Design Generation Using Stable Diffusion and Questionnaire Survey

Masakazu Kobayashi[1] and Katsuyoshi Kume [2]

[1] Toyota Technological Institute, kobayashi@toyota-ti.ac.jp
[2] Toyota Technological Institute, sd19036@toyota-ti.ac.jp

Corresponding author: Masakazu Kobayashi, kobayashi@toyota-ti.ac.jp

Abstract. Stable Diffusion, a deep learning-based image generation technique, has recently gained significant attention for its ability to generate detailed images conditioned on text descriptions called "prompts." By inputting appropriate text descriptions, Stable Diffusion can create various images, ranging from fictional animals to realistic photographs and illustrations. In this research, a novel method for generating product designs by combining Stable Diffusion and questionnaire surveys is proposed. In the proposed method, Kansei words representing specific impressions are input to Stable Diffusion as prompts to generate product designs with desired characteristics. To identify customer preferences and the impressions they want the product to have, questionnaire surveys are conducted. Based on the results, images of products that customers like and products with impressions that customers desire are selected, and additional training is conducted using Stable Diffusion. This enables Stable Diffusion to generate even more suitable product images for customers. In a case study, the proposed method is applied to chair design to verify its ability to generate product designs that convey specific impressions and are preferred by subjects.

Keywords: Kansei Engineering, Aesthetic Design, Deep Learning, Stable Diffusion.
DOI: https://doi.org/10.14733/cadaps.2024.859-868

1 INTRODUCTION

As science and technology continue to mature, differentiating products based on performance, functional features, or price becomes increasingly challenging. Consequently, companies must distinguish their products by focusing on subjective and abstract qualities such as aesthetics and comfort, which are assessed based on customers' feelings. In Japanese, this concept is referred to as "Kansei." The qualities evaluated by customer Kansei are known as "Kansei quality."

In the field of Kansei engineering (referred to as affective or emotional engineering), the methods for measuring customer Kansei or the impression of products have been developed and applied to many case studies. In these methods, the semantic differential (SD) method [8] is widely used. Based on the measurement and analysis methods of customer Kansei, various aesthetic design methods have also been developed. These methods generate a new aesthetic design that a customer prefers best by revealing the relationships between the results of customers' Kansei evaluation of the same type of existing products as the design target and their aesthetic features. In these methods, various analysis methods such as artificial neural network [4] [5], fuzzy set theory [3], interactive reduct evolutionary computation [20], multidimensional scaling [5], rough set theory [6] [8-9] [11] [19] self-organizing map [5], etc. are used.

In recent years, deep learning has attracted a lot of attention. Deep learning is a class of machine learning algorithms based on artificial neural networks. Deep learning uses multiple layers to extract higher-level features progressively and automatically from the raw input. In the case of image classification, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits, letters, or faces. Various types of deep learning models have been developed and applied to computer vision, speech recognition, natural language processing, audio recognition, self-driving cars, board games, etc. Deep learning has also been used in research on Kansei engineering. For example, Quan et al. proposed the Knsei engineering-based neural style transfer for product innovation (KENPI) framework [12]. Dai et al. proposed the approach for automatic design scheme generation based on generative adversarial networks (GAN)[1][2]. Schmitt and Weiss designed innovative chairs inspired by the chair images generated by GAN [15]. Kobayashi et al. confirmed that product images generated using GAN with only product images that customers preferred as input were highly likely to be preferred by customers [6].

In 2022, Stable Diffusion [17], another deep learning-based image generation method, has been proposed and has attracted much attention. Stable Diffusion uses a variant of the diffusion model (DM) [16], called the latent diffusion model (LDM) [13], and generates detailed images conditioned on text descriptions called "Prompt." By inputting appropriate text descriptions as prompts, Stable Diffusion can generate a variety of images, including fictional animals that look like a mixture of real animals, detailed images indistinguishable from actual photographs, and cartoon-like illustrations.

