diverse plating methods requires plates of various colors, patterns, and materials. Additionally, we limited both the dataset and query images to "flat plates" with shallow depth, excluding deep bowls and cups. No limitations were set for shape, allowing plates of all geometries, including round, square, flower-shaped, and others.
3. Contain sets of plated-food images and images with the plated food cropped out.
   This enables searching the dataset for photos using plates similar to those owned by users through similarity searches targeting only the plates.

First, as the source of dish-image information, we targeted recipe sites developed by food companies. Compared to user-generated recipe sites (such as Cookpad[4]), these sites exhibit more refined plating approaches. The photography tends to capture complete plate sections in dish images, thus meeting Requirement 1. Additionally, diverse flat plates with various colors, patterns, and materials are used, satisfying Requirement 2. In this research, we targeted 'Kikkoman Home Cooking[5]' (hereafter, Kikkoman) and 'Ajinomoto PARK Recipe Encyclopedia[6]' (hereafter, Ajinomoto), operated by the Kikkoman and Ajinomoto seasoning manufacturers. On these sites, recipes can be searched using filters and various parameters. Using these we searched for dishes filtered by 'main dishes' from Kikkoman and dishes filtered by 'main dishes, side dishes, staple foods' from Ajinomoto (which allowed multiple category selection), because these categories contain many dishes using 'flat plates.' From these search result pages, we collected 13,510 images using web scraping. Next, we manually excluded images containing non-plate kitchenware, multiple plates, and severely unclear images. After selection, 9,780 images were retained as "dish images usable as reference

[4] https://cookpad.com/jp (confirmed on 8 June, 2025).
[5] https://www.kikkoman.co.jp/homecook/(confirmed on 8 June, 2025).
[6] https://park.ajinomoto.co.jp/recipe/(confirmed on 8 June, 2025).

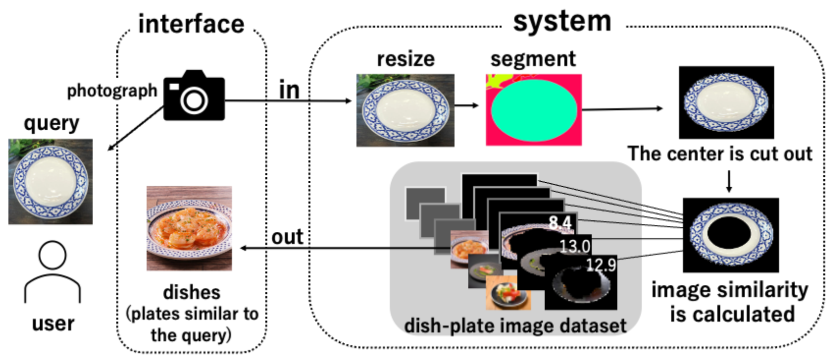


Fig. 2. Overview of the proposed method

sources." At this point, the corporate logo in the lower right of the images covered parts of the plates. Because this could impact similarity calculations, we extracted the corresponding rectangular regions from all images.

Since our proposed system aims to perform similarity searches that focus solely on the plates featured in dish images, it is necessary to link these dish images with images that have only the plate portions cropped from them. The cropping method is shown below. First, we performed segmentation[1]to divide the image into parts corresponding to plates and parts that do not. In this research, we performed semantic segmentation using Oneformer[9]. In Oneformer, color-codes through segmentation using the large-scale ADE20k image dataset [25] developed for scene understanding. Through this process, all pixels in the input image are represented by colors corresponding to each object class(e.g., plate, table). In this process, the segment images were assigned an opacity of 0 with labels hidden. For the class-color correspondence table, we referenced a GitHub discussion about color labels used in ControlNet's semantic segmentation[7]. This study designated regions corresponding to the "plate"(#00FFB8) and "tray"(#29FF00) classes as plate areas. Next, we created a mask where regions represented by colors corresponding to the above color codes among pixels in the segment image were designated as white (#FFFFFF). . All other regions were represented as black (#000000). Applying this mask to preserve only pixel information corresponding to plate areas in the input image and representing everything else as black (#000000), created images with the non-plate information removed. For images where plate areas were segmented as classes other than the two classes above, we manually specified the plate area colors and performed the same cropping process. Additionally, there were 144 cases where plate areas in the segment images became indistinguishable from the background table or plated dishes. These were excluded from the dataset as segmentation failures. Finally, we resized the Kikkoman images to match the 800×533 -pixel Ajinomoto images, which had a larger number of cases. All images were saved in 3 chan-

[7] https://cookpad.com/jp (confirmed on 8 June, 2025).

nels without transparency information. Through this process, we constructed a dataset comprising 9,636 [16] pairs linking dish images with cropped images of only the plate regions.

