Image Aesthetics and its Effects on Product Clicks in E-Commerce Search

Alexandre Maros, Fabiano Belém, Rodrigo Silva, Sérgio Canuto
Jussara M. Almeida, Marcos A. Gonçalves
Universidade Federal de Minas Gerais - Computer Science Department
{alexandremaros,fmuniz,rmsilva,sergiodaniel,jussara,mgoncalv}@dcc.ufmg.br

ABSTRACT

Product search engines are a key factor for online business. Retrieving relevant products in an E-Commerce (EC) website is of the utmost importance, as a single EC website can have millions of products with very similar features. One aspect that is not widely studied in this scenario is the effect that the image of the product (notably aesthetic properties) shown to the customer has on the customer's interest. Previous studies have been able to link certain characteristics of images to our innate interest. For instance, it is known that bright images with several colors are more likely to attract one's attention than dark ones. However, these issues have been understudied in the EC context. In this context, we conduct experiments on real-world EC to analyze the effects that the product's image aesthetic has on the user interest (expressed in clicks) in the product. Experimental results show that this relationship exists and that it is more visible in some categories of products.

KEYWORDS

E-Commerce, product images, search, aesthetic properties

ACM Reference Format:
Alexandre Maros, Fabiano Belém, Rodrigo Silva, Sérgio Canuto, Jussara M. Almeida, Marcos A. Gonçalves. 2019. Image Aesthetics and its Effects on Product Clicks in E-Commerce Search. In *Proceedings of the SIGIR 2019 Workshop on eCommerce (SIGIR 2019 eCom),* 4 pages.

1 INTRODUCTION

E-Commerce (EC) platforms have become a popular means to bring greater shopping convenience to costumers around the globe. These platforms bring a series of social and technical challenges, which have not been extensively studied. Examples of these challenges include problems related to the trust and familiarity that these websites convey to the customers [7] as well as new computational challenges, specially in the area of information retrieval, such as new ways to increase revenue given specific searches and user profiles [19].

After stumbling upon a product in an EC website, as a result of an explicit search or associated with an ad, the user's decision to click on such product (which, in turn, may lead to a purchase) may be influenced by the image of the product shown to the user. Thus, even though this is commonly overlooked or taken for granted, the product's image, alongside with its title, description and tags, may be deciding factors between buying the product or not. The quality of the product must be clearly conveyed through these features in order to convince the customer to buy the product [2, 10]. In such context, it is known in other domains that certain image characteristics (e.g., brightness, colorfulness) are related to one's innate interest [5]. However, this issue has not been fully investigated in the context of EC search. Thus, we here investigate the hypothesis that *aesthetics properties of the product's image shown to the customers of EC websites does influence the product clicks in searches, and possibly the amount of purchases*, since clicking and visualizing a product is an important step to guide the user decision for purchasing the product or not.

The challenge of evaluating the aesthetics, beauty, or quality of an image, music or any artistic work has been studied for quite some time [1, 6]. The main problem involving this field of study is that beauty is often considered as personal; what may be pretty for one person may not be liked by another. With product image classification these factors can vary even more, e.g., the image does not necessarily needs to be pretty, but it must be clearly visible. However, our driving hypothesis is that products that have more attractive images (have better brightness, contrast, saturation) may have a higher probability of being clicked when they appear.

Even though this is a complex and noisy problem to model, there are few general characteristics of artistic works that determine their visual quality. For instance, in photography, exposure, rule of thirds, contrast, and other characteristics are often carefully planned and chosen by the photographer. A few authors, such as [1], define beauty as a ratio between harmony and complexity of a work.

The present study tackles the aforementioned hypothesis by analyzing the relationship between the *quality* of a product's image and the probability of it being clicked when presented as a result for a search query in a large EC website focused on crafts and personalized products. Our goal is to verify whether the image features, including features related to aesthetics, can explain, at least partially, those clicks. Our experimental results show that this relationship does exist, even though it seems to be noisy. The applied methodology could be used to improve the search results of EC, presenting more attractive products to the customers or to guide sellers towards improving the attractiveness of their products.

