Discovering Pictorial Brand Associations from Large-Scale Online Image Data

Gunhee Kim
Disney Research Pittsburgh
gunhee@cs.cmu.edu

Eric P. Xing
Carnegie Mellon University
epxing@cs.cmu.edu

Abstract

In this paper, we study an approach for discovering brand associations by leveraging large-scale online photo collections contributed by the general public. Brand Associations, one of central concepts in marketing, describe customers' top-of-mind attitudes or feelings toward a brand. (e.g. *what comes to mind when you think of Burberry?) Traditionally, brand associations are measured by analyzing the text data from consumers' responses to the survey or their online conversation logs. In this paper, we go beyond textual media and take advantage of large-scale photos shared on the Web. More specifically, we jointly achieve the following two fundamental tasks in a mutually-rewarding way: (i) detecting exemplar images as key visual concepts associated with brands, and (ii) localizing the regions of brand in images. For experiments we collect about five millions of images of 48 brands crawled from five popular online photo sharing sites. We then demonstrate that our approach can discover complementary views on the brand associations that are hardly obtained from text data. We also quantitatively show the superior performance of our algorithm for the two tasks over other candidate methods.*

1. Introduction

The *brand equity*, one of core concepts in marketing, describes a set of values or assets linked to a brand [1, 11]. It is a key source of bearing the competitive advantage of a company over its competitors, boosting efficiency and effectiveness of marketing programs, and attaining the price premium due to increased customer satisfaction and loyalty, to name a few. One central component of the brand equity is *brand associations*, which are the set of associations consumers perceive with the brand [11]. Its significance lies in that it is a *customer-driven* brand equity. In other words, the brand associations are directly connected to customers' *top-of-mind* attitudes or feelings toward the brand over certain products or services, which provoke the reasons to preferentially purchase the products or services of the brand. For example, if a customer strongly associates *Burberry* with *men's coats*, he may tend to first consider *Burberry* products over other competitors' ones when he needs it.

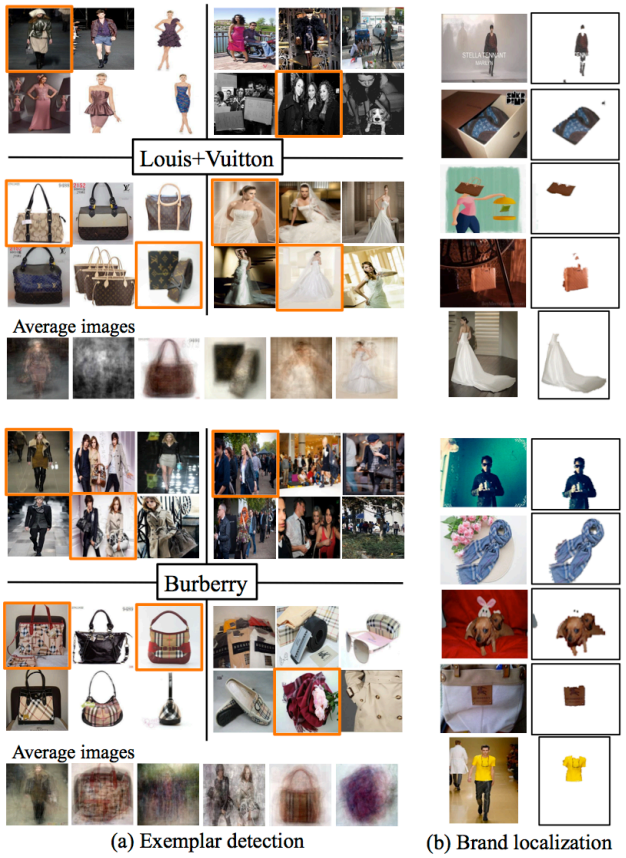


Figure 1. Motivation for discovering brand associations from Web community photos with two examples of *Louis+Vuitton* and *Burberry*. (a) Task 1: we perform exemplar detection and clustering to discover the key visual concepts of brands. We choose 24 exemplars (*i.e.* cluster centers) per brand. In the bottom, we also show the average images of six sampled clusters, whose exemplars are bounded in the orange color. (b) Task 2: we identify the most likely regions of brand in each image.

Traditionally, brand associations are built from direct consumer responses to carefully-designed questionnaires [4, 5, 11, 22, 24]. The survey with human subjects are usually time-consuming is prone to suffer from sam-

pling bias and common methods bias. To overcome these issues, with the recent emergence of online social media, it has become popular to indirectly leverage consumer generated data on online communities. Beneficially, the resources on such social media are obtainable inexpensively and almost instantaneously from a large crowd of potential customers. One typical example of such endeavor is the *Brand Association Map* developed by Nielsen Online [2, 17], in which central concepts and themes correlated with a given brand name are automatically extracted from billions of online conversations over Weblogs, boards, and Wiki.

