Impression Estimation Model of 3D Objects Using Multi-View Convolutional Neural Network

Keisuke Sakashita[1], Kensuke Tobitani[2], Koichi Taguchi[3], Manabu Hashimoto[3], Iori Tani[4], Sho Hashimoto[5], Kenji Katahira[6], and Noriko Nagata[1](✉)

[1] Kwansei Gakuin University, 2-1 Gakuen, Sanda-shi 669-1337, Hyogo, Japan
nagata@kwansei.ac.jp
[2] University of Nagasaki, 1-1-1 Manabino, Nagayo-cho, Nishi-Sonogi-gun, Nagasaki 851-2195, Japan
[3] Chukyo University, 101-2 Yagoto Honmachi, Showa-ku, Nagoya-shi 466-8666, Aichi, Japan
[4] Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe-shi 657-0013, Hyogo, Japan
[5] Seinan Gakuin University, 6-2-92 Nishijin, Sawara-ku, Fukuoka 814-8511, Japan
[6] Waseda University, 1-104 Totsukamachi, Shinjuku-ku, Tokyo 169-8050, Japan

Abstract. The ultimate goal of this study is to provide intuitive design support for 3D objects. As a first attempt, we propose a method for estimating impressions of common 3D objects with various characteristics. Although many studies have been conducted to estimate objects' aesthetics, not enough research has been conducted to estimate the various impressions of objects necessary for design support. The data set of human impressions of 3D objects is constructed based on psychological methods. To account for the variability in people's ratings, the distribution of ratings is represented by a histogram. By learning the distribution of impression ratings, with the estimation model, we can realize an impression estimation model with high estimation accuracy. In the accuracy validation experiment, the proposed method's estimated results (estimated impression distribution) showed a moderate to high positive correlation with the distribution of human impressions. In addition, we confirmed that the proposed method has greater estimation accuracy than previous studies and that it captures the tendency for variation in people's impression evaluations (the global tendency of impression distribution). Furthermore, visual confirmation of the relationship between the estimation results of the constructed impression estimation model and 3D objects suggests that the proposed method is capable of identifying the main physical features associated with impression words, confirming the proposed method's validity.

Keywords: DNN · Impression estimation model · Multi-viewpoint images · Kansei · Aesthetic concepts

1 Introduction

Since the fourth industrial revolution, the rapid development of 3D printer technology and the spread of 3D model databases have created an environment for outputting a

K. Sumi et al. (Eds.): IW-FCV 2022, CCIS 1578, pp. 343–355, 2022.
https://doi.org/10.1007/978-3-031-06381-7_24

variety of 3D modeling objects. These changes in the environment provide a wide range of users with opportunities for personal manufacturing activities. However, there are challenges in the spread of personal manufacturing activities because many users do not have specialized knowledge or skills in manufacturing. Even if users have latent needs for design, it is not easy to reflect them in the form of an object. To provide a wide range of users with opportunities for manufacturing activities in the future, it is necessary to support users so that they can engage in manufacturing intuitively and easily. As one such support, a design support system that searches for and recommends 3D objects based on words that express sensory impressions (e.g., "gay," "hard-looking," etc.) is considered effective because impression words are intuitive and sensory to the user and easily express the user's latent needs. To realize these design support systems, it is necessary to map the relationship between 3D objects' physical features and the impressions the objects evoke. Recently, many studies have been conducted on estimating objects' aesthetics, but these studies have dealt with preferences and feelings toward objects, and not enough research has been conducted on estimating the impressions of objects.

The ultimate goal of this study is to provide intuitive design support for 3D objects, and as a first attempt, we propose a method for estimating the impression of a 3D object. Here, the impression of a 3D object is only based on shape and does not take into account other factors, such as texture or color. For estimation, we use a multi-view convolutional neural network (MVCNN), which takes as input a set of images rendered from multiple viewpoints of a 3D object. The data set of people's impressions of 3D objects is created based on psychological methods. To account for the variation in people's ratings, the distribution of ratings is represented by a histogram. By learning the distribution of ratings, the impression estimation model can achieve an estimation model that has a high correlation with the human impression ratings. Finally, we confirm this study's effectiveness by conducting experiments to verify the proposed method's accuracy.