2 RELATED WORK

There are several works focused on the definition and evaluation of the aesthetics of an image. In one of the earliest studies [1], the aesthetic measure is formalized as the ratio between order and complexity. In [18], the artistic process is modeled from an

In: J. Degenhardt, S. Kallumadi, U. Porwal, A. Trotman (eds.):
Proceedings of the SIGIR 2019 eCom workshop, July 2019, Paris, France, published at http://ceur-ws.org

Information Theory perspective to allow the quantification of these properties. In particular, the Shannon Entropy and Kolmogorov Complexity are used to estimate values for the order and complexity of some famous paintings.

A more recent work [14] proposes a method to extract features from images in the Photo.net dataset in order to use machine learning techniques such as Support Vector Machines (SVMs) to classify images into aesthetically *good* or *bad* categories. Some of these features, originally proposed by [5, 11], include characteristics such as *brightness*, *contrast*, *saturation*, *central saturation* and *image ratio*. Other characteristics such as Bags of Visual Words (BOVW) [4], Fisher Vectors (FV) [17] and GIST descriptors [16] have also been used to improve learning accuracy. The BOVW and FV algorithms work by clustering Scale-Invariant Feature Transform (SIFT) vectors [12] that capture local properties of an image (e.g., “Does this patch contain sharp edges?”, “Is the color of this patch saturated?”). GIST in turn uses histograms to capture information from images.

The use of Convolutional Neural Networks (CNN) is explored in [13]. The objective is to build a robust way to automatically extract features and predict whether an image is considered *good* or *bad*. The main advantage is that there is no need to propose nor calculate predefined features, which can be expensive. However, there are two main disadvantages: (a) the models must have a fixed-size input, implying the need to scale the images, causing information loss (this can be minimized by using random crops of the same image) and; (b) the ability to interpret and understand the produced model is affected. The CNNs are trained and tested using the AVA dataset[1], containing over 1.5 million images.

There are very few works that try to correlate the quality of the image of a product with the product popularity (e.g., click rate). In [8], a study of the impact of the image on user clicks was conducted. The authors used a limited number of image features and a stochastic gradient boosting model to predict CTR of the randomly selected products. The results indicated significant correlation between the images features and CTR. In [2], the authors find that the “Perceived Product Quality”, whose definition relies directly on information available to the consumer, such as the product images, description, title and reviews directly affects the interest in a product. In here, we take a different perspective based on the aesthetics of the image associated with the product and on features that can be automatically extracted from them based on this perspective.

3 IMAGE FEATURES

We use use three sets of of automatically extracted features to capture the aesthetic of the product images. The first one corresponds to state-of-the-art aesthetic base features that capture the most fundamental aspects of the image (e.g. exposure and saturation). The second one is the GIST descriptor that captures scene categorization and image layout. The third and last set is the BOVW, a generic content-based set of features which describe the distribution of local patches within the image.

3.1 Base Features

The base features consists of 66 attributes extracted from the product’s image. These attributes are directly related to the main aspects of the image, including features that capture the exposure, contrast, saturation, ratio, depth, sharpness and composition [3, 5, 15]. Most of these features were successfully exploited in aesthetic classification in other domains [14], thus being be a good starting point. Table 1 presents these features with a brief description .

Table 1: Base Features for the images

Feature	Description
$f1 - f4$	Sharpness [3, 15]
$f5$	Exposure
$f6, f7$	Contrast and Interval Contrast
$f8$	Average Intensity (Datta 1)
$f9, f10$	Colorfulness Color/Black & White (Datta 2)
$f11, f12$	Average Saturation and Hue (Datta 3, 4)
$f13 - f15$	Central Hue, Saturation and Intensity (Datta 5–7)
$f16 - f19$	Hue Texture (Datta 10, 11, 12, 19)
$f20 - f23$	Saturation Texture (Datta 13, 14, 15, 20)
$f24 - f27$	Texture Value (Datta 16, 17, 18, 21)
$f28 - f30$	Height, Width and Sum of Height and Width
$f31, f32$	Composition (Datta 24, 25)
$f33 - f47$	HSV Segmentation (Datta 26–40)
$f48 - f52$	Segmentation Sizes (Datta 41–45)
$f53 - f57$	Segmentation Codes (Datta 48–52)
$f58 - f60$	HSV Depth ((Datta 53–55)
$f61$	Convexity (Datta 56)
$f62, f63$	Dominant Hue and Hue Compression
$f64$	Ratio of pixels next to dominant Hue
$f65$	Average distance to dominant Hue
$f66$	Color Dispersion