In this paper, we propose to take advantage of large-scale online *photo* collections toward discovering brand associations. Digital images are rapidly gaining popularity as a form of communicating information online, but have been not explored so far for this purpose. With handy availability of digital cameras and smartphones, people can freely take the pictures on their favorite products. Moreover, many online tools enable users to easily bookmark, comment, or share the images of products that they wish to buy.

More specifically, our objective in this paper is to develop novel methods to jointly perform the following two basic tasks as an initial step toward the study of photo-based brand associations. (See the examples in Fig.1). The two tasks can be regarded as *image level* and *sub-image level* discovery of brand associations, respectively.

(1) *Detecting key image clusters and their exemplars associated with brands*: One of core problems in brand association research has been to identify important concepts associated with brands, and visualize them in a form of networks or maps [2, 4, 22, 24]. Therefore, as shown in Fig.1.(a), our first task is to identify a small number of exemplars and image clusters as core visual concepts of brands from large-scale and ever-growing online image data that are tagged and organized by general users.

(2) *Localizing the regions of brand in images*: Our second task aims to localize the regions that are most relevant to the brand in each image in an unsupervised way, as shown in Fig.1.(b). In other words, we detect the regions that frequently recur across the image set of the brand. We perform pixel-level image segmentation to delineate the regions of brand. Therefore, this task helps reveal typical interactions between users and products in natural social scenes, which can lead various potential benefits, ranging from content-based image retrieval to online multimedia advertisement.

In this paper, we propose to close a loop between solving the two tasks since they are *mutually rewarding*. The exemplar detection and clustering can group the similar images, which can promote the brand localization since we can leverage the recurring foreground signals. At the same time, localizing the regions of brands can enhance the similarity measurement between images, which subsequently contributes better exemplar detection and clustering.

For evaluation, we collect about five millions of images of 48 brands of four categories (*i.e*. *sports*, *luxury*, *beer*, and *fastfood*) from five popular photo sharing sites, including FLICKR, PHOTOBUCKET, DEVIANTART, TWITPIC, and PINTEREST. In our experiments, we present the examples of picture-based brand associations for some selected brands. We also demonstrate compelling quantitative results of our exemplar detection/clustering and brand localization approach over other candidate methods.

1.1. Relations to Previous work

Measuring brand associations: In almost all previous studies on brand associations, surveys on customers are the main approach for collecting source data [2, 4, 22, 24]. Therefore, from a viewpoint of brand association research, our contribution is to introduce a novel source of data for the study. In this category of research, the brand association map of Nielsen Online [2, 17] is closely related to our work since both approaches explore online data. However, the uniqueness of our research lies in leveraging online image collections, which convey complementary views on the associations that are hardly captured by texts. In addition, we localize the regions of brand in every photo, which is another novel feature of our work.

Computer vision for product images: Recently, with the exploding interests in electronic commerce, computer vision techniques have widely applied to analyze product images for commercial applications. Some notable examples include the product image search and ranking [9], the logo and product detection in natural images [10, 15], the attribute discovery in product images [3], and clothing parsing in fashion photos [26]. However, the objectives of our work fundamentally differ from those of previous work; we aim to extract and visualize the core concepts of the brands from extremely diverse online pictures, whereas most of past research has focused on detecting a fixed number of specified product models or logos in the images. Therefore, in our work, it is important to mine the visual topics that do not explicitly contain the products but reflect general public's thoughts, feelings, or experiences over the brands (*e.g*. sponsored yacht competition scenes in the *Rolex* image set).

1.2. Contributions

The contributions of our work are outlined as follows:

(1) We study the problem of discovering image and sub-image level brand associations from large-scale online photos. As far as we know, our work is the first attempt so far on such photo-based brand association analysis. For marketing research, our work can provide another novel and complementary way to visualize general public's impressions or thoughts on the brands. For computer vision research, our work can widen its applicability to new areas of electronic commerce.

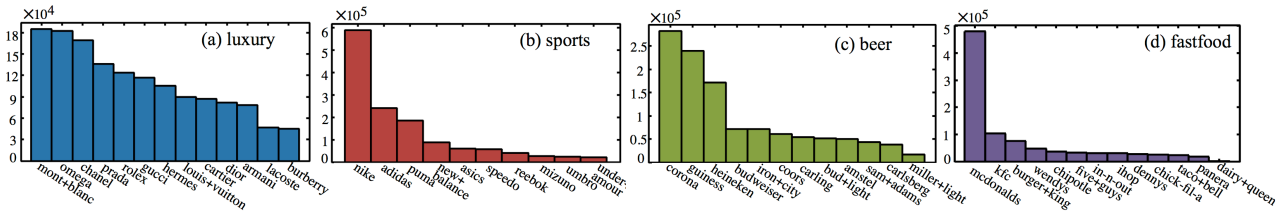


Figure 2. The dataset of 48 brands crawled from five photo sharing sites: FLICKR, PHOTOBUCKET, DEVIANTART, TWITPIC, and PINTEREST. The total number of images is 4,720,724.