2 Previous Research

Examples of research utilizing impressions of objects include retrieval technologies that use words expressing impressions (hereafter referred to as "impression words") as queries [1–4]. All of these techniques are realized using the relationship between the object's physical characteristics and the impressions it evokes. Research on such impression estimation has been conducted mainly on two-dimensional images, but there are some reported cases of its application with 3D objects [5, 6]. However, these studies used geometrically simple 3D objects, and no systematic results have been reported on the impression evaluation of general 3D objects with various features.

On the other hand, research has been actively conducted to model human sensibilities toward objects, such as preferences, aesthetic scores, and aesthetic values, using deep-learning techniques [7–9]. However, in many fields, such as psychology, design, and Kansei engineering, it is assumed that a clear distinction is made between the evaluator's emotional response to an object, such as aesthetics, and the properties of the object itself [10]. For example, aesthetic evaluations such as "beautiful" and "preference" are mediated by more emotional reactions in the processing of information about the object's specific properties, such as "gay" and "hard-looking." These properties of objects are called aesthetic concepts [11–13]. In an intuitive design support system, the grasp of aesthetic concepts is indispensable for design creation using the object's properties as a query. Only in a few studies have researchers used deep-learning techniques to model aesthetic concepts. In this study, we define an aesthetic concept as an impression and propose a method for estimating the impression of a 3D object by modeling the relationship between its physical features and the impression of it using deep-learning techniques.

Estimating impressions from 3D objects can be described as a recognition problem based on classification and regression, in which the objective variable is the impression evaluation value and the explanatory variable is the physical-feature value of the 3D object. Although research on 3D-object recognition was initially focused on the design of 3D features [14], recently, end-to-end learning and recognition methods using deep neural networks (DNNs) have become mainstream. Recognition methods based on this DNN can be roughly classified as RGBD-based [15, 16], Point Cloud-based [17, 18], Voxel-based [19, 20], and multi-view-based [21, 22]. Currently, multi-view-based methods are considered more accurate than others in large-scale 3D-object recognition [23]. Studies of aesthetic estimation of 3D objects using MVCNN [21], a multi-view-based method [9], have been reported. However, there have been no reports on impression estimation of 3D objects. Therefore, we propose constructing an impression estimation model for 3D objects.

In addition, for the selection of impression words to be used in the construction of the impression estimation model, it is necessary to select impressions that are necessary and sufficient for intuitive design support. We have conducted impression evaluation experiments using various abstract shapes based on psychological methods and clarified the structure of impression evaluation that can be common to various 3D shapes [10, 24]. In this study, too, the quantification of impressions of 3D objects is based on psychological methods and takes into account the structure of people's evaluations of 3D objects. This effort yields valid impression evaluation data. Furthermore, by using a multi-view-based DNN, this data set is trained in an end-to-end manner to achieve impression estimation of various 3D objects. The estimation model is constructed using 3D-model data sets of multiple object categories with human impression evaluation data, and we will determine whether the MVCNN, which has shown high-quality performance in the 3D-object recognition task, is also effective in the 3D-object impression estimation task.

3 Modeling the Relationships Between Visual Impressions and Physical Characteristics

In this study, we propose an impression estimation method for 3D objects using a multi-view convolutional neural network (MVCNN) [21]. Figure 1 shows the proposed method's basic design. First, we conduct impression evaluation experiments on 3D objects to quantify the impressions 3D objects evokes. We use the resulting impression distribution as a supervisory signal for the model. The distribution of impressions includes the variability of evaluations caused by differences in individuals' sensory evaluation tendencies. Next, we create a multi-view image of a 3D object rendered from multiple viewpoints and used it as an input signal for the model. Finally, we use a MVCNN to model the relationship between the aforementioned input signal and the supervisory signal to achieve impression estimation of the 3D object. The above basic design enables us to verify the task of impression estimation of 3D objects using MVCNN.