In this research, a new method for generating product designs based on Stable Diffusion and questionnaire surveys. A little more specific, to generate a product design with certain impressions, Kansei words representing those impressions are input to Stable diffusion as prompts. To identify the customer's preferences and the impressions that the customer wants the product to have, and to generate the most favorable product design, questionnaire surveys are combined with stable diffusion. Then, questionnaire surveys are conducted to identify what impression the customers want the product to have and what kind of product they prefer. Based on the results, fine-tuning called "embedding" is performed in Stable diffusion to generate images of products that customers would most likely prefer. In the case study, the proposed method is applied to a chair design to verify whether it is possible to generate product designs that give specific impressions to subjects and whether it is possible to generate product designs that subjects prefer.

2 STABLE DIFFUSION:

Stable Diffusion (SD) is one of the deep learning-based image generation methods. SD can generate detailed images conditioned on text descriptions called "Prompt". By inputting appropriate text descriptions as prompts, SD can generate a variety of images, including fictional animals that look like a mixture of real animals, detailed images indistinguishable from actual photographs, and cartoon-like illustrations. SD is based on a variant of the diffusion model (DM) called the latent diffusion model (LDM). DM consists of two processes, named forward and reverse diffusion processes. In the forward diffusion model, gaussian noise is iteratively applied to image X_0 to obtain a fully noisy image X_T, as shown in Figure 1. In the reverse diffusion process, gaussian noise is conversely removed iteratively from the complete noise image X_T to obtain an image equivalent to X_0. By training these processes using existing images, it becomes possible to generate images from random noise. The difference between DM and LDM lies in whether the diffusion process is performed in the image pixel space or the latent space.

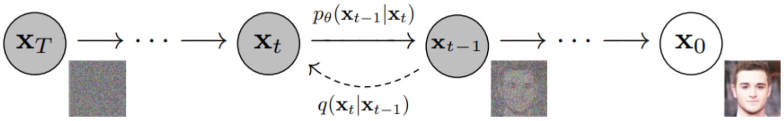

Figure 1: Markov chain of forward and reverse diffusion processes [16].

SD consists of U-Net [14] that mimics the reverse diffusion process of LDM, a variational autoencoder (VAE), and a text encoder. During image generation using SD, U-Net denoises randomly generated noise in the latent space while conditioning it on a text sequence called a prompt, obtaining the latent representation of the image. A VAE decoder then converts this laten representation back into pixel space, resulting in the final image. SD has various functions, such as t2i, which generates images from text input; i2i, which generates new images from image input; and embedding, which fine-tunes using images provided by the user.

3 PROPOSED METHOD:

The proposed method consists of the following 4 steps.

Step1: Generation of product images with specific impressions
Step2: Verification of customer impressions of the products through questionnaire surveys
Step3: Evaluation of customer preferences for the products through questionnaire surveys
Step4: Fine-tuning and product image generation

3.1 Step1: Generation of Product Images with Specific Impressions

In Step 1, product images for the questionnaire surveys in Steps 2 and 3 are generated by giving prompts to Stable Diffusion. The prompts that input into stable diffusion consist of the words that describe the target product, e.g., “chair” or “office desk”, and the Kansei words that describe the impressions that the product should have, e.g., “soft” or “gorgeous”. For the Kansei words, pairs of Kansei words that are suitable for describing the target product, e.g., hard-soft, are collected and combined for the prompt. In the example, four pairs of Kansei words are collected and two of them are combined. Product images are generated for 24 combinations of Kansei words except for paired, conflicting Kansei words. Multiple images are generated for each combination of Kansei words.

3.2 Step2: Verification of Customer Impressions of the Products Through Questionnaire Surveys

A questionnaire survey is conducted to confirm whether a customer receives the same impressions from the product images generated in Step 1 as the Kansei words used in the prompts. In the questionnaire, a customer rates the product images on a 5-point scale using Kansei words that are given as the prompts. The questionnaire results are analyzed, and if the product images generated have the impressions as prompted, then go to the next step. If not, return to Step 1, and product images should be re-generated. In this case, it is advisable to change the Kansei words to synonyms, add words that identify the part to which the Kansei words apply, or increase the number of words that describe the target product.

