Abstract

Understanding the style of furniture items is a well known challenging problem in E-commerce. The task of determining which style a furniture piece belongs to, e.g., Scandinavian or Mid-century Modern, can be subjective, and the boundary between two different furniture styles can be vague. Moreover, to our knowledge, there is no existing dataset that is publicly available. In this work we introduce a new large image dataset collected from web, which is composed of different furniture styles across a diverse set of furniture types (e.g. couch, table, ...), we then apply recent deep learning method to tackle the problem of classifying style of furniture item images. We benchmark a multi-task algorithm to the problem of classifying style and we propose the problem of learning furniture style across furniture types that can serve as a benchmark for transfer learning algorithms.

1. Introduction

E-commerce is one of the fastest growing sectors in the industry. In 2017, sales of physical goods through E-commerce amounted to 453.5 billion US dollars and increased by 16% year over year according to U.S. Census Bureau report [2]. And it has been shown that Furniture and Home Furnishing have been one of the key areas in E-commerce. In 2017, around 12% of online sales are in the Furniture and Home Furnishing category [14]. The importance of the Furniture and Home Furnishing category in E-commerce can also be verified by recent emphasis of this category on Walmart.com [11] and Amazon.com [1].

Customers often have implicit or explicit preference of style, e.g., scandinavian, mid-century modern, and etc., when buying furniture. Thus understanding furniture style is an essential task in the Furniture and Home Furnishing category for an E-commerce website such as Walmart.com. Understanding furniture style enables browsing experiences centered around different furniture styles. E.g., we can group furniture items based on style and create dedicated shelf space related to each unique furniture style so that user can explore items which belong to their interest. Also style annotation enables better search experience by adding the capability of matching user queries on furniture style to the corresponding items. E.g., when user submits a query such as “mid-century modern couch”, we can quickly locate couches which have mid-century modern as their tagged style even if the title and description of these items do not contain the style phrase. Finally, style annotation of furniture items can further enhance search experience by adding the functionality of style facets on search result page so that customers can pick relevant styles to filter search result items they are interested in. In Figure 2 we show some examples of above mentioned applications on Walmart.com website.

Classifying furniture style is a very challenging problem, boundary between different furniture styles can be vague and potentially overlapping. E.g., in Figure 1 we show three different chairs: on the left the tapiiovaara rocking chair is mostly scandinavian, but not likely mid-century modern; on the right the gio ponti via dezza chair is mostly mid-century modern, but not scandinavian; in the middle, we have the hans wegner papa bear chair belongs to both mid-century model and scandinavian [12]. Moreover depending on the furniture type (E.g. chairs, tables, couches), different details of the item can determine its style. E.g. for tables, material and legs are important to identify style and
Manually classifying styles of furniture items is obviously not scalable for the Walmart.com product catalog which contains more than a million furniture items. Style data provided by sellers is usually sparse, many of the style tags provided are too broad, e.g., everything is tagged with Modern or Classic, or simply wrong. And extracting style information from textual information, e.g., from title or description of furniture item, can also be unreliable, as textual data can be noisy and incomplete. On the other hand, our catalog has a rich set of images for each product, usually each furniture item we will have one primary image and a few secondary images which are all professionally taken. Considering that style is conceived more from visual perspective, in this work, we focus on classifying furniture style directly from the furniture item image using advanced image classification algorithm.

Inspired by ImageNet [3], we believe that a large-scale dataset focusing on furniture style is a critical resource for developing and benchmarking furniture style classification algorithms. Thus in this work, we introduce a new dataset collected from web which was tagged manually by in-house furniture style experts. The dataset contains 20,890 images on 35 furniture types, and 16 furniture styles. Based on this furniture style dataset we benchmark a few state-of-the-art deep image classification networks to classify style of a furniture item. We also benchmark a multi-task algorithm that leverages the furniture type data to improve style classification and we also evaluate if style understanding can be transferred across furniture types.

The novelty of this paper is the introduction of the first image dataset on furniture style and furniture type. Since this new dataset contains both style and furniture type information for each item, it can also be used to benchmark multi-task classification algorithms as well as transfer learning algorithms. This paper also provides an initial benchmark for three different problems. We hope that our initial effort can spark more research around these interesting problems.

The rest of the paper is organized as follows: We first discuss the collection of the furniture style dataset in Section 2. In Section 3, we explore how state-of-the-art deep image classification networks can be leveraged to classify styles of furniture item image. In Section 4, we present the empirical results of classifying furniture style on the furniture style dataset. And finally, we discuss related works in Section 5.