(2) We propose an algorithm to jointly achieve exemplar detection/clustering and brand localization in a mutually-rewarding way. With experiments on about five million images of 48 brands from five popular photo sharing sites, we quantitatively demonstrate that our approach outperforms other candidate methods for the two tasks.

2. Problem Setting

2.1. Image Data Crawling

Since we are interested in consumer-driven views on the brands, we use the online photos that are contributed and organized by general Web users. As source data, we crawl images from the five popular photo sharing sites, in which the characteristics of the pictures are different from one another. The FLICKR and PHOTOBUCKET are the two most popular photo sharing sites in terms of volumes of photos. The PINTERST hosts the pictures bookmarked by users, and the DEVIANTART contains the artwork created by users (*e.g.* not only photos but also digital art). The TWITPIC includes the pictures shared via the Twitter. We exclude the GOOGLE IMAGE SEARCH because much of the pictures are originated from online shopping malls or news agencies.

We query the brand names via the built-in search engines of the above sites to search for the pictures tagged with brand names. We download all retrieved images without any filtering. We also crawl meta-data of the pictures (*e.g.* time stamps, titles, user names, texts), if available.

Fig.2 summarizes our dataset of 4,783,345 images for 48 brands, which can be classified into four categories: *luxury*, *sports*, *beer*, and *fastfood*. The number of images per brand varies much according to the popularity of the brand.

2.2. Overview of Our Approach

The input of our algorithm is a set of images for a brand of interest, which is denoted by $\mathcal{I} = \{I_1, \cdots, I_N\}$. We let N to be the number of images. The first step is to build a K-nearest neighbor (KNN) graph $\mathcal{G} = (\mathcal{I}, \mathcal{E})$, in which each image is connected with its K most similar images in $\mathcal{I}$. We will present our image descriptors, similarity measures, and KNN graph construction in section 3.1.

The next step is to perform exemplar discovery and clustering. We find M number of exemplars $\mathcal{A}(\subset \mathcal{I})$, which are representative images that are distinctive from one another. The exemplar discovery is important in two aspects. First, the exemplars comprise of a concise but comprehensive set of key visual concepts associated with brands. Second, for brand localization, the exemplars are used as references to detect the most brand-related regions in cluttered images. We formulate the exemplar detection as diversity ranking on $\mathcal{G}$, which aims to rank and choose the M best nodes in the graph to reduce redundancy while maintaining their centrality. Then, the image clustering partitions $\mathcal{I}$ into a cluster set $\mathcal{C} = \{\mathcal{C}_1, \cdots, \mathcal{C}_M\}$ so that each image is associated with its most prototypical exemplar. The details of our approach will be discussed in section 3.2.

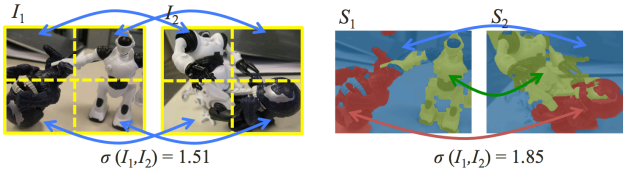
Next, we carry out the brand localization, whose goal is to search for the regions that are most relevant to the brand in an image. The brand localization is achieved by applying the *cosegmentation* technique [14, 12, 19], in which we simultaneously segment multiple images that share common objects or foregrounds, given that the recurring objects across the images can be leveraged as a high-level clue to the regions of interest to be segmented. We summarize the procedure of cosegmentation in section 3.3.

The exemplar detection/clustering and the cosegmentation are closed in a loop so that they can be mutually rewarding. The clustering helps discover the coherent groups of images from extremely diverse Web images. Conversely, segmentation can enhance the exemplar finding and clustering by promoting a more accurate image similarity measure, which will be justified in section 3.1. After finishing the cosegmentation step, we can return to the KNN graph construction and repeat the whole algorithm again with the new segmentation-based image similarity metric.

3. Approach

3.1. Constructing Similarity Graph of Images

Image description: We use one of standard image description methods as follows. We densely extract HSV color SIFT and histogram of oriented edge (HOG) feature on a regular grid of each image at steps of 4 and 8 pixels, respectively. Then, we form 300 visual words for each feature type by applying K-means to randomly selected features. Finally, the nearest word is assigned to every node of the grid. As the image and region descriptor, we build L_1 normalized spatial pyramid histogram to count the frequency



(a) Similarity by spatial histogram (b) Similarity by segmentation

Figure 3. The benefit of segmentation for image similarity measurement. (a) Before segmentation, the spatial pyramid histograms on the whole images may not correctly reflect the location and scale variations. (b) After segmentation, the image similarity is computed as the mean similarity of best assigned segments.

of each visual word in the two levels of regular grids [16].

