The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Deep Ranking for Style-Aware Room Recommendations (Student Abstract)

İlkay Yıldız,[1] Esra Ataer-Cansızoğlu,[2] Hantian Liu,[2]
Peter Golbus,[2] Ozan Tezcan,[3] Jae-Woo Choi[2]
[1]Northeastern University EECE, 14 Palace Road Apt.9, Boston, MA 02115, yildizi@ece.neu.edu, 8572655701
[2]Wayfair LLC., 4 Copley Pl 7th floor, Boston, MA 02116, {ecansizoglu, hliu2, pgolbus, jchoi}@wayfair.com
[3]Boston University EECE, 8 St Marys St 324, Boston, MA 02215, mtezcan@bu.edu

Abstract

We present a deep learning based room image retrieval framework that is based on style understanding. Given a dataset of room images labeled by interior design experts, we map the noisy style labels to comparison labels. Our framework learns the style spectrum of each image from the generated comparisons and makes significantly more accurate recommendations compared to discrete classification baselines.

Introduction

Interior design and home decoration heavily rely on guesswork. Seeing a product in a room context is important for customers to make more confident decisions. Moreover, it is extremely challenging for customers to search through massive online catalogs and find pieces that match with their room design and stylistic preferences. Therefore, it is crucial to show products in a room context tailored to a customer's taste. There exist different room styles defined by designers; we focus on 4 major ones that are highly popular: *modern, traditional, cottage, coastal*. Each style is described with certain criteria about fabric, color scheme, material, furniture style and flooring, and labelling a scene with a style is a highly complicated task. Thus, using designer labels for a style-based shopping experience brings several challenges.

We collect a dataset of room images labeled by interior design experts. Each image receives class labels from 10 experts, where each label corresponds to one of the 4 major room styles. Our dataset exhibits two major challenges: (i) Due to significant subjectivity in style assessment, there exists high inter-expert variability in class labels. (ii) Distribution of labeled images over different style classes is very imbalanced. Training a neural network over such a noisy and imbalanced dataset is prone to inaccurate style estimates, and leads to ineffective room recommendations to customers.

We overcome these challenges by introducing *comparisons* into training. A comparison label indicates the *relative order* between a pair of data samples. Incorporating comparisons into training has two advantages. First, unlike class labels, comparisons reveal both intra-class and inter-class relationships between samples. Second, comparisons are often less variable compared to class labels, i.e. experts disagreeing in determining class labels often agree on the relative order of data samples. This has been extensively documented in many domains, references are available upon request. Motivated by these observations, we map the noisy style labels generated by multiple experts to comparison labels. Particularly, given a room style and a pair of images, a comparison outcome is determined by the relative order of the number of labels each image receives. Our architecture is inspired by siamese networks (Bromley et al. 1994) and learns the style spectrum of each image from the generated comparison labels. It extends the Bradley-Terry (Bradley and Terry 1952) model to learn from comparisons and predict both class and comparison labels indicating room style.

Experiments

As baselines, we separately train the base network of the siamese architecture on set of all noisy multi-expert class labels and set of only clean high agreement class labels; we denote the resulting models as *RoSE v1n* and *RoSE v1c* , respectively. We evaluate our architecture (*RoSE v2*), as well as the baselines both qualitatively and quantitatively.

Table 1 shows images with *RoSE v2* scores and number of true labels for modern, traditional, cottage, and coastal, respectively. For each style, score is the base network output and label is the number of experts declaring the image as that style. *RoSE v2* predictions closely align with the multi-expert class labels. For correctly classified images, scores and expert labels highly agree on style assignment, e.g. score is $1/1$ and label is $10/10$ for the predicted style. Even for incorrectly classified images, *RoSE v2* and experts disagree on the same subset of styles, i.e. the styles receiving nonzero scores from *RoSE v2* and nonzero labels from experts match. Table 2 shows *RoSE v2* recommendations with the corresponding number of true labels. Recommendations indicate the 5 closest images to the seed image w.r.t. Euclidean distance of base network embeddings, ranked from left to right. Given a seed where all experts agree on the style, *RoSE v2* recommends images with the correct style and high agreement $(10/10)$ expert labels. Meanwhile, given a seed where experts disagree on the style, *RoSE v2* recommends images with labels disagreeing on the same subset of styles as the seed. These results validate that comparisons can successfully capture noise in class labels and accurately model the style spectrum.