2. Furniture Style Dataset

We identified 16 styles that are common across different types of furniture items. The list of furniture styles are shown in Figure [3] The style of a piece of furniture can be identified by looking at different aspect of the item, like its shape, color and so on. Depending on the type of furniture the style can be inferred by looking at different parts of an item. For example for chairs, the back is very discriminative of style and for tables the legs are also very telling. Since the type of furniture changes dramatically how style is identified, in this dataset we collect both the furniture type and the style. This allows for some very interesting research.
3. Furniture Style Classification

The creation of a large scale image classification dataset such as ImageNet [3] has fueled the last decade of computer vision research progress on deep neural network architecture for image classification. We have seen the network evolving from the initial AlexNet [7], to VGG [13], ResNet [5], Inception [16], Inception-Resnet V2 [15], and more recently NASNet [19], with deeper and deeper network, and higher and higher accuracy in terms of classifying images on the ImageNet dataset.

Another reason why ImageNet-based deep neural network architecture got huge amount of attention, is because the resulting network and its corresponding weights can be transferred to new classification tasks to bootstrap the training process [17], as the network has already learnt from ImageNet basic low level image features.

In this work, we consider two very recent network architectures Inception V3 (Iv3), and Inception-Resnet V2 (IRv2) to classify type and style of the furniture item independently.

The ImageNet dataset has only one single label for
each image, so the architecture of most image classification model look like the top module in Figure 5. In our dataset, we have two orthogonal labels for each image, furniture type and style. So we explore a network that consider both labels in a joint fashion, which is motivated by recent study on Multi-task Learning [10].

Figure 5. Model Architecture: Top - Single Task Model; Bottom - Multi-Task Model

In the bottom module of Figure 5, we show our model architecture for the furniture style classification problem, where the CNN network can be either Iv3 or IRv2 in our work. Let \( x \) be a furniture product image, \( L(x) \) be the overall loss, \( L_{style}(x) \) be the loss on style and \( L_{cat}(x) \) be the loss on category, we have the following loss which is used in our model,

\[
L(x) = wL_{style}(x) + (1 - w)L_{cat}(x)
\]  

where \( w \) is a weight which can be used to tune the relative importance of the two labels in our multi-task setting.

4. Experiment

In this section, based on the furniture style dataset, we study how state-of-the-art image classification networks are able to classify furniture images. Our experiments focus on the following three aspects: first, accuracy w.r.t. understanding furniture type and style; second, difficulty of differentiating different styles; finally, how style learnt by the model can be transferred to new furniture types.

We implemented all the experiments using Python 2.7 and TensorFlow 1.10. For this paper only, we start the training from a pre-trained set of weights created using ImageNet and downloaded from TensorHub [1], so that our results can be reproduced easily. Our optimization is done using RMSProp, we set decay to 0.9, momentum to 0.9, epsilon to 1.0, no locking for updating, and non-centered version for the optimizer. We use a simple grid search for tuning the learning rate (on a subset of the training data). In training, we use starting learning rate \( 5e-5 \), and exponential decaying rate 0.94 per every 2 epochs. We limit our training iterations to be 300 epochs over the dataset. We deployed our experiment on a GPU server with four Xeon E5-2660 v4 14-core CPU, four NVidia Tesla V100 GPU cards, 500GB DDR4 RAM, and 6.4 TB NVMe SSD Drive.

4.1. Classification Results

To understand the benefit of the multi-task model, we consider the following baseline models which are trained only on one of the two tasks: Iv3-Style which is Inception V3 model trained on furniture style labels only; Iv3-Type which is Inception V3 model trained on furniture type labels only; IRv2-Style which is Inception-Resnet V2 model trained on furniture style labels only; IRv2-Type which is Inception-Resnet V2 model trained on furniture type labels only. We use accuracy on each label to measure the performance of different models, where accuracy on a particular label, e.g., furniture style, is defined as the percentage of test furniture images which has been correctly classified to the corresponding label value in the test dataset.

We split the dataset between train and test by sampling each type and style combination individually. We assign 1/10 of the data to test but we guarantee at least two data points in the test for each type and style combination.