3.3 Step3: Evaluation of Customer Preferences for the Products Through Questionnaire Surveys

Questionnaire surveys are conducted to measure customer’s preferences for the product images generated in Step 1 and validated in Step 2. Preference for the product images is rated on a 5-point scale.

3.4 Step4: Fine Tuning and Product Image Generation

Product images used for embedding are selected in the following two ways.

(a) The average score of the preference is calculated for each group of product images with the same prompt, and the images of the group with the highest preference are selected.

(b) Product images with a 5 rating on a 5-point scale of preference are selected, regardless of their group.

In method (a), the most favorable impressions perceived by the customer are first identified, and product images with these impressions are used for embedding. As a result, relatively similar product images are expected to be obtained. In method (b), product images that the customer evaluates as favorable are used for embedding, regardless of their impressions. Since the products that customers rate as favorites do not always have the same impression (as customer’s preferences can vary greatly), a wide variety of product images may be selected. If SD can generate images of products that customers like, even in this scenario, it demonstrates that SD can learn information about customers' preferences from diverse images.

Embedding is then performed using these images to generate product images that reflect the subject's preferences.

4 CASE STUDY

To confirm the effectiveness of the proposed method, it is applied to a chair design. In the case study, Stable Diffusion Web UI [18] shown in Figure 2, which is a kind of GUI of Stable Diffusion and runs on a browser, was used. In the case study, the txt2img tab, which generates images from prompts, and the Train tab, which performs additional training called Embedding using images, were used. Figure 2 shows the txt2img tab. By entering a prompt in the text area labeled "Prompt" and pressing the Generate button on the right, an image is generated and displayed in the area where the chair image is shown in Figure 2. The settings that affect the generated images include the Stable Diffusion model (written as Stable Diffusion checkpoint in the GUI), Sampling method, Sampling steps, CFG Scale, and Seed.

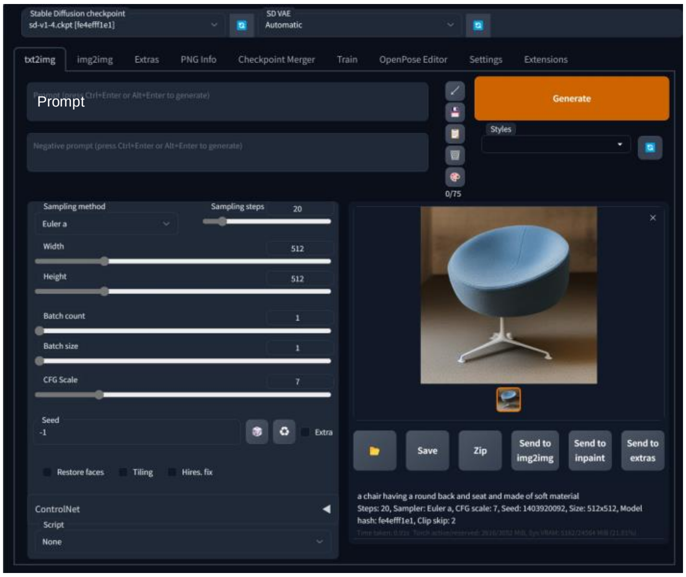

Figure 2: Stable Diffusion Web UI.

In the case study, Stable Diffusion 1.4 [17] and Euler A were used for the model and sampling method, respectively. The sampling step and CFG scale were set to 20 and 7, respectively. Seed values are used to generate initial noise images, and the same seed value always generates the same image if the prompt and all other settings are the same. In the case study, the seed value was set randomly for each time. The GPU used here was NVIDIA RTX2080Ti. 4 male undergraduates participated as subjects. The proposal was applied to each subject individually. Eight Kansei words or four pairs of Kansei words (round/square, cool/cute, luxury/simple, and soft/hard) were used to describe the impression of the chair. Two of these eight Kansei words were selected for each prompt. Since conflicting combinations (e.g., round/square) were excluded, 24 combinations were possible. These combinations were named Groups 1-24 and shown in Tab.1. In the first attempt, 10 product images were generated for each group using two Kansei words plus "wooden chair" as the prompt. Several examples are shown in Figure 3.

