



Image Analytics in Marketing

Daria Dzyabura, Siham El Kihal, and Renana Peres

Contents

Introduction	2
Top Research Questions by Area Using Image Analytics	5
Product Design	5
Advertising	7
Branding	8
Online Shopping Experience	10
Consumer Perspective	11
Data: Consumer Vs. Firm Images	12
Consumer-Generated Images	12
Firm-Generated Images	15
Methods	17
Feature Extraction	18
Model Training	20
Model Evaluation and Validation	21
Model Application	21
Integrating It All Together	22
Conclusion	24
Cross-References	25
References	25

D. Dzyabura

New Economic School and Moscow School of Management SKOLKOVO, Moscow, Russia

e-mail: ddzyabura@nes.ru

S. El Kihal (✉)

Frankfurt School of Finance & Management, Frankfurt, Germany

e-mail: s.elkihal@fs.de

R. Peres

School of Business Administration, Hebrew University of Jerusalem, Jerusalem, Israel

e-mail: renana.peres@mail.huji.ac.il

Abstract

Recent technical advances and the rise of digital platforms enhanced consumers' abilities to take and share images and led to a tremendous increase in the importance of visual communication. The abundance of visual data, together with the development of image processing tools and advanced modeling techniques, provides unique opportunities for marketing researchers, in both academia and practice, to study the relationship between consumers and firms in depth and to generate insights which can be generalized across a variety of people and contexts.

However, with the opportunity come challenges. Specifically, researchers interested in using image analytics for marketing are faced with a triple challenge: (1) To which type of research questions can image analytics add insights that cannot be obtained otherwise? (2) Which visual data should be used to answer the research questions, and (3) which method is the right one?

In this chapter, the authors provide a guidance on how to formulate a worthy research question, select the appropriate data source, and apply the right method of analysis. They first identify five relevant areas in marketing that would benefit greatly from image analytics. They then discuss different types of visual data and explain their merits and drawbacks. Finally, they describe methodological approaches to analyzing visual data and discuss issues such as feature extraction, model training, evaluation, and validation as well as application to a marketing problem.

Keywords

Image analytics · Visual information · Image processing · Image tagging · Firm images · Consumer images · Feature extraction · Deep neural networks · High-level features · Low-level features · Human-coded features · Color histograms · Gabor filters

Introduction

“The drawing shows me at one glance what might be spread over ten pages in a book.”

Ivan S. Turgenev, Fathers and Sons, 1862.

In the past two decades, images have been playing an increasing role in the marketing arena. Social media outlets have become more image rich, new versions of mobile phones have enhanced ability to take, store, and share photos, and storage and communication infrastructures have become more accessible. These processes have immensely increased the significance of images in consumer life in general, and in marketing in particular.

Images have always been an important part of firms' marketing efforts. Visuals convey a sense of proximity and closeness, and thus, are able to represent objects

better than words (Amit et al. 2009). Relative to text, visual information was found to be better processed and better remembered by humans (MacInnis and Price 1987). Therefore, many of the components of product design, packaging, brand elements, advertising, and design of shopping outlets, use visuals.

These visuals impact consumer response and purchase. For example, Raghubir and Greenleaf (2006) found that certain geometrical ratio of rectangular packaging and print ads influence consumers' relative preferences and purchase intentions. Meyers-Levy and Zhu (2008) showed that various visual elements of store design, such as architecture, freestanding in-store structures, display surfaces, type and arrangement of display cases, mirror orientation, and artwork, relate to consumers' choice and shopping behaviors.

In marketing communications, images are dominant in brands' paid media, such as print, outdoor, and television advertising. Over time, the proportion of pictorial content in a typical magazine ad has grown, while the proportion of text has been gradually shrinking (McQuarrie 2008). A higher proportion of pictorial content in an ad is more efficient in attracting attention (Pieters and Wedel 2004), and is associated with more positive attitude toward the ad (Radach et al. 2003) and with a better memory of the advertised brand (Wedel and Pieters 2000). The layout of the ad, and the size and color combination of its elements have a strong impact on consumers' perceptions and attitudes toward the ad as well as the brand (Janiszewski 1998; Wedel and Pieters 2014; Cho et al. 2008). For example, studies have found that ads containing colors with lighter shade (high "value" in color theory terms) lead to greater liking of the ad. This effect is mediated by stronger feelings of relaxation induced by higher value colors. Higher levels of chroma (a color dimension that relates to the intensity of the color) induce feelings of excitement, which in turn also increase the likeability of the ad (Gorn et al. 1997). Finally, upward looking angles are aligned with perception of potency while photos taken from a downward looking angle were found to lead to a more detailed recall of the brand (Peracchio and Meyers-Levy 2005).

In addition to product design and marketing communications, images are prominently used in marketing research. Consumers express their thoughts, perceptions, and emotions through images. Images have been demonstrated to successfully disrupt people's well-rehearsed narratives and reflect authentic thoughts and deep metaphors. For this reason, visual research methods often better reflect emotions, cultural practices, and attitudes compared to verbal methods (Reavey 2012). Qualitative visual methods are used to arrange brand associations on a map (John et al. 2006), to create brand collages (Zaltman and Coulter 1995; Zaltman and Zaltman 2008), and to elicit brand associations. Other studies used lab experiments (Peracchio and Meyers-Levy 2005) or user-generated digital content (Liu et al. 2017; Klostermann et al. 2018; Pavlov and Mizik 2019; Dzyabura and Peres 2021) to create visual representation of brand associations and connect them with brand characteristics.

The development of digital platforms has further increased the role of images in consumers' lives. Pictures became an important part of brands' owned media – websites, apps, and social media outlets, as well as of brands' earned media – that is,

brand content posted by users on social media. Industry reports indicate that 74% of the content generated by firms contains some form of visual elements, including photos, illustrations, videos, and data visualization (Venngage 2020). Much of the visual activity happens through social media outlets: every minute 136,000 photos are uploaded to Facebook. Every day five Million photos are uploaded to Instagram, added to its corpus of 50 billion photos, which are viewed and receive 3.5 billion likes per day by Instagram's one billion users (ibid.). Consumers use images to communicate with each other and share their experiences, feelings, and impressions. Brands use visual data to learn about consumers' needs and perceptions, to create and communicate value, to shape consumers' attitudes and drive them into action. This ongoing activity has created a rich, dynamic, and vibrant visual ecosystem, which provides a fertile ground for marketing research and marketing activity.

The abundance of visual data, together with the development of image processing tools and advanced modeling techniques, provides unique opportunities for marketing researchers, in both academia and practice, to study the relationship between consumers and firms in depth and to generate insights which can be generalized across a variety of people and contexts.

However, with the opportunity come challenges. Specifically, researchers interested in using image analytics for marketing are faced with a triple challenge. First is the formulation of the research question. Since working with visuals requires elaborate data collection and elaborate analysis, one should identify a research question to which image analytics can add insights that are difficult to obtain in other, more conventional ways.

The second challenge is the choice of data. Visual data sources include user-generated content on brands' web pages (e.g., comments on the brand's Facebook page), data from consumer interactions with other consumers (e.g., one's own Instagram), firm-generated content, general photo repositories (e.g., Flickr), visual product presentation in shopping outlets (eBay, Airbnb), or directly elicited visuals (e.g., online collages). Each of these data sources has its own merits as well as limitations and sometimes, once the research question has been identified, the right data source needs to be carefully chosen. Sometimes, none of the existing data sources contains all the information of interest and the researcher must find ways to combine several sources or supplement the dataset with additional data collection.

The third challenge is the choice of method – most methods used in image analytics were developed in engineering and computer science and were not necessarily optimized for marketing questions. Therefore, using image analytics in marketing requires tailoring existing methods to better fit the data and the research question, or developing new methods altogether.