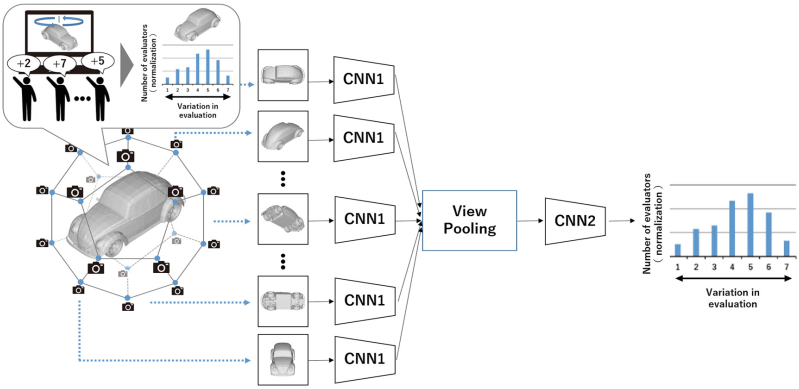

Fig. 1. Overview of impression estimation of 3D objects

4 Building an Impression Estimation Model

4.1 Data Set

In this section, we show how to create the dataset needed to build the impression estimation model.

Collection and Selection of 3D Model Data. We collected and selected 3D model data. To ensure the diversity and comprehensiveness of shape expressions, we collected 3D models from ShapeNet [23], ModelNet40 [25], and CG DATA BANK [26], which are large-scale databases of 3D models. As a result, we collected 3D model data from the Car (632 models), Vase (575 models), and Chair categories (985 models).

Rendering a 3D Model Data for Multi-viewpoints Images and Stimulus Presentation Video. We used the 3D model to create experimental stimuli. There are two types of experimental stimuli: multi-view images used as input signal for the estimation model and images for impression evaluation experiments. We used the experimental images to present to the subjects when they evaluated the impressions the 3D model evokes. The common pre-processing for creating both experimental stimuli is shown below. First, we obtained each 3D model's size, and we standardized their sizes among the models by scale conversion based on the values. Next, we used the local reference frame (LRF) to unify the 3D models' postures. We describe the rendering method below. We used Phong's reflection model for rendering. We rendered the multi-view image from each vertex of the dodecahedron in the direction of the center of gravity of the 3D model encapsulated in the dodecahedron. In this way, we created a set of 20 multi-view images. To render the experimental images, we set the camera position at 18° vertical to the ground to capture the 3D model rotating in the horizontal direction.

Assigning Impression Evaluation Values. To quantify the impressions of 3D objects, we conducted impression evaluation experiments using the semantic differential method (SD) based on the findings of previous studies [24]. We used three adjective pairs, "soft-hard," "gay-sober," and "stable-stable," as evaluation words. We selected these adjective pairs based on the findings of a previous study [24] that the three factors (evaluation, activity, and potency) [27] that frequently appear in SD studies also constitute the main criteria for impressions of 3D shapes. In the evaluation of impressions, we presented the images for the aforementioned impression evaluation experiment, and we evaluated the impressions from various directions comprehensively. As a result of the experiment, we obtained 7 levels (-3 to $+3$) of impression evaluation values for the three adjective pairs. Next, we numerically assigned the seven ratings (-3 to $+3$) to class labels from 1 to 7, and we used the discrete probability distributions as the impression distributions in the 3D model. Because the number of raters differed among the samples, we normalized the impression distribution by the number of raters in each sample. We used the normalized impression distribution as the supervisory signal for the estimation model. Table 1 shows the specification of the created data set.

In this study, we conducted the impression evaluation experiment by crowdsourcing, so the evaluation data was cleaned from the content and response time of the subjects. We describe the cleaning procedure below. (1) We designated experimental participants who could not successfully complete the evaluation due to malfunction of the response system or network conditions as unevaluated respondents and eliminated all of their evaluation data. (2) We checked experimental participants who answered 0 (middle of the 7-point scale: neither) to one or more models, and considered insincere respondents, and all of their evaluation data were eliminated. (3) We checked the distribution of the insincere respondents' response time, and we designated the experimental participants who finished their responses in 156 s or less, which was the most frequent value, as short-time respondents and eliminated all of their evaluation data. As a result, of the 6101 total respondents, 388 were unevaluated respondents, 628 were dishonest respondents, and 307 were short-time respondents, so the final number of valid respondents was 4778.