Image similarity: We compute the image similarity differently according to whether image segmentation is available or not. We argue that even imperfect segmentation helps enhance the measurement of image similarity, which can justify closing a loop between the exemplar detection/clustering and the segmentation for brand localization. Fig.3 shows a typical example, in which the two images are similar in that both include the two same toys of the *Burger+king*. When images are not yet segmented, the image similarity is calculated by applying the Gaussian kernel to the two-level spatial pyramid histograms, which are not robust against location, scale, and pose variation as shown in Fig.3.(a). On the other hand, segmentation can largely alleviate this issue as shown in Fig.3.(b). Given the two segment sets of the two images, we find the best one-to-one matches between them by solving the linear assignment problem. Then, we compute the mean of similarities between corresponding segments as an image similarity metric. For the segment similarity, we use the same Gaussian kernel on the spatial pyramids of the segments.

K-nearest neighbor graph: We use the descriptors and similarity measures defined above, in order to construct a KNN graph between images. For a large $\mathcal{I}$, comparing all pairwise similarity by brute-force, which takes $\mathcal{O}(N^2)$, can be overly slow. In such cases, we exploit the idea of multiple random divide-and-conquer [25] using meta-data of images, which allows to create an approximate KNN of high accuracy within $\mathcal{O}(N \log N)$ time.

3.2. Exemplar detection and clustering

Given a KNN graph $\mathcal{G}$, we then perform the exemplar detection. As a base algorithm, we use the diversity ranking algorithm of [14], which solve submodular optimization on the similarity graph $\mathcal{G}$, and choose L number of exemplars that are not only most central but also distinctive one another. Since L exemplars are discovered in a decreasing order of ranking scores, one can set L to an arbitrary large number. In this paper, we denote the exemplar detection procedure by $\mathcal{A} = \texttt{SubmDiv}(\mathbf{G}, L)$ where $\mathcal{A}$ is the set of exemplars and $\mathbf{G} \in \mathbb{R}^{N \times N}$ is the adjacency matrix of the graph $\mathcal{G}$. The pseudocode is summarized in the step **1**–**2** of Algorithm 1, and its detailed procedures and theoretic analyses are referred to [14].

Algorithm 1: Exemplar detection and clustering.

Input: (1) Image graph $\mathbf{G}$. (2) Number of exemplars L.
Output: (1) Exemplar set $\mathcal{A}$ and cluster set $\mathcal{C}$.

1: Append a constant vector $\mathbf{z} \in \mathbb{R}^{(N+1)\times 1}$ to the end column of $\mathbf{G}$ and $\mathbf{z}^T$ to the end row of $\mathbf{G}$. ($N = |\mathbf{G}|$).
2: $\mathcal{A} = \texttt{SubmDiv}(\mathbf{G}, M)$.
3: $\{\mathcal{C}_l\}_{l=1}^{L} = \texttt{ClustSrc}(\mathbf{G}, \mathcal{A})$.

/* Select M number of central and diverse exemplars $\mathcal{A}$.
Function $[\mathcal{A}] = \texttt{SubmDiv}(\mathbf{G}, M)$
  1: $\mathcal{A} \leftarrow \emptyset$. $\mathbf{u} = \mathbf{0} \in \mathbb{R}^{N\times 1}$.
  while $|\mathcal{A}| \leq L$ **do**
    2: **for** $i = 1 : N$ **do** $\mathbf{u}(i) = \texttt{TempSrc}(\mathbf{G}, \{\mathcal{A} \cup i\})$.
    3: $\mathcal{A} \leftarrow \mathcal{A} \cup \text{argmax}_i\, \mathbf{u}$. Set $\mathbf{u} = \mathbf{0}$.

/* Get marginal gain u from the $\mathbf{G}$ and the node set $\mathcal{P}$.
Function $[u] = \texttt{TempSrc}(\mathbf{G}, \mathcal{P})$
  1: Solve $\mathbf{u} = \mathbf{L}\mathbf{u}$ where $\mathbf{L}$ is the Laplacian of $\mathbf{G}$ under constraints of $\mathbf{u}(\mathcal{P}) = 1$ and $\mathbf{u}(N+1) = 0$.
  2: Compute the marginal gain $u = |\mathbf{u}|_1$.

/* Get cluster set $\mathcal{C}$ from the graph $\mathbf{G}$ and exemplars $\mathcal{A}$.
Function $\mathcal{C} = \texttt{ClustSrc}(\mathbf{G}, \mathcal{A})$
  1: Let $L = |\mathcal{A}|$ and $L = |\mathbf{G}|$. $\mathcal{V}$ is vertext set of $\mathbf{G}$.
  2: Compute the matrix $\mathbf{X} \in \mathbb{R}^{(L-L)\times L}$ by solving $\mathbf{L}_u\mathbf{X} = -\mathbf{B}^T\mathbf{I}_s$ where if we let $\mathcal{X} = \mathcal{V}\backslash\mathcal{A}$, $\mathbf{L}_u = \mathbf{L}(\mathcal{X}, \mathcal{X})$, $\mathbf{B} = \mathbf{L}(\mathcal{A}, \mathcal{X})$, and $\mathbf{I}_s$ is an $L \times L$ identity matrix.
  3: Each vertex $v \in \mathcal{V}$ is clustered $c_v = \text{argmax}_k\, \mathbf{X}(j,k)$.