These three challenges are not independent of each other – the data and method need to be congruent with the research question. For example, a research which seeks to elicit brand associations (Dzyabura and Peres 2021) will need to use interpretable features (such as objects in the pictures), rather than low-level image patterns, and consequently can use image tagging methods and tag-based classifiers to extract high-level features. On the other hand, forecasting the success of a brand based on consumer reviews (Zhang and Luo 2019) has more freedom in choosing

the features, but requires showing that pictorial content contains information which cannot be retrieved by a straightforward sentiment or content analysis of the review text. Finding the right combination of research question, data source, and method is the key to producing meaningful image analytics research in marketing.

Our goal in this chapter is to provide guidance as to how to approach this triple challenge: formulate a worthy research question, select the appropriate data source, and apply the right method of analysis. We start by identifying research questions in five areas that would greatly benefit from using image analytics. We then discuss the different types of visual data including firm-generated and user-generated data and explain their merits and limitations. While data sources constantly change, we suggest guidelines for their characterization and evaluation. We further describe the methodological approaches to analyzing visual data, discussing issues such as feature extraction, model training, classification, and deep learning. We conclude with a decision matrix which can be used as a tool to assist in matching the data and method to the problem at hand. To provide the novice researcher with a gateway to start implementing the ideas, we provide a hands-on tutorial (available on https://github.com/dariasil/image_tutorial), which contains code implementation of several fundamental image analytics tasks and explanations on the required software tools and libraries. We hope that the set of research questions, data sources, richness of methods, code and examples, and guidelines as to how to bring them all together, will help marketing researchers to maximize the tremendous potential of image analytics methods in order to expand the understanding of important research problems and gain meaningful, valuable insights for the benefit of the field.

Top Research Questions by Area Using Image Analytics

The rapid evolution of the visual ecosystem has created unprecedented opportunities to obtain new perspectives on enduring marketing questions. At the same time, it also evoked a large number of new managerial decisions and consumer behaviors which need to be studied. We are just beginning to scratch the surface of this fascinating realm. We outline below the five major areas in marketing that have been most affected by this ecosystem and offer, within each of them, a set of research questions that could lead the further research using image analytics. These questions are summarized in Fig. 1.

Product Design

In many product categories, design is a dominant factor in consumer choices (Bloch 1995; Rubera 2015). Firms use product design and aesthetics to differentiate themselves (Crilly et al. 2004) and to strategically position their brand among competitors (Keller 2003).

Research has demonstrated how specific elements of product and package design impact consumer perception (Greenleaf and Raghubir 2008). Studies of product



Fig. 1 Summary of future research questions for image analytics in marketing

aesthetics are mostly focused on one or several specific visual aspects of a design, such as characteristic lines, silhouettes, ornamentation, color, or texture (Orsborn et al. 2009; Eisenman et al. 2016; Chan et al. 2018). Image analytics, on the other hand, allows taking a more holistic view and study the joint, synergetic effect of the overall product design to customer decision or product performance. It also makes it possible to automatically compare a large number of designs and derive quantitative insights and predictions.

Characterizing Designs: How Can Designs Be Characterized Above and Beyond Their Specific Visual Elements?

Images can be used to classify and characterize product designs without the need to break them down into specific predefined visual elements. Such classification can help to:

1. Measure similarity and differences between designs. Specifically, quantify the distance of a focal design from the “average” design, to evaluate how unique the focal design is. This distance could be used to construct a metric measuring design differentiation and design innovativeness.
2. Map designs to brand perceptual dimensions. For example, whether a car design looks family friendly, or a shoe looks rugged, or a sofa looks modern.
3. Match the product design to the customer’s personal style. Such matching can be used to identify and assemble the products to recommend to customers (see stitchfix.com).
4. Creating new designs. Models of image analytics can be used to augment the creative process of product design by suggesting novel and unexpected combinations of existing design elements. Algorithms of generative adversarial networks (GAN) that use computer vision to assist in the process of product design. Burnap et al. (2019) demonstrate how image analytic algorithms of GAN to generate models of cars for the design team to consider.

Quantifying the Value of Designs: How Can We Assess and Predict Consumer Attitudes Toward Various Product Designs?

Traditionally, demand models are based on quantifiable product attributes (e.g., miles per gallon, battery life, screen size, brand name, safety rating, price). The design of a product, that is, its overall appearance, vibes, emotion, and symbolism, is hard to be decomposed into quantifiable product attributes, yet they are critical factors in consumer choice. When a model is estimated using only these traditional functional attributes, these design characteristics end up in the error term.

Image analytics can improve the accuracy of such models by incorporating product images alongside the traditional characteristics in the demand model. For example, they can incorporate information about the success of previous designs in order to forecast various aspects of demand such as product liking, purchase in different channels, product returns or word-of-mouth. Combining image analytics with traditional models requires the development of new models and estimation methods. Specifically, the challenge is to retain the interpretability and un-biasness of some of the traditional coefficients such as price.

Advertising

Image analytics opens new possibilities for taking a systematic, quantitative approach to selecting, adjusting, and optimizing the visual composition of print and video advertisements.

Assessing Ad Creativity: How Can Print and Video Advertisements Be Rated According to Their Level of Creativity? What Are the Combinations of Visual Elements That Make an Ad Perceived as Creative?

Creativity is an important property of advertising messages that is associated with ad recall and effectiveness (Ang et al. 2007). It is sometimes defined as being composed of two factors: divergence and relevance. Divergence is the originality of the ad, and relevance is the extent to which at least some ad or brand elements are meaningful or valuable to the consumer (Smith et al. 2007).

Identifying the visual qualities which construct creativity is challenging. Therefore, creativity of visual ads is typically evaluated by human judges, using research tools such as surveys (Yang and Smith 2009; Sheinin et al. 2011), creativity awards (Lehnert et al. 2014), or crowdsourcing platforms (Kireyev et al. 2020).

Image analytics can allow for automated and scalable assessment of ad creativity. This can be done, for example, by comparing the focal ads to award winning ads, or to a corpus of candidate ads. Toubia and Netzer (2017) developed a prototypicality-based measure of text creativity. Based on their approach, a similar measure can also be constructed for images, either over predefined attributes or as a self-emergent arrangement of the visual space.

Linking Visuals to Emotional and Cognitive Effects: What Visuals Should Be Included in an Ad in Order to Achieve a Desired Outcome? What Objects, Colors, Shades, or Visual Structures Can Be Used to Spark Laughter, Fear, Urgency, Attention, Long-Term Recall, or Other Effects?

Image analytics could address these questions by taking large repositories of photos and their corresponding consumer reactions, and identifying images with certain emotional and cognitive effects.

Initial steps in this direction were taken by Rietveld et al. (2020), who extracted emotional information (i.e., arousal and valence levels) from Instagram photos of different brands and combined them with text analysis to predict customer engagement with the brand. However, there is need for further research in order to achieve a more complete visual-emotional-cognitive mapping.

Monetizing the Value of Images: What Is the Value of Images in Various Stages of the Customer Journey?

For the first time, differential effectiveness of visuals throughout the purchase funnel can be quantified by image analytics. Specifically:

1. Which visual features are most appropriate for various stages of the purchase funnel – what visuals get consumers' attention? Enhance awareness? Increase consideration, liking, and purchase intentions?
2. Which visuals should a firm present to customers at various stages in the customer life cycle? For example, are different images effective for customer acquisition vs. repeat purchase, upgrade, development, and retention?
3. How should visual features in ads be priced? Image analytics could revolutionize the way ads are priced. While media outlets are priced according to their reach, creative advertising is priced based on the effort invested and the reputation of the creative team. Quantifying the value of various visual components of an ad can lead to a differential pricing scheme. For example, measuring the relative value of a face in an ad versus a white space, or scenery, enables value-based pricing of ad creatives.

Branding

Images play a key role in consumer brand perception, recall, and associations (Peracchio and Meyers-Levy 1994, 2005). Image analytics opens new opportunities for brands to execute their desired positioning through visuals, manage their brand portfolio, and foster brand collaborations.

Visual Brand Representation: What Is the Visual Representation of Brand Associations? How Does It Align with Brand Characteristics? What Is the Role of the Visual Brand Elements and Brand Communications in Shaping Brand Perception and Associations?