Table 1. Data set specifications

3D object category	Car Vase Chair
Number of sample	632 575 985
Database	ShapeNet, ModelNet40, CG DATA BANK
Number of evaluations per model	20-40 people
Experiment participants	Men and women (ages 20-60)
Presentation format of the 3D model	Model rotation video
Evaluated word	3 adjective pairs
Rating scale	7-step SD

Table 2. Detail of DNN architecture

Layer		Input size	Output size	Kernel	Stride
CNN1	conv1	227 × 227 × 3	55 × 55 × 96	11 × 11	4
	maxpool1	55 × 55 × 96	27 × 27 × 96	3 × 3	2
	conv2	27 × 27 × 96	27 × 27 × 256	5 × 5	1
	maxpool2	27 × 27 × 256	13 × 13 × 256	3 × 3	2
	conv3	13 × 13 × 256	13 × 13 × 384	3 × 3	1
	conv4	13 × 13 × 384	13 × 13 × 384	3 × 3	1
	conv5	13 × 13 × 384	13 × 13 × 256	3 × 3	1
	maxpool3	13 × 13 × 256	6 × 6 × 2526	3 × 3	2
View pooling		20 × 9216	9216	–	–
CNN2	fc1	9216	4608	–	–
	fc2	4608	4608	–	–
	fc3	4608	7	–	–

4.2 Training

In this study, we used a 3D model database (632 car categories, 575 vase categories, and 985 chair categories) with the impression distribution shown in Sect. 4.1 for model training and evaluation. We constructed the model using a total of nine combinations of object categories and adjective pairs. For cross-validation, we divided the data set into train, validation, and test (8:1:1) and adopted a 9-fold cross-validation.

Next, we describe the structure of the DNNs we used in the model (Table 2). The structure of the CNN1 layer is based on AlexNet. The CNN1 layer shares the weights in the CNN to be optimized with each viewpoint. The view-pooling layer smooths the image features of each viewpoint output from the CNN1 layer into one dimension, combines them in the row direction, and extracts the values of the viewpoint with the largest value one column at a time. In other words, the view-pooling layer is responsible for selecting the viewpoints that are effective for impression estimation. The activation function for the output layer is the softmax function. We used Adam as the optimization algorithm for training. To avoid the gradient vanishing problem that is a concern in DNNs, we used rectified linear units as the activation function. The loss function is the cross-entropy error. We set the learning rate to 0.001, the number of epochs to 300 for the Car and Vase category and 200 for the Chair category.

5 Results and Discussions

To verify the proposed method's effectiveness, we conducted a validation experiment using several comparison methods, 3D ShapeNets [19], a Voxel-based method, and single-view CNN (SVCNN), which inputs a single-view image to the proposed method.

We used correlation coefficients and mean squared errors as evaluation measures of estimation accuracy. We calculated the correlation coefficient by converting the distribution of human impressions and the estimation results (estimated impression distribution) into expected values. We calculated the mean squared error from the sum of the errors of each class in the human impression distribution and the estimated impression distribution.

Table 3. Average of Correlation coefficient (LCC) and Mean squared error (MSE)

Car

Method	Input Format	soft-hard		gay-sober		stable-unstable	
Index	-	LCC	MSE	LCC	MSE	LCC	MSE
3D ShapeNets	voxel	-0.10	0.06	0.07	0.09	0.01	0.07
SVCNN	single-image	0.54	0.05	0.72	0.05	0.21	0.05
MVCNN	multi-image	0.60	0.05	0.78	0.05	0.31	0.05

Vase

Method	Input Format	soft-hard		gay-sober		stable-unstable	
Index	-	LCC	MSE	LCC	MSE	LCC	MSE
3D ShapeNets	voxel	0.01	0.08	0.07	0.09	-0.06	0.14
SVCNN	single-image	0.32	0.05	0.48	0.05	0.65	0.06
MVCNN	multi-image	0.46	0.04	0.54	0.04	0.74	0.05

Chair

Method	Input Format	soft-hard		gay-sober		stable-unstable	
Index	-	LCC	MSE	LCC	MSE	LCC	MSE
3D ShapeNets	voxel	0.10	0.08	-0.28	0.11	0.16	0.09
SVCNN	single-image	0.58	0.05	0.59	0.05	0.62	0.05
MVCNN	multi-image	0.60	0.05	0.60	0.05	0.62	0.05

5.1 Overall Performance

Table 3 shows the average of the correlation coefficients and mean squared errors obtained in each verification. Table 3 shows that the proposed method showed a strong positive correlation in two of the nine conditions and a moderate positive correlation in six conditions. We thus confirmed the proposed method's practical effectiveness.