Next, we run the clustering using the random walk model [8], in which each image i is associated with the exemplar that a random walker starting at i is most likely to reach first. Then, we cluster the images that share the same exemplar as the most probable destination. This procedure is implemented as a function `ClustSrc` of Algorithm 1.

3.3. Brand Localization via Cosegmentation

The clustering output is the groups of coherent images $\mathcal{C} = \{\mathcal{C}_l\}_{l=1}^{L}$. The brand localization is achieved by separately applying the cosegmentation algorithm to each cluster. The separate cosegmentation scheme is beneficial for both scalability and performance. For scalability, it can promote parallel computation trivially. For performance, it prevents cosegmenting the images of no commonality, which contradicts the basic assumption of cosegmentation. Obviously, the images in the same cluster are likely to share similar product types or themes of the brand, which can be detected by the cosegmentation approach.

The goal of cosegmentation is to partition each image into foreground (*i.e.* the regions recurring across the images) and background (*i.e.* the other regions). We select the MFC method [12] as our base cosegmentation algorithm, since it is scalable and has been successfully tested

Algorithm 2: Brand localization via cosegmentation.

Input: (1) Cluster set $\mathcal{C} = \{\mathcal{C}_l\}_{l=1}^{L}$. (2) Image graph $\mathbf{G}$.
Output: (1) Set of segmented images $\mathcal{F}$ for each $i \in \mathcal{I}$.

foreach $\mathcal{C}_l \in \mathcal{C}$ **do**
- **1**: Find central image $c = \texttt{SubmDiv}(\mathbf{G}_l, 1)$ where $\mathbf{G}_l = \mathbf{G}(\mathcal{C}_l)$ is the subgraph of $\mathcal{C}_l$.
- **2**: Apply the unsupervised MFC algorithm [12] to $\{c \cup \mathcal{N}_c\}$ where $\mathcal{N}_c$ is the neighbor of c in the graph $\mathbf{G}_l$. As a result, we obtain segmented images $\mathcal{F}_{c \cup \mathcal{N}_c}$.
- **3**: Let $\mathcal{U}_l \leftarrow \mathcal{C}_l \backslash \{c \cup \mathcal{N}_c\}$. $\mathcal{F} \leftarrow \mathcal{F}_{c \cup \mathcal{N}_c}$.
- **while** $\mathcal{U}_l \neq \emptyset$ **do**
  - **4**: Sample an image i from $\{\mathcal{U}_l \cap \mathcal{N}_{\mathcal{F}}\}$.
  - **5**: Get foreground model $\{v_i\} = \texttt{FM}\,(\{\mathcal{N}_i \cap \mathcal{F}\})$.
  - **6**: Segment the image $\mathcal{F}_i = \texttt{RA}\,(i, \{v_i\})$.
  - **7**: $\mathcal{U}_l \leftarrow \mathcal{U}_l \backslash i$. $\mathcal{F} \leftarrow \mathcal{F} \cup \mathcal{F}_i$.

/* $\{v_i\} = \texttt{FM}\,(\mathcal{F}_i)$ is the function to learn foreground model $\{v_i\}$ of MFC [12] from the segmented images $\mathcal{F}_i$.
/* $\mathcal{F}_i = \texttt{RA}\,(i, \{v_i\})$ is the function to run region assign- ment of MFC [12] on image i using $\{v_i\}$.

with Flickr user images. The MFC algorithm consists of two procedures, which are *foreground modeling* and *region assignment*. The foreground modeling step learns the appearance models for foreground and background, which are accomplished by using any region classifiers or their combinations. We use the Gaussian mixture model (GMM) on the RGB color space. The foreground models compute the values of any given regions with respect to the foregrounds and background, based on which the region assignment allocates the regions of an image via a combinatorial-auction style optimization to maximize the overall allocation values. More details of the algorithm can be found in [12].

For each cluster $\mathcal{C}_l$, we perform the cosegmentation by iteratively applying the foreground modeling and region assignment steps under the guidance of the subgraph $\mathcal{G}(\mathcal{C}_l)$ whose vertex set is $\mathcal{C}_l$. Inspired by [13], its basic idea is that the neighboring images in $\mathcal{G}(\mathcal{C}_l)$ are visually similar, and thus they are likely to share enough commonality to be segmented together. Therefore, we iteratively segment each image i by using the learned foreground models from its neighbors in the graph. Then, the segmented image i is subsequently used to learn the foreground models for its neighbors' segmentation. That is, we iteratively run foreground modeling and region assignment by following the edges of $\mathcal{G}(\mathcal{C}_l)$. The overall algorithm is summarized in Algorithm 2. For initialization, as shown in step **1**–**2** of Algorithm 2, we run the unsupervised version of the MFC algorithm to the exemplar of C_l and its neighbors, from which the iterative cosegmentation starts.

4. Experiments

In our evaluation, we first present the exemplars for the core concepts of several competing brands that are detected by our method. Then, we quantitatively evaluate the proposed approach from two technical perspectives: exemplar detection/clustering in section 4.2, and brand localization via image cosegmentation in section 4.3. Since the main goal of this paper is to achieve two technical tasks for brand associations, we focus on the validation of algorithms instead of user study.