In Table 1 we show top-1 and top-5 accuracy results on both tasks with different model architecture and different values of \( w \) in Equation (1). For \( w = 1.0 \), the model is trained on style only, whereas for \( w = 0.0 \), the model is trained on furniture type only. Compared with models which have been trained only on one of the two tasks, the multi-task model can achieve a better accuracy by letting the model see more information from the other task. E.g., for the style classification task, Iv3-MultiTask can achieve a better top-1/top-5 accuracy on style when \( w \geq 0.6 \) in Equation (1). Similarly for Inception-Resnet V2, IRv2-MultiTask can achieve a better top-1/top-5 accuracy on style when \( w \geq 0.5 \) compared with IRv2-Style. This shows that having the two labels learnt together helps the model differentiating both the type and style of the piece of furniture. In Table 1 we can see that by setting \( w = 0.5 \), we can achieve the best test accuracy result on both tasks, defined as the sum of the accuracies. Finally from Table 1 we can easily observe that the more advanced model IRv2 can in general achieve a better performance compared to the simpler model Iv3.

As we discussed earlier, the difficulty of classifying style of furniture lies in the fact that the boundaries between some styles are vague. To verify this is the case, we show in Figure 6 the confusion matrix of style classification task, where entries in the matrix represent the percentage of misclassified test examples for the corresponding category on the x-axis, entries representing correct predictions are set to 0. It can be seen from this figure that there are pairs of styles which the models struggle to differentiate, e.g., modern vs. contemporary, and as discussed in the introduction, scandi-
<table>
<thead>
<tr>
<th>Model Arch</th>
<th>Model</th>
<th>w</th>
<th>Type Top-1</th>
<th>Style Top-1</th>
<th>Sum Top-1</th>
<th>Type Top-5</th>
<th>Style Top-5</th>
<th>Sum Top-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iv3</td>
<td>Iv3-Style</td>
<td>1.0</td>
<td>0.5726</td>
<td></td>
<td></td>
<td>0.9202</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Iv3-Type</td>
<td>0.0</td>
<td>0.6000</td>
<td></td>
<td></td>
<td>0.9362</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Iv3-MultiTask</td>
<td>0.2</td>
<td>0.5813</td>
<td>0.5192</td>
<td>1.1097</td>
<td>0.9317</td>
<td>0.8903</td>
<td>1.8219</td>
</tr>
<tr>
<td></td>
<td>Iv3-MultiTask</td>
<td>0.4</td>
<td>0.5601</td>
<td>0.5461</td>
<td>1.1562</td>
<td>0.9132</td>
<td>0.9117</td>
<td>1.8249</td>
</tr>
<tr>
<td></td>
<td>Iv3-MultiTask</td>
<td>0.5</td>
<td>0.5561</td>
<td>0.5686</td>
<td>1.1347</td>
<td>0.9082</td>
<td>0.9177</td>
<td>1.8259</td>
</tr>
<tr>
<td></td>
<td>Iv3-MultiTask</td>
<td>0.6</td>
<td>0.5327</td>
<td>0.5805</td>
<td>1.1132</td>
<td>0.9002</td>
<td>0.9247</td>
<td>1.8249</td>
</tr>
<tr>
<td></td>
<td>Iv3-MultiTask</td>
<td>0.8</td>
<td>0.5017</td>
<td>0.5975</td>
<td>1.0993</td>
<td>0.8783</td>
<td>0.9272</td>
<td>1.8055</td>
</tr>
<tr>
<td>Irv2</td>
<td>Irv2-Style</td>
<td>1.0</td>
<td>0.5850</td>
<td></td>
<td></td>
<td>0.9227</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Irv2-Type</td>
<td>0.0</td>
<td>0.6065</td>
<td></td>
<td></td>
<td>0.9431</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Irv2-MultiTask</td>
<td>0.2</td>
<td>0.6150</td>
<td>0.5117</td>
<td>1.1267</td>
<td>0.9411</td>
<td>0.8923</td>
<td>1.8334</td>
</tr>
<tr>
<td></td>
<td>Irv2-MultiTask</td>
<td>0.4</td>
<td>0.5935</td>
<td>0.5656</td>
<td>1.1591</td>
<td>0.9352</td>
<td>0.9242</td>
<td>1.8594</td>
</tr>
<tr>
<td></td>
<td>Irv2-MultiTask</td>
<td>0.5</td>
<td>0.5920</td>
<td>0.5950</td>
<td>1.1870</td>
<td>0.9392</td>
<td>0.9377</td>
<td>1.8718</td>
</tr>
<tr>
<td></td>
<td>Irv2-MultiTask</td>
<td>0.6</td>
<td>0.5840</td>
<td>0.5920</td>
<td>1.1761</td>
<td>0.9327</td>
<td>0.9307</td>
<td>1.8633</td>
</tr>
<tr>
<td></td>
<td>Irv2-MultiTask</td>
<td>0.8</td>
<td>0.5257</td>
<td>0.6090</td>
<td>1.1347</td>
<td>0.9042</td>
<td>0.9312</td>
<td>1.8354</td>
</tr>
</tbody>
</table>

Table 1. Accuracy results on furniture type and style.