Next, we confirm the superiority of the proposed method over the comparison methods. The estimation accuracy of the proposed method is much better than that of 3D ShapeNets. In the comparison between the proposed method and SVCNN, we confirmed that the accuracy of the proposed method is higher than that of SVCNN, although the improvement is not as high as that of 3D ShapeNets. In particular, the correlation coefficients of the proposed method improved by about 0.1, compared to SVCNN in the "soft-hard" condition of the Vase category. In the next section, we will discuss the comparison of methods and object categories.

5.2 Comparison of the Proposed Method with Each Comparison Method

To compare the methods' accuracy, we conducted a multiple comparison test using the Dunnett method with the proposed method (MVCNN) as the control group and each compared method as the treatment group. The alternative hypothesis is $\mu c > \mu i$ (μc: mean value of the correlation coefficient or mean square error of the control group, μi: mean value of the correlation coefficient or mean square error of the treatment group).

First, we compare MVCNN to 3D ShapeNets. The results of the significance test between the proposed method (MVCNN) and 3D ShapeNets showed that both indices were significantly different ($p < 0.05$) in 9 out of 9 conditions. This result can be attributed to the fact that MVCNN's input signal has a higher resolution than that of 3D ShapeNets. On the other hand, 3D ShapeNets is a method that shows a high recognition rate of about 77% in the 3D object recognition task [19]. When we applied this method to the impression estimation task, the estimation accuracy decreased significantly. This result suggests that the impression estimation task requires a higher-resolution representation of the object shape than the 3D object recognition task.

Next, we compare the MVCNN and the SVCNN. The results of the significance test between the proposed method (MVCNN) and SVCNN differed in the evaluation indices. The correlation coefficient showed a significant difference ($p < 0.05$) in five conditions. In contrast, the mean squared error was not significantly different in all nine conditions. These results can be attributed to the different nature of the two evaluation indices. In this study, we calculated the correlation coefficient from the expected value of the separation probability distribution. Because the expected value is weighted by the probability of belonging to each class in the impression distribution, it is more suitable for evaluating the global trend of the distribution than the mean square error. The fact that the proposed method showed significant differences only in the correlation coefficient indicates that the proposed method captures the global tendency of the impression distribution, i.e., the tendency of the variation of people's impression evaluation, better than SVCNN. In other words, we suggest that the proposed method captures the tendency of the variability of human impression evaluation better by using the shape information from multiple viewpoints rather than from a single viewpoint.

5.3 Comparison of Impression Estimates for Each Object Category

We focus on three object categories. In all object categories, the estimation accuracy of the proposed method is better than that of SVCNN. However, there are some conditions where the effect of using MVCNN is stronger and some conditions where it is weaker, depending on the inherent properties and structures related to the shapes of the object categories. Specifically, the estimation accuracy of the Car category and the Vase category tends to be improved by using MVCNN, while the accuracy of the Chair category remains unchanged. From these results, we can confirm that the proposed method is superior to SVCNN, especially in the object category condition where the appearance and impression of the shape changes depending on the viewpoint. We believe that quantifying the impressions of 3D objects for each viewpoint and learning the distribution of impressions will be effective in improving the estimation accuracy. Furthermore, the introduction of an EMD-based loss function can be used to improve the accuracy [8].

5.4 Relationship Between Estimated Evaluation Value and 3D Models

In this section, we will visually check the relationship between the model's estimation results and the 3D object's shape. We assigned estimated evaluation values to the data for test and converted them to expected values. For each combination of object category and evaluation term, we adopted the model with the highest accuracy in the cross-validation for estimation. For each object category, we sorted the expected values in descending order and identified the top and bottom 15 samples. Figure 2, 3, and 4 show the results.