4.1. Examples of Brand Associations

Fig.4, shows high-ranked visual concepts of three competing brands of the *sports* and *beer* category, respectively. For each brand, we first find the 30 top-ranked exemplars, from which we manually select 20 ones. Such manual selection is done due to remove highly redundant ones. On the right, we also show six sampled average images for the exemplars with the orange-colored boundaries. Each average image is obtained from the 40 closest neighbors to each exemplar in the same cluster.

We make several interesting observations as follows. First of all, all brands show their own characteristic visual themes, which are distinctive between competing brands. (*e.g.* *basketball* in the *Nike*, *American football* in the *Reebok*, and *swimming* in the *Speedo*). Second, since we use general users' photos on the Web, more of highly ranked exemplars attribute to users' experiences on the brands rather than the products themselves. For example, one can see many *jogging* scenes in the *sport* category and the *party* scenes in the *beer* category. It may result from that people prefer to take pictures on personal memorable moments. Third, the exemplar output suggests that, as one of future work, it is important to correctly deal with polysemous brand names. For example, the *Coors* is also the name of a baseball stadium, and thus many baseball scenes are detected as dominant exemplars.

4.2. Results on Clustering

Task: We evaluate the performance of our algorithm for the exemplar detection/clustering task, by comparing with several candidate methods. For quantitative evaluation, we first choose 20 brands (*i.e.* five brands per category), and generate 100 sets of groundtruth per brand as follows. We randomly sample three images (i, j, k) from the image set of a brand, and manually label which of j and k is more similar to image i. We denote $j \succ k|i$ if j is more similar to i than k. Although the labeled sets are relatively few compared to the dataset size, in practice this sampling-based annotation is commonly adopted in standard large-scale benchmark datasets such as ImageNet [6] and LabelMe [21].

After applying each algorithm, suppose that C_i, C_j, and C_k denote the clusters that include image i, j, and k, respectively. Then, we compute the similarity between clusters $\sigma(C_j, C_i)$ and $\sigma(C_k, C_i)$ by using the *random walk with restart* (RWR) algorithm [23], where $\sigma(C_j, C_i)$ is propor-

Figure 4. Examples of top-ranked exemplars associated with three competing brands in the sports and beer category, respectively. We show 20 exemplars on the left, and 6 sampled average images for the exemplars with the orange boundary.

tional to the probability that a random walker stays in cluster C_j when the walker follows the edges of the graph with probability λ and randomly restarts from cluster C_i with probability $1-\lambda$. Finally, we compute the accuracy of the algorithm using the Wilcoxon–Mann–Whitney statistics:

$$ACC := \frac{\sum_{(i,j,k)} \mathbb{I}(j \succ k | i \ \wedge \ \sigma(C_j, C_i) > \sigma(C_k, C_i))}{\sum_{(i,j,k)} \mathbb{I}(j \succ k | i)}$$

where $\mathbb{I}$ is an indicator function. The accuracy increases only if the algorithm can partition the image set into coherent clusters, and the similarities between clusters coincide well with human's judgment on the image similarity.

Baselines: We compare our algorithm with four baselines. The (`KMean`) and the (`Spect`) are the two popular clustering methods, K-means and spectral clustering, respectively. The (`LP`) is a label propagation algorithm for community detection [18], and the (`AP`) is the *affinity propagation* [7], which is a message-passing based clustering algorithm. Our algorithm is tested in two different ways, according to whether image segmentation is in a loop or not. The (`Sub`) does not exploit the image cosegmentation

Methods	*Sports*					*Fastfood*					*Beer*					*Luxury*				
	AD	AS	NK	RB	SP	BG	IN	MC	TB	PN	BD	CA	CO	GU	HN	AM	BB	HM	LV	RL
Sub-M	**69**	**67**	**67**	68	**67**	**58**	58	**68**	**65**	**64**	**53**	**53**	**56**	**55**	**53**	**61**	**64**	**64**	**65**	**64**
Sub	64	62	62	**69**	65	53	**60**	63	**65**	61	48	**53**	48	48	46	56	56	60	63	53
Kmean	56	58	54	60	57	46	47	56	56	51	43	42	43	43	41	47	53	51	55	51
Spect	55	55	46	61	52	42	44	57	55	55	41	43	41	41	41	50	52	49	52	51
LP	57	57	55	64	57	48	47	61	55	53	43	42	43	45	43	50	54	48	56	50
AP	61	59	57	63	56	50	47	56	59	54	43	42	43	42	47	51	50	49	55	49

Sports: AD (*adidas*), AS (*asics*), NK (*nike*), RB (*reebok*), SP (*speedo*). **Fastfood**: BG (*burger+king*), IN (*in-n-out*), MC (*mcdonalds*), TB (*taco+bell*), PN (*panera*). **Beer**: BD (*budweiser*), CA (*carlsberg*), CO (*coors*), GU (*guinness*), HN (*heineken*). **Luxury**: AM (*armani*), BB (*burberry*), HM (*hermes*), LV (*louis+vuitton*), RL (*rolex*).