3.2 GIST

The GIST descriptor is a low-dimensional scene descriptor that captures a set of characteristics, such as naturalness, roughness, expansion and ruggedness. These characteristics are estimated using spectral information and coarse localization. The image is partitioned into a 4×4 regular grid and a histogram of gradients (with 20 bins) is computed for each of the 16 regions and 3 color channels. Finally, all histograms are concatenated to form a 960D vector [16].

3.3 Bag of Visual Words (BOVW)

BOVW represents an image by a histogram of local features [4]. First, an unordered set of local patches are extracted and described by SIFT descriptor. A visual vocabulary is learned by clustering these descriptors with K-Means. The local features are then extracted from an image by counting the number of local descriptors assigned to each visual word in a fixed-length histogram. This algorithm has been very successful in image classification. [14].

4 DATASET

Our evaluation exploits real data from Elo7, the largest Brazilian e-marketplace focused on creative and personalized products[2]. In this EC platform, sellers register their own products, uploading pictures and providing descriptive text data (e.g. title, description, price, tags). Product searches return a grid of products where the photos are prominent and textual data include only the product’s title and price. Since the sellers are responsible for this information, there is a lot of heterogeneity in terms of quality of the products’ text features and, specially, their pictures, making it an ideal scenario for this study. The data contains over four million products in 42

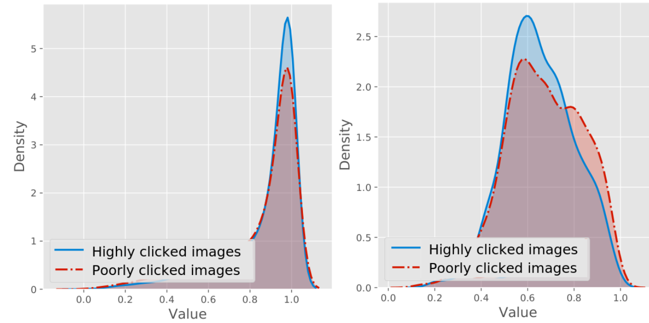
[1]https://research.google.com/ava/

[2]https://www.elo7.com.br

Figure 1: Examples of images considered good and bad from the Elo7 website

(a) Image with a high click score **(b) Image with a low click score**

Figure 2: Distribution of the feature value from the most and least relevant product per query



(a) $f5$ – Exposure **(b) $f8$ – Average Intensity**

different categories and a hundred thousand unique queries across two months (October and September of 2018).

In this work, departing from our main hypothesis, we associate the interest of users in a product with the quality of its image. More formally, given a search query q and a product p, we have the number of times in which p was retrieved after a customer searched for q (impressions) and how many times it was clicked. The interest score of a product with relation to a query is then calculated as the ratio between the number of clicks and impressions times the log of the number of clicks: $clickscore = \frac{clicks}{impressions} \times \log_2 clicks$.

The addition of the logarithmic term is to give an additional priority for products that have a high click count and are more popular. We only considered products that have more than a thousand impressions to try to eliminate eventual noise coming from products that appeared very few times to the customers. For example, a high click rate on a product that appeared very few times over many queries may suggest a customer looking for a very specific product that does not otherwise show up in other searches and therefore is not a clear indication that the product has a good image as perceived by larger set of users.

Based on these scores, we then label images that are above the 80th percentile as "highly clicked" images. Images that are bellow this percentile were considered "poorly clicked". The idea is to check if image features, including features related to aesthetics, influence the amount of product clicks in e-commerce search.

Figure 1 shows two examples of product's images from the website. Figure 2a and 2b are from products that have a high and low score, respectively. The first image has more colors, better lighting and is overall more attractive than the first one.

5 EXPERIMENTS AND ANALYSIS

Two experiments are conducted. First, we verify whether there is a significant difference in the values of image features between highly relevant (clicked) products and the less clicked products. Second, we test the capacity of a machine learning model to predict whether or not a product will be highly clicked based on its image features.

