Room Style Estimation for Style-Aware Recommendation

Esra Ataer-Cansizoglu*, Hantian Liu, Tomer Weiss, Archi Mitra, Dhaval Dholakia, Jae-Woo Choi and Dan Wulin
Wayfair
Boston, MA
*ecansizoglu@ieee.org

***Abstract*—Interior design is a complex task as evident by multitude of professionals, websites, and books, offering design advice. Additionally, such advice is highly subjective in nature since different experts might have different interior design opinions. Our goal is to offer data-driven recommendations for an interior design task that reflects an individual's room style preferences. We present a style-based image suggestion framework to search for room ideas and relevant products for a given query image. We train a deep neural network classifier by focusing on high volume classes with high-agreement samples using a VGG architecture. The resulting model shows promising results and paves the way to style-aware product recommendation in virtual reality platforms for 3D room design.**

***Index Terms*—style estimation, neural networks, recommendation**

I. Introduction

Interior design and home decoration involve a high amount of guesswork. In addition to visual appearance of each individual item, their group composition is also highly important. Therefore the context where a product is placed provides valuable information to understand customer's style. Although a room's style can be predefined and categorized, these are typically hard to verbalize by non-experts. To add to the confusion, a room's design might be influenced by multiple styles rather than a single discrete style. Hence even experts might disagree on the primary style of a room. Despite such ambiguities, our goal is to suggest users images of rooms that are compatible with their taste.

Understanding user preferences lies at the heart of e-commerce websites such as Wayfair. An accurate representation of customer's interests and style enables better product recommendations and personalization. Although there has been recent work on understanding subjective attributes related to aesthetic and fashionablity [1]–[6], few focus on interior design [3], [4], [8]. Lun et al. [3] focused on similarly shaped, salient, geometric elements as an indication of style similarity between furniture. Weiss et al. [8] developed a method for placement of a curated set of furniture in a room. Similarly, Pan et al. [4] approached the problem of compatibility between furniture with a similarity model. These approaches lack a holistic understanding of room style, since they focus on furniture-level style similarity and geometry. Additionally, they focus on certain visual features and disregard subjectivity.

The contributions of this paper are (i) an image retrieval framework to inspire customers by finding room ideas and products that have similar style with a given query image, (ii) a multi-expert data labeling approach to handle subjectivity, and (iii) a deep learning-based classification method focusing on high volume classes and high-agreement samples in order to depict distinctive visual features reflecting style.

II. Method

Our method consists of two steps. First, we collect data and gather labels for each image from multiple experts. Second, we train a deep neural network to classify each image with a style.

A. Data Collection and Labeling

We gathered a data set of $800K$ room images by scraping through housing websites and using images created by our in-house designers. Due to high subjectivity, each image is tagged by 10 experts using one of the primary style. These stylistic terms are loosely defined and reflect trends that tend to dominate the home goods space over time. In this work we focus on the following 7 primary styles: *modern*, *traditional*, *coastal*, *cottage*, *eclectic*, *rustic* and *industrial*. Each style is described with certain criteria about fabric, color scheme, material, furniture style and flooring, as shown in Figure 1.

We faced with two major challenges at this stage. First, we observed a high amount of disagreement among experts. Second, the data displayed a large class imbalance. Figure 2(a) shows the confusion matrix where each cell reports the number of images with 4 tags from each style normalized by total number of images with 4 tags from either of the styles. Figure 2(b) shows histogram of number of images per style

Fig. 1: Examples of major room styles.

where at least 7 experts agreed on the label. As can be seen 70% of our data is defined as *modern* and *traditional* styles, which can make learning harder for others.

B. Training Room Style Estimator (RoSE)

To classify room styles, we utilized deep neural networks. We transfer learned from a VGG network architecture [7] that is trained for place classification on *Places365* data set [9]. We used previous last layer of the network as a feature extractor in our retrieval experiments. Given the challenges with data imbalance and inter-expert variability, we focused on high volume classes with high agreement samples. We carried out two different training schemes:

RoSE v1.0. Initially, we added a fully connected layer of 8 dimensions to the network and trained it for classification of 2 styles: *modern* and *traditional* styles. Both are prominent in our data (Figure 2(b)).