Table 1. Clustering accuracies of two variants of our approach (Sub-$\star$) and four baselines for the 20 selected brands. The average accuracies over the 20 brands are (Sub-M): **62.0**%, (Sub): 57.8%, (Kmean): 50.5%, (Spect): 49.2%, (LP): 51.4%, and (AP): 51.7%.

output, whereas the (Sub-M) is our fully geared approach. That is, this comparison can justify the usefulness of our alternating approach between clustering and cosegmentation. We set $L = 300$, and use the same image features in section 3.1 for all the algorithms.

Quantitative results: Table 1 reports the results of our algorithm and four baselines across 20 brand classes. In most brand classes, the accuracies of our method (Sub-M) are better than those of all the baselines. The average accuracy of our (Sub-M) is 62.0%, which is much higher than 51.7% of the best baseline (AP). In addition, the average accuracies of the (Sub-M) are notably better than (Sub), which implicates that the cosegmentation for brand localization can improve the clustering performance as expected.

4.3. Results on Brand Localization

Task: The brand localization task is evaluated as follows. As groundtruths, we manually annotate 50 randomly sampled images per brand, for the same 20 brands in the previous experiments. We do not label too obvious images depicting products on white background, since we here are interested in measuring the localization capabilities of the algorithms for natural images. The accuracy is measured by the intersection-over-union metric $(GT_i \cap R_i)/(GT_i \cup R_i)$, where GT_i is the groundtruth of image i and R_i is the regions segmented by the algorithm. It is a standard metric in segmentation literature [12, 14]. We compute the average accuracy from all annotated images.

Baselines: We select two baselines that can discover and segment the regions of objects from a large set of images in an unsupervised manner (*i.e.* with no labeled seed images). The (LDA) [20] is an LDA-based unsupervised localization method, and the (COS) [14] is a state-of-art submodular optimization based cosegmentation algorithm. Our algorithm is tested in three different versions, according to whether exemplar detection/clustering is in a loop or not. The (MFC) runs our cosegmentation without involving our clustering output (but using a random partitioning instead), in order to show the importance of the clustering step when segmenting highly diverse Web images. The (MFC-S) is a single loop of our exemplar detection/clustering and cosegmentation, and (MFC-M) iterates this process more than twice. In almost all cases, it converges in two iterations. Hence, this comparison can quantify the accuracy increase by the iterative algorithm. We run all algorithms in an unsupervised way for a fair comparison. Since it is hard to know the best number of foregrounds K in advance (*e.g.* multiple foregrounds may exist in each image), we repeat each method by changing K from one to five, and report the best results.

Quantitative results: Table2 shows that our method outperforms other candidate methods in almost all classes. Especially, our average accuracy is 49.5%, which is notably higher than 36.7% of the best baseline (COS). In addition, the average accuracy of the (MFC-M) is also higher than those of (MFC-S) and (MFC), which demonstrates that the clustering and cosegmentation are mutually-rewarding.

Qualitative analysis: Fig.5 shows six groups of brand localization examples. The images of each group belong to the same cluster, and thus are cosegmented. The online user images contain extremely diverse topics and their appearances, even though they are associated with the same brand. (*e.g.* The horse images in Fig.5.(c) are seemingly irrelevant to the *reebok*). Our approach can detect a small set of representative exemplars, cluster images accordingly, and segment common regions in an unsupervised and bottom-up way. Consequently, our approach shows a potential to be a useful building block for various Web applications (*e.g.* brand detection for online multimedia advertisement).

5. Conclusion

In this paper, we proposed a novel approach for discovering the brand associations by leveraging large-scale photo collections shared online. With the experiments of about five millions of images for 48 brands, we have found several novel observations from the picture-based brand associations. We also demonstrated superior clustering and brand localization performance over other candidate methods.

Acknowledgement: This work is supported by NSF IIS-1115313 and AFOSR FA9550010247.

Methods	*Sports*					*Fastfood*					*Beer*					*Luxury*				
	AD	AS	NK	RB	SP	BG	IN	MC	TB	PN	BD	CA	CO	GU	HN	AM	BB	HM	LV	RL
MFC-M	44.6	53.1	**52.8**	57.7	43.7	**42.4**	**47.8**	52.9	**55.5**	**50.1**	43.9	**47.7**	**45.0**	**54.9**	**53.9**	**46.1**	49.3	49.1	52.9	**47.2**
MFC-S	**55.1**	45.8	47.1	**63.8**	39.9	31.9	31.8	**58.0**	53.0	47.2	38.1	47.5	42.8	45.5	48.0	42.8	**51.3**	**49.8**	54.9	41.9
MFC	44.7	50.6	42.9	52.1	**54.9**	24.3	29.4	40.9	48.1	43.8	**44.0**	45.3	31.6	37.5	37.6	40.5	36.7	38.0	**55.7**	35.9
COS	23.8	**66.1**	49.3	45.1	48.8	25.4	31.7	34.3	31.1	40.5	34.9	32.5	30.4	37.6	41.3	23.6	38.5	26.8	31.1	40.4
LDA	28.8	39.4	29.8	39.6	23.2	21.9	22.5	38.8	33.2	41.0	25.9	31.5	25.1	39.2	32.1	25.1	34.6	26.4	30.0	24.7