5.1 Feature Distributions

To check whether there is a significant difference between products that are more frequently clicked against their counterparts, we selected two thousand queries that had products with over a thousand impressions. For each query, we compare image features of the highest and lowest ranked products (most and least "relevant" products, respectively). For each feature, we performed the Kolmogorov-Smirnov test to check whether the two distributions (associated with the lowest and highest ranked products according to their click scores) are significantly different. 42 out of the 66 features had *p-values* under the 0.05 threshold, meaning that they are statistically different. Figure 2 shows the density plot of two features with a *p-value* under 0.05.

By looking at the density plot (Figure 2) of the features under the two categories (most and least relevant) we can see that the distributions are slightly different, corroborating the idea that image features do influence users in the click decision. For instance, in Figure 3a we can see that images that are more clicked have a higher exposure, which makes sense, since brighter images tend to attract more attention.

5.2 Quality Prediction

We performed a second experiment to check how accurately we can predict the quality of the product's image based on its click data. Given that the products have a category, and since the images vary vastly from category to category, we built different models for each category. Thus, we divided the prediction into 10 sets of products corresponding to the ten most common categories. A thousand images labeled as "highly clicked" and a thousand image labeled as "poorly clicked" were put into each one of these ten groups. Finally, we use three sets of features (Base, Base + GIST and Base + BOVW) to train an SVM [9] with a 10-Fold cross validation procedure.

Table 2 shows the metrics F1, Precision and Recall for the predictions (in the test sets of the cross-validation procedure) of the ten categories and the three sets of features. Firstly, we can see that, although predictions are hard, as the categories have a mean F1 of 56.7% for the Base features (the worst being 53% and the best 61%), for most categories they are still statistically better than random (F1=0.5). Thus, there is indeed some predictive power in the set of Base features. Secondly, we can see that the GIST features only

Table 2: Classification results for different categories of products using different sets of features.

	Base			Base + GIST			Base + BOVW		
Category	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
Birthdays and Parties	0.56 ± 0.04	0.55	0.58	0.57 ± 0.04	0.55	0.59	0.58 ± 0.03	0.56	0.60
Party favors	0.56 ± 0.05	0.52	0.61	0.50 ± 0.04	0.50	0.49	0.55 ± 0.06	0.52	0.60
Decoration	0.59 ± 0.03	**0.57**	0.62	0.58 ± 0.04	**0.57**	0.59	0.60 ± 0.02	0.57	0.64
Baby	0.53 ± 0.04	0.54	0.52	0.54 ± 0.04	0.53	0.54	0.55 ± 0.03	0.55	0.55
Children's	$\mathbf{0.61 \pm 0.05}$	0.52	**0.72**	0.55 ± 0.03	0.51	0.60	$\mathbf{0.61 \pm 0.04}$	0.53	**0.72**
Invitations	0.54 ± 0.03	0.52	0.56	0.45 ± 0.07	0.52	0.41	0.54 ± 0.03	0.53	0.55
Home	0.55 ± 0.04	0.56	0.54	0.58 ± 0.03	0.55	0.58	0.58 ± 0.04	**0.59**	0.57
Clothes	0.57 ± 0.03	0.56	0.59	0.57 ± 0.04	0.55	0.60	0.58 ± 0.03	0.54	0.63
Paper & Co	0.55 ± 0.02	0.55	0.55	$\mathbf{0.59 \pm 0.03}$	0.53	0.66	0.57 ± 0.03	0.55	0.58
Candies	$\mathbf{0.61 \pm 0.02}$	0.55	0.69	$\mathbf{0.59 \pm 0.03}$	0.54	**0.67**	0.60 ± 0.02	**0.59**	0.64

helped to improve the F1 score in four out of the ten categories, making only the recall higher whilst reducing the precision. The BOVW features improved the results in six of the ten categories, increasing both, recall and precision. This means that local features have an interesting potential in this application.