RoSE v1.1. Our next goal was to increase class coverage. Based on our discussion with designers and the analysis on our dataset, we merged *coastal* and *cottage* classes. With these 3 styles: *modern*, *traditional* and *coastal-cottage*, we trained the model with a classification goal. Our goal was not only to learn a better style representation by increasing class coverage, but also to avoid overfitting due to small training set size. To that end, we modified VGG's network architecture. We removed all fully connected layers from VGG architecture and reduced the number of channels to 16 for the last two convolutional layers. They were followed by an 8-dimensional fully connected layer and softmax layer.

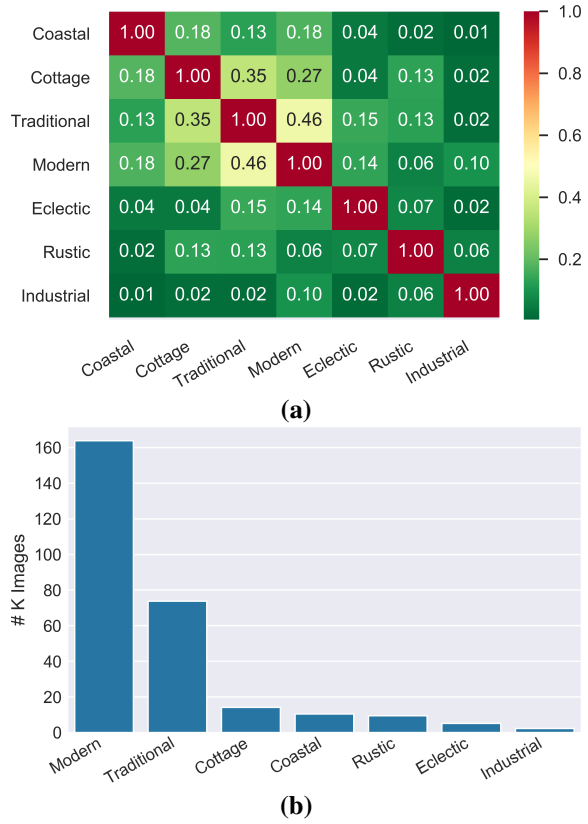


Fig. 2: Challenges faced during data collection: (a) confusion matrix among experts depicting high inter-expert variability where each cell reports the number of images with 4 tags from each style normalized by total number of images with 4 tags from either of the styles, (b) distribution of number of images per style where each tag is agreed by at least 7 experts showing the class imbalance problem.

III. Experiments and Results

A. Data Preparation

Considering class imbalance problem, we used samples that are tagged by at least n experts for each style: *modern* ($n = 10$), *traditional* ($n = 8$), *coastal* and *cottage* ($n = 7$). We carried out data augmentation for *coastal* and *cottage* samples by randomly applying horizontal flip, noise addition or rotation. Our final data set contains $160K$ images that are split into 80%, 10% and 10% as train, test and validation sets respectively.

B. Implementation

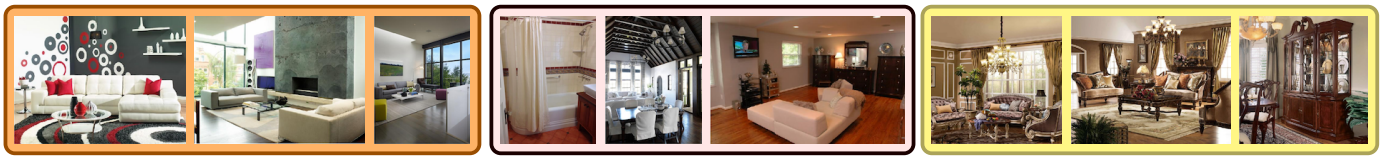
We implemented our approach in Python, using Keras and Tensorflow. For training, we used RMSprop with a learning rate of 0.0001. We froze all layers until the last fully connected layer for `RoSE v1.0` and until last two convolutional layers for `ROSE v1.1`. Training took about 8 hours on NVIDIA Tesla GPU with 16GB memory for each version.