A recently proposed tool to explore the visual representation of brand perception and elicit brand associations was described in the work of Dzyabura and Peres (2021). They developed a platform for eliciting brand associations through creating and analyzing online collages of images and showed how these collages can be used to retrieve a visual representation of brand associations and to connect it to brand personality and brand equity metrics. Such approaches have the potential to address many additional questions relating to the nature of these associations, their dynamics over time, their representation in brand communications, and their connection to various brand metrics.

Every brand has a unique set of visual brand elements (logo, colors, fonts, etc.) created by designers in collaboration with brand managers. These brand elements reflect the brand positioning, foster the desired associations and differentiate the brand from its competitors. Through image analytics, marketing scholars and brand managers can evaluate to what extent a proposed design achieves these goals (Dew et al. 2019).

Brand Hierarchy: What Are the Optimal Relationships Between the Visual Elements of Brands in a Brand Portfolio?

Sub-brands within a brand hierarchy require identities which are distinct from one another and yet convey the identity of the master brand. Brands vary in the extent the master brand dominates these sub-brand identities. For example, Fig. 2 shows the brand hierarchy of FedEx and Gillette. For FedEx, the master brand visual elements are clearly dominant, while for Gillette, the sub-brands have distinct visual elements of their own with the master brand being represented to a much lesser degree.



Source: FedEx (2020)

Fig. 2 Examples of the brand hierarchies of FedEx and Gillette

Image analytics can assist in achieving the desired balance between these two extremes. First, image analytics methods can be used to measure the level of visual coherence within the brand hierarchy. Second, it identifies the visual elements that create the perception of similarity. Third, it can connect the overall visuals of the hierarchy to brand performance metrics.

Brand Strategic Collaborations: When Brands Collaborate with Each Other, What Is the Right Mix of Their Visual Elements Which Will Ensure that Both Brands Are Fairly Represented?

Creating a visual identity for a collaboration of brands is often challenging and complicated for the collaborating parties to agree which visual elements should be taken from each brand and how to combine them together. Consider, for example, the two designs of the joint Philadelphia-Milka brand illustrated in Fig. 3. Design A contains more Milka colors, but a larger Philadelphia logo than Design B. Do they manage to achieve parity? Image analytics can help address such dilemma by evaluation to what extent a proposed design represents the desired collaborative identity.

Online Shopping Experience

Image analytics can help firms make better decisions with respect to physical store design and the visual elements of online shopping outlets.



Fig. 3 Comparison of collaborative design packages with different mix of visual elements

The Role of Visuals in Online Product Display: How Does the Composition of Visual Elements, Objects, Size, Background, and Relative Location Impact the Search, Click, and Purchase Propensity?

When photographing a shirt for the online shop, the retailer has numerous options as to how to present the item: folded neatly on a flat surface, laying more carelessly, hanging against the wall, worn by a model, photographed against a solid color background, in an outdoor or indoor location, etc. All these factors influence consumer reactions and expectations from the product.

For example, Zhang et al. (2019) demonstrate that the photographic properties of homes displayed on Airbnb, such as diagonal dominance and rule of thirds, influence demand. Li et al. (2019a, b) show, on the same platform, that the order and layout in which the photos are presented also influence demand. Peng et al. (2020) show that the facial attractiveness of hosts on such rental platforms also influences the occupancy of these homes. More research is needed to explore a larger variety of situations, context, product categories, and consumer behaviors and understand their underlying mechanisms.

The Role of Visuals in Ecommerce Website Design: How Do the Visual Components of an Ecommerce Website Contribute to Profitability?

Designing an ecommerce website is a visual challenge. Designers must make decisions on the sizes and colors of the items that are displayed on the website (e.g., product images or buttons) and create a design that helps users to search for products, explore assortments, get inspired, and discover new products. At the same time, the items should be presented in a way that will match their brand identity. This raises several practical questions which can be answered by employing image analytics methods:

1. How do images change/affect consumers' propensity to keep searching on the website? How does this propensity change at different stages of the search process?
2. How to create more personalized website layouts? For example, Hauser et al. (2009, 2014) use a multi-armed bandit approach that balances exploration and exploitation to automatically match the look-and-feel of the website to customers' cognitive styles.

Consumer Perspective

Studies have shown that consumers use photos to express emotions and attitudes as well as to document their experiences (Van House et al. 2005). This usage has greatly increased with the abundance of mobile phone cameras, storage space, and sharing apps. While traditionally photos were taken on special occasions, people have moved to continuously documenting and sharing their daily routines.

Uncovering Consumer Attitudes: What Are the Hidden Consumer Traits and Attitudes That Can Be Revealed Through Images and Go Beyond the Standard Metrics?

The rich body of consumer-generated photos can be used by researchers to gain a deeper understanding of the consumer experience, to profile and characterize a wide range of experiences, and, in addition, to segment consumers based on dimensions that could not be revealed otherwise.

For example, photos taken by consumers (either posted on social media or collected directly through mobile diaries) can show what is the actual choice set that consumers face when walking around the supermarket; what their environment looks like when sitting in a restaurant; what food brands are served at the same meal; what is their personal style and how it relates to the brands they buy, etc.

Data: Consumer Vs. Firm Images

Once the research question has been formulated, constructing the appropriate data is the next key step. A good dataset for image analysis should satisfy the following criteria: First, it should capture the specific constructs being studied. In many cases this involves the combination of images and additional data. For example, photos that users post on restaurant reviews and the corresponding restaurant financial performance (Zhang and Luo 2019), or photos of Airbnb properties accompanied by property price, location, history, and host characteristics (Zhang et al. 2017, 2019; Li et al. 2019a, b).

Second, the dataset should be large enough to allow drawing insights. The state-of-the-art deep neural network models are trained on the ImageNet dataset, a freely available dataset which contains over 1.2 million images, organized into one thousand categories (<http://image-net.org/>). In marketing such big sizes are rare, but the datasets still have to contain thousands of images for researchers to be able to draw meaningful insights. Third, the dataset should contain minimal biases that could interfere with the main constructs. For example, using user-generated content to understand brand perception should be done carefully, since the sample is not controlled, and since users often post strategically to signal something about themselves (rather than about the brand) to their peers.

Below we describe the main data sources for image analytics in marketing, as summarized in Fig. 4. As illustrated in the figure, the data sources can be classified into consumer-generated images and firm-generated images.

Consumer-Generated Images

Consumer images include all the images created by consumers for different purposes: as a part of their own documentation of experiences and memories, for the purpose of sharing with other consumers, and for sharing with firms. They can be

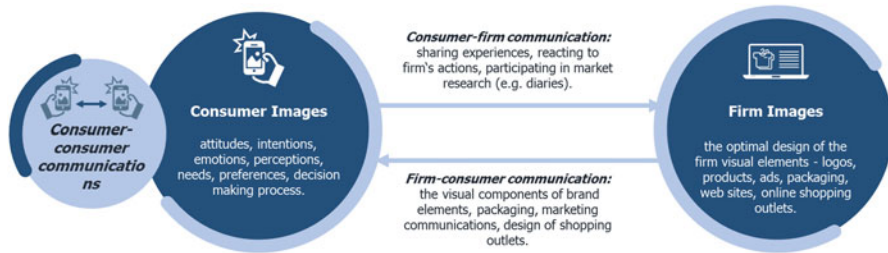


Fig. 4 Sources for image data in Marketing

retrieved by researchers either through mining Internet and social media outlets, or directly elicited through surveys, diaries, panels, and collage making tasks.

Images from Internet and Social Media

Consumers increasingly share images on social media platforms such as Instagram or Facebook, and also on review platforms such as Yelp or [Booking.com](#). For example, both Rietveld et al. (2020) and Liu et al. (2020) use Instagram images to monitor how brands are portrayed by consumers, and compare their perception to firm-generated visuals. Zhang and Luo (2019) use consumer-posted images on Yelp as a leading indicator of restaurant survival. They show that photos are more predictive of restaurant survival than reviews. Jalali and Papatla (2016) use brand images posted by users on Instagram to see how the color composition of the photo influences its click-through rate, when a photo is curated by the website of the brand. They found that click-rates are higher for photos that include higher proportions of green and lower proportions of red and cyan. They also found that photos with higher click-rates are characterized by higher chroma of red and blue. Klostermann et al. (2018) use brand-related Instagram posts to derive insights on how consumers think and feel about a brand (McDonald's) in different brand-related situations.