Table 2. Brand localization accuracies of three variants of our approach (MFC-$\star$) and two baselines. The average accuracies are (MFC-M): **49.5**%, (MFC-S): 46.8%, (MFC): 41.7%, (COS): 36.7%, and (LDA): 30.6%.

(a) speedo (b) reebok (c) reebok

(d) taco+bell (e) armani (f) burger+king

Figure 5. Six groups of brand localization examples. The images of each group belong to the same cluster, and thus are jointly segmented.

References

[1] D. A. Aaker. Measuring Brand Equity Across Products and Markets. *Cal. Manag. Rev.*, 38(3):102–120, 1996. 1

[2] N. Akiva, E. Greitzer, Y. Krichman, and J. Schler. Mining and Visualizing Online Web Content Using BAM: Brand Association Map. In *ICWSM*, 2008. 2

[3] T. L. Berg, A. C. Berg, and J. Shih. Automatic Attribute Discovery and Characterization from Noisy Web Data. In *ECCV*, 2010. 2

[4] A. C.-H. Chen. Using Free Association to Examine the Relationship between the Characteristics of Brand Associations and Brand Equity. *J. Product Brand Management*, 10(7):439–451, 2001. 1, 2

[5] J. E. Danes, J. S. Hess, J. W. Story, and J. L. York. Brand Image Associations for Large Virtual Groups. *Qualitative Market Research*, 13(3):309–323, 2010. 1

[6] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009. 5

[7] B. J. Frey and D. Dueck. Clustering by Passing Messages Between Data Points. *Science*, 315:972–976, 2007. 6

[8] L. Grady. Random Walks for Image Segmentation. *IEEE PAMI*, 28:1768–1783, 2006. 4

[9] Y. Jing and S. Baluja. PageRank for Product Image Search. In *WWW*, 2008. 2

[10] H. Kang, M. Hebert, A. A. Efros, and T. Kanade. Connecting Missing Links: Object Discovery from Sparse Observations Using 5 Million Product Images. In *ECCV*, 2012. 2

[11] K. L. Keller. Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *J. Marketing*, 57(1):1–22, 1993. 1

[12] G. Kim and E. P. Xing. On Multiple Foreground Cosegmentation. In *CVPR*, 2012. 3, 4, 5, 7

[13] G. Kim and E. P. Xing. Jointly Aligning and Segmenting Multiple Web Photo Streams for the Inference of Collective Photo Storylines. In *CVPR*, 2013. 5

[14] G. Kim, E. P. Xing, L. Fei-Fei, and T. Kanade. Distributed Cosegmentation via Submodular Optimization on Anisotropic Diffusion. In *ICCV*, 2011. 3, 4, 7

[15] J. Kleban, X. Xie, and W.-Y. Ma. Spatial Pyramid Mining for Logo Detection in Natural Scenes. In *ICME*, 2008. 2

[16] S. Lazebnik, C. Schmid, and J. Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *CVPR*, 2006. 4

[17] NielsenOnline. Brand Association Map, 2010. 2

[18] U. N. Raghavan, R. Albert, and S. Kumara. Near Linear Time Algorithm to Detect Community Structures in Large-Scale Networks. *Phys Rev E*, 76(036106), 2007. 6

[19] C. Rother, T. Minka, A. Blake, and V. Kolmogorov. Cosegmentation of Image Pairs by Histogram Matching Incorporating a Global Constraint into MRFs. In *CVPR*, 2006. 3

[20] B. C. Russell, A. Efros, J. Sivic, W. T. Freeman, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In *CVPR*, 2006. 7

[21] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: A Database and Web-based Tool for Image Annotation. *IJCV*, 77:157–173, 2008. 5

[22] O. Schnittka, H. Sattler, and S. Zenker. Advanced Brand Concept Maps: A New Approach for Evaluating the Favorability of Brand Association Networks. *I. J. Research in Marketing*, 2012. 1, 2

[23] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos. Neighborhood Formation and Anomaly Detection in Bipartite Graphs. In *ICDM*, 2005. 5

[24] B. D. Till, D. Baack, and B. Waterman. Strategic Brand Association Maps: Developing Brand Insight. *J. Product Brand Management*, 20(2):92–100, 2011. 1, 2

[25] J. Wang, J. Wang, G. Zeng, Z. Tu, R. Gan, and S. Li. Scalable k-NN Graph Construction for Visual Descriptors. In *CVPR*, 2012. 4

[26] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Parsing Clothing in Fashion Photographs. In *CVPR*, 2012. 2