Mathematical and Computational Applications

MDPI

Article

Furniture Style Compatibility Estimation by Multi-Branch Deep Siamese Network

Ayumu Taisho [1,*], Keiko Ono [2], Erina Makihara [2], Naoya Ikushima [1] and Sohei Yamakawa [1]

[1] Graduate School of Science and Engineering, Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe 610-0394, Kyoto, Japan
[2] Department of Science and Engineering, Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe 610-0394, Kyoto, Japan
* Correspondence: taisho.ayumu@mikilab.doshisha.ac.jp

Abstract: As demands for understanding visual style among interior scenes increase, estimating style compatibility is becoming challenging. In particular, furniture styles are difficult to define due to their various elements, such as color and shape. As a result, furniture style is an ambiguous concept. To reduce ambiguity, Siamese networks have frequently been used to estimate style compatibility by adding various features that represent the style. However, it is still difficult to accurately represent a furniture's style, even when using alternate features associated with the images. In this paper, we propose a new Siamese model that can learn from several furniture images simultaneously. Specifically, we propose a one-to-many ratio input method to maintain high performance even when inputs are ambiguous. We also propose a new metric for evaluating Siamese networks. The conventional metric, the area under the ROC curve (AUC), does not reveal the actual distance between styles. Therefore, the proposed metric quantitatively evaluates the distance between styles by using the distance between the embedding of each furniture image. Experiments show that the proposed model improved the AUC from 0.672 to 0.721 and outperformed the conventional Siamese model in terms of the proposed metric.

Keywords: Siamese network; CNN; furniture; style; compatibility; ambiguity

Citation: Taisho, A.; Ono, K.; Makihara, E.; Ikushima, N.; Yamakawa, S. Furniture Style Compatibility Estimation by Multi-Branch Deep Siamese Network. *Math. Comput. Appl.* **2022**, *27*, 76. https://doi.org/10.3390/mca27050076

Academic Editor: Leonardo Trujillo

Received: 27 July 2022
Accepted: 31 August 2022
Published: 4 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Understanding visual styles in interior scenes has attracted enormous interest in various domains, such as art, advertising, and e-commerce [1–3]. Visual styles are typically estimated based on reference images, so the dynamics of user preferences by assessing visual styles when user-preferred images are given [4,5]. Specifically, the furniture includes various kinds of information, such as color, shape, size, material, and texture; thus, detecting appropriate visual features is challenging for visual style detection [6,7]. In image processing, various visual features related to HOG, SIFT, and SURF features have been proposed [8–10]. A more recent trend is automatically using deep visual features to extract complex features [11,12]. To extract the deep visual features, the Flatten layer is used before the dense layer in a Deep Convolutional Neural Network. The deep visual features better measure visual similarities between furniture images than conventional image processing methods.

Meanwhile, style congruence is essential for furniture. A few studies have focused on furniture style compatibility [13–15], utilizing visual embeddings in Euclidian space to classify complex boundaries by outputs from multiple networks. In these studies, the Siamese network, a Deep Metric Learning method, was used to evaluate the compatibility of furniture styles. When mapping to a feature space, Siamese networks optimize the Euclidean distance between items, so similar items are close and vice versa [16]. Thus, Siamese networks can not only estimate the similarity but also estimate the degree of similarity.

However, furniture style is a concept that is an ambiguous concept that is difficult to define. For example, while some people may describe a chair as modern, others can describe it as traditional. Therefore, to accurately estimate the similarity, the ambiguity must be alleviated. Aggarwal et al. improved furniture style compatibility by combining classification loss when training a regular Siamese network [13]. Weiss et al. also accurately represented furniture styles by assigning multiple possible applicable style labels to furniture [17]. Each of these studies addressed furniture style ambiguity using additional information when training the Siamese network. However, adequately representing furniture styles is difficult with supplemental information. Therefore, the conventional structure of the Siamese network, which compares two furniture images, does not facilitate accurate evaluation.

In this paper, we propose a model that improves the ambiguous conventional Siamese network and utilizes multiple furniture images. Specifically, multiple furniture images are utilized in a one-to-many ratio, and all images in "many" are of the same style to evaluate similarity. Consequently, the model learns the style's characteristics more accurately, even when ambiguous furniture images are given as input. Moreover, the proposed Simaese network, that learns in a one-to-many ratio, can infer the compatibility between a furniture image and other same-style furniture images. Although furniture of the same style is not necessarily interchangeable, as a first step in proposing a new model, this study assumes that furniture of the same style is interchangeable and verifies the performance.

We conducted an evaluation experiment using the Bonn Furniture Styles Dataset [13] and used the area under the ROC curve (AUC), a commonly used metric in compatibility estimation, to verify the effectiveness of the proposed method. However, the AUC evaluation does not indicate the distance between styles in the feature space. Therefore, to analyze distance in the feature space, we propose a metric called Styles Difference Distance (SDD) that represents the distance between each style. The results show that the proposed method improves the accuracy of determining style similarity among furniture and better maps the distance between different styles in the feature space. That is, furniture items with the same style are placed closer together in the feature space and vice versa. Our main contributions are summarized as follows:

(1) We propose a method that learns multiple furniture images in a one-to-many ratio that expands a Siamese network. Compared to the conventional Siamese network, which uses two images in a one-to-one ratio, it can better estimate the similarity of furniture.
(2) We proposed SDD as a new evaluation scale for compatibility assessment. Using the test dataset, we analyzed the Euclidean distance of each style for both the proposed and conventional methods. As a result, the different styles were successfully placed farther apart in the feature space.
(3) The proposed method can recommend furniture that fits well with multiple pieces of furniture; because of its input method, it can search for furniture that fits the style of all items.

2. Related Work

Because furniture style is an ambiguous concept that is difficult to define quantitatively, the degree to which a piece of furniture belongs to a style differs. For example, while there could be a classic modern chair, there could be a chair that is only slightly modern. Therefore, it is necessary to go beyond conventional style-based furniture classification to evaluate style compatibility among furniture quantitatively. Thus, Siamese networks are often used in this field. In addition, since the ambiguity of furniture styles makes accurate learning difficult, there is a need to mitigate ambiguity. For this reason, research has been conducted to train a Siamese network using additional information representing furniture styles along with its usual image features.

Aggarwal et al. exploited classification loss when training a Siamese network [13]. Specifically, a softmax layer was added to the subnetwork of the Siamese network to simultaneously learn image features and the classification loss associated with the clas-

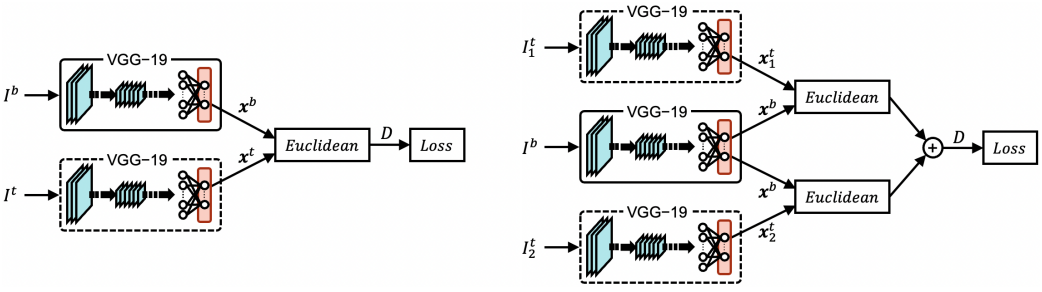
sification results. Consequently, they succeeded in improving the accuracy of furniture style compatibility evaluation. Weiss et al. also gave multiple style labels to furniture to provide a detailed representation of ambiguous furniture styles [17]. Specifically, ten interior designers assigned a style to each piece of furniture to ensure accurate learning. Bell et al. proposed a method for learning the similarity between symbolic furniture images on a white background and actual room images [18]. Li et al. proposed learning the joint embedding of images and 3D models [19].

In these studies, while distinctive features were applied to the Siamese network, the network's structure was a conventional Siamese network that compares two images. As a result, even if features are devised, it is difficult to learn accurately if either image is ambiguous. Therefore, we propose a Siamese network that learns in a one-to-many ratio between multiple furniture images. Learning in a one-to-many ratio alleviates style ambiguity and brings compatible items closer together in the feature space. Compared to the conventional method, which uses two images, our method maintains accurate learning even when ambiguous images are mixed.

3. Siamese Network

A Siamese network is a type of Deep Metric Learning that consists of two sub-networks and a distance function. Due to its structure, the Siamese network uses two input images and is trained by comparing the compatibility of two input images. In other words, embeddings of similar images are closer together, and dissimilar images are farther apart in the feature space. When the model is trained, the model is optimized by inputting a pair of images into a sub-network with shared weights. It then adjusts the distance between the paired images using a distance function. Therefore, the Siamese network can classify images and quantitatively measure the similarity of images. Because it learns the embedding method, the Siamese network can robustly respond to unknown classes, thus making it effective for few-shot and one-shot learning [16,20]. In addition, even when the training data is insufficient, the Siamese network can ensure sustainable training because it is an excellent model for discriminating visual similarities. Therefore, the Siamese network is effective for simple classifications and clustering of images, but it can also be used to estimate compatibility regarding the affinity between items because it can measure the degree of similarity [21–24].

The conventional Siamese model used in this study is shown in Figure 1a. As this study focuses on images, VGG-19 [25], a CNN that excels in image classification, was used as the sub-network. VGG-19 is pretrained on Imagenet [26]. In addition, this study assumes a dataset with multiple styles (e.g., Modern and Asian) and categories (e.g., chair and table) to learn and estimate style compatibility.



(**a**) (**b**)

Figure 1. Structure of Siamese networks. (**a**) Conventional Siamese model. (**b**) Proposed Siamese model. $J = 2$ in I_j^t ; $j = 1, 2, \cdots, J$.

In the conventional Siamese model, (I^b, I^t) is the input to VGG-19 with shared weights, with I^b as the base image and I^t as the target image. Note that the furniture category C of the input (I^b, I^t) is defined as $C(I^b) \neq C(I^t)$ since the goal is to estimate style compatibility independent of the furniture category. Then, the 512-dimensional output $(\boldsymbol{x}^b, \boldsymbol{x}^t)$ from

VGG-19 is embedded in the feature space, and the Euclidean distance D between $(\boldsymbol{x}^b, \boldsymbol{x}^t)$ is calculated. The Euclidean distance D is defined by the as follows:

$$D(I^b, I^t) = \| \boldsymbol{x}^b - \boldsymbol{x}^t \|_2 \tag{1}$$

To achieve optimal D, the conventional Siamese model employs Contrastive Loss [16] as the loss function. Contrastive Loss can learn embeddings so that D is smaller when (I^b, I^t) are in the same style and D is farther away when (I^b, I^t) are in different styles. First, in the style compatibility estimation, when (I^b, I^t) are in the same style, i.e., when they are positively compatible, the loss L_p is expressed as

$$L_p(I^b, I^t) = \frac{1}{2} D^2 \tag{2}$$

From (2), the smaller D is, the smaller L_p becomes, which converges to 0. Therefore, it is possible to make the distance between positively compatible (I^b, I^t) smaller. Meanwhile, when (I^b, I^t) are in different styles, i.e., negatively compatible, the loss L_n as

$$L_n(I^b, I^t) = \frac{1}{2} \max[m - D, 0]^2 \tag{3}$$

In (3), the margin m is a hyper-parameter, and L_n converges to 0 when D is greater than m. As a result, it is possible to move negatively compatible items (I^b, I^t) apart. Contrastive Loss is the combination of L_p and L_n and is expressed by the following equation:

$$L_{con}(I^b, I^t, Y) = \frac{1}{2}(YD^2 + (1 - Y)\max[m - D, 0]^2) \tag{4}$$

In (4), label Y is equal to 1 when (I^b, I^t) are positively compatible and equal to 0 when (I^b, I^t) are negatively compatible. This allows L_{con} to apply L_p and L_n depending on the input image to achieve optimal D.

4. Proposed Model

4.1. Proposed Siamese Architecture

To improve the accuracy of estimating furniture style compatibility, we propose a Siamese architecture that learns using three or more input images by increasing I^t to multiple images. The structure of the proposed Siamese model is shown in Figure 1b. As in the conventional Siamese model, the sub-network of the proposed model uses VGG-19, which is pretrained with ImageNet. Moreover, as stated in Section 3, the proposed model assumes a dataset with multiple styles and categories to learn and estimate style compatibility.

In the proposed Siamese model, the multiple input images are defined as $(I^b, \boldsymbol{I}^t)$, $\boldsymbol{I}^t = \{I_j^t \; ; \; j = 1, 2, \cdots, J\}$, where J denotes the number of $\boldsymbol{I}^t$ pieces to be compared with I^b. The input images are each input to a VGG-19 with shared weights. Note that because the proposed Siamese model aims to estimate style compatibility independent of the furniture category (e.g., chair, table), the furniture category C of the input $(I^b, \boldsymbol{I}^t)$ is $C(I^b) \neq C(I_j^t) \neq C(I_i^t)$, $j \neq i$. Next, the 512-dimensional outputs $(\boldsymbol{x}^b, \boldsymbol{\chi}^t)$, $\boldsymbol{\chi}^t = \{\boldsymbol{x}_j^t \; ; \; j = 1, 2, \cdots, J\}$ from each VGG-19 are embedded in the feature space. Finally, the sum D of Euclidean distances between $\boldsymbol{x}^b$ and each $\boldsymbol{x}_j^t$ is calculated, where D is defined as follows:

$$D(I^b, \boldsymbol{I}^t) = \sum_{j=1}^{J} \| \boldsymbol{x}^b - \boldsymbol{x}_j^t \|_2 \tag{5}$$

From (5), the proposed Siamese model learns compatibility between I^t and multiple $\boldsymbol{I}^t$ simultaneously. In other words, when $(I^b, \boldsymbol{I}^t)$ are positively compatible, I^b is moved closer to multiple $\boldsymbol{I}^t$, and when $(I^b, \boldsymbol{I}^t)$ are negatively compatible, I^b is placed farther away from

multiple $\boldsymbol{I}^t$. To achieve optimal D, the proposed Siamese model also employs Contrastive Loss as the loss function and learns in the same way as the conventional Siamese model.

The most distinctive feature of the proposed Siamese model is that it optimizes the network in a one-to-many ratio using three or more input images. This structure is of great significance when evaluating the compatibility of furniture styles that contain ambiguity. Furniture styles consider a composite of various factors, resulting in variations in embedding in the feature space. That is, furniture images often deviate from the distribution of embeddings for each style. Therefore, the conventional Siamese model has difficulty learning accurately when two input images contain ambiguous furniture images, i.e., embeddings far from the center of the distribution. Moreover, in the proposed Siamese model that utilizes three or more input images, the impact per image becomes smaller by increasing the number of $\boldsymbol{I}^t$, according to (5). Therefore, even if $\boldsymbol{I}^t$ contains furniture images far from the center of the distribution, the presence of other $\boldsymbol{I}^t$ reduces the influence of such furniture images, and the style can be learned. In other words, the proposed Siamese model can guarantee style learning by increasing the number of $\boldsymbol{I}^t$ to reduce the possibility that only ambiguous furniture is included in $\boldsymbol{I}^t$. In addition, at the time of inference, the proposed Siamese model infers compatible furniture for multiple $\boldsymbol{I}^t$. Therefore, compared to the conventional Siamese model, which can only infer furniture compatible with one piece of I^t, the proposed Siamese model, which can consider compatibility with multiple pieces of furniture, is more convenient.

4.2. Style Difference Distance

To quantitatively evaluate the ability of the proposed Siamese model to estimate style compatibility, we propose the SDD evaluation measure. SDD indicates the distance between $S(\boldsymbol{I}^t)$ and other styles, which is calculated by computing the Euclidean distance between $\boldsymbol{I}^t$ and a large number of $\boldsymbol{I}^b$. For example, when the style of $S(\boldsymbol{I}^t)$ is Modern, the difference of the average Euclidean distance between $\boldsymbol{I}^t$ and all Modern $\boldsymbol{I}^b$, and $\boldsymbol{I}^t$ and all Traditional $\boldsymbol{I}^b$ is the SDD of Traditional when $S(\boldsymbol{I}^t)$ is Modern. In other words, SDD can clarify the distance between $S(\boldsymbol{I}^t)$ and other styles; the larger the SDD, the greater the distance is between $S(\boldsymbol{I}^t)$ and other styles. This means that furniture that is farther away is less likely to be recommended during inference, and furniture of the same style as $S(\boldsymbol{I}^t)$ can be inferred with high accuracy; accordingly, the larger the SDD is, the higher the performance becomes. Thus, the approach to the distance between styles is a practical and essential factor in inference. Therefore, SDD is an effective evaluation measure that approaches the distance between styles, which is not considered by the AUC, a conventionally used evaluation measure.

SDD was used when evaluating the proposed method with a furniture image dataset of multiple furniture categories and styles. In particular, the style compatibility was quantitatively estimated by analyzing the distance D between I^b and $\boldsymbol{I}^t$. Because there are multiple candidates of $\boldsymbol{I}^b$ for $\boldsymbol{I}^t$, we define all possible furniture image sets as $\boldsymbol{I}^{b'} = \{I_k^{b'}; k = 1, 2, \cdots, K\}$ from the dataset. This means that the distance to all furniture images that are different from the furniture category of $\boldsymbol{I}^t$ is measured. The calculation of SDD is shown below. Let $S = \{s_v; v = 1, 2, \cdots, V\}$ be the total style and $S(\boldsymbol{I}^t) = r$ be the style of $\boldsymbol{I}^t$. $(\boldsymbol{I}_{s_v}^{b'}, \boldsymbol{I}_r^{b'})$ denotes the $\boldsymbol{I}^{b'}$ whose style is (s_v, r).

$$SDD_{s_v} = \frac{1}{|\boldsymbol{I}_{s_v}^{b'}|} \sum_{k=1}^{|\boldsymbol{I}_{s_v}^{b'}|} D(\boldsymbol{I}_{s_v,k}^{b'}, \boldsymbol{I}^t) - \frac{1}{|\boldsymbol{I}_r^{b'}|} \sum_{k=1}^{|\boldsymbol{I}_r^{b'}|} D(\boldsymbol{I}_{r,k}^{b'}, \boldsymbol{I}^t) \tag{6}$$

SDD is the distance between both the style different from $\boldsymbol{I}^t$ and the style same as $\boldsymbol{I}^t$. Therefore, the derived value of SDD is equal to the number of styles in the dataset, and the output is a histogram. In (6), SDD_{s_v} is the SDD of one out of all the styles, where $(|\boldsymbol{I}_{s_v}^{b'}|, |\boldsymbol{I}_r^{b'}|)$ denotes the total number of $(\boldsymbol{I}_{s_v}^{b'}, \boldsymbol{I}_r^{b'})$. Thus, (6) calculates the distance between (s_v, r), i.e., the average difference between a style and a style identical to $\boldsymbol{I}^t$. Note that

when calculating the SDD of the same style as $\boldsymbol{I}^t$, i.e., $s_v = r$, $SDD_{s_v=r} = 0$ because it is the distance between the same styles. Finally, the histogram created by calculating SDD for r for all styles is given as

$$\boldsymbol{G}_{SDD} = \sum_{v=1}^{V} SDD_{s_v} \tag{7}$$

From (7), SDD is calculated for the number of styles V, and the histogram $\boldsymbol{G}_{SDD}$ is created. The calculations of SDD clarify the distance between $\boldsymbol{G}_{SDD}$ and the other styles and enables quantitative style compatibility estimation. Note that the conventional Siamese model can also be evaluated by SSD when J is set to 1 in (5).

5. Experiments

The experiment has two aims: (i) evaluation of the compatibility between furniture styles and (ii) evaluation of the distance between styles. Regarding (i), we trained the conventional and proposed Siamese models on the furniture image dataset and evaluated the compatibility of furniture styles by AUC. Note that for the proposed Siamese model in this experiment, we set $J = 2$ and $J = 3$ in (5), where the three furniture images were learned in a one-to-two ratio, and the four furniture images were learned in a one-to-three ratio, respectively. For (ii), we evaluated the distance between styles by analyzing the embedding of furniture images in the feature space using the proposed SDD of each model learned in (i).

5.1. Furniture Image Dataset

In this experiment, the Bonn Furniture Styles Dataset [13] was used. The Bonn Furniture Styles Dataset contains 90,298 images, which are classified into six categories: beds, chairs, dressers, lamps, sofas, and tables. The images in each category are classified into 17 different styles, which include Asian, Beach, and Contemporary.

In addition, when training the conventional and proposed Siamese models, referring to [13], the Bonn Furniture Styles Dataset was divided into training, validation, and test data at a ratio of 68:12:20.

5.2. Creation of Input Image Sets

The conventional and proposed Siamese models used in this experiment were trained based on multiple input images. Therefore, it was necessary to create a set of input images to train each model. The Siamese network trains and optimizes models so that the distance between images is closer for similar input image sets and vice versa. For each model, a set of positively compatible input images (positive inputs) and a set of negatively compatible input images (negative inputs) are created. The input image set used for each model utilizes the segmented Bonn Furniture Styles Dataset. Then, 24,000 pairs were created for training, 4800 pairs for validation, and 4800 pairs for testing, with the same ratio of positive and negative inputs.

In the conventional Siamese model, two sets of positive and negative inputs are created to implement the learning method with two input images. Examples of positive and negative inputs in the conventional Siamese model are shown in Figure 2. Positive inputs define furniture images of the same style, and negative inputs define those of different styles. However, within each set of input images, the same furniture category (e.g., two chairs) should not be included.

To train the proposed Siamese model with three input images in a one-to-two ratio, three sets of positive and negative inputs are created. Examples of positive and negative inputs in the proposed Siamese model are shown in Figure 3. A positive input is defined as a furniture image with all the same styles. Negative inputs learn negative style compatibility in a one-to-two ratio by defining 2 out of 3 furniture images as the same style. As in the conventional Siamese model, the same furniture category is not included in each input image set. Similar to the one-to-two learning, in the one-to-three learning, 3 out of 4 furniture images are in the same style for the negative inputs.

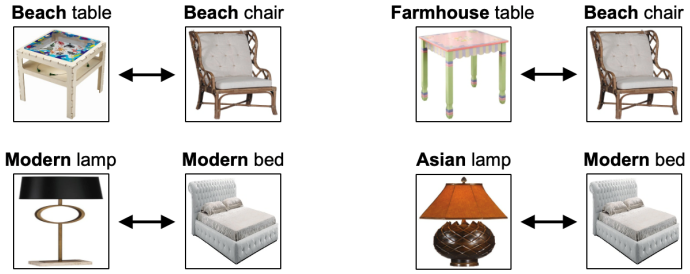


Figure 2. Example of input image set for the conventional Siamese model. (**a**) positive inputs. (**b**) negative inputs.

Figure 3. Example of input image set for the proposed Siamese model. (**a**) positive inputs. (**b**) negative inputs.

5.3. Performance Evaluation

Furniture style compatibility was learned using the training and validation data from the input image set for both the conventional and proposed Siamese models. Then, for each pretrained model, we evaluated its ability to estimate style compatibility by the AUC using the test data. In addition to the two Siamese models, the ability of VGG-19 to estimate the AUC was validated as a baseline. Specifically, features were extracted from the final fully-connected layer of VGG-19 pretrained to classify the 17 styles, and the AUC was calculated based on the Euclidean distance between two paired images, referring to [13]. After the evaluation by AUC, the distance between styles was quantitatively evaluated by SDD, the newly proposed evaluation measure.

5.3.1. Parameter Settings

When training the conventional and proposed Siamese models, RMSprop was employed as the optimizer. The batch size was set to 150 and the epoch number was set to 8. Referencing to [13], the learning rate was changed in the learning process, and two-stage learning was performed. Specifically, the learning rate was set to 0.0001 in the first five epochs and 0.00001 in the last three epochs to fine-tune the parameters of the model. We also set the hyper-parameter, margin m, in the Contrastive Loss given by (4) to 10. Then, using the test data, each model trained on furniture images was evaluated by AUC. The AUC outputs a value between 0 and 1, with 1 indicating a completely successful identification, 0.5 indicating a completely random identification, and 0 indicating a completely failed identification. In other words, the base value of AUC is 0.5, and the closer the value is to 1, the higher the performance of the furniture style compatibility evaluation. To accurately analyze the results, ten trials of the procedure were conducted, and the average AUC was calculated.

Next, the distance between styles was calculated with SDD for the pretrained conventional and proposed Siamese models. Specifically, 100 $\boldsymbol{I}^t$ were prepared for each of the 17 styles in this experiment, and SDD_{s_v} and $\boldsymbol{G}_{SDD}$ were calculated for each style.

5.3.2. Evaluation by AUC

The ROC curves and AUCs of all models using the test data are shown in Figure 4. Figure 4 shows that the AUCs of all Siamese models were higher than those of VGG-19. The AUC of the proposed one-to-two Siamese model was 0.721, and that of the proposed one-to-three Siamese model was 0.719, while the conventional Siamese model was 0.672. Therefore, compared to a simple CNN or conventional Siamese model, the proposed Siamese models can improve the accuracy of estimating the compatibility of ambiguous furniture styles. However, the proposed one-to-three model did not improve the AUC compared with the one-to-two model. Therefore, although the number of compared images requires optimization, maintaining a higher performance than the conventional Siamese model is possible.

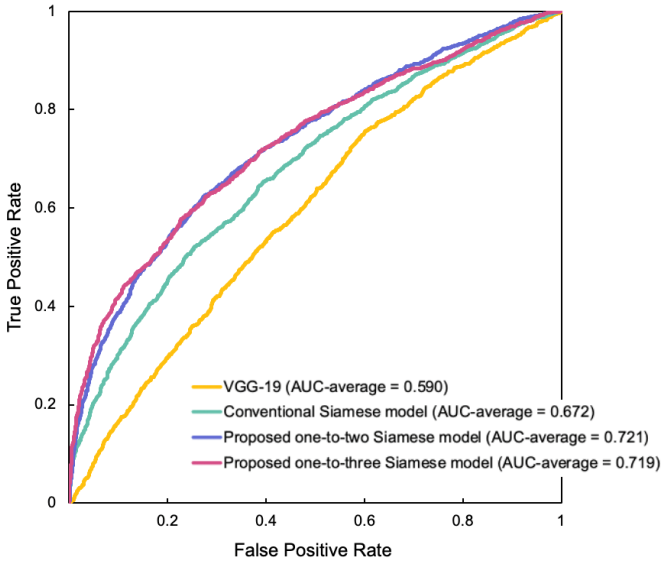


Figure 4. ROC curves and AUCs for the conventional and proposed models with test data (the AUC is the average over ten trials).

The results show that increasing the number of input images reduced the ambiguity of furniture styles, thus proving the effectiveness of the proposed Siamese models. In addition, the proposed Siamese model successfully learned furniture items that are compatible with either of the multiple pieces of furniture due to its one-to-many ratio learning method.

5.3.3. Evaluation by SDD

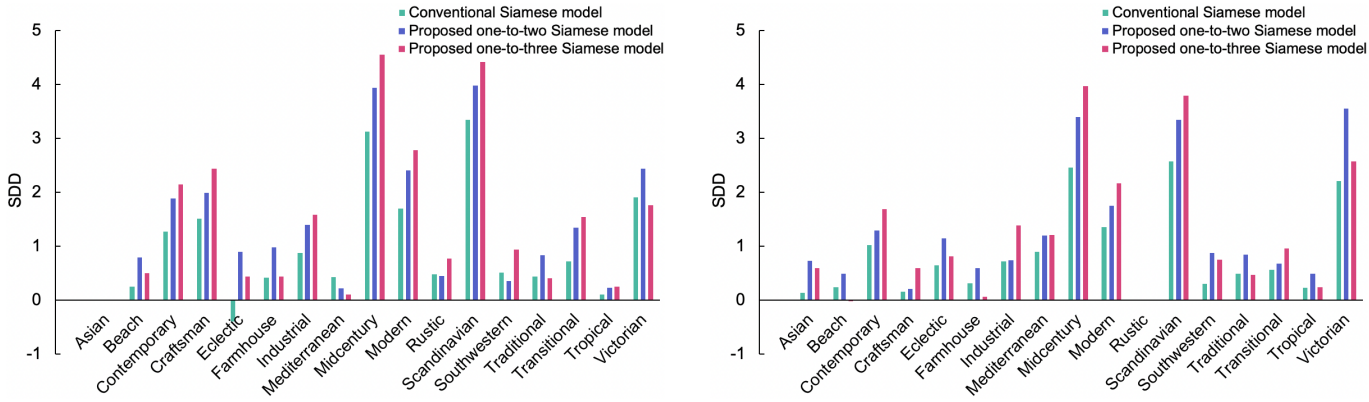
The results of the SDD evaluation for each model show that the proposed Siamese model tends to have a larger SDD than the conventional Siamese model. In this section, we focus on Asian, Rustic, Traditional, and Tropical as $S(\boldsymbol{I}^t)$, which highlight the differences in SDD.

First, the average value of D is shown in Table 1 for each of the four $S(\boldsymbol{I}^t)$. From Table 1, when $S(\boldsymbol{I}^t)$ is Rustic, the average value of $D_{Rustic,Rustic}$ is the smallest for the conventional and proposed one-to-two Siamese models. Meanwhile, $D_{S(\boldsymbol{I}^t),S(\boldsymbol{I}^t)}$ is sometimes not the smallest D for both the conventional model and proposed models. For example, when $S(\boldsymbol{I}^t)$ is Traditional, the average values for $D_{Traditional,Asian}$ and $D_{Traditional,Beach}$ are smaller than those for $D_{Traditional,Traditional}$ in the conventional Siamese model. However, in these cases, the largest difference is 0.41, which is in an acceptable error range compared with the difference of D with other styles. Therefore, the D for these cases can be disregarded. These events are uncommon regardless of the model. The results indicate that each model can reduce $D_{S(\boldsymbol{I}^t),S(\boldsymbol{I}^t)}$.

Table 1. Average of distance D to each $S(\boldsymbol{I}^{b'})$ and the four $S(\boldsymbol{I}^{t})$ is derived using the conventional (Convl), proposed one-to-two (1-to-2), and proposed one-to-three (1-to-3) models. The four $S(\boldsymbol{I}^{t})$ are the following: Asian, Rustic, Traditional, and Tropical.

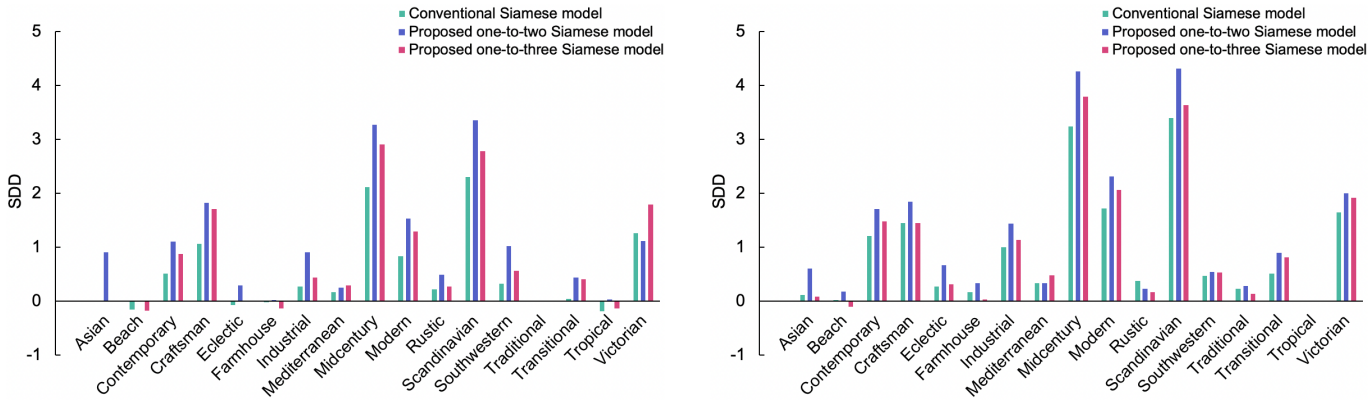
	$S(\boldsymbol{I}^{t})$: Asian			$S(\boldsymbol{I}^{t})$: Rustic			$S(\boldsymbol{I}^{t})$: Traditional			$S(\boldsymbol{I}^{t})$: Tropical		
	Convl	**1-to-2**	**1-to-3**	**Convl**	**1-to-2**	**1-to-3**	**Convl**	**1-to-2**	**1-to-3**	**Convl**	**1-to-2**	**1-to-3**
Asian	4.72	5.03	4.68	5.43	5.74	5.43	5.48	6.11	5.37	4.88	5.37	5.10
Beach	4.98	5.83	5.20	5.53	5.50	4.80	5.32	5.20	5.17	4.78	4.95	4.89
Contemporary	6.00	6.92	6.84	6.31	6.30	6.52	6.01	6.31	6.24	5.97	6.48	6.50
Craftsman	6.24	7.03	7.13	5.45	5.22	5.43	6.56	7.03	7.08	6.21	6.61	6.46
Eclectic	4.31	5.94	5.13	5.94	6.15	5.65	5.41	5.49	5.37	5.03	5.44	5.33
Farmhouse	5.14	6.02	5.13	5.61	5.60	4.90	5.46	5.22	5.21	4.92	5.10	5.04
Industrial	5.60	6.43	6.27	6.01	5.74	6.22	5.77	6.11	5.81	5.76	6.20	6.15
Mediterranean	5.15	5.26	4.80	6.19	6.20	6.04	5.67	5.45	5.66	5.10	5.11	5.49
Midcentury	7.85	8.97	9.24	7.75	8.40	8.80	7.61	8.48	8.28	8.00	9.02	8.80
Modern	6.43	7.45	7.48	6.64	6.76	7.00	6.34	6.73	6.66	6.48	7.09	7.07
Rustic	5.20	5.49	5.46	5.28	4.99	4.83	5.72	5.69	5.64	5.13	5.00	5.17
Scandinavian	8.06	9.02	9.10	7.87	8.35	8.63	7.80	8.56	8.16	8.16	9.08	8.64
Southwestern	5.23	5.40	5.63	5.60	5.88	5.59	5.83	6.22	5.94	5.23	5.31	5.55
Traditional	5.16	5.88	5.10	5.78	5.85	5.30	5.49	5.19	5.36	4.99	5.06	5.15
Transitional	5.45	6.38	6.24	5.85	5.69	5.79	5.54	5.64	5.78	5.27	5.67	5.82
Tropical	4.83	5.27	4.94	5.52	5.49	5.07	5.29	5.23	5.22	4.75	4.76	5.00
Victorian	6.63	7.47	6.46	7.49	8.55	7.41	6.76	6.32	7.16	6.40	6.78	6.93

Next, the SDD for each of the four styles is shown in Figure 5. The SDD of the proposed Siamese model tends to be larger than that of the conventional Siamese model when each of the four styles is $S(\boldsymbol{I}^{t})$. A larger SDD means that each style that differs from $S(\boldsymbol{I}^{t})$ is farther away in the feature space. Therefore, the proposed Siamese model is more successful in distancing $S(\boldsymbol{I}^{t})$ and the different styles of furniture than the conventional model. In particular, the SDD of the proposed one-to-two Siamese model is larger for all styles when Rustic is $S(\boldsymbol{I}^{t})$, and the one-to-three Siamese model exceeds the SDD of the conventional Siamese model for most styles. Thus, the superiority of SDD size depends on the $S(\boldsymbol{I}^{t})$ rather than the number of comparison images. Therefore, the performance of the proposed Siamese model, which learns with a one-to-many ratio, is independent of the number of furniture images compared, but both exhibit larger SDDs than the conventional Siamese model. The results of the SDD evaluation indicate that the proposed Siamese model brings furniture of the same style closer to $S(\boldsymbol{I}^{t})$ and moves furniture of different styles farther from $S(\boldsymbol{I}^{t})$, thereby clarifying the distance between each style.



(**a**) (**b**)

Figure 5. *Cont.*



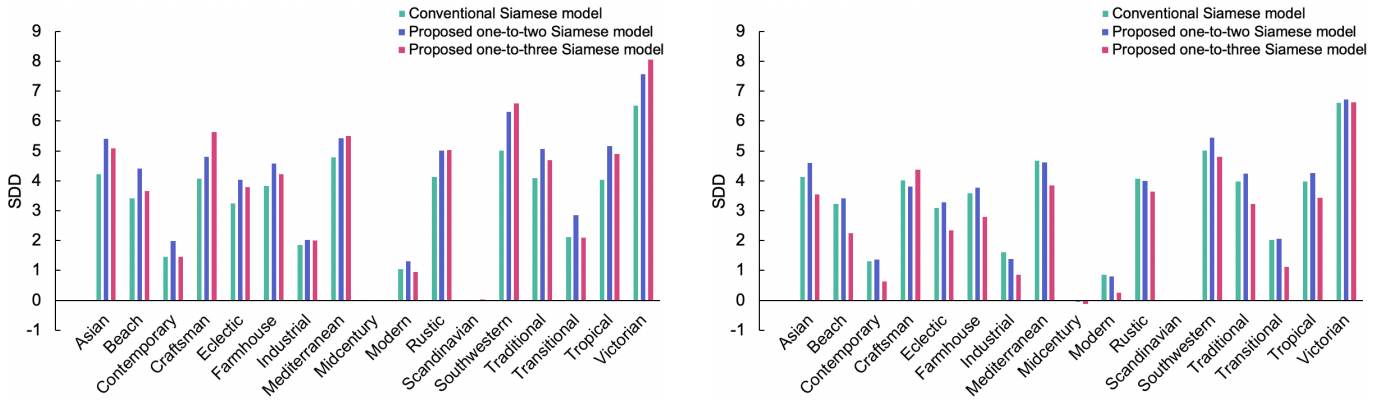
(**c**) (**d**)

Figure 5. SDD results of four $S(\boldsymbol{I}^t)$. The four $S(\boldsymbol{I}^t)$ are the following: Asian, Rustic, Traditional, and Tropical. (**a**) Asian. (**b**) Rustic. (**c**) Traditional. (**d**) Tropical.

6. Discussion

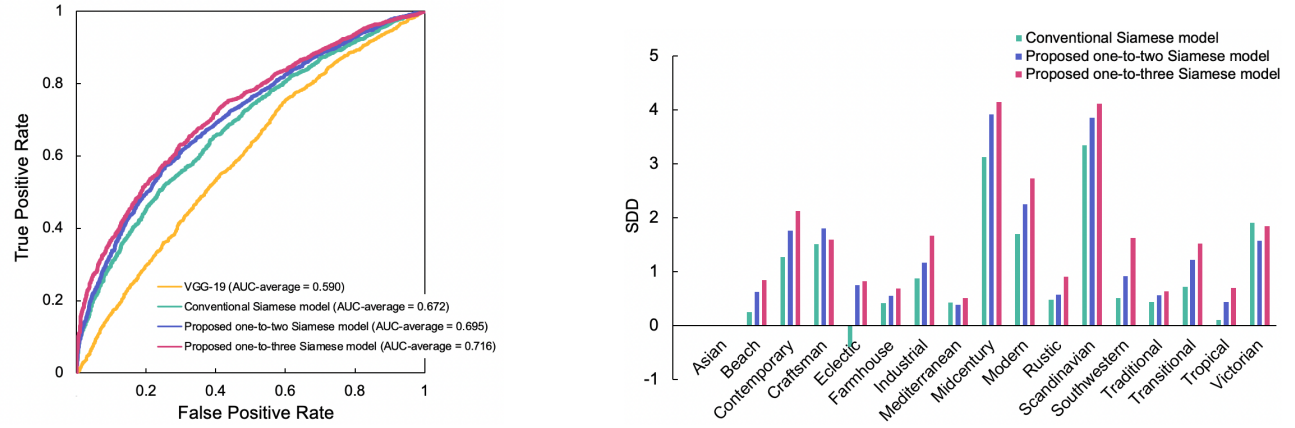
Here, the proposed SDD is discussed in detail. In Figure 5, Midcentury and Scandinavian SDDs are particularly large for all four $S(\boldsymbol{I}^t)$. The SDDs when $S(\boldsymbol{I}^t)$ is Midcentury and Scandinavian are shown in Figure 6. It can be seen that the SDDs of Midcentury and Scandinavian are very small, regardless of which is $S(\boldsymbol{I}^t)$. Therefore, Midcentury and Scandinavian are very similar styles and are considered compatible in the Siamese models. In Figure 6, there is no difference between the SDD of the conventional Siamese and proposed one-to-two models, and the one-to-three model often has a smaller SDD when Scandinavian is $S(\boldsymbol{I}^t)$. However, the results in Figure 5 and those for Midcentury in Figure 6 indicate that the SDDs of both proposed models are more significant for most styles. Thus, although our proposed Siamese model is not perfect, it improved the accuracy of style compatibility for the majority of the 17 styles when evaluated with AUC and SDD.

In this study, because we aimed to estimate style compatibility independent of furniture category, we conducted experiments by defining the category C of furniture images $(I^b, \boldsymbol{I}^t)$ to be input to the proposed Siamese model as $C(I^b) \neq C(I_j^t) \neq C(I_i^t)$, $j \neq i$. In practice, we will discuss the case where $C(I^b) \neq C(I_j^t) = C(I_i^t)$, $j \neq i$ since we believe there is a demand to infer compatible furniture for multiple pieces of furniture in the same category, e.g., checking the compatibility of a lamp and two chairs. As in Section 5, experiments were conducted on the one-to-two and one-to-three proposed Siamese models, and the evaluation results using AUC and SDD are shown in Figure 7. Figure 7a shows that the AUC of the proposed one-to-two Siamese model was 0.695, and that of the proposed one-to-three Siamese model was 0.716, whereas that of the conventional Siamese model was 0.672. These results suggest that increasing the number of compared furniture images may be effective since the one-to-three model was more accurate than the one-to-two model. Next, Figure 7b shows that when $S(\boldsymbol{I}^t)$ is Asian, the proposed Siamese model outperforms of the conventional Siamese model for most styles in terms of the SDD. Therefore, the proposed method can learn the distance between styles even when $\boldsymbol{I}^t$ are in the same category. In conclusion, the proposed Siamese model effectively learns the style compatibility of furniture items in the same and different categories, and the AUC and SDD are effective in learning the distance between styles.



(**a**) (**b**)

Figure 6. SDD results of two $S(\boldsymbol{I}^t)$. The two $S(\boldsymbol{I}^t)$ are the following: Midcentury and Scandinavian. (**a**) Midcentury. (**b**) Scandinavian.



(**a**) (**b**)

Figure 7. Experimental results of AUC and SDD (Category C is $C(I^b) \neq C(I_j^t) = C(I_i^t)$, $j \neq i$, for the input to the proposed Siamese model). (**a**) ROC curves and AUCs (AUC is the average over ten trials). (**b**) SDD result where $S(\boldsymbol{I}^t) =$ Asian.

7. Conclusions

To improve the accuracy of estimating furniture style compatibility, we proposed a new Siamese network that evaluates the compatibility of three or more furniture images by increasing the number of $\boldsymbol{I}^t$. In addition, as a quantitative measure for style compatibility, we proposed a new metric, SDD, based on the Euclidean distance between images. To verify the usefulness of the proposed Siamese model in estimating furniture style compatibility, we conducted an evaluation experiment using the Bonn Furniture Styles Dataset, which contains 17 different styles of furniture. The results were analyzed under two metrics: AUC and SDD. The results show that the AUC of the proposed Siamese model was 0.721 in the one-to-two model and 0.719 in the one-to-three model, whereas the VGG-19 and the conventional Siamese model obtained AUCs of 0.590 and 0.672, respectively. The results of the SDD evaluation show that the proposed Siamese model better increases the distance between different styles than the conventional Siamese model. From these two points, the proposed Siamese model is an effective method for estimating furniture style compatibility.

In addition, the proposed Siamese model learns based on the compatibility between three or more images. Therefore, it can estimate furniture images that are style-compatible with any piece of furniture in $\boldsymbol{I}^t$, i.e., furniture images with good compatibility. In conclusion, the proposed Siamese model is more efficient than the conventional Siamese model both in terms of performance on the evaluation scale and in practical use, and it is an effective model for the task of style compatibility estimation.

Author Contributions: Conceptualization, K.O.; data curation, A.T.; formal analysis, A.T.; funding acquisition, K.O. and E.M.; investigation, A.T.; methodology, A.T. and K.O.; project administration, K.O.; software, A.T.; supervision, E.M.; visualization, A.T.; writing—original draft, A.T.; writing—review editing, N.I. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI Grant Number 21K12097 and JSPS KAKENHI Grant Number 20K14101.

Data Availability Statement: The dataset (Bonn Furniture Styles Dataset) used in this paper was provided by https://cvml.comp.nus.edu.sg/furniture/index.html, accessed on 26 July 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kim, J.; Heo, W. Interior Design with Consumers' Perception about Art, Brand Image, and Sustainability. *Sustainability* **2021**, *13*, 4557. [CrossRef]
2. Shiau, R.; Wu, H.Y.; Kim, E.; Du, Y.L.; Guo, A.; Zhang, Z.; Li, E.; Gu, K.; Rosenberg, C.; Zhai, A. Shop the look: Building a large scale visual shopping system at pinterest. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual, 6–10 July 2020; pp. 3203–3212.
3. Mu, C.; Zhao, J.; Yang, G.; Zhang, J.; Yan, Z. Towards practical visual search engine within elasticsearch. *arXiv* **2018**, arXiv:1806.08896.
4. Kim, J.; Lee, J.K. Stochastic Detection of Interior Design Styles Using a Deep-Learning Model for Reference Images. *Appl. Sci.* **2020**, *10*, 7299. [CrossRef]
5. Pan, T.Y.; Dai, Y.Z.; Tsai, W.L.; Hu, M.C. Deep model style: Cross-class style compatibility for 3d furniture within a scene. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 4307–4313.
6. Hu, Z.; Wen, Y.; Liu, L.; Jiang, J.; Hong, R.; Wang, M.; Yan, S. Visual classification of furniture styles. *ACM Trans. Intell. Syst. Technol. (TIST)* **2017**, *8*, 1–20. [CrossRef]
7. Yoon, S.Y.; Oh, H.; Cho, J.Y. Understanding furniture design choices using a 3D virtual showroom. *J. Inter. Des.* **2010**, *35*, 33–50.
8. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; IEEE: Piscataway, NJ, USA, 2005; Volume 1, pp. 886–893.
9. Ke, Y.; Sukthankar, R. PCA-SIFT: A more distinctive representation for local image descriptors. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, Washington, DC, USA, 27 June–2 July 2004; IEEE: Piscataway, NJ, USA, 2004; Volume 2, p. II.
10. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [CrossRef]
11. Hu, T.; Qi, H.; Huang, Q.; Lu, Y. See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification. *arXiv* **2019**, arXiv:1901.09891.
12. Khan, H.; Shah, P.M.; Shah, M.A.; ul Islam, S.; Rodrigues, J.J. Cascading handcrafted features and Convolutional Neural Network for IoT-enabled brain tumor segmentation. *Comput. Commun.* **2020**, *153*, 196–207. [CrossRef]
13. Aggarwal, D.; Valiyev, E.; Sener, F.; Yao, A. Learning style compatibility for furniture. In Proceedings of the 40th German Conference on Pattern Recognition, Stuttgart, Germany, 9–12 October 2018; Springer: Cham, Switzerland, 2018; pp. 552–566.
14. Polania, L.F.; Flores, M.; Nokleby, M.; Li, Y. Learning furniture compatibility with graph neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 366–367.
15. Liu, B.; Zhang, J.; Zhang, X.; Zhang, W.; Yu, C.; Zhou, Y. Furnishing Your Room by What You See: An End-to-End Furniture Set Retrieval Framework with Rich Annotated Benchmark Dataset. *arXiv* **2019**, arXiv:1911.09299.
16. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In Proceedings of the ICML Deep Learning Workshop, Lille, France, 6–11 July 2015; Volume 2.
17. Weiss, T.; Yildiz, I.; Agarwal, N.; Ataer-Cansizoglu, E.; Choi, J.W. Image-Driven Furniture Style for Interactive 3D Scene Modeling. *Comput. Graph. Forum* **2020**, *39*, 57–68. [CrossRef]
18. Bell, S.; Bala, K. Learning visual similarity for product design with convolutional neural networks. *ACM Trans. Graph. (TOG)* **2015**, *34*, 1–10. [CrossRef]
19. Li, Y.; Su, H.; Qi, C.R.; Fish, N.; Cohen-Or, D.; Guibas, L.J. Joint embeddings of shapes and images via cnn image purification. *ACM Trans. Graph. (TOG)* **2015**, *34*, 1–12. [CrossRef]
20. Simo-Serra, E.; Trulls, E.; Ferraz, L.; Kokkinos, I.; Fua, P.; Moreno-Noguer, F. Discriminative learning of deep convolutional feature point descriptors. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 118–126.

21. Veit, A.; Kovacs, B.; Bell, S.; McAuley, J.; Bala, K.; Belongie, S. Learning visual clothing style with heterogeneous dyadic co-occurrences. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4642–4650.
22. Polanía, L.F.; Gupte, S. Learning fashion compatibility across apparel categories for outfit recommendation. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 4489–4493.
23. Yuan, H.; Liu, G.; Li, H.; Wang, L. Matching recommendations based on siamese network and metric learning. In Proceedings of the 2018 15th International Conference on Service Systems and Service Management (ICSSSM), Hangzhou, China, 21–22 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.
24. Gao, G.; Liu, L.; Wang, L.; Zhang, Y. Fashion clothes matching scheme based on Siamese Network and AutoEncoder. *Multimed. Syst.* **2019**, *25*, 593–602. [CrossRef]
25. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
26. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 248–255.