C. Results

Our model achieved a classification accuracy of 88.7% in `v1.0` and 81.2% in `v1.1`. Figure 3 shows predictions for some sample images from our test set using `RoSE v1.0`. The leftmost 3 images were predicted as modern, while rightmost 3 images were predicted as traditional. Images in the middle are uncertain, since the confidence score of the model is around 0.5. One possible explanation is that the images do not contain enough context due to limited furniture or decor. Another possibility is that they have visual features that align with multiple styles. For example, images that show traits of both traditional and modern.

Since our goal is to employ our model for image retrieval, we carried out retrieval experiments using the output of second last layer as visual embeddings for each image. We carried out retrieval experiments in two different use cases: (i) finding room ideas similar to a given room image and (ii) finding complementary products given the image of a product in a room setting.

1) Room Image Retrieval: We formed a retrieval test set covering all 7 styles. In addition to the test split that includes samples from the 4 covered styles, we gathered images tagged as either *rustic*, *eclectic* and *industrial* by at least $n = 7$ experts. We queried each room image from the set and retrieved the most similar 5 room images according to the Euclidean distance between the embeddings. Our goal is to retrieve images consistent with the room type. For example,



(a) Modern (b) Mixed Classification (c) Traditional

Fig. 3: Classification results: leftmost 3 images are classified as modern, while rightmost 3 images are classified as traditional. The middle panel shows images that have the highest uncertainty, where the classifier's output probability is around 0.5.

TABLE I: Room retrieval results: each row shows a different query result, where leftmost image is the query image.

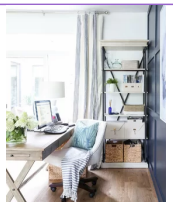
Query Image	*Ranked Results*				
					
					

TABLE II: Retrieval results for complementary items for querying sofas from accent tables (top) and querying accent tables from sofas (bottom).

Query Image	*Ranked Results*				
					
					

if a bedroom is queried, we want to retrieve bedroom images with a similar style. To that end, we search for images with the same room type as the query image[1]. Figure 4 shows the recall rate for each model along with the random baseline where 5 random samples were selected from the set of images that has the same room type with the query image. Although model is trained on only 4 styles, it still performs better than baseline on the uncovered styles. The figure also displays the average recall rate at k per style in order to better see the representative power of our model for each style. As expected, `RoSE v1.0` performs better for *modern* and *traditional*, while `RoSE v1.1` performs better on *coastal* and *cottage*. Hence, depending on the style distribution of data, one of them can be preferred over the other. Table I displays qualitative results from room retrieval experiments.

2) Cross-Class Product Recommendation: Finding complementary items is a challenging task in interior design. A big portion of Wayfair's product imagery consists of environmental images. In environmental images, a main product is situated in an stylist-curated environment that includes other comple-

[1]Note that retrieving images without any filtering on room type gives comparable results in terms of recall rate.