The appeal of Internet and social media as sources of consumer-generated images is that they are abundant, free, unaided, and cover a broad range of topics. However, for many relevant research questions, these data will not capture the constructs of interest. First, social media data are available for only certain categories and brands: while the brand Nike generates a lot of social media commentary, finding social media posts on other brands such as Colgate, is difficult. Second, it is difficult to control the characteristics of the content contributors. For example, users who have a stronger relationship with the brand (Labrecque 2014), or hold a particularly strong positive or negative opinion, may contribute more than those who have only mild opinions (Lovett et al. 2013). Finally, it is important to carefully interpret the content, since consumers' posts may serve a self-signaling or other purpose (Han et al. 2010).

Social media resources are also valuable in constructing the visual representation of concepts. In many cases the images are tagged and labeled by the users, and these tags can be used as a means of describing the content of the picture. This labeling goes beyond object detection. It can be used to interpret the visual representation of emotions (e.g., happiness), abstract characteristics (e.g., glamorous), and general

concepts (e.g., big-city life). For example, several researchers have used Flickr to gather an annotated dataset of images (e.g., Dhar et al. 2011; McAuley and Leskovec 2012; Zhang et al. 2012; Dzyabura and Peres 2021). Flickr lends itself well to gathering an annotated image dataset, because it provides a search engine that returns the most relevant images for a keyword. The search is based on text labels provided by users, image content, and clickstream data (Stadlen 2015). An image ranked at the top for a particular query has often been validated by tens of thousands of users who clicked on the image, reflecting a large population consensus regarding a strong association between the image and the query term.

Directly Elicited Images

Another approach for retrieving visual data from consumers is direct elicitation, namely, asking respondents to provide, create, rate, or choose images according to certain criteria.

Elicitation can go in one of two directions – one is presenting the respondents with an image and asking them to indicate the properties of interest in this image. For example, does the image look fun (Liu et al. 2020), does a clothing item in an image has asymmetry (Dzyabura et al. 2020), does a logo look modern (Dew et al. 2019), etc. Typically, several human judges are required to rank each image. Management of such tasks can be done using commercially available software tools such as Amazon MTurk or Appen.

The other direction is to provide respondents with a concept (a brand, an emotion, a mood, etc.) and ask them to select or create the images that best represent this concept to them. Such a technique is used by Dzyabura and Peres (2021) who developed a brand visual elicitation platform that allows firms to ask consumers to create collages of images that they associate with a brand. Collage making is a projective technique that has long been used for qualitative research by psychologists (Koll et al. 2010) and brand researchers (Zaltman and Coulter 1995; Zaltman and Zaltman 2008). Typically, participants select images representing the concept in focus, and then explain to the moderator why they chose these images. Dzyabura and Peres (2021) have used image analytics to transform this task into a quantitative market research method. Using online data collection, image processing, and machine learning techniques, collage making now allows researchers to retrieve a large number of images for any concept of interest, over a large number of respondents. Figure 5 describes examples of collages (with some verbal descriptions) created using this method for the brand Starbucks.

Another useful method to directly elicit visuals from consumers is mobile diaries. Mobile diaries are a trending tool to collect repeated self-reports about experiences. They have been used as a research tool in variety of domains including psychology, geography, health, medicine (e.g., Hektner et al. 2007; Heinonen et al. 2012; Hensel et al. 2012; Hofmann and Patel 2015), marketing practice, and recently also by academic researchers (Lovett and Peres 2018). In mobile diary visual studies, respondents are usually asked to take photos of certain experiences – for example, photograph what they see on the shelf, windows of stores they stop by, or the content

by professional teams to meet the firm positioning goals. Thus, they constitute a visual representation of the firm strategy and can be used to study market structure and competitive landscape. The reactions to these images by consumers can, in turn, be used to study consumer response to various marketing actions. We list below several main sources for firm-generated data.

Product Images on Retail Websites

Retail websites are a great source for product images, since they contain many products from various vendors. The images are typically of good quality, focused on the product itself, and often capture the product from various angles. Many retail sites also require standardization of the images. For example, the shoe retail website [Zappos.com](https://www.zappos.com) photographs all shoes in the same way: from seven angles, against a solid white background. The uniformity makes it easier for image processing algorithms to focus on the product image. Retail sites typically provide other relevant product information such as price, materials, size, manufacturer, and brand, which can complement the analysis.

The challenge with using image data from retail websites is that they often lack data on many dependent variables of interest, such as clicks, likes, purchases, profitability, product returns, and repeat purchase rates. Answering research questions regarding these variables requires collaboration with the firm. For example, Dzyabura et al. (2020) used product images from an apparel retailer's online shop and collaborate with the firm to obtain information on the corresponding products' online and offline purchases and product return rates. They found that incorporating the image in the prediction model in addition to the nonvisual attributes (e.g., price, category, season, size) greatly improved its accuracy.

Some multi-vendor retail platforms such as Alibaba, eBay, Airbnb, and Amazon provide consumer reviews and information about the product popularity, which can serve as a proxy for some of these variables. Zhang et al. (2017), for example, use property images posted by Airbnb hosts and combine it with the occupancy rate of the property.

Images on the Firm Social Media Pages

The social media pages of firms are also a rich source of image data. They contain, in addition to the products themselves, other components of the firm's visual representation – endorsers, users and usage scenarios, print and video ads, brand elements, sceneries, locations, activities, and all the visuals curated by the firms to create and enhance its brand associations. These data can be used to study the competitive landscape by identifying differences and similarities between the visual representation of competing brands, as well as between the brand self-representation and consumer perceptions of the firm. Unlike the retail sites, social media outlets also contain more dependent variables, such as consumer likes, shares, reactions, and comments. For example, Li and Xie (2020) used photos of major airlines and sport utility vehicle brands collected from Twitter and Instagram, and measured the engagement they created through retweets and likes. One of their findings is that

the presence of a human face and the fit between the image and the textual content of the post can induce higher user engagement on Twitter but not on Instagram.

The Firm Brand Communications

Researchers can use visual elements of brands of interest in order to explore questions related to brand associations and brand image. This often requires the research team to assemble their own dataset. For example, Dew et al. (2019) assembled a dataset consisting of logos, textual description of firms, industry labels, and brand personality rating of 706 major brands. Then they used image analytics to explore the visual elements of logos they assembled and show how they can be used to create new brand identities and spark ideation.

Advertising Databases

Advertising visuals are continuously being generated by firms, displayed in social media outlets, websites, magazines, TV channels, and billboards. Interestingly, central repositories of advertising images are hard to find. Studies on advertising design effectiveness are mostly behavioral (Wedel and Pieters 2008) and use specific manipulations to test theories. A notable exception is the paper of Pieters et al. (2007), who used data from advertisements by several chains of grocery retailers in the Netherlands to measure the relative importance of the pictorial content in the ad in getting consumer attention.

Large-scale advertising image data across multiple firms, if assembled, could be used to study aspects of parity and differentiation between similar offerings and explore how the competitive landscape is reflected in the visual space. Combined with brand perception, consumer responses, and advertising expenditure, these data can be used to study advertising effectiveness and provide guidance for the optimal design for an ad. Such data repositories are already available in many domains such as fashion (Xiao et al. 2017), autonomous driving (Caesar et al. 2020), and medical imaging. A common general repository is ImageNet, a public dataset of 1.2 million images labeled by humans, which was used to train many state-of-the-art models (e.g., Krizhevsky et al. 2017). A joint data collection effort in advertising could lead to many new and impactful insights on the visual aspects of advertisements.

Methods

Marketing researchers who study visual data have an unprecedented opportunity of access to state-of-the-art methods and analysis techniques. These methods are rapidly improving due to the increased computational power and ongoing efforts of the machine-learning community to broaden the scope of the analysis tools and make them publicly available and user-friendly. An image analytic process is typically composed of the four stages: feature extraction, model training, model evaluation and validation, and model application to the marketing problem.