From the results in Fig. 2, we can see that the impression estimation model we have developed is mainly based on the "soft-hard" condition of the Car category in terms of shapes such as corners, straight lines, curves, and curved surfaces; the "gay-sober" condition in terms of shapes such as overlapping surfaces and edges; the "stable-unstable" condition was mainly characterized by characteristics such as height and length. Next, with the results in Fig. 3, we confirmed that the "soft-hard" condition of the Vase category was characterized mainly by the shapes of corners, straight lines, curves, and curved surfaces and that the "gay-sober" condition was characterized mainly by the shapes of overlapping surfaces and edges. Finally, from the results in Fig. 4, we confirmed that

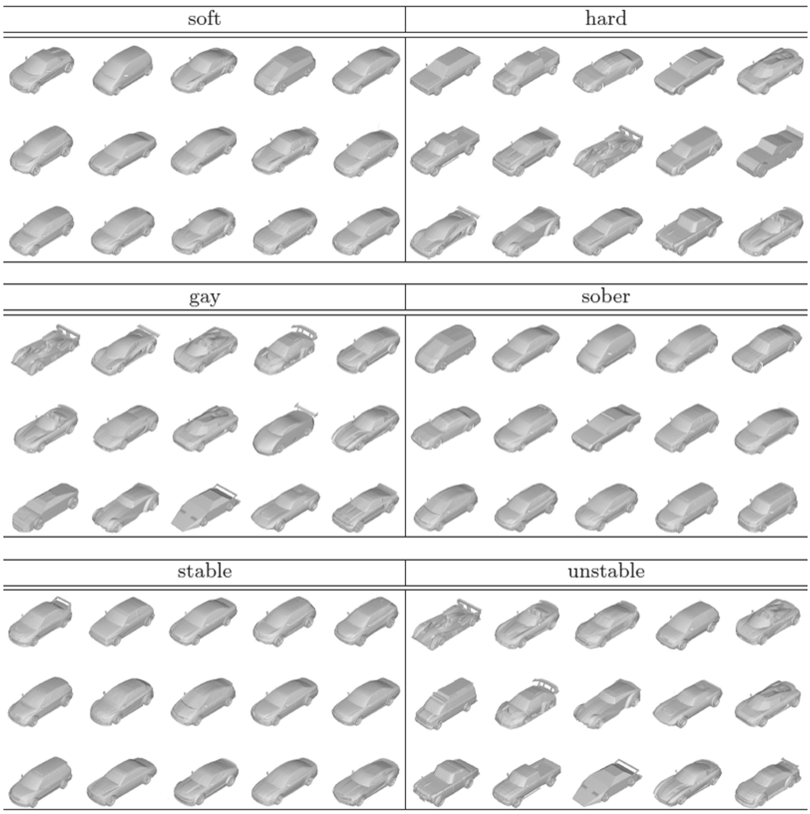

Fig. 2. Top samples of estimation score for each impression word in car. The figure is sorted from top left to bottom right in the order descending of estimation score for each impression word.