Figure 3: Generated chair images for subject1 (Left: round & cool, Right: simple & soft).

An initial questionnaire survey was conducted on these chair images. Table 1 displays the questionnaire results, presenting the Kansei words and average scores for each group.

Group	1	2	3	4	5	6	7	8	9	10	11	12
Kansei words	round	round	round	round	round	round	square	square	square	square	square	square
	cool	cute	luxury	simple	hard	soft	cool	cute	luxury	simple	hard	soft
Average score	4.10	3.96	4.00	3.91	3.09	3.30	4.03	3.46	3.15	4.39	4.91	3.19

13	14	15	16	17	18	19	20	21	22	23	24
cool	cool	cool	cool	cute	cute	cute	cute	luxury	luxury	simple	simple
luxury	simple	hard	soft	luxury	simple	hard	soft	hard	soft	hard	soft
3.98	3.76	4.10	2.99	4.05	3.75	3.70	3.78	3.25	4.46	4.73	4.00

Table 1: Results of the initial questionnaire survey.

Although the average scores for each group are not low, there are some groups with scores below 4. Therefore, the prompts were improved, and new images were generated. In this attempt to obtain better chair images, "round" and "square" were described as "having a round back and seat." "Soft" and "hard" were described as "made of soft material." As for the words for chair,

"chair" was used. 10 product images were generated for each group, with some of the resulting images shown in Figure 4.

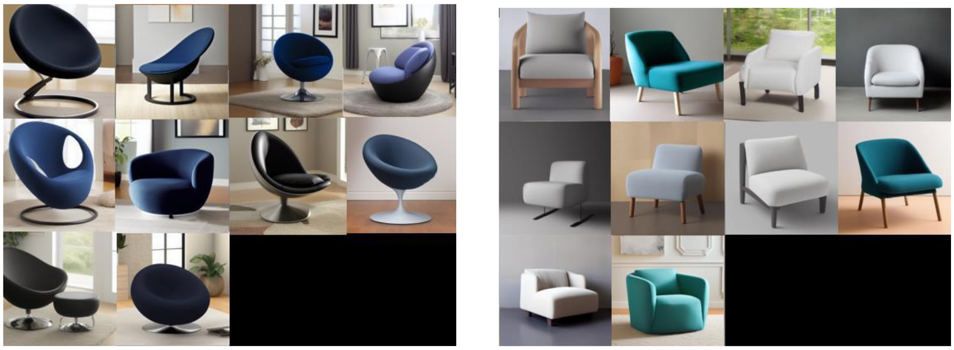

Figure 4: Generated chair images for subject1 (Left: round & cool, Right: simple & soft).

A first questionnaire survey was conducted one more time on these chair images. Table 2 shows the Kansei words and the average score for each group. This table confirms that the use of Kansei words as prompts can generate a product image with specific impressions.

Group	1	2	3	4	5	6	7	8	9	10	11	12
Kansei words	round cool	round cute	round luxury	round simple	round hard	round soft	square cool	square cute	square luxury	square simple	square hard	square soft
Average score	4.68	4.69	4.23	4.56	4.25	4.75	3.71	3.65	3.85	4.76	4.78	3.93

13	14	15	16	17	18	19	20	21	22	23	24
cool luxury	cool simple	cool hard	cool soft	cute luxury	cute simple	cute hard	cute soft	luxury hard	luxury soft	simple hard	simple soft
4.24	4.26	4.38	3.84	4.56	4.1	4.43	4.68	3.83	4.63	4.71	4.53

Table 2: Results of the initial questionnaire survey.

A second questionnaire was then conducted to measure the preference for these chairs. Table 3 shows the groups with the higher average scores for the preferences. The table shows that each subject has different preferences.

	Subject1		Subject2		Subject3		Subject4	
Rank	Group	Average score	Group	Average score	Group	Average score	Group	Average score
1	13	4.6	22	5	22	4.8	3	5
2	24	4.4	16, 13, 3	4.6	24	4.5	1	4.7
3	1	4.3			13, 17	4.4	6, 14, 24	4.6

Table 3: Results of the second questionnaire survey.