Feature Extraction

The first step of the image analytics process is determining what feature space to work in. A key challenge of working with images is that the raw input elements – the pixels that make up the images – are not suitable features. A single pixel in isolation does not lend itself to meaningful interpretation. Compared to text, for example, this challenge is particularly pronounced. In text, the basic unit of analysis is words, which carry a meaning, a positive or negative valence, and can be grouped by topic. Pixels, on the other hand, have none of these properties. Therefore, a critical step in any modeling of image data is generating features which will provide a meaningful representation of the images.

There are multiple approaches to feature generation. One is predefined feature extraction. Researchers have developed a variety of predefined features. Perhaps the simplest are **color** histograms, which capture the distribution of the color composition of the image. Such histograms are created by discretizing the colors in the image into bins, based on a color space, and counting the number of image pixels in each bin. The most common color spaces are RGB (red, green, blue) and HSV (hue, saturation, value). Another dimension of interest is **shape**, where the features are line directions, corners, and curves. A third common property is **texture** – defining the repeating patterns in the image, such as line and color intensity. Texture is most commonly measured by a Gabor filter, which detects repeating frequencies of color in certain parts of the image. For videos, Li et al. (2019b) add a dynamic component by defining a measure of visual variation, calculated by decomposing a video into a number of static frames and then computing the visual distance between consecutive frames.

The role of the feature extraction step has been revolutionized with the development of deep neural networks (NN). In an NN, the feature selection and the model training are done simultaneously, so the network automatically extracts the features that are optimal for the specific research problem. For example, it would extract different features for determining whether there is a pedestrian or a traffic light in an image, versus determining whether an image appears fun or serious. A deep NN does such simultaneous training by applying several “layers” of nonlinear transformations on the raw pixel data. The outputs of each transformation serve as the features, or predictor variables, for the next layer. Through multiple layers of such transformations, the network extracts higher- and higher-level representations of the data, allowing the final layer to easily classify the data. For example, lower layers of a deep-learning model may extract edges and textures, whereas higher layers detect motifs, object parts, and complete objects (Goodfellow et al. 2016). The final layer maps the resulting features onto the target variables with a classification function.

There are many neural network architectures that are used in different applications that differ in the types of functions captured by their nodes, their depth, data flow, etc. – together, these make up the network architecture. The networks that work best for image analytics problems are Convolutional Neural Networks, or ConvNets. ConvNets are characterized by the first several layers of the network being *convolutional layers*: each neuron applies a particular transformation to a small

part of the image. Rather than applying a transformation to the entire image, it processes small “batches” of the image separately, to detect shapes and edges in different parts of the image. Two commonly used ConvNet architectures are ResNet and VGG19.

Feature extraction using deep NN almost always results in higher predictive accuracy than human-coded features. However, it has two major caveats. One is that the features are not interpretable – they are complex nonlinear transformations of pixels, which have no meaning on their own. While this is not a disadvantage if the main task is prediction, it is, if interpretable insights are desired. For example, interpretability is important if the goal is to understand what people associate with a brand (Klostermann et al. 2018; Dzyabura and Peres 2021), what kind of image content gets most engagement on social media (Li and Xie 2020), to give recommendations for photographing a home for rent on Airbnb (Zhang et al. 2019), or to create a promotional video for a project in a crowd funding platform (Li et al. 2019b). In such cases, the modeling must be done on interpretable features. One way to obtain them is by using tagging software or a tagged dataset from sources such as Flickr, in order to identify the objects, activities, sceneries, and themes presented in the image. Thus, the image is described by a set of words or tags, which serve as the features, and the image analysis task is transformed into a text analysis task. This opens a wide range of options for the analysis: using word embeddings, various dictionaries such as LIWC, sentiment analysis techniques, and topic modeling (chapter ► “Automated Text Analysis”).

The other caveat is that training a deep neural network from scratch is extremely challenging: it requires a very large annotated dataset, massive memory and computational power, and complex engineering. Additionally, a lot of modeling choices, such as the number, type, and order of layers, how much regularization to use, and learning rate, are made through trial and error. To simplify the training stage, most deep learning applications in marketing use *transfer learning*: rather than designing and training a new neural network for every task, they use a network that has already been trained by someone else for a different purpose. The reasoning behind it is that knowledge gained from performing one task can be used to perform another. For example, knowledge gained from recognizing image aesthetics could be applied to forecast product demand or recognize brand perceptual attributes (Bengio et al. 2011; Bengio 2012).

The transfer can be done either by taking the final layers of the trained network as is, or by fine tuning the trained NN. Dzyabura et al. (2020) took the first approach – they used the second-to-last pre-output of ResNet, which is trained on the ImageNet data, as features in a random forest model to predict the return rates of clothing items.

The fine tuning approach does train the NN, but instead of initializing the model parameters with random numbers, the model is initialized with parameters learned from another NN. The idea is similar to using an informed prior in Bayesian estimation. Relative to training the model from scratch, fine tuning significantly increases model performance and avoids overfitting (Donahue et al. 2014; Girshick et al. 2014; Yosinski et al. 2014). Li et al. (2019a, b) employ fine tuning by using ResNet50 to train their model and learn picture quality and room type (e.g.,

bedroom, bathroom) for images from Airbnb postings. The resulting features are used to predict occupancy rates of the properties. Interestingly, Zhang et al. (2019) also predict image quality on Airbnb, but they use a different pretrained model, VGG16 (Simonyan and Zisserman 2015), also pretrained on ImageNet. The paper uses the results to explain the decision-making process of the hosts who use pictures of lower quality even when a high-quality option is free and available.

Model Training

Regardless of what approach was chosen for feature extraction, the resulting feature space for an image problem will be very large, often larger than the number of observations. Standard statistical methods which assume linear models and estimate their coefficients cannot be applied. The large number of coefficients makes it impossible to identify every single coefficient without bias. Therefore, researchers apply machine learning methods which are tailored for working in very large feature spaces.

Many image analytics problems can be formulated as a *supervised classification task* – determining whether an image belongs to one or multiple predefined categories or classes. For example, in Liu et al. (2020), image classification is used to determine whether an image exhibits a brand perceptual attribute: does the image look fun? rugged healthy? glamorous? A classifier is trained on an annotated dataset of images labeled with the desired classes, that is, images that are known to belong or not to belong to the classes (e.g., images which are glamorous and not glamorous). As explained in Sect. 3, the annotated datasets can be taken from publicly available sources, company data, or collected by the researchers. After training, the classifier can assign a class to a new, unlabeled image.

Some research problems do not involve classifying images into predefined categories. Instead, the researcher is looking to identify patterns in the data – such as recurring objects, colors, shapes, and themes. This is most commonly done by representing the images in a feature space (either using predefined features or with deep neural networks), and then using *unsupervised methods* (e.g., clustering using K-means or nearest neighbors) to group them together (Dew et al. 2019; Peng et al. 2020; chapter ► “Cluster Analysis in Marketing Research”). If the feature extraction is based on tagging, then one could use unsupervised methods for text analysis – such as topic modeling (Dew et al. 2019; Nanne et al. 2020; Peng et al. 2020). In some cases, the image analytic task aims at creating novel combinations of existing patterns, for example – to create new designs of the product, using generative models (Burnap et al. 2019), or predicting “design gaps” in a certain market (Burnap and Hauser 2018).

Model Evaluation and Validation

Once the model has been trained, it is important to evaluate and validate it. Evaluation establishes its performance and validation ensures that the output obtained from the images measures the construct of interest.

A proper evaluation is done by testing the model performance on a different sample than the one it was trained on. In most machine learning algorithms, a portion of the sample is held out and used to test the model. This out-of-sample test is important since the large size of the feature space can easily lead to overfitting. This is only true for supervised learning. For unsupervised learning, model accuracy is hard to evaluate, because there is no specific independent target variable.