3.3 Processing flow

The overview of the proposed system's processing flow is as follows(Fig.2). The system mainly comprises four functions: input/output, plate cropping, central portion removal, and similarity search. First, users photograph the plate for which they want to reference plating examples using a camera. The photographed image is input at a size of 800×533 pixels. Next, semantic segmentation using Oneformer[9] is performed to remove the background within the photographed image. For regions represented by colors corresponding to "plate"(#00FFB8) or "tray"(#29FF00) among pixels in the segmented image, cropping processing is performed using the same method as in dataset creation. This completes the input image preprocessing. Next, using the preprocessed input image as a query, similar ones are searched from among the plate images in the dataset. Similarity scores (distances between image features) are calculated separately for plate images in the dataset. The top 100 images with highest-ranking distances (approximately 1% of the dataset) are selected as "plates similar to the query." As output, the dish images linked to those plate images are displayed. Hereby, plating examples can be pseudo-referenced for any plate.

3.4 Preliminary experiment

In the proposed system, we used imgsim[8] as the similarity calculation library. Imgsim uses AugNet[4], a deep learning paradigm, for its training. This model performs self-supervised learning[19] that does not require labels as in conventional methods. This creates augmented versions by changing the angle, brightness, and position of images, and learns a feature extraction method[2] such that these are treated as the same image. In the inference stage, images are embedded into a 192-dimensional vector space[24] using that feature extractor. This enables identification of "essentially identical images even when there are differences in lighting conditions or angle conditions during dish photography." In other words, AugNet is "a model that learns similarity of visual patterns robust to physical appearance variations," and its understanding of composition and semantics is weak. Therefore, for the plate images handled in this study, by combining it with OneFormer[9] to supplement compositional features, retrieval close to "human perceptual similarity[22]" can be achieved. The algorithm calculated the similarity between plate images in the dataset and the query plate image. In this process, the former were prepared by cropping only the plate portion from the dish images, resulting in a void in the center of the plates. On the other hand, the latter is a plate without a dish served on it, so there is no void in the center.

[8] https://github.com/chenmingxiang110/AugNet (confirmed on 9 June, 2025).

Table 1. The image distance between the query and the ground truth images

id	1-A	1-B	1-C	2-A	⋯	10-B	10-C	average
non-void	13.93	15.50	12.56	17.45	⋯	19.54	14.21	16.53
void	13.25	12.12	8.52	11.05	⋯	16.80	11.19	12.83
delta	-0.67	-3.38	-4.04	-6.40	⋯	-2.74	-3.02	-3.70

This section reports on preliminary experiments conducted to establish whether the differences in voids affect similarity calculations.

First, ten plate images were collected from an e-commerce site as query images. After removing background from the images, two versions were prepared: one with a central circular void and one without In this process, the void radius was defined as 0.65 of the plate region's bounding box size. Next, three plate images similar to each query plate were collected from e-commerce sites (30 images in total), identified as A, B, and C, and the central region of each was cropped. In this process, the cropped regions were manually shaped into irregular circles, consistent with those observed in the dataset, resembling the areas where foods were originally placed. These images were then used as reference images for measuring the distance to the queries. For the two prepared query images (with and without voids) the image distances were calculated against the reference images. A shorter distance for the void queries compared to the non-void queries would indicate the effectiveness of introducing voids.

For a subset of the query–reference image pairs, the computed distances are presented in a Fig.1. The top row shows the query IDs and the corresponding reference image IDs, the middle two rows indicate the computed distances. The bottom row represents the differences, denoted as $\Delta = distance\_\{void\} - distance\_\{non\text{-}void\}$. The average calculated distance was 16.53 for the non-void case and 12.83 for the void case, indicating an improvement of 3.7 points. For 26 of the 30 reference images, the computed distance was smaller with voids than without. These results suggested that the presence of central voids in query images contributes to decreasing the calculated distance between images. Therefore, in the processing flow mentioned in Section 3.3, we added a process that applies elliptical voids to the center of the plate images after cropping the plates. A circular void with a radius of 65 % of the bounding box size is applied, centered at the bounding box centroid of the plate region, and filled with black (#000000). This void processing is considered to not only align the shape of the query images more closely with the plate images in the dataset, but also mitigate the risk of the color information in the central region of the query plates being treated as noise.