Table 1: Example images with *RoSE v2* scores and no. of true labels for modern, traditional, cottage, and coastal, respectively.

	Correct Classification			Incorrect Classification		
Image						
Scores	0, 1.0, 0, 0	1.0, 0, 0, 0	0, 0, 0, 1.0	0, 0.34, 0.63, 0.03	0, 0, 0.65 , 0.35	0.1, 0.25, 0.57, 0.08
Labels	0, 10, 0, 0	10, 0, 0, 0	0, 2, 0, 8	0, 6, 4, 0	0, 0, 2, 8	1, 8, 1, 0

Table 2: *RoSE v2* recommendations with no. of true labels for modern, traditional, cottage, and coastal, respectively.

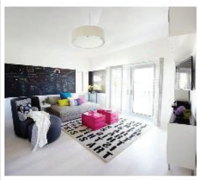
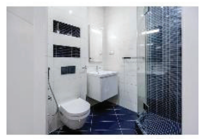
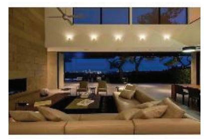
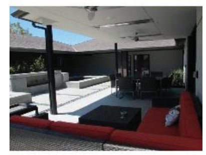
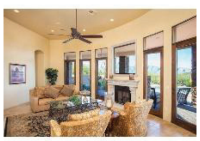
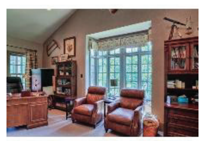
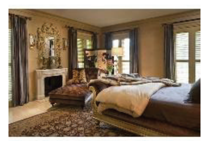
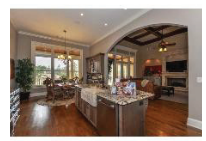
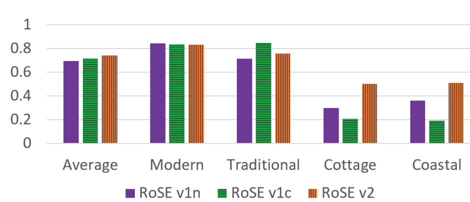
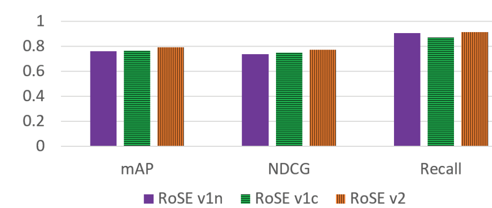
	Seed Image	Recommendations Ranked From Left to Right				
						
Labels	10, 0, 0, 0	10, 0, 0, 0	10, 0, 0, 0	10, 0, 0, 0	10, 0, 0, 0	10, 0, 0, 0
						
Labels	2, 8, 0, 0	0, 10, 0, 0	0, 10, 0, 0	0, 9, 1, 0	0, 10, 0, 0	0, 9, 1, 0

Figure 1: *RoSE v1n* , *RoSE v1c* , vs. *RoSE v2* evaluations



(a) Recall rate at 1



(b) Retrieval at 5

Figure 1 shows retrieval performances of *RoSE v1n* and *RoSE v1c* vs. *RoSE v2* , based on per-class and average recall rate at 1, along with average mAP, NDCG, and recall rate at 5 for predicting high agreement class labels. Learning from comparisons via *RoSE v2* leads to 0.739 recall rate at 1, 0.792 mAP at 5, 0.772 NDCG at 5, and 0.913 recall rate at 5. *RoSE v1n* and *RoSE v1c* perform significantly poorly in retrieval on low sample classes, i.e. cottage and coastal, with up to 0.362 recall rate at 1. With almost equally successful retrievals on all 4 classes, *RoSE v2* significantly improves over *RoSE v1n* and *RoSE v1c by* 21% *recall rate at 1 on cottage,* 15% *recall rate at 1 on coastal, and* 3% *w.r.t. recall rate at 1, mAP at 5, and recall rate at 5.*

References

Bradley, R. A., and Terry, M. E. 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika* 39(3/4):324–345.

Bromley, J.; Guyon, I.; LeCun, Y.; Säckinger, E.; and Shah, R. 1994. Signature verification using a "siamese" time delay neural network. In *Advances in Neural Information Processing Systems*, 737–744.