Validation depends greatly on the nature of the task. Basically, validation ensures that the model was successful in capturing the construct it intended to measure. For example, if the analysis was done to assess how happy a face is, the results should be validated by testing that faces that were identified as happy are indeed perceived as happy by people. Validation is particularly challenging and particularly important in the unsupervised case. Since predictive accuracy cannot be shown, how can one prove that one clustering is better than another clustering, or that the identified patterns are true? For example, Dzyabura and Peres (2021), used images to extract brand associations from collages of images created by users. They used two layers of validation to demonstrate that the extracted associations are the correct ones: users were first asked to match extracted associations to a collage, and second, to guess the brand based on the associations. Dew et al. (2019) built a model to predict the visual features of a logo based on the verbal descriptions of the brand from the company website, and validated it by taking the brand ShakeShack, using the model to predict the visual elements of its logo and comparing it against the existing logo. Peng et al. (2020) studied face attractiveness and whether it can predict product sales in ecommerce platforms. After extracting facial features, they validated the model using a group of coders that were asked to rank the attractiveness of the faces. In Zhang and Luo (2019), user-posted images were used to predict survival of restaurants. In such tasks, when looking at the examples the algorithm misclassified, one can notice which types of restaurant the classifier fails, and add information accordingly. In addition, by calculating the proportion of mistakes of each type, we can have a better understanding of the precision rate that is possible.

Model Application

The final step of the analysis is applying the model to the research problem, whether it be computing a brand metric (Liu et al. 2020), forecasting demand (Peng et al. 2020), or optimizing the visual communication (Li and Xie 2020; Li et al. 2019b). This stage is important, since in marketing, clustering or classifying images is rarely the end goal. The images are a manifestation of a more fundamental underlying construct, and their analysis is typically an intermediate step in deriving meaningful

insights with respect to this construct and its relationship with perceptual, behavioral, and economic variables.

Integrating It All Together

Image analytics could very easily go wrong. The researcher is faced with numerous data sources, code packages, constantly improving methods, and pre-trained models. All of these open a broad range of research opportunities, yet they often create confusion as to the right choice of the model components. Specifically, the researcher has to carefully match the research problem, data, and method. This is a challenging task: the data, although very rich, might not contain the variables of interest; the model might be good in classifying images but incapable of yielding interpretable insights; the data can suffer from various biases and confounds, such as user strategic posting and self-signaling. Many failures in image analytics tasks are caused by incorrect matching between the various components, leading to none, or even worse – misleading insights.

To ensure an optimal match between the research question, data, and method in order to produce the highest quality analytics with meaningful insights, the researcher should ask herself two questions: first, whether or not there is a single dependent variable that is the crux of the research question. Such a variable could be demand (Zhang et al. 2017, 2019), crowdfunding success (Li et al. 2019b; Peng et al. 2020), business survival (Zhang and Luo 2019), ad recall (Rosbergen et al. 1997), or product return rates (Dzyabura et al. 2020). Second, whether interpretability of the features is important for the task. That is, do the desired insights involve interpretation of specific elements of the image? The answers to these questions determine the appropriate methods and data type. They can be described in the following 2×2 matrix presented in Fig. 6.

Most computer vision tasks fall into the *top right quadrant* of the matrix: there is a specific target variable of interest and interpretation of the features is not necessary. This is the quadrant where most engineering computer vision problems belong. A typical computer vision task is to identify, for instance for a self-driving car, whether an image contains a pedestrian or a traffic light. A good algorithm for such problems is engineered to detect the objects of interest with a low probability of error. It does not need to be able to say what about the picture forms a pedestrian. The set of methods in this quadrant are by definition supervised, and are typically based on deep neural networks.

Thanks to the rapid growth and development of the computer vision field, research questions in marketing that fall into this quadrant have a rich and constantly improving set of methods to choose from. The choice of method depends on the nature of the problem. Nanne et al. (2020) compared different computer vision algorithms to monitor user-generated content, and found that they have different strengths: Google Cloud Vision is more accurate in object detection, whereas Clarifai provides more useful labels to interpret the portrayal of a brand and YOLOV2 did not prove to be useful to analyze visual brand-related UGC. This

		Does the task require interpretability of the features	
		Yes	No
Is there a single target variable?	Yes	Interpretable features – theory based, or tagging. Econometric models.	Dataset that is labelled with the target variable. Supervised – deep learning models.
	No	Interpretable features – theory based, or tagging. Unsupervised classification	Structure the problem – either by finding target variables that capture the construct, or identifying a relevant feature space

Fig. 6 A classification matrix for combining data, features, and image analytic model

should be taken into account when conducting the analysis, and one might need to use several methods and assess their performance for the specific research problem.

In order to apply these methods, the researcher needs to obtain good data from a reliable source. The image data must be annotated with the target variable and it must be very large: contain thousands of annotated images in the training set. In marketing, this is often challenging to accomplish. Most brands, for example, do not produce thousands of images. To get a large enough dataset, one may need to pool data across multiple brands or over time, risking introducing noise to the data.

Unlike engineering, many marketing questions do involve some sort of interpretation of the image characteristics and their relationship with the target variable. For example, in the case of forecasting demand based on product images, one likely wants to know what it is about the product or the way it was photographed that leads to high or low demand. Such cases belong to the *top left quadrant*. In this quadrant, one faces a tradeoff, because the most accurate predictive models are not based on interpretable features. Once one introduces a constraint on the types of features, predictive accuracy will be compromised.

One way in which interpretable features can be generated is using image tagging software which extracts the image content and produces a list of the objects, sceneries, activities, moods, and other themes in the image (Klostermann et al. 2018; Rietveld et al. 2020; Nanne et al. 2020). Another way is to apply domain expertise to identify the relevant features. In photography, these may be diagonal dominance and rule of thirds (Zhang et al. 2019); in apparel, these may be types of prints, graphics, collar type, sleeve length, symmetry or metallic details (Vilnai-Yavetz and Tifferet 2015). These can be extracted using either machine learning classifiers or using human judges. Once the features were extracted, one can use

econometric methods, such as regression, to obtain insights as to how they relate to the target variable.

If both interpretable insights and accurate predictions are important, the analysis should include both: a deep learning model trained to optimize the features and the model for maximum predictive accuracy, and regression analyses over interpretable features. For example, Dzyabura et al. (2020) use deep learned features to predict the return rate of a product based on its image. Then, they have four independent judges manually label the images with respect to industry standard design elements such as symmetry, pattern (solid, floral, striped, geometric/abstract), and additional details (text, metallic/sequin, graphic, lace). The authors then analyze which of these are associated with higher return rates in a regression.

The **bottom left quadrant** comprises situations in which there are no dependent variables of interest. Instead, we ask an open-ended question such as – how is a brand perceived by consumers? How do consumers use the product? What visual features of logos are associated with what brand perceptions?

These questions often require the combined use of interpretable features with unsupervised learning algorithms. The features create a meaningful and managerially relevant space, into which all the relevant observations can be mapped. The unsupervised algorithms, in turn, detect patterns and identify data-driven classifications in this space. For example, Dew et al. (2019) use features taken from theories of logo semantics to form a “visual dictionary” that describes logos in a way that is meaningful to designers (e.g., the amount of white space, corners, and edges). They then use a probabilistic modeling framework to flexibly capture the linkages between the brand descriptions, logo features, industry labels, and brand personality metrics. Dzyabura and Peres (2021) used tagging to identify image content, and then use unsupervised topic modeling to reveal latent topics. Klostermann et al. (2018) tag objects and situations using object detection software, and then employ unsupervised clustering algorithms to form associative networks connecting image content to consumer sentiment. The resulting map of associations is informative for brand management, communications, and monitoring the response of consumers to new products and features.

Finally, questions that belong to the **bottom right quadrant** do not have a specific target variable and do not require interpretability of the features. This quadrant is challenging since it applies structure neither to the features nor to the dependent variable. In order to impose structure on these problems one might consult with domain experts to identify either some relevant dependent variables that are of interest to managers or a set of features that represent the space in a meaningful way.