4 Experiment

In the proposed system, the results of computerized similarity searches must match human similarity evaluations. We conducted experiments to verify the

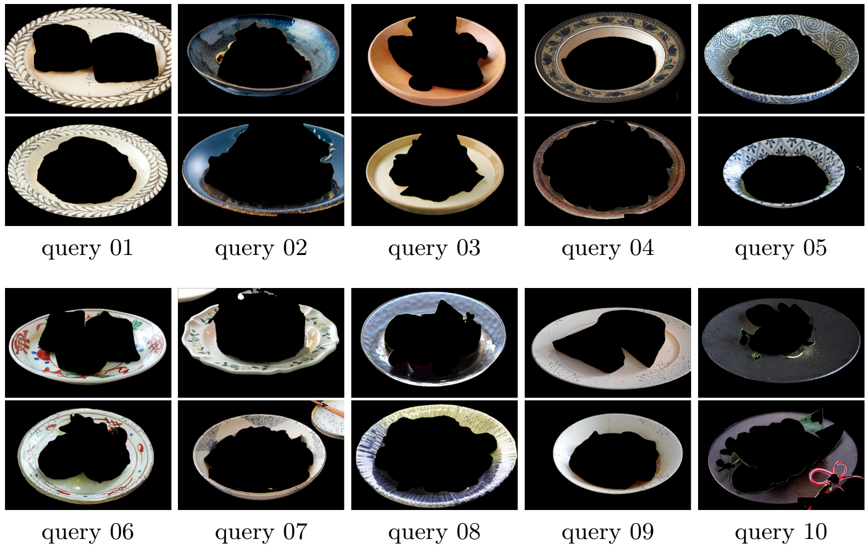


Fig. 3. 10 pairs of query plates (top) and plates with the highest similarity scores (bottom)

accuracy of the proposed similarity search method. The experiment calculated similarity scores for the query image against each of the 9,636 plate images in the dataset. Subsequently, another experiment had humans evaluate the similarity rankings of plates sampled from those results. The degree of correlation between the similarity scores calculated by computers and those evaluated by humans demonstrated the validity of the similarity calculation method if no correlation was observed, the similarity was reconsidered.

4.1 Preparation

First, reference query images were collected and processed as follows. Ten plates with varied colors, patterns, and materials were selected from plate product pages on the e-commerce site Rakuten Ichiba[9], under the condition that they were photographed from an oblique overhead angle similar to the dataset images. In this process, to match the same conditions as the system processing that masks the center of the query images, we adopted photos with plated food from among the dish images on the product pages. These query images were segmented and cropped in the same manner as during dataset creation, and were resized to 800×533 pixels.

Next, we calculated similarity scores between these query images and each plate image in the dataset. The similarity calculation using imgsim was per-

[9] https://www.rakuten.co.jp/(confirmed on 27/June/2025)

formed on Google Colaboratory with an Intel Xeon CPU @ 2.20GHz (1 physical core, 2 logical cores) and 12.7 GB RAM. The computation took approximately 7 seconds per pair (one query to one plate image in the dataset), and approximately 200 minutes for all 9,636 pairs. For each query, the plate images with the highest similarity are shown in the Fig.3.

The images used in the evaluation experiments were sampled as follows from the similarity calculation results. The dataset of approximately 9,000 items was divided into 4 tiers on a logarithmic scale (10^0, 10^1, 10^2, 10^3), and representative ranks “d” (d={2,5,8}) that evenly divided each tier into thirds were used as the standard. The reference rank “r” is expressed as follows.

$$r = d \times 10n(n = 0, 1, 2, 3) \tag{1}$$

Additionally, considering sampling bias, neighboring ranks of $r\pm1$ for each reference rank r ware also evaluated. Therefore, we conducted stratified sampling with a total of 36 samples per query image (12 reference ranks $\times$ 3 neighboring positions).

4.2 Human evaluation

To measure human similarity evaluation for “sampled plate and the query pairs,” we used a questionnaire. For plates sampled using the reference rank r, we created questionnaires (Google Forms[10]) to be answered by groups A, B, and C. In this setup, A was assigned rank $r-1$, B was assigned rank r, and C was assigned rank $r+1$ respectively.

The questionnaire presented pairs of query plate images and their similar plate images (both with plated food before cropping). Each group evaluated a total of 120 pairs comprising 10 queries $\times$ 12 reference ranks. The sections were divided by d values, and image pairs within each section (40 pairs) were displayed in random order for each respondent. For evaluation, we used a 5-point Likert scale ranging from “1 Not similar” to “5. Similar.” The responses were remotely collected by posting the forms on Yahoo Crowdsourcing[11]. The questions comprised two parts: 1) a link to navigate to Google Forms for responses, and 2) for entering the password displayed after completing the form. This ensured that the participants would respond to the forms. The experiment targeted approximately 200 people per group, with the gender balance maintained.

4.3 Result

The experimental results were analyzed as follows. First, for data cleansing, responses that met the following conditions were removed as insincere responses: responses with identical values (such as “1” for all items in the form) and low variance responses (such as all “2” and “3”). As a result, the following numbers

[10] https://workspace.google.com/intl/ja/products/forms/ (confirmed on 9/June/2025)
[11] https://crowdsourcing.yahoo.co.jp/ (confirmed on 9 June, 2025)

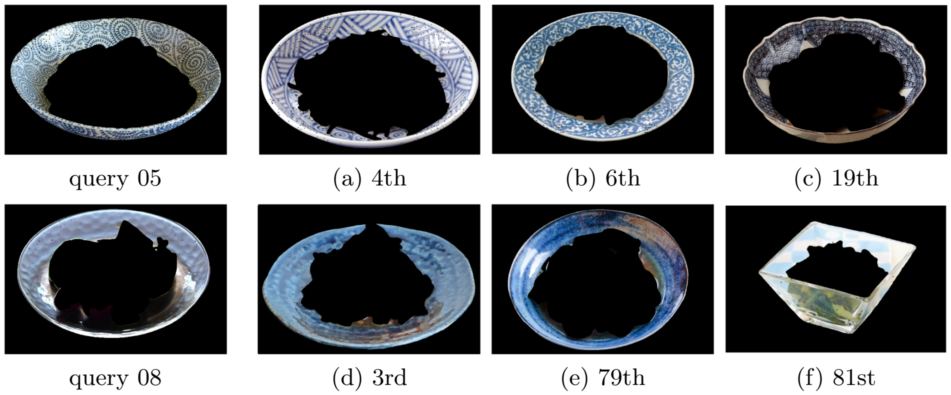


Fig. 4. Queries (leftmost) and plates where similarity evaluations between humans and computers showed consistent trends

of valid responses were obtained: 209 for Group A (out of 211 total responses), 213 for Group B (out of 223 total responses), and 192 for Group C (out of 200 total responses). We averaged all the responses of the respondents for each pair and used this as the evaluation value e for that pair. Next, to measure the correlation between human evaluation values and computer-calculated similarity for each pair, we calculated the correlation coefficients. In this process, the computer-calculated similarity values decrease as the distance between images decreases (more similar), while human evaluation scores increase as similarity increases. To resolve this discrepancy, we calculated correlation coefficients using inverted scores $e' = 6 - e$ for human evaluation values "e." For each query, the correlation coefficients of scores e' were calculated by respondent group (each with 12 reference ranks).

Table 2 summarizes the correlation coefficients between computer-calculated similarity and human evaluation e' for all queries. For queries 03 and 05, the calculated similarity and Group C evaluation e' are shown in scatter plots(Fig.6). Generally, the correlation coefficient values were approximately 0.5 to 0.8, showing moderate to strong correlation.

Table 2. The correlation coefficients between computer-calculated distance and human evaluation e' for each group (values ≥ 0.50 are shown in bold)

query id	01	02	03	04	05	06	07	08	09	10
GroupA	**0.55**	**0.93**	0.48	**0.65**	**0.77**	**0.75**	**0.55**	**0.83**	0.39	**0.78**
GroupB	**0.64**	**0.80**	**0.64**	0.11	**0.64**	**0.70**	0.40	**0.73**	0.18	**0.70**
GroupC	0.48	**0.84**	0.28	0.19	**0.84**	0.32	**0.64**	**0.74**	-0.16	**0.72**

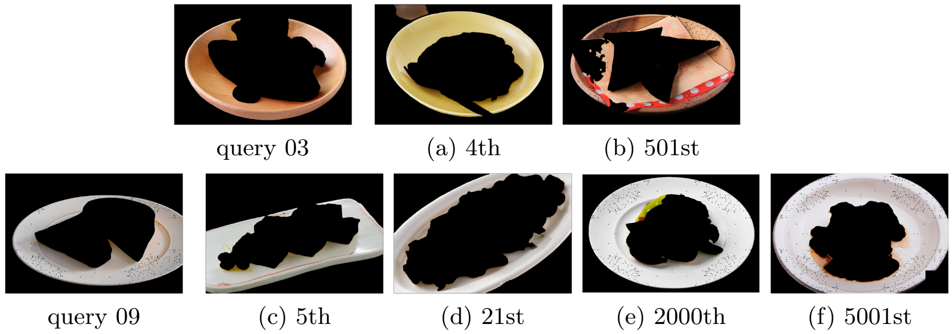


Fig. 5. Queries (leftmost) and plates where similarity evaluations between humans and computers showed different trends

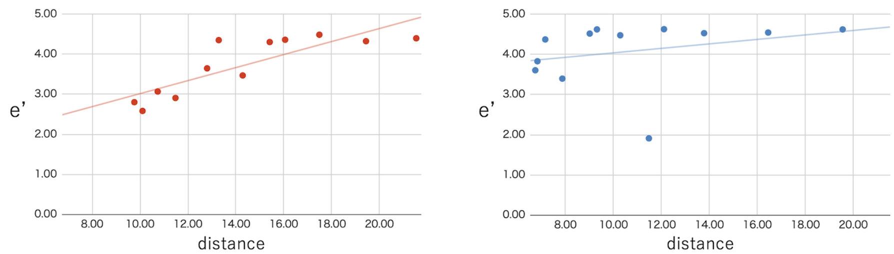
Queries 02, 05, 08, and 10 have high correlation coefficients, generally exceeding 0.70.

The top similarity group for query 05 contained many plates with blue geometric patterns (Fig.4-(a),(b),(c)), which are a major characteristic of query 05, and the evaluation values tended to be high (Fig.6-(a)) . Since imgsim extracts local color-contrast features from the image, it can be inferred that plates with regular color arrangements due to patterns are likely to create clear differences between those with matching features and those without.

Despite being transparent and thus expected to pose difficulties in color extraction and similar plate detection, query 08 showed a high correlation. In the top similarity group (within the top 100), as expected, transparent plates were only partially detected (Fig.4-(f)). However, many plates with commonalities to the query plate ware included. For example, those with shallow indentation-like surfaces (Fig.4-(d)) and those inducing a perception of coolness (Fig.4-(e)) due to glossiness or material type. This is considered to be one of the reasons why the top group was evaluated as having relatively high similarity. However, if the most distinctive feature (being a transparent glass plate) did not match, it would be difficult to use it as a reference source for plating. Similarity search methods should be considered that use both color and material characteristics.

In contrast, queries 03, 04, and 09 frequently produced correlation coefficients under 0.4, indicating minimal correlation.

In the top similarity group for query 03, the characteristics of a wooden query plate are not perceived, with many yellow ceramic plates detected(Fig.5-(a)) . Conversely, a plate matching the wood material characteristics ranked 501st(Fig.5-(b)) and was not included in the top group. In the cropping process of this plate image, segmentation of the serviette has failed. The red margin of the serviette was probably rendered as noise, resulting in a relatively low similarity being calculated. The similarity evaluation value for this plate e' is 1.91, indicating that it evidences very high similarity(Fig.6-(b)). While humans can recognize wood texture, computers cannot, which is also considered a major



(a) query 05 (b) query 03

Fig. 6. Scatter plot of calculated distance and human evaluation e' (Group C)

factor in the evaluation gap between computers and humans. For query 03, this discrepancy is considered to be a factor contributing to the low correlation.