Finally, the product images were selected in the ways described in (a) and (b) of Step 4. In method (a), images from the rank 1 group shown in Table 3 were selected. In method (b), all images with a score of 5 were selected. The number of images selected by method (b) was 28 for subject 1, 79 for subject 2, 47 for subject 3, and 79 for subject 4. Fine tuning was performed using these images. Chair images generated for each subject are shown in Figures 5 - 8. Since the images selected using method (b) exhibit greater diversity compared to those selected using method (a), the images obtained through embedding using these images also exhibit a similar trend.

Figure 5: Generated chair images for subject1 (Left: Method (a), Right: Method (b)).

Figure 6: Generated chair images for subject2 (Left: Method (a), Right: Method (b)).

Figure 7: Generated chair images for subject3 (Left: Method (a), Right: Method (b)).

Figure 8: Generated chair images for subject4 (Left: Method (a), Right: Method (b)).

4.1 Discussion

The impressions that the subject received from the chair images generated for each subject and preferences for them were measured by questionnaire survey. Table 4 shows the results. This table shows the average scores for the 20 generated images for each condition.

	Subject1		Subject2		Subject3		Subject4	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Preference	4.4	3.9	4.7	4.9	4.8	4.9	3.1	3
Cute	2.4	1.5	4.5	4.8	4	2.3	3.4	3.5
Cool	4.2	4.5	1.5	1.2	2	3.7	2.6	2.5
Square	1.7	2.5	1.7	1.5	1.6	1.8	1	1.3
Round	4.3	3.5	4.3	4.5	4.4	4.2	5	4.7
Simple	1.6	2.1	1	4.1	1	2.9	1.3	4.2
Luxury	4.4	3.9	5	1.9	5	3.1	4.7	1.8
Soft	4.5	3.8	5	4.7	5	5	4.9	3.4
Hard	1.5	2.2	1	1.3	1	1	1.1	2.6

Table 4: Results of the second questionnaire survey.

When focusing on preferences, the results show that both methods (a) and (b) can generate the product images preferred by the subjects except subject4. When focusing on the impressions received from the product, the results show that the image generated after fine-tuning with method (a) has the same impression as the image of the group used for fine-tuning. On the other hand, in method (b), the subject's preference is the only factor to be considered when selecting images for fine tuning, so the impressions of the selected images are diverse. Therefore, the impressions of the images generated after fine tuning are also diverse. The method (a) is considered to identify impressions that are closely related to the customer's preferences first, and then to generate product images with the same impressions. On the other hand, method (b) is considered to use the customer's favorite product for fine tuning to allow Stable Diffusion to learn the customer's favorite design elements and then to generate products that include those elements. Therefore, despite the high preference for the generated images, their impressions are diverse. Either way, the results show that the proposed method can generate product images preferred by subjects using Stable Diffusion.

5 CONCLUSION

In this research, a new method for generating product images using stable diffusion, which has been attracting attention in recent years, is proposed. By combining stable diffusion and questionnaire surveys, the proposed method can generate products that have specific impressions and that customers prefer from texts called prompts. In the case study, the effectiveness of the proposed method is demonstrated by applying it to the chair design. The proposed method can generate product images based on customer desires expressed in the form of words (Kansei words in the proposed method). Furthermore, when a customer makes a Kansei evaluation of the generated product images, the proposed method can generate product images more suitable for the customer based on their results. From these features, it can be said that the proposed method is suitable for designing tailor-made products rather than mass-produced products.

A future research direction involves adjusting the weights of the words used in the prompts. In SD, the weight of a word can be adjusted using formats such as (hard), ((hard)), or (hard:1.2). These all emphasize the influence of the word "hard" on the output image. Conversely, it is also possible to weaken the effect. By utilizing this feature, an attempt can be made to generate product images that are more suitable for individual customers. If the weights can be appropriately set using methods such as questionnaires, it might be possible to generate product images that better suit individual customers.

6 ACKNOWLEDGMENT

This research was supported by JSPS KAKENHI Grant Number 20K12540.