Conclusion

In the past two decades, major technological advances and the popularity of digital platforms made taking and sharing images a crucial part of consumers’ daily lives. In addition to the abundance of visual data, image processing tools and advances in modeling techniques created unprecedented opportunities to obtain new

perspectives on important marketing questions. We are now able to study new phenomena, investigate the relationship between consumers and firms and obtain insights that would have been difficult or impossible to obtain otherwise.

Using image analytics to generate insights is not trivial though. Researchers are faced with different sources of data, various analysis techniques, and continuously improving methods. In order to benefit from implementing image analytics in solving relevant marketing problems, matching a good research question with the right visual data and appropriate method comes with many challenges. However, once the researcher is able to surmount these challenges, many marketing areas can benefit from image analytics to gain new insights. In the area of **Product Design**, researchers can for example explore how to characterize designs above and beyond their specific elements. Moreover, they can use image analytics to quantify the value of designs by incorporating product images in traditional consumer demand models. In the area of **Advertising**, image analytics can allow for a holistic quantitative approach to selecting, adjusting, and optimizing the visual composition of print and video advertisements. In **Branding**, image analytics opens new perspectives for firms to strategically position their brands, manage their brand portfolio, and identify new collaborations. Image analytics can also help firms make well-grounded decisions to enhance consumers' **Online Shopping Experience** by identifying the role of visuals in ecommerce websites for example. Finally, from a **Consumer Perspective**, image analytics has the potential to reveal through images more about consumers than we knew so far. For example, firms can understand how consumers see brands, how they think about consumption, and how they perceive and evaluate their environments.

Cross-References

- ▶ [Automated Text Analysis](#)
- ▶ [Cluster analysis in Marketing Research](#)

References

- Amit, E., Algom, D., & Trope, Y. (2009). Distance-dependent processing of pictures and words. *Journal of Experimental Psychology: General*, 138(3), 400.
- Ang, S. H., Lee, Y. H., & Leong, S. M. (2007). The ad creativity cube: Conceptualization and initial validation. *Journal of the Academy of Marketing Science*, 35(2), 220–232.
- Bengio, Y. (2012). Deep learning of representations for unsupervised and transfer learning. In I. Guyon, G. Dror, V. Lemaire, G. Taylor, & D. Silver (Eds.), *Proceedings of the ICML workshop unsupervised transfer learn* (pp. 17–36). Bellevue.
- Bengio, Y., Bergeron, A., Boulanger-Lewandowski, N., Breuel, T., Chherawala, Y., Cisse, M., & Erhan, D. (2011). Deep learners benefit more from out-of-distribution examples. In G. Gordon, D. Dunson, & D. Miroslav (Eds.), *Proceedings of the 14th international conference artificial intelligence statist* (pp. 164–172). Fort Lauderdale, FL.
- Bloch, P. H. (1995). Seeking the ideal form: Product design and consumer response. *Journal of Marketing*, 59(3), 16–29.

- Burnap, A., & Hauser, J. (2018). Predicting “design gaps” in the market: Deep consumer choice models under probabilistic design constraints. *arXiv preprint arXiv*, 1812.11067.
- Burnap A., Hauser, J., & Timoshenko A. (2019). *Design and evaluation of product aesthetics: A human-machine hybrid approach*. Available at SSRN 3421771.
- Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., & Beijbom, O. (2020). NuScenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11621–11631).
- Chan, T. H., Mihm, J., & Sosa, M. E. (2018). On styles in product design: An analysis of US design patents. *Management Science*, 64(3), 1230–1249.
- Cho, H., Schwarz, N., & Song, H. (2008). Images and preferences: A feelings-as-information analysis. In M. Wedel & R. Pieters (Eds.), *Visual marketing: From attention to action* (pp. 259–276). New York: Lawrence Erlbaum Associates.
- Crilly, N., Moultrie, J., & Clarkson, P. J. (2004). Seeing things: Consumer response to the visual domain in product design. *Design Studies*, 25(6), 547–577.
- Dew, R., Ansari, A., & Toubia, O. (2019). *Letting logos speak: Leveraging multiview representation learning for data-driven logo design*. Available at SSRN 3406857.
- Dhar, S., Ordonez, V., & Berg, T. L. (2011, June). High level describable attributes for predicting aesthetics and interestingness. In *CVPR 2011* (pp. 1657–1664). IEEE.
- Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014, January). Decaf: A deep convolutional activation feature for generic visual recognition. In *International conference on machine learning* (pp. 647–655).
- Dzyabura, D., & Peres, R. (2021). Visual elicitation of brand perception. *Journal of Marketing*, forthcoming.
- Dzyabura, D., El Kihal, S., Ibragimov, M., & Hauser J. (2020). *Leveraging the power of images in managing product return rates*. Available at SSRN 3209307.
- Eisenman, M., Frenkel, M., & Wasserman, V. (2016). Toward a theory of effective aesthetic communication. In *Academy of management proceedings* (Vol. 2016, p. 12822). Briarcliff Manor: Academy of Management.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580–587).
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT Press.
- Gorn, G. J., Chattopadhyay, A., Yi, T., & Dahl, D. W. (1997). Effects of color as an executional cue in advertising: They’re in the shade. *Management Science*, 43(10), 1387–1400.
- Greenleaf, E., & Raghubir, P. (2008). Geometry in the marketplace. In M. Wedel & R. Pieters (Eds.), *Visual marketing: From attention to action* (pp. 113–143). New York: Lawrence Erlbaum Associates.
- Han, Y. J., Nunes, J. C., & Drèze, X. (2010). Signaling status with luxury goods: The role of brand prominence. *Journal of Marketing*, 74(4), 15–30.
- Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2019). *The power of brand selfies in consumer-generated brand images*. Columbia Business School Research Paper Forthcoming.
- Hauser, J. R., Urban, G. L., Liberali, G., & Braun, M. (2009). Website morphing. *Marketing Science*, 28(2), 202–223.
- Hauser, J. R., Liberali, G., & Urban, G. L. (2014). Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph. *Management Science*, 60(6), 1594–1616.
- Heinonen, R., Luoto, R., Lindfors, P., & Nygård, C. H. (2012). Usability and feasibility of mobile phone diaries in an experimental physical exercise study. *Telemedicine and e-Health*, 18(2), 115–119.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks: Sage Publications, Inc.