Query 09 (a simple white plate) also resulted in low correlation. The plates ranked 5th ($e' = 4.46$) and 21st ($e' = 3.77$) (hereinafter called low-e') exhibited high computer-calculated similarity but low human evaluation values(Fig.5-(c),(d)). Conversely, the plates ranked 2000th ($e' = 1.86$) and 5001st ($e' = 2.43$) (hereafter referred to as high-e') showed low computer-calculated similarity but high human evaluation values (Fig.5-(e),(f)). First, in terms of color, low-e' appears to be white plates with slight shadows, resulting in relatively dark colors similar to the query, whereas high-e' appears to be pure white plates with high brightness. Humans do not differentiate between the two types, categorizing both as "white plates." The computer, similarly, is generally robust to variations in lighting conditions. However, white plates are relatively abundant in the dataset, which likely caused their similarity scores to be diluted, resulting in lower rankings. Next, in terms of shape, low-e' are oval and rectangular, whereas high-e' are nearly circular plates, similar to the query. Presumably, the computer could not evaluate similarity based on the exact contour of the plate region. Therefore, it is presumed that if computers can more precisely capture shape information and recognize colors more robustly under varying lighting conditions, they could achieve closer alignment with human evaluation values.

From these results, it became clear that the proposed similarity search method tends to detect plates with similar characteristics in terms of color and patterns. However, there is room for further consideration regarding shape, material, and lighting-invariant color recognition. For a more rigorous similarity search, it would be necessary to combine similarity search methods that incorporate these characteristics.

5 Conclusion

This research aimed to realize new dining experiences by creating diverse and contextually appropriate dining table presentations by introducing novel food plating styles. We proposed a method to support decision-making regarding food and plate combinations and plating approaches by referencing plating examples of plates similar to those available to the user. This paper verified the accuracy of the proposed "plate similarity search" in the proposed method. In the results, the computer-calculated similarity and human similarity evaluation somewhat matched. Furthermore, the preliminary experiments demonstrated that the presence of a circular void at the center of the query plates leads to reduced distances between images in similarity search. By introducing methods that measure similarity not only in color and pattern, but also in shape, material, and color recognition invariant to lighting conditions, it will become possible to search for plates with higher similarity from a human perspective. Additionally, we plan to develop a plate plating a reference interface and conduct user experiments to measure the system's contribution to improving dining experiences.

References

1. Aguilar, E., Remeseiro, B., Bolaños, M., Radeva, P.: Grab, pay, and eat: Semantic food detection for smart restaurants. IEEE Transactions on Multimedia **20**(12), 3266–3275 (2018). https://doi.org/10.1109/TMM.2018.2831627
2. Ahmed, A.S., Ibraheem, I.N.: Recent advances in content based image retrieval using deep learning techniques: A survey. In: AIP Conference Proceedings (2024), https://api.semanticscholar.org/CorpusID:274249880
3. Chang, Y.J., Lo, H.Y., Huang, M.S., Hu, M.C.: Representative photo selection for restaurants in food blogs. In: Proc. IEEE International Conference on Multimedia & Expo Workshops. pp. 1–6 (2015). https://doi.org/10.1109/ICMEW.2015.7169814
4. Chen, M., Chang, Z., Lu, H., Yang, B., Li, Z., Guo, L., Wang, Z.: AugNet: End-to-end unsupervised visual representation learning with image augmentation (2021). https://doi.org/10.48550/arXiv.2106.06250
5. Cui, R., Chen, W., Peng, D., Minamizawa, K., Pai, Y.S.: FoodMorph: Changing food appearance towards less unhealthy food intake. In: SIGGRAPH Asia 2023 Posters (2023). https://doi.org/10.1145/3610542.3626149
6. Deldjoo, Y., Nazary, F., Ramisa, A., McAuley, J., Pellegrini, G., Bellogin, A., Noia, T.D.: A review of modern fashion recommender systems. ACM Computing Surveys **56**(4) (2023). https://doi.org/10.1145/3624733
7. Ha, S.W., Nurain, N., Agapie, E., Chung, C.F.: Preparing and experiencing food during life events: Implications for technology supporting social and value changes. In: Proc. ACM Conference on Human Factors in Computing Systems (2025). https://doi.org/10.1145/3706598.3713183
8. Han, S.Y., Kang, E.J.: ChilDish: The smart plate and cup for children. In: Proc. Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (2017). https://doi.org/10.1145/3029798.3034946
9. Jain, J., Li, J., Chiu, M.T., Hassani, A., Orlov, N., Shi, H.: Oneformer: One transformer to rule universal image segmentation. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2989–2998 (2023). https://doi.org/10.48550/arXiv.2211.06220
10. Javed, M., Malik, F.A., Awan, T.M., Khan, R.: Food photo posting on social media while dining: An evidence using embedded correlational mixed

methods approach. Journal of Food Products Marketing **27**(1), 10–26 (2021). https://doi.org/10.1080/10454446.2021.1881861

11. Kassab, H., Mahmoud, A., Bahaa, M., Mohamed, A., Hamdi, A.: MMIS: Multimodal dataset for interior scene visual generation and recognition (2024), https://arxiv.org/abs/2407.05980
12. Liao, S., Ding, Y., Mok, P.: Recommendation of mix-and-match clothing by modeling indirect personal compatibility. In: Proc. ACM International Conference on Multimedia Retrieval. p. 560–564 (2023). https://doi.org/10.1145/3591106.3592224
13. Liu, H., Tang, X., Chen, T., Liu, J., Indu, I., Zou, H.P., Dai, P., Galan, R.F., Porter, M.D., Jia, D., Zhang, N., Xiong, L.: Sequential llm framework for fashion recommendation (2024), https://arxiv.org/abs/2410.11327
14. Lv, P., Fan, J., Nie, X., Dong, W., Jiang, X., Zhou, B., Xu, M., Xu, C.: User-guided personalized image aesthetic assessment based on deep reinforcement learning. IEEE Transactions on Multimedia **25**, 736–749 (2023). https://doi.org/10.1109/TMM.2021.3130752
15. Nishimura, T., Ishiguro, K., Higuchi, K., Kotera, M.: Multimodal dish pairing: Predicting side dishes to serve with a main dish. In: Proc. 1st International Workshop on Multimedia for Cooking, Eating, and Related Applications. p. 1–9 (2022). https://doi.org/10.1145/3552485.3554934
16. Okamoto, K., Yanai, K.: UEC-FoodPix Complete: A large-scale food image segmentation dataset. In: Pattern Recognition. ICPR International Workshops and Challenges. pp. 647–659 (2021). https://doi.org/10.1007/978-3-030-68821-9_51
17. Peng, Y., Jemmott, J.B.: Feast for the eyes: Effects of food perceptions and computer vision features on food photo popularity. International Journal of Communication **12**, 24 (2018), https://api.semanticscholar.org/CorpusID:148794751
18. Sakurai, S., Narumi, T., Ban, Y., Tanikawa, T., Hirose, M.: CalibraTable: tabletop system for influencing eating behavior. In: SIGGRAPH Asia 2015 Emerging Technologies (2015). https://doi.org/10.1145/2818466.2818483, 10.1145/2818466.2818483
19. Schall, K., Barthel, K.U., Hezel, N., Jung, K.: Optimizing CLIP models for image retrieval with maintained joint-embedding alignment (2024), https://arxiv.org/abs/2409.01936
20. Takahashi, C., Matsushita, M., Yamanishi, R.: Exploration cycle finding a better dining experience: a framework of meal-plates. Procedia Computer Science **225**, 2902–2911 (2023). https://doi.org/10.1016/j.procs.2023.10.283
21. Takahashi, R., Takahashi, C., Matsushita, M.: Estimation of plates impression by restaurant information on the gourmet website. Procedia Computer Science **246**, 3868–3877 (2024). https://doi.org/10.1016/j.procs.2024.09.160
22. Tizhoosh, H.R., Pantanowitz, L.: On image search in histopathology (2024), https://arxiv.org/abs/2401.08699
23. Xu, J., Chen, Z., Yang, S., Li, J., Wang, W., Hu, X., Hoi, S., Ngai, E.: A survey on multimodal recommender systems: Recent advances and future directions (2025), https://arxiv.org/abs/2502.15711
24. Zhang, H., Duo, Y.: Improving Deep Hashing Performance for Image Retrieval Using Multi-Dimensional Attention . In: Proc. 5th International Conference on Artificial Intelligence, Big Data and Algorithms. pp. 700–703 (2025). https://doi.org/10.1109/CAIBDA65784.2025.11183133
25. Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba, A.: Scene parsing through ade20k dataset. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. pp. 633–641 (2017). https://doi.org/10.1109/CVPR.2017.544