REFERENCES

[1] Dai, Y., Li, Y., Liu, L.: New Product Design with Automatic Scheme Generation, Sensing and Imaging, 20(29), 2019. http://dx.doi.org/10.1007/s11220-019-0248-9

[2] Goodfellow, I J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks, NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2, 2014, 2672-2680. http://dx.doi.org/10.1145/3422622

[3] Hsiao, S.W.; Huang, H.C.: Applying the semantic transformation method to product form design, Design Studies, 19(3), 1998, 309-330. http://dx.doi.org/10.1016/S0142-694X(98)00009-X

[4] Hsiao, S.W.; Huang, H.C.: A neural network based approach for product form design, Design Studies, 23(1), 2002, 67-84. http://dx.doi.org/10.1016/S0142-694X(01)00015-1

[5] Kobayashi, M.; Kinumura, T.; Higashi, M.: A Method for Supporting Aesthetic Design Based on the Analysis of the Relationships Between Customer Kansei and Aesthetic Element, Computer-Aided Design & Applications, 13(3), 2015, 281-288. http://dx.doi.org/10.1080/16864360.2015.1114385

[6] Kobayashi, M., Niwa, K.: Method for Grouping of Customers and Aesthetic Design Based on Rough set theory, Computer-Aided Design & Applications, 15(4), 2018, 565-574. http://dx.doi.org/10.1080/16864360.2017.1419644

[7] Kobayashi, M.; Thongpramoon, P.: Generation of Product Design Using Gan Based on Customer's Kansei Evaluation, 9th International Conference on Kansei Engineering and Emotion Research 2022, Barcelona, Spain, September 6-9, 2022.

[8] Kuramaru, K.; Takanashi, R.; Mori, N.: The Method of Design Project Based on the Declaration of Checking Preference and Non-preference with the Product, Journal of Japan Society of Kansei Engineering, 1(1), 2001, 65-72. http://doi.org/10.5057/jjske2001.1.65

[9] Ohki, M.; Harada, T.; Inuiguchi, M.: Decision Rule Visualization System for Knowledge Discovery by Means of the Rough Set Approach, Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 24(2), 2012, 660-670. http://doi.org/10.3156/jsoft.24.660

[10] Osgood C.E., Suci G.J., Tannenbaum P.: The Measurement of Meaning, University of Illinois Press, 1967.
[11] Pawlak, Z.: Raugh Sets, International journal of Information Computer Science, 11(5), 1982, 341-356. http://doi.org/10.1007/BF01001956
[12] Quan H, Li S, Hu J.: Product Innovation Design Based on Deep Learning and Kansei Engineering, Applied Sciences, 8(12), 2018, 2397-2415. http://dx.doi.org/10.3390/app8122397
[13] Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B.: High-Resolution Image Synthesis with Latent Diffusion Models, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, June 18-24, 2022. http://dx.doi.org/10.1109/CVPR52688.2022.01042
[14] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, 2015, 234-241. http://dx.doi.org/10.1007/978-3-319-24574-4_28
[15] Schmitt, P. Weiss, S.: The Chair Project, https://philippschmitt.com/work/chair
[16] Ho, J.; Jain, A.; Abbeel, P.: Denoising Diffusion Probabilistic Models, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada, December 6 – 12, 2020. https://dl.acm.org/doi/pdf/10.5555/3495724.3496298
[17] Stable diffusion, https://github.com/CompVis/stable-diffusion , CompVis.
[18] Stable Diffusion Web Ui, https://github.com/AUTOMATIC1111/stable-diffusion-webui, AUTOMATIC1111.
[19] Yamada, K.; Moroga, U.; Unehara, M.: Design Support for Generating Novelty with Rough Sets Theory and Conceptual Hierarchy, Transactions of Japan Society of Kansei Engineering, 11(1), 2012, 17-26. http://dx.doi.org/10.5057/jjske.11.17
[20] Yanagisawa, H.; Fukuda, S.: Kansei Design by Interactive Reduct Evolutionary Computation: With Attention Paid to Favored Feature of Design, Transactions of the Japan Society of Mechanical Engineers. C, 70(694), 2004, 1802-1809. http://dx.doi.org/10.1299/kikaic.70.1802