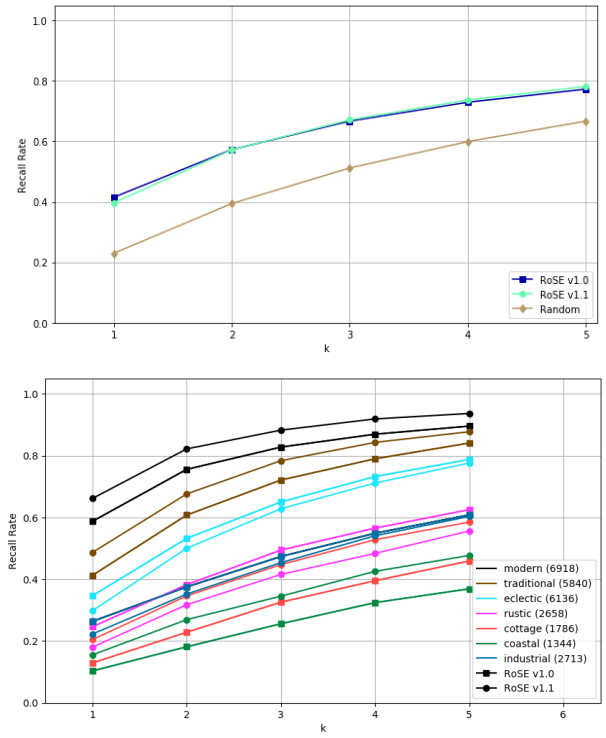


Fig. 4: Results of retrieval experiments: recall rate at k in retrieval experiments. Numbers in parenthesis show number of images per style.

mentary products. Therefore such images act as a ground-truth context for recommendations. We use our model to find style similarity between products using associated environmental imagery. We provide qualitative results by focusing on two classes: *sofas* and *accent tables*. Accent table is a category that includes coffee, cocktail, side and end tables. We randomly selected 800 images from each class by taking into account style distribution of products in our catalog. We query sofa images among accent table images and vice versa. Table II displays two example query results. Even though accent tables and sofas do not occur in most of the images at the same time, our model recommends visually complementary items showing its strength in context-based style understanding.

IV. Conclusion and Future Work

We presented our work on image retrieval that aims to inspire customers with design ideas similar to their styles. We trained a deep neural network classifier by focusing on high volume classes and high agreement samples as a workaround for inter-expert variability and high class-imbalance. The resulting models show over 80% classification accuracy and promising results for style-based product recommendations.

We provided quantitative analysis on the retrieval performance of our framework by considering a query result relevant if it has the same style with the seed image. However, since our goal is to apply our method for customer recommendations, we also evaluate it based on customer experience. We started testing our framework on a customer style quiz, where customers can like and dislike room images to get recommendations according to their style preferences. Initial readings of our test show that our model increases customer engagement significantly compared to random ranking of images with style filters. Similarly, we plan to test our model for complementary product recommendation on our website in order to assess the alignment of our solution with the customer use case.

Integration of our framework into a virtual reality platform for 3D room design is an important extension of our work which will require 3D object models. Wayfair is dedicated to creating 3D models of products for improving customer experience in visualization. As a next step, we would like to integrate our recommendation framework with Wayfair's 3D room design platform[2], which allows to seamlessly design and create rooms (See Fig. 5). In addition to style-aware product recommendation, 3D room design involves many geometry problems related to dimension and pose of the products. Hence, automated item placement is also in the scope of our future work.

Fig. 5: Wayfair Room Planner 3D, an interior design platform.

Acknowledgment

We thank Brendan Sullivan and Ama Edzie for discussions on customer use cases and their support on data collection.

References

[1] S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1657–1664. IEEE, 2011.

[2] W.-L. Hsiao and K. Grauman. Learning the latent look: Unsupervised discovery of a style-coherent embedding from fashion images. pages 4213–4222. IEEE, 2017.

[3] Z. Lun, E. Kalogerakis, and A. Sheffer. Elements of style: learning perceptual shape style similarity. *ACM Transactions on Graphics (TOG)*, 34(4):84, 2015.

[4] T.-Y. Pan, Y.-Z. Dai, M.-C. Hu, and W.-H. Cheng. Furniture style compatibility recommendation with cross-class triplet loss. *Multimedia Tools and Applications*, 78(3):2645–2665, 2019.

[5] R. Schifanella, M. Redi, and L. M. Aiello. An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures. In *International AAAI Conference on Web and Social Media*, 2015.

[6] E. Simo-Serra, S. Fidler, F. Moreno-Noguer, and R. Urtasun. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 869–877, 2015.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[8] T. Weiss, A. Litteneker, N. Duncan, M. Nakada, C. Jiang, L.-F. Yu, and D. Terzopoulos. Fast and scalable position-based layout synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 2018.

[9] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[2] Available at www.wayfair.com/RoomPlanner3D, as of Sep. 2019.