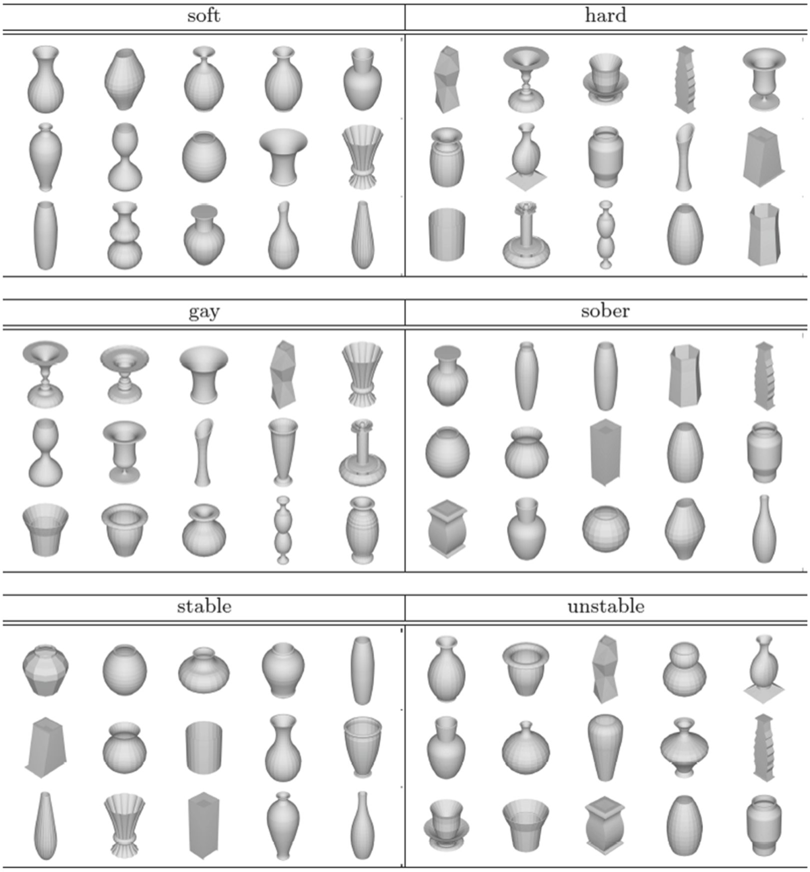

Fig. 3. Top samples of estimation score for each impression word in vase. The figure is sorted from top left to bottom right in the order descending of estimation score for each impression word.

the "soft-hard" condition of the Chair category was characterized mainly by the shape of the corners, straight lines, curves, curved surfaces, etc. and that the "stable-unstable" condition was characterized mainly by the bottom surface's shape.

These results suggest that the "soft-hard" conditions are distinguished by the bottom surface's shape and the shape of the corners, straight lines, curves, and curved surfaces, and the "gay-sober" condition is characterized by the shape of the overlapping surfaces and edges. The fact that these impression evaluation criteria are common among object categories suggests that the proposed method is capable of identifying the main physical features associated with the impression words. The interpretation of the relationship between these physical features and impressions was also reported in a study on the quantification of the structure of human impression evaluation of 3D objects [28], confirming our claim's validity.

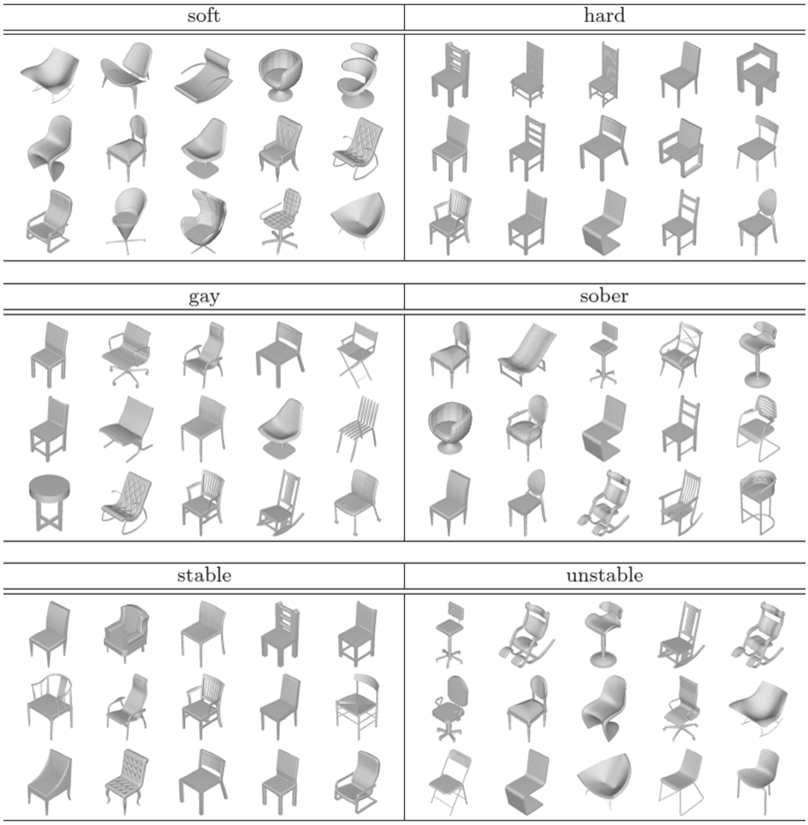

Fig. 4. Top samples of estimation score for each impression word in chair. The figure is sorted from top left to bottom right in the order descending of estimation score for each impression word.