...navian vs. mid-century. The confusion matrix also shows that the boundaries between country-cottage and modern farmhouse are also hard to identify. On the other hand, from the confusion matrix, we can also see that certain styles are easily differentiable from other style. E.g., the model achieves a high degree of accuracy for Bohemian, Traditional, and Mission furniture styles. This intuitively makes sense as these styles are more visually unique compared with other styles, and powerful models such as Iv3 and IRv2 can learn good visual features to differentiate them.

4.2. Style Understanding Across Category

We have discussed in the previous section that by learning from both tasks simultaneously, we can improve the classification accuracy of both furniture type and style. However, it is not clear whether the style information can be “transferred” between different furniture types, e.g., assume we hold out a few furniture types from the dataset, and train our model using only the remaining furniture types, can the learnt model still correctly identify style on the hold out types which were not seen during training?

To answer this question, we randomly picked five furniture types, bench, coffee table, dining room set, loveseat, and pub set, train our model using only remaining furniture types, and test how the learnt model can predict style for these five furniture types which the model has not seen before.

In Table 2, we show the style prediction accuracy for these five furniture types under settings where they have been seen in the training compared with the accuracy numbers under settings where they have not. As can be seen from the table, in general if the model has seen these furniture types during training, then we will have a better performance. For some furniture types, e.g., coffee table, the style learnt from other furniture types like table, dining table can help predict coffee table style. Thus we can still achieve a reasonable style prediction accuracy even though we have not seen coffee table examples during training. On the other hand for bench, we see the learnt model makes more style prediction mistakes. We believe this is because bench has a more unique shape compared with other furniture types in the dataset, thus making it more difficult to learn a good style prediction without seeing any examples.

5. Related Work

The field of image classification has seen a significant shift towards deep learning based approaches in the past 10 years. One of the earliest work in this domain is AlexNet [7], which used a much deeper and wider architecture and explored the usage of Rectified Linear Units (ReLU) and Dropout technology which have become extremely popular for later work. Later works such as VGG [13] and Inception [16], improved upon AlexNet by considering much smaller filters in convolutional layers, and $1 \times 1$ convolutional blocks. ResNet [5] revolutionized existing work in the field by introducing the novel identity link in-between...
different convolutional layers. This idea was later leveraged and incorporated into Inception to become the basis of Inception-ResNet v2 \cite{15}. Recently researchers from Google have proposed model architectures which instead of being manually designed, got learn through reinforcement learning \cite{19}.

Similar to other areas in Machine Learning and Computer Vision, the advancement of the image classification field has been mostly fueled by the availability of large-scale image dataset such as ImageNet \cite{3}, which provides the community with a standard to benchmark and improve existing techniques. However, ImageNet is a general purpose dataset, thus when comes to particular application we usually need dataset which is more tailored to the domain. E.g., recently researchers have proposed fine-granularity image classification dataset on different domains, such as dataset for vegetables and fruits \cite{6}, dataset for animals \cite{18}, dataset for aircrafts \cite{9}, and dataset for fashion \cite{8}, and dataset for furniture categorization \cite{4}.

6. Conclusion

In this work, we introduce a large scale image dataset of furniture style collected from web across a diverse set of furniture types. The dataset has been collected through general image search engine, and tagged by our in-house furniture style specialist. We benchmarked the two tasks of furniture type and style classification using state-of-the-art convolutional neural networks, and observed that by taking both labels into consideration at the same time through multi-task learning, we can achieve a better performance on both tasks. We study the challenge of classifying furniture styles, and we also discuss how learnt style understanding model can be leveraged to predict style information for a new furniture type. We plan to release the dataset and our benchmark to the public. We believe the release of this dataset will provide the community with a high quality data source which can be leveraged to study and benchmark new algorithms for classification, multi-task learning, and transfer learning.

References