Finally, we note that clearly some categories are easier to predict than others. The categories "Candies" and "Children's" had the highest F1 while categories such as "Invitations" and "Baby" had the lowest. This could be explained by the types of images that are presented in these categories. In the "Candies" category, we usually have products that have bright colors, with different shapes, sizes and textures that can attract attention and clicks. In this category, images that do not have many colors or that are darker may have a lower click rate. However, in the "Invitations" category, where we have very similar images to one another, mostly white paper invitations with similar texture, the image may not be a deciding factor for a product's click rate.

Although there are some categories that are easier to predict than others, the precision is still low. This indicates that the image alone is not completely responsible for the user's decision to click on a product. Other factors such as the price range and the title of the product may influence customers actions. But, as we can see from the results, a product's image quality does have some influence and could possibly be used to improve search results for an EC platform. The question of how much the quality of a product's image influences its click rate as compared to its other properties (such as title and price) is left as future work.

6 CONCLUSIONS & FUTURE WORK

In this work we investigated how product images influence product clicks in e-commerce platforms. The detection of aesthetically "bad" or "good" images can be used to improve e-commerce search engines, and, consequently, customers' satisfaction and revenue for e-commerce companies. They may also provide feedback to help sellers to make their product more attractive to customers. Our experiments show that image attributes such as brightness, colorfulness and contrast can influence product clicks. First, we analyzed how some of these features vary when comparing more frequently clicked products with less clicked ones. We then used machine learning to try to predict product clicks based on image features. The performed experiments show that there is potential in this technique, specially for some specific categories of products.

As future work, we intend to test other machine learning methods and to add other features, such as Fisher Vectors. We can also propose specialized models for different categories, since the influence of the images in product clicks seem to vary according to the product category. We plan to study in more detail why and how product categories differ in order to produce more accurate models to better predict product clicks from image quality. Finally, we intend to run additional comparative experiments (e.g., comparing images with product's titles and prices) to deepen our understanding of a customer's motivation to click on a product.

ACKNOWLEDGEMENTS

This work was partially supported by Elo7, the FAPEMIG-PRONEX-MASWeb project – Models, Algorithms and Systems for the Web, process APQ-01400-14, as well as by the National Institute of Science and Technology for the Web (INWEB), CNPq and FAPEMIG.

REFERENCES

[1] G. D. Birkhoff. *Aesthetic measure*, volume 38. Harvard University Press Cambridge, 79 Garden St, Cambridge, MA 02138, USA, 1933.
[2] Z. Chen and A. J. Dubinsky. A conceptual model of perceived customer value in e-commerce: A preliminary investigation. *Psychology & Marketing*, 20(4), 2003.
[3] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In *Human vision and electronic imaging XII*, volume 6492. International Society for Optics and Photonics, 2007.
[4] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision*, 2004.
[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV*, 2006.
[6] L. Frost. *What makes a painting good?* PhD thesis, Rhodes University Grahamstown, 1987.
[7] D. Gefen. E-commerce: the role of familiarity and trust. *Omega*, 28(6), 2000.
[8] A. Goswami, N. Chittar, and C. H. Sung. A study on the impact of product images on user clicks for online shopping. In *WWW*, 2011.
[9] M. A. Hearst. Support vector machines. *IEEE Intelligent Systems*, 13(4), July 1998.
[10] S. E. Kaplan and R. J. Nieschwietz. A web assurance services model of trust for b2c e-commerce. *Int. J. of Accounting Information Systems*, 4(2):95 – 114, 2003.
[11] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *CVPR '06*, pages 419–426, 2006.
[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, 60(2):91–110, 2004.
[13] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang. Rating image aesthetics using deep learning. *IEEE Transactions on Multimedia*, 17(11):2021–2034, 2015.
[14] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *ICCV*, 2011.
[15] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. A no-reference perceptual blur metric. In *Image processing'02*, volume 3, pages III–III. IEEE, 2002.
[16] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. of Computer Vision*, 42(3):145–175, 2001.
[17] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *CVPR*, pages 1–8, June 2007.
[18] J. Rigau, M. Feixas, and M. Sbert. Informational aesthetics measures. *IEEE Computer Graphics and Applications*, 28(2), 2008.
[19] L. Wu, D. Hu, L. Hong, and H. Liu. Turning clicks into purchases: Revenue optimization for product search in e-commerce. In *SIGIR*, pages 365–374, 2018.