- Hensel, D. J., Fortenberry, J. D., Harezlak, J., & Craig, D. (2012). The feasibility of cell phone based electronic diaries for STI/HIV research. *BMC Medical Research Methodology*, 12(75), 1–12.
- Hofmann, W., & Patel, P. V. (2015). Survey signal a convenient solution for experience sampling research using participants' own smartphones. *Social Science Computer Review*, 33(2), 235–253.
- Jalali, N. Y., & Papatla, P. (2016). The palette that stands out: Color compositions of online curated visual UGC that attracts higher consumer interaction. *Quantitative Marketing and Economics*, 14(4), 353–384.
- Janiszewski, C. (1998). The influence of display characteristics on visual exploratory search behavior. *Journal of Consumer Research*, 25(3), 290–301.
- John, D. R., Loken, B., Kim, K., & Monga, A. B. (2006). Brand concept maps: A methodology for identifying brand association networks. *Journal of Marketing Research*, 43(4), 549–563.
- Keller, K. L. (2003). Brand synthesis: The multidimensionality of brand knowledge. *Journal of Consumer Research*, 29(4), 595–600.
- Kireyev, P., Timoshenko, A., & Yang, C. L. (2020). *Scaling human effort in idea screening and content evaluation*. INSEAD Working Paper No. 2020/42/MKT, HEC Paris Research Paper No. MKG-2020-1384, Available at SSRN: <https://ssrn.com/abstract=3685882> or <https://doi.org/10.2139/ssrn.3685882>
- Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2018). Extracting brand information from social networks: Integrating image, text, and social tagging data. *International Journal of Research in Marketing*, 35(4), 538–556.
- Koll, O., Von Wallpach, S., & Kreuzer, M. (2010). Multi-method research on consumer–brand associations: Comparing free associations, storytelling, and collages. *Psychology & Marketing*, 27(6), 584–602.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Labrecque, L. I. (2014). Fostering consumer–brand relationships in social media environments: The role of Parasocial interaction. *Journal of Interactive Marketing*, 28(2), 134–148.
- Lehnert, K., Till, B. D., & Ospina, J. M. (2014). Advertising creativity: The role of divergence versus meaningfulness. *Journal of Advertising*, 43(3), 274–285.
- Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research*, 57(1), 1–19.
- Li, H., Simchi-Levi, D., Wu, M. X., & Zhu, W. (2019a). *Estimating and exploiting the impact of photo layout in sharing economy*. Available at SSRN.
- Li, X., Shi, M., & Wang, X. S. (2019b). Video mining: Measuring visual information using automatic methods. *International Journal of Research in Marketing*, 36(2), 216–231.
- Liu, X., Burns, A. C., & Hou, Y. (2017). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising*, 46(2), 236–247.
- Liu, L., Dzyabura, D., & Mizik, N. (2020). Visual listening in: Extracting brand image portrayed on social media. *Marketing Science*, 39(4), 669–686.
- Lovett, M. J., & Peres, R. (2018). Mobile diaries – Benchmark against metered measurements: An empirical investigation. *International Journal of Research in Marketing*, 35(2), 224–241.
- Lovett, M. J., Peres, R., & Shachar, R. (2013). On brands and word of mouth. *Journal of Marketing Research*, 50(4), 427–444.
- MacInnis, D. J., & Price, L. L. (1987). The role of imagery in information processing: Review and extensions. *Journal of Consumer Research*, 13(4), 473–491.
- McAuley, J., & Leskovec, J. (2012, October). Image labeling on a network: Using social-network metadata for image classification. In *European conference on computer vision* (pp. 828–841). Berlin/Heidelberg: Springer.
- McQuarrie, E. F. (2008). Differentiating the pictorial element in advertising – A rhetorical perspective. In M. Wedel & R. Pieters (Eds.), *Visual marketing: From attention to action* (pp. 91–112). New York: Psychology Press.

- Meyers-Levy, J., & Zhu, R. (2008). Perhaps the store made you purchase it: Toward an understanding of structural aspects of indoor shopping environment. In M. Wedel & R. Pieters (Eds.), *Visual marketing: From attention to action* (pp. 193–224). New York: Psychology Press.
- Nanne, A. J., Anthunis, M. L., van der Lee, C. G., Postma, E. O., Wubben, S., & van Noort, G. (2020). The use of computer vision to analyze brand-related user generated image content. *Journal of Interactive Marketing*, 50, 156–167.
- Orsborn, S., Cagan, J., & Boatwright, P. (2009). Quantifying aesthetic form preference in a utility function. *Journal of Mechanical Design*, 131(6), 061001.
- Pavlov, E., & Mizik, N. (2019). *Increasing consumer engagement with firm-generated social media content: The role of images and words*. Working Paper, University of Washington.
- Peng, L., Cui, G., Chung, Y., & Zheng, W. (2020). The faces of success: Beauty and ugliness premiums in e-commerce platforms. *Journal of Marketing*, 84(4), 67–85.
- Peracchio, L. A., & Meyers-Levy, J. (1994). How ambiguous cropped objects in ad photos can affect product evaluations. *Journal of Consumer Research*, 21(1), 190–204.
- Peracchio, L. A., & Meyers-Levy, J. (2005). Using stylistic properties of ad pictures to communicate with consumers. *Journal of Consumer Research*, 32(1), 29–40.
- Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing*, 68(2), 36–50.
- Pieters, R., Wedel, M., & Zhang, J. (2007). Optimal feature advertising design under competitive clutter. *Management Science*, 53(11), 1815–1828.
- Radach, R., Lemmer, S., Vorstius, C., Heller, D., & Radach, K. (2003). Eye movements in the processing of print advertisements. In R. Radach & H. Deubel (Eds.), *The mind's eye* (pp. 609–632). Amsterdam: Elsevier Science Publishers.
- Raghubir, P., & Greenleaf, E. A. (2006). Ratios in proportion: What should the shape of the package be? *Journal of Marketing*, 70(2), 95–107.
- Reavey, P. (Ed.). (2012). *Visual methods in psychology: Using and interpreting images in qualitative research*. Routledge. London.
- Rietveld, R., van Dolen, W., Mazloom, M., & Worring, M. (2020). What you feel, is what you like influence of message appeals on customer engagement on Instagram. *Journal of Interactive Marketing*, 49, 20–53.
- Rosbergen, E., Pieters, R., & Wedel, M. (1997). Visual attention to advertising: A segment-level analysis. *Journal of Consumer Research*, 24(3), 305–314.
- Rubera, G. (2015). Design innovativeness and product sales' evolution. *Marketing Science*, 34(1), 98–115.
- Sheinin, D. A., Varki, S., & Ashley, C. (2011). The differential effect of ad novelty and message usefulness on brand judgments. *Journal of Advertising*, 40(3), 5–18.
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In Proceedings of *International Conference on Learning Representations (ICLR)*. Available at <https://arxiv.org/abs/1409.1556>
- Smith, R. E., MacKenzie, S. B., Yang, X., Buchholz, L. M., & Darley, W. K. (2007). Modeling the determinants and effects of creativity in advertising. *Marketing Science*, 26(6), 819–833.
- Stadlen, A. (2015). *Find every photo with Flickr's new unified search experience*. Available at <https://blog.flickr.net/en/2015/05/07/flickr-unified-search/>
- Toubia, O., & Netzer, O. (2017). Idea generation, creativity, and prototypicality. *Marketing Science*, 36(1), 1–20.
- Van House, N., Davis, M., Ames, M., Finn, M., & Viswanathan, V. (2005). The uses of personal networked digital imaging: An empirical study of cameraphone photos and sharing. In *CHI'05 extended abstracts on human factors in computing systems* (pp. 1853–1856). ACM.
- Venngage. (2020). *14 Visual content marketing statistics to know for 2020*. Available at <https://venngage.com/blog/visual-content-marketing-statistics/>
- Vilnai-Yavetz, I., & Tifferet, S. (2015). A picture is worth a thousand words: Segmenting consumers by Facebook profile images. *Journal of Interactive Marketing*, 32, 53–69.

- Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science*, 19(4), 297–312.
- Wedel, M., & Pieters, R. (2008). A review of eye-tracking research in marketing. *Review of Marketing Research*, 4(2008), 123–147.
- Wedel, M., & Pieters, R. (2014). *Looking at vision* (p. 2014). Abingdon: Routledge.
- Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv*, 3.
- Yang, X., & Smith, R. E. (2009). Beyond attention effects: Modeling the persuasive and emotional effects of advertising creativity. *Marketing Science*, 28(5), 935–949.
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 3320–3328). Available at <https://arxiv.org/abs/1411.1792>
- Zaltman, G., & Coulter, R. H. (1995). Seeing the voice of the customer: Metaphor-based advertising research. *Journal of Advertising Research*, 35(4), 35–51.
- Zaltman, G., & Zaltman, L. H. (2008). *Marketing metaphoria: What deep metaphors reveal about the minds of consumers*. Boston: Harvard Business Press.
- Zhang, M., & Luo, L. (2019). *Can User-posted photos serve as a leading indicator of restaurant survival? Evidence from Yelp*. Available at SSRN 3108288.
- Zhang, H., Korayem, M., You, E., & Crandall, D. J. (2012). Beyond co-occurrence: Discovering and visualizing tag relationships from geo-spatial and temporal similarities. In *Proceedings of the fifth ACM international conference on web search and data mining* (pp. 33–42). Available at <https://doi.org/10.1145/2124295.2124302>
- Zhang, S., Lee, D. D., Singh, P. V., & Srinivasan, K. (2017, May 25). How much is an image worth? *Airbnb property demand estimation leveraging large scale image analytics*. Available at SSRN 2976021
- Zhang, S., Mehta, N., Singh, P. V., & Srinivasan, K. (2019). *Can lower-quality images lead to greater demand on AirBnB?* Technical report, working paper, Carnegie Mellon University.