On the other hand, some conditions made it difficult to interpret the relationship between impressions and shape, such as Chair's "gay-sober" and Vase's "stable-unstable" conditions. For such conditions, it is considered possible to capture the relationship visually between shape and impression by using visualization techniques such as Grad-Cam [29], which can explain CNN's decision criteria.

6 Conclusion

The ultimate goal of this study was to provide intuitive design support for 3D objects, and as a first attempt, we proposed an impression estimation method for 3D objects. We used a multi-view convolutional neural network to estimate impressions of various 3D objects. For training and evaluation of the estimation model, we used 3D model data, to which we added impression evaluation data. To ensure the validity of the impression evaluation data, quantification of the impressions evoked by the 3D objects was based on psychological methods.

13. Chen, Y., Chen, D., Han, X., Huang, X.: Generic and specific impression estimation of clothing fabric images based on machine learning. In: 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 1753–1757. IEEE, Zhangjiajie (2015)
14. Hashimoto, M., Akizuki, S., Takei, S.: A survey and technology trends of 3D features for object recognition. Electron. Commun. Japan **100**(11), 31–42 (2017)
15. Lai, K., Bo, L., Ren, X., Fox, D.: Sparse distance learning for object recognition combining RGB and depth information. In: 2011 IEEE International Conference on Robotics and Automation, pp. 4007–4013. IEEE, Shanghai (2011)
16. Song, X., Herranz, L., Jiang, S.: Depth cnns for RGB-D scene recognition: Learning from scratch better than transferring from RGB-CNNs. In: Thirty-first AAAI Conference on Artificial Intelligence, vol. 28, Issue 2, pp. 4271–4277. San Francisco (2017)
17. Li, J., Chen, B.M., Lee, G.H.: SO-Net: Self-organizing network for point cloud analysis. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9397–9406. IEEE, Salt Lake (2018)
18. Deng, H., Birdal, T., Ilic, S.: PPFNet: Global context aware local features for robust 3d point matching. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 195–205. IEEE, Salt Lake (2018)
19. Wu, Z., Shuran, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D ShapeNets: a deep representation for volumetric shapes. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1912–1920. IEEE, Boston (2015)
20. Sedaghat, N., Zolfaghari, M., Amiri, E., Brox, T.: Orientation-boosted voxel nets for 3D object recognition. arXiv:1604.03351 (2017)
21. Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: Multi-view convolutional neural networks for 3D shape recognition. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 945–953. IEEE, Santiago (2015)
22. Kanezaki, A., Matsushita, Y., Nishida, Y.: RotationNet: joint object categorization and pose estimation using multiviews from unsupervised viewpoints. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5010–5019. IEEE, Salt Lake City (2018)
23. Large-scale 3D shape retrieval from ShapeNet Core55 Homepage. https://shapenet.cs.stanford.edu/shrec17/
24. Katahira, K., et al.: Major factor in kansei evaluation of 3D objects. Trans. Japan Soc. Kansei Eng. **15**(4), 563–570 (2016)
25. Modelnet40 Homepage. https://modelnet.cs.princeton.edu/
26. Cgdata bank Homepage. https://cgdatabank.com/
27. Osgood, C., Suci, G., Tannenbaum, P.: The Measurement of Meaning. University of Illinois Press (1957)
28. Miyai, S., Katahira, K., Sugimoto, M., Nagata, N., Nikata, K., Kawasaki, K.: Hierarchical structuring of the impressions of 3D shapes targeting for art and non-art university students. In: Stephanidis, C. (ed.) HCII 2019. CCIS, vol. 1032, pp. 385–393. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23522-2_50
29. Selvaraju, R.R., Das, A., Vedantam, R.: Grad-CAM: why did you say that? arXiv:1611.07450 (2016)