Available online at www.sciencedirect.com
ScienceDirect
Journal of Interactive Marketing 50 (2020) 156–167
JOURNAL OF INTERACTIVE MARKETING
www.elsevier.com/locate/intmar

The Use of Computer Vision to Analyze Brand-Related User Generated Image Content

Annemarie J. Nanne [a,*] & Marjolijn L. Antheunis [a] & Chris G. van der Lee [a] & Eric O. Postma [b] & Sander Wubben [a] & Guda van Noort [c]

[a] Tilburg Center for Cognition and Communication, Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, The Netherlands
[b] Department of Cognitive Science and Artificial Intelligence, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, The Netherlands
[c] Amsterdam School of Communication Research, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, The Netherlands

Available online 05 February 2020

Abstract

With the increasing popularity of visual-oriented social media platforms, the prevalence of visual brand-related User Generated Content (UGC) have increased. Monitoring such content is important as this visual brand-related UGC can have a large influence on a brand's image and hence provides useful opportunities to observe brand performance (e.g., monitoring trends and consumer segments). The current research discusses the application of computer vision for marketing practitioners and researchers and examines the usability of three different pre-trained ready-to-use computer vision models (i.e., YOLOV2, Google Cloud Vision, and Clarifai) to analyze visual brand-related UGC automatically. A 3-step approach was adopted in which 1) a database of 21,738 Instagram pictures related to 24 different brands was constructed, 2) the images were processed by the three different computer vision models, and 3) a label evaluation procedure was conducted with a sample of the labels (object names) outputted by the models. The results of the label evaluation procedure are quantitatively assessed and complemented with four concrete examples of how the output of computer vision can be used to analyze visual brand-related UGC. Results show that computer vision can yield various marketing insights. Moreover, we found that the three tested computer vision models differ in applicability. Google Cloud Vision is more accurate in object detection, whereas Clarifai provides more useful labels to interpret the portrayal of a brand. YOLOV2 did not prove to be useful to analyze visual brand-related UGC. Results and implications of the findings for marketers and marketing scholars will be discussed.

Keywords: Visual brand-related UGC; Computer vision; Pre-trained computer vision; Image mining; Automated content analysis

The Use of Computer Vision to Analyze Brand-Related User Generated Image Content

Image-oriented social media platforms such as Instagram or Pinterest have grown quickly in popularity, with 35% of the adults and 72% of the adolescents in the United States using Instagram in 2018 (Pew Research Center, 2018a; Pew Research Center, 2018b). Via these platforms, consumers not only share important moments in life but also share experiences with their favorite products and/or brands (Chari, Christodoulides, Presi, Wenhold, & Casaletto, 2016). This is called visual brand-related User Generated Content (UGC; Muntinga, Moorman, & Smit, 2011). Consumers have always shared experiences with brands either offline or online (Ismagilova, Slade, & Williams, 2016). However, with the rise of social media, the impact of this phenomenon has increased tremendously. As it became easier to create brand-related content and distribute it to a large audience, consumers have become active branding agents (Hennig-Thurau, Hofacker, & Bloching, 2013).

The portrayal of a brand by consumers can seriously affect a brand's performance, both positively and negatively (Erkan & Evans, 2016; Gensler, Völckner, Liu-Thompkins, & Wiertz,

* Corresponding author at: Warandelaan 2, 5037 AB Tilburg, The Netherlands.
E-mail address: a.j.nanne@uvt.nl (A.J. Nanne).

https://doi.org/10.1016/j.intmar.2019.09.003
1094-9968

2013). Consumers can produce brand-related UGC to declare their love for a brand but also to openly complain when a brand does not meet their expectations (Gensler et al., 2013). Moreover, brand-related UGC can provide brands with an opportunity to learn more about their consumers. Analyzing this content can, for example, give insights into who is using their product and when they are using it (Fan & Gordon, 2014). This can be both valuable information to improve customer segmentation, and hence advertisement targeting, as well as the stepping stone towards the creation of new products inspired by the needs of the target group (Fan & Gordon, 2014; Vilnai-Yavetz & Tifferet, 2015).

As the amount and diversity of visual brand-related UGC increases, the interest in automatic coding, as opposed to the now usual manual coding (e.g., De Vries, Gensler, & Leeflang, 2012; Hollenbeck & Kaikati, 2012), grows. Computer vision seems a promising tool to automatically analyze the content of visual brand-related UGC (Cheng, Han, & Lu, 2017). With computer vision, images are automatically analyzed and classified in terms of their content (i.e., object classification or scene classification: LeCun, Bengio, & Hinton, 2015). Because of its automatic character, computer vision allows for the identification of many content characteristics (e.g., objects, color, or brand logos; Bakhshi & Gilbert, 2015; Bianco, Buzzelli, Mazzini, & Schettini, 2017) in a large and diverse set of images, within a short period of time. Computer vision models have developed quickly and nowadays are already able to detect thousands of objects with accuracy rates of up to 97% (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017).

In principle, such powerful computer vision algorithms could support the analysis of visual brand-related UGC for marketers and scholars. Unfortunately, the use of computer vision as a tool in marketing research is not straightforward. Most marketing researchers do not have enough data or computational resources at their disposal *to develop* or train a computer vision model, and are hence dependent on one of the so-called *pre-trained* models made available by scientists (e.g., YOLOV2: Redmon & Farhadi, 2017) or commercial organizations (e.g., Clarifai, 2019a; Google Cloud, 2019). The application of these ready-to-use models to analyze visual brand-related UGC brings several challenges.

These pre-trained computer vision models are typically not trained on brand-related UGC but on images of everyday objects and scenes. As a result, the output of these models consists of labels that describe the content of the picture, unrelated to common marketing outcomes, such as attitude or brand image. Up until this point it is unknown if computer vision models can be used in a marketing context. Additionally, models that are not trained on visual brand-related UGC might not be accurate in analyzing this content as the training images determine what objects can be recognized by a model (Lin et al., 2014). Brand-related UGC differs from any other content as it often contains objects that are only partly visible or very unclear as opposed to staged images that have one clear subject. Moreover, the representation of different brands results in a large diversity of content and hence many different types of objects to be detected (e.g., a coffee brand might be displayed on a sunny terrace and a shoe brand in a wild forest). As a result, in order to start using computer vision in a marketing context, it is crucial to know whether pre-trained models are accurate when applied to visual brand-related UGC and if the output of computer vision models can result in relevant marketing insights.

Therefore, the aim of this paper is to examine the suitability of computer vision in general, as well as the performance of individual pre-trained models to analyze visual brand-related UGC. In order to do so, the accuracy of the output labels is tested via a label evaluation procedure and is complemented with concrete examples of how to use computer vision in a marketing context. The findings 1) provide methodological insights into the use of computer vision models in a marketing context, 2) provide a stepping stone for the use of these models in future research, 3) add to computer vision comparisons in the computer science domain by taking visual brand-related UGC as the subject of analysis and providing insights into how the models behave and perform with visual brand-related UGC instead of the merely staged images that are used in previous studies (e.g., Szegedy et al., 2017), and 4) allow us to examine whether and how marketers can use these pre-trained models to monitor the performances of their brands.

Theoretical Framework

The Growth and Importance of Visual Brand-Related UGC

The rise of smartphones and global mobile data infrastructure enables consumers to create brand-related UGC anywhere, anytime (Serrano & Ramjaun, 2018). Not only is the creation of brand-related UGC now easier than ever before, but the distribution of such content through large mobile networks also has become effortless (Kaplan & Haenlein, 2010).

The increasing amount of brand-related UGC is not without effect (Erkan & Evans, 2016). As opposed to Marketer Generated Content (MGC), UGC is considered more trustworthy (e.g., Chu & Kim, 2011). As a result, exposure to brand-related UGC can change consumers' brand attitudes and, consequently, buying behavior (Erkan & Evans, 2016). These effects might be even larger for *visual* brand-related UGC as visual content draws attention more quickly and is often remembered better than textual content (Chau, Au, & Tam, 2000; Hernández-Méndez & Muñoz-Leiva, 2015).

Therefore, especially the impact of visual brand-related UGC urges insights into its content. By analyzing the large amount of brand-related UGC, marketers have the opportunity to learn about consumers' brand experiences. As a result, visual brand-related UGC can give important insights into how consumers see their brand and identify important moments of use. These insights can in turn help to target advertisements more specifically to people who are interested in a brand, leading to more effective marketing strategies (e.g., McDonald & Dunbar, 2012). Previous research has shown already that it is possible to identify different consumer groups based on Facebook profile pictures (Vilnai-Yavetz & Tifferet, 2015). As a result, it might also be possible to use visual brand-related

UGC to segment consumer groups. However, because there is so much visual brand-related UGC, monitoring can be time consuming. Computer vision might therefore be an interesting option to automatically analyze visual brand-related UGC.

Computer Vision to Analyze Visual Brand-Related UGC

Contrary to what the name suggests, computer vision models cannot actually *see* the content of an image (Marr, 1982). Instead, such models make use of mathematical algorithms to deduce what content is shown (Szeliski, 2011). In order for a model to recognize an object, it has to be trained on an extensive dataset of labeled examples. Training a model is done by feeding the model a bulk of example images (e.g., an image of a dog) and associated labels (i.e., dog). During training, the model adapts millions of parameters that define the multi-layer mapping from image to label (LeCun et al., 2015). Depending on the training procedure and example images, computer vision models can be trained for many different tasks, for example recognizing the style of a specific artist or faces of celebrities in a picture (Guo, Zhang, Hu, He, & Gao, 2016; Van Noord, Hendriks, & Postma, 2015).

In the field of visual brand-related UGC some first steps have been taken with the use of computer vision, resulting in practical implications for marketing. Liu, Dzyabura, and Mizik (2017) trained a computer vision model to analyze brand image portrayal on Instagram. The model analyzes Instagram pictures that are associated with a particular brand (by means of a hashtag) and derives overall impressions that are conveyed about a brand, such as "rugged" or "romantic." In a similar vein, Tous et al. (2018) built a model that can curate the visual identity of a brand. They use a two-factor approach. In the first stage, concrete objects in the images are recognized (e.g., car). In the second stage, the model compares the content of the image with images that are representative for a brand, to identify which images are suitable for marketing outcomes (Tous et al., 2018).

These first promising attempts involve the training of models, which prevents computer vision from becoming widely available for marketing researchers and practitioners. The process of model training is cumbersome and time consuming and marketing researchers are usually not familiar with training such a model and lack the resources for doing so. Therefore this study examines the applicability of pre-trained computer vision models.

Pre-Trained Computer Vision Models to Analyze Visual Brand-Related UGC

Computer vision models that are already trained on a large dataset of labeled examples are called pre-trained models. Such models do not require training and can be directly applied to novel images, provided that these images are representative for the distribution of images the model was originally trained on. Pre-trained models are ready to use for everyone, without requiring detailed knowledge of the underlying mechanisms. The model is not flexible and hence outputs the same labels regardless of who uses the system. Multiple pre-trained models have been released in recent years, both by scientists as well as commercial organizations. Though the exact training procedure differs for every model, the problems with regard to the analysis of visual brand-related UGC outlined above apply to all pre-trained computer vision models. Therefore, the current research focuses on a representative subset of computer vision models.

The current research considers one freely available model called YOLOV2 (Redmon & Farhadi, 2017) and two commercial models: Clarifai (Clarifai, 2019a) and Google Cloud Vision (Google Cloud, 2019). We include both freely available and commercial models because they each have their own benefit. Freely available models are more transparent in the algorithm used and training data. Commercial models, on the other hand, can often distinguish between more labels, which might be beneficial to gain a more comprehensive idea of the content of the image. The YOLOV2 model was chosen because it makes use of a unique data analysis architecture and manages to reach very high accuracy levels (Redmon & Farhadi, 2017). Google Cloud Vision and Clarifai were chosen, because they have a large variety of labels. All three models are widely used in scientific research (e.g., Aker & Kalkan, 2017; Jaakonmäki, Müller, & Brocke, 2017; Mazloom, Rietveld, Rudinac, Worring, & Van Dolen, 2016).

YOLOV2

The You Only Look Once model version 2 (YOLOV2: Redmon & Farhadi, 2017) was first released in 2016 and is characterized by its ability to analyze the whole image by using one single convolutional network (Redmon, Divvala, Girshick, & Farhadi, 2016). The model was trained on the MS COCO dataset (Lin et al., 2014). A dataset of 200,000 labeled images of 80 different objects, photographed in context (e.g., a bike on the road). The YOLOV2-model can recognize these 80 objects with accuracy levels of up to 95%. Previous research applied this model to detect drones in the air (Aker & Kalkan, 2017).

Google Cloud Vision

Google Cloud Vision was released in 2016 as part of the Google Cloud Platform and offers multiple models to analyze images (Google Cloud, 2019). Apart from the general label detection model, they offer, amongst others, models to detect explicit content, logos, and faces. It is also possible to find images similar to the ones entered by using Google Search (Google Cloud, 2019). Google Cloud claims that the platform can differentiate between thousands of different labels (Google Cloud, 2019); however, the exact list of labels is not publicly available. Also, the training algorithm and training images have not been revealed. Nevertheless, various researchers have made use of Google Cloud Vision. For example, Ferwerda and Tkalcic (2018) used Google Cloud Vision to relate content characteristics of Instagram pictures to personality traits of the post-owner and Mazloom et al. (2016) identified content characteristics, such as the presence of a brand or person that could predict the popularity of Instagram pictures.

Clarifai

Clarifai started in 2013. It is characterized by the use of intermediate layers in a deep convolutional neural network to improve visual recognition (Zeiler & Fergus, 2013). Like Google Cloud Vision, Clarifai offers a variety of models to analyze visual content. These include the general model to detect objects and scenereies but also more specialized models focusing on food, apparel, or celebrities (Clarifai, 2019a). The general model alone is claimed to be able to recognize over 11,000 different labels (Clarifai, 2019b). The dataset that was used to train the model is unknown. Clarifai has been used in various scientific research. For example, Chen and Dredze (2018) used Clarifai to determine how vaccinations are portrayed in pictures on social media and Jaakonmäki et al. (2017) used Clarifai to analyze what content characteristics made users more engaged with brand-related visual UGC.

Challenges in the Use of Pre-Trained Computer Vision Models

The fact that the models are pre-trained brings three possible problems for the analysis of visual brand-related UGC. First, these models are not trained on visual brand-related UGC, and therefore it is unknown how well the models will perform on such content. There is reason to doubt the accuracy of computer vision models for visual brand-related UGC because previous research has shown that computer vision models are not robust to noise (Hosseini, Xiao, & Poovendran, 2017; Papernot et al., 2015). Minor imperfections in images can already negatively affect the label that computer vision models assign to them. For example, Hosseini et al. (2017) tested Google Cloud Vision by adding impulse noise to the image (i.e., adding light pixels to dark spaces and the other way around). A noise rate of on average 14% was enough to get the Google Cloud Vision API to return completely different labels. An image previously correctly identified as an airplane was suddenly categorized as bird and as a teapot was labeled as biology.

Second, because the available models are all trained in a different way, different models might therefore give a different analysis of the same image. This applies both to the training images that were used as well as to the labels that the model is trained to output. For example, a database primarily trained on faces, will not be able to correctly classify other objects in images, and therefore may incorrectly classify other objects as faces or miss the presence of objects altogether because the models can only detect labels they were trained for (LeCun et al., 2015).

Third, the output of the pre-trained computer vision models is not marketing-related. The models focus on the recognition of objects in an image, for example a person or a car. It is unclear whether such general labels can be used for marketing purposes.

Because it is unknown if computer vision is useful to analyze visual brand-related UGC and how computer vision models behave with this content, it is hard to choose the correct computer vision model for marketing research. Therefore, the current paper analyzes the suitability of three different computer vision models, YOLOV2, Google Cloud Vision, and Clarifai, to examine visual brand-related UGC.

Method

To examine the suitability of computer vision models for the examination of visual brand-related UGC, the study is split up in three phases. In the first phase of the research, a visual brand-related UGC database was constructed. In the second phase of the research, the images in the database were run through the three computer vision models providing a set of labels for every image in the dataset. In the third phase, a label evaluation procedure was conducted, in which human coders judged the accuracy of the labels that were attached to the images in the second phase.

Phase 1: Database Construction

A database was constructed containing the visual brand-related UGC of 24 different brands. The brands were selected based on the “100 most loved brands on social media” brand list (Netbase, 2018). The brands selected had to sell a physical product, have an unambiguous brand name (e.g., Bodyshop was excluded because it yielded a lot of pictures of garages, while it is also the name of a cosmetics brand), and together represent a variety of branches. The requirement of a physical product was necessary because it excluded platform brands that are hard to represent in a picture (e.g., Instagram itself). The final list of brands can be found in Table 1.

The images in the database were collected from Instagram, because of its visual character and current popularity (Pew Research Center, 2018a; Pew Researcher Center, 2018b). For each brand, the 1,000 most recent posts tagged with the brand name, were collected in December 2018. We chose brand-tagged pictures over non-tagged pictures because they provide many opportunities to analyze a brand's image. Moreover, because these are the pictures that appear when searching for a specific brand, they might be more influential to a brand's image than non-tagged pictures.

The initial database consisted of 24,000 images. There were 91 duplicate images that appeared for multiple brands. These images were excluded because it was considered noise in the data. Another 57 images were excluded because they were posted by official Instagram accounts of brands or official retailers of the brand.[1] Also sponsored content was omitted from the database, by excluding any posts that contain the hashtags “#spon” or “#ad,” hashtags used to indicate that a social media post contains sponsored content (Advertising Standards Authority UK, 2019).[2] This excluded another 133

[1] The list of excluded accounts was created by inspecting all accounts that contributed a picture on their name, profile picture, biography, and the presence of a “verified batch” that Instagram provides to official accounts of famous people and brands.

[2] Further inspection of these account found that pictures with #spon belonged mostly to influencers, pictures hashtagged with #ad are mostly posted by advertising agencies involved in the production of the ad.

Table 1
Overview of brands included in the sample.

Brand	Number of images	Branch
BMW	803	Automotive
Ferrari	855	Automotive
Ford	893	Automotive
Adidas	936	Consumer goods
Chanel	948	Consumer goods
Lacoste	966	Consumer goods
Listerine	781	Consumer goods
L'Oréal Paris	876	Consumer goods
Nike	951	Consumer goods
Tiffany & Co	913	Consumer goods
Tommy Hilfiger	984	Consumer goods
Yves Saint Laurent	962	Consumer goods
Burger King	870	Food & beverage
Coca-Cola	892	Food & beverage
Gordon's	929	Food & beverage
Heineken	917	Food & beverage
McDonalds	853	Food & beverage
Moët	898	Food & beverage
Nestlé	915	Food & beverage
Starbucks	929	Food & beverage
Apple	913	Technology
LG	904	Technology
Nintendo	839	Technology
Samsung	942	Technology

images. Finally, there were 2,050 images that could not be analyzed because they could not be downloaded. Therefore, the final database used in the analyses consisted out of 21,669 images (see Table 1). This research was approved by the Research Ethics and Data Management Committee of the School for Humanities and Digital Sciences at Tilburg University.

Phase 2: Computer Vision Models

To determine the computer vision output, all images in the dataset were run through the three different computer vision models. The models include the object detection model of Google Cloud Vision (Google Cloud, 2019), the general model of Clarifai (Clarifai, 2019b), and the complete YOLOV2 model (Redmon & Farhadi, 2017). See Fig. 1 for an impression of the output of the various computer vision models.

Phase 3: Label Evaluation

Sample Preparation

A stratified random sample of 600 images (25 images per brand) was drawn from the total database to be included in the label evaluation. Each image was already labeled by the three computer vision models, which resulted in three sets of labels (one for each model). Three separate lists were created, within each list the labels detected by one of the models, up to a maximum of 10 labels per model (see Fig. 1). If a model had detected more than 10 labels, the first 10 with the highest accuracy levels (as determined by the model) were selected. The final dataset consisted of 1,800 image-label combinations (600 images × 3 computer vision models). Each image-label combination was shown to five individual coders in order to enhance validity.

Procedure

The evaluation was conducted via Figure Eight (2019), a commonly used crowdsourcing platform for label evaluation that allows for online allocation of tasks to human coders (e.g., Tran et al., 2016). During the evaluation procedure, each coder was assigned a maximum of 30 image-label combinations that were shown to the coder in random order. For every label, the coders had to indicate to what extent they found the label accurate on a 7-point scale (ranging from 1 = *not accurate at all* to 7 = *very accurate).* Before starting the actual coding, each coder had to read the instructions and pass a test of 10 training images, to make sure they understood the coding rules. Once in every 10 images, another training question was inserted to ensure the coders were still paying attention.

Analysis

To analyze the computer vision output we make use of multiple qualitative and quantitative assessments. First, descriptive statistics are provided. Second, the accuracy of the labels is analyzed by means of linear mixed-effect analysis. Third, four concrete examples of possible marketing insights are presented. More specifically, we provide 1) a t-SNE plot (Van der Maaten & Hinton, 2008) to give insights into the clustering of different brands on an abstract level, 2) a classifier to examine if the pictures related to different brands are distinct enough to be automatically recognized as such, 3) word clouds to display the most frequent words per brand, and 4) topic modeling to identify co-occurring words.

Results

Descriptive Statistics

General statistics of the output showed that YOLOV2 recognized 78 different labels, Google recognized 3,418 different labels, and Clarifai recognized 3,577 different labels. The average number of labels per image was the highest for Clarifai, which always outputs 20 labels per image. Google recognized on average 7.87 labels and YOLOV2 recognized an average of 3.69 labels per image.

Evaluation of the Accuracy of Computer Vision Output

To assess the accuracy of the labels, we performed a linear mixed-effect analysis with the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2019). The mean accuracy, as indicated by the coders during the label evaluation procedure, served as dependent variable. The model further included a fixed effect of computer vision model and a random effect of coder to account for the repeated measures design. To maximize interpretability, YOLOV2 was used as reference category because this model had the lowest overall accuracy score (M = 4.31, SD = 2.10). With means of 5.03 (SD_{clarifai} = 0.97) and 5.69 (SD_{Google} = 0.97) respectively both Clarifai (t(2,223.52) = 9.33, $p < .001$) and Google Cloud Vision (t(1975.97) = 19.25, $p < .001$) scored significantly higher than YOLOV2. Post hoc comparisons also showed a significant

Google Cloud Vision	Clarifai	YOLOV2	Google Cloud Vision	Clarifai	YOLOV2
Denim	Fashion	Person	Bottle	No person	Bottle
Footwear	Street	Handbag	Carbonated Soft drinks	Drink	Car
Girl	Woman		CocaCola	Beer	Car
Jeans	Urban		Cola	Travel	Cup
Outerwear	Girl		Drink	Vehicle	Diningtable
Road	City		Glass	Recreation	Vase
Shoe	Model		Glass bottle	Wine	
Shoulder	People		Soft drink	Outdoors	
Snapshot	Step		Water	Glass	
Trousers	Adult			Bottle	

Fig. 1. An example of the output of the three computer vision models.

difference between Google Cloud Vision and Clarifai (t (1870.40) = 12.45, $p < .001$).

To check whether the accuracy differences are not the result of the different labels outputted by the various models, we ran the same analysis again, this time only including the labels that were recognized by both Google Cloud Vision and Clarifai. There were 413 identical labels for Google Cloud Vision and Clarifai. Sixteen of those were also recognized by YOLOV2. Therefore YOLOV2 was also included in the analysis. The results showed a similar outcome as with the full database. YOLOV2 scored with an average of 4.73 (SD = 2.18) significantly lower than Clarifai (M = 5.18, SD = 1.54, t(1,378.37) = 8.07, $p < .001$) and Google Cloud Vision (M = 5.69, SD = 1.23, t(4,228.94) = 18.07, $p < .001$). Also the difference between Google Cloud Vision and Clarifai was still significant (t(1,274.60) = 8.59, $p < .001$).

Four Concrete Examples of Applying Computer Vision to Gain Marketing Insights

To gain more insights into how the output labels can be used to analyze visual brand-related UGC, we provide four concrete examples, 1) a t-SNE plot, 2) a classifier, 3) frequency word clouds, and 4) a topic modeling procedure will be presented.

t-SNE

In marketing research it can be interesting to see how one brand is positioned as compared to other brands. To assess this, a t-SNE analysis was performed (Van der Maaten & Hinton, 2008). The t-Distributed Stochastic Neighbor Embedding (t-SNE) is a dimensionality reduction technique that is very effective for the visualization of high-dimensional data. The labels generated by our vision models are represented as so-called bag-of-word vectors. For N labels, each image is represented by an N-dimensional label vector which contains 0s, except for those labels that are generated by the computer vision models, these are assigned 1s. If two images give rise to very similar labels outputted by the vision models, the two label vectors will be similar as well. t-SNE maps these similar high-dimensional label vectors on nearby points in a 2D scatter plot, whereas it will map dissimilar label vectors onto points with a larger separation. The maps generated by t-SNE are typically more reliable than those generated by linear methods such as multidimensional scaling. Fig. 2 shows the t-SNE plot for three car brands, BMW, Ford, and Ferrari. Each dot in the plot represents the labels generated for a single image. The distance between dots (images) is proportional to the dissimilarity of their label vectors. Each color of the dot is determined by the brand depicted in the corresponding image. It is important to note that the two axes of the plot do not have a clear interpretation due to the nonlinear mapping performed by the t-SNE algorithm. Only the distances between dots and their clustering are meaningful.

The t-SNE plot shows a large cluster in the middle, and a small cluster of BMW images (red dots) on the left. The presence of all three brands (colors) in the large cluster means that most of the associated images are similarly labeled. In other words, in terms of image labels these brands are not distinct. This is not the case for the small cluster on the left. These BMW images stand out in terms of the labels. Evidently, these images are associated with a rather unique subset of BMW images. Further analysis of the labels related to the cluster showed that, whereas the big cluster mostly consists of labels such as vehicle, car, or transportation system, the left cluster is associated with motor-related labels, with words such as motorcycle, sunglasses, and landscape as unique words. Our plot reveals the known fact that, from the three brands, BMW is the only motorcycle vendor. In conclusion, a t-SNE plot can give insights into the clustering of different brands, but there is also considerable overlap because brands in the same product branch have many similarities (i.e., all car brands are likely to have a car in the picture).

Classification

To further examine whether the label output for different brands actually differs, a multiclass classification task was performed to predict to which of the 24 brands an image belongs. The labels of Google Cloud Vision, Clarifai, and YOLOV2 were transformed into word uni-grams (e.g., "denim"), bi-grams (e.g., "denim footwear"), and tri-grams (e.g., "denim footwear girl") and served as input features for the classification task. To improve the generalizability, features that occurred in less than 10 cases and feature appearing in more than 25% of the cases were removed. XGBoost (Chen & Guestrin, 2016) was used as classification algorithm and results were obtained using 10-fold cross-validation. The results of this classifier is compared to a baseline that predicts brands in a stratified way: it guesses the brand based on the distribution of images (e.g., when 70% of the images belongs to Adidas, the chance that a particular brand belongs to Adidas is 70%). Using this stratified random baseline, the classifier would predict the right brand about 1.5% of the time. After training the classifier, YOLOV2 labels helped to classify the image correctly in only 4% of the cases,

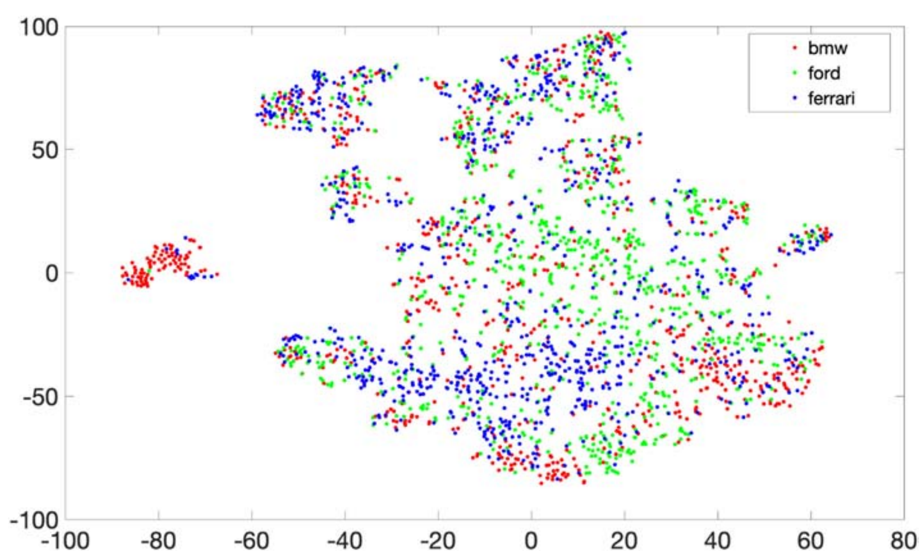

Fig. 2. t-SNE plot to assess output label similarities of images for BMW, Ford, and Ferrari. The plot was generated with the tsne function in Matlab (version R2018b; The Mathworks, 2018) with the following t-SNE parameters: # of PCs = 20, perplexity = 30, and maximum number of optimization iterations = 1,000.

barely better than chance. For Clarifai and Google Cloud Vision the classifier did much better. Based on Clarifai labels, the classifier managed to guess the correct brand in 15% of the cases, for Google Cloud Vision this percentage is even higher with 23.7% of the images correctly classified. Though 23.7% might not seem that much, it still shows that there is something in the images that makes the output different for different brands.

Frequency Word Clouds

Frequency word clouds can give more in depth insights into how the output differs per brand for the three computer vision models. A word cloud is a visual representation of a text, with words that occur more often represented in a larger font size (Word it Out, 2019). Fig. 3 shows word clouds for three different brands in the same branch. For this example we used brands in the category alcoholic drinks (i.e., Heineken, Gordon's, and Moët). We constructed the word clouds by counting the frequencies of all labels that were detected for the specific brand and by a specific model. Consequently, we plotted the 75 most frequent words in a word cloud, with larger fonts representing higher frequencies. For each brand there are three word clouds, one for every computer vision model, to examine how the different models perform. When comparing the word clouds, it is clear that YOLOV2 produces the same kind of output for all brands. Google and Clarifai produce more distinct word clouds (see Fig. 3). This is in line with previous results of the classifier. Google identifies a variety of objects

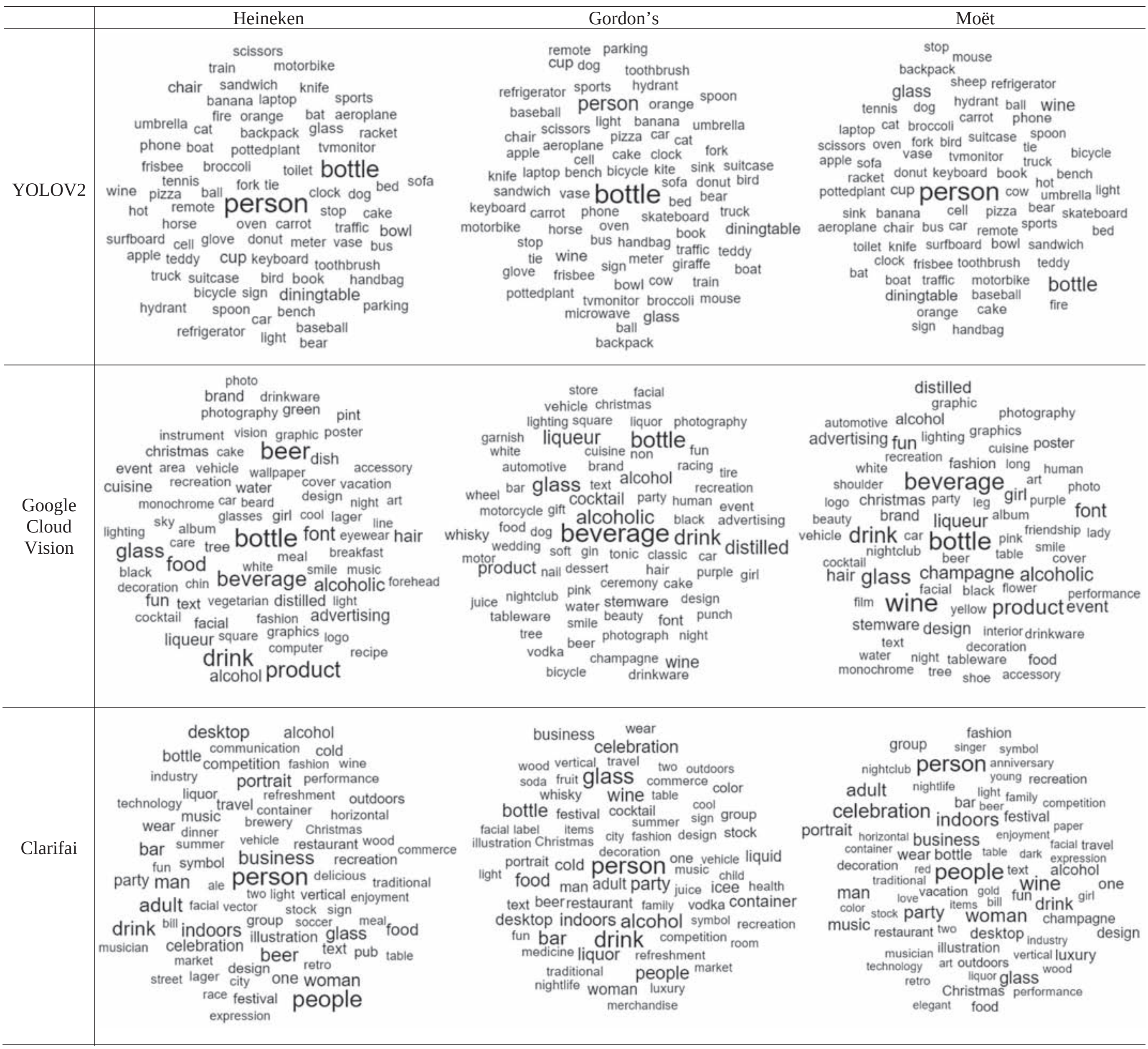

Fig. 3. Word clouds produced by the three different computer vision models for Heineken, Gordon's, and Moët consecutively. *Note.* Word clouds created via online tool worditout.com/word-cloud *Note.* The number of words to display was limited to 75 for all word clouds.

present in the image, such as glass, liqueur, bottle, whisky, and food. Furthermore, Google always recognized the word “product.” Hence, Google is more focused on products, whereas Clarifai focuses on the presence of people. Frequently recognized words include people, person, man, woman, and portrait. But in addition to this, Clarifai also attaches more subjective labels to the images, such as celebration, party, anniversary, or luxury. When comparing the output of the different brands, Clarifai therefore seems the most insightful. Looking at the Clarifai word clouds one can deduce that person, drink, bottle, alcohol, glass, people, bar, and party are typical for Gordon's. Person, people, adult, drink, business, beer, and man are most present for Heineken, and people, person, adult, woman, indoors, party, wine, drink, and celebration are most present in Moët. This yields insights regarding brand use and perception as, apparently, Gordon's and Moet are often consumed at a party. Moreover, for Heineken there are more men present whereas for Moët women are more prominent. These insights can be used to adjust marketing campaigns or channels and inspire the development of new products.

Topic Modeling

The word clouds showed which words often occur for specific brands. However, it does not yet indicate which words occur together. As this can be useful to interpret the image, we further examined the clustering of words by means of topic modeling. A Latent Dirichlet Allocation (LDA) procedure was used to automatically cluster words that often appear simultaneously in an image. Similar to Thompson and Mimno (2018), hyperparameter optimization occurred every 20 intervals after the first 50. Furthermore, the number of topics that were identified by the topic modeling procedure was set to 3 for each of the 24 brands. Due to space limitations, we again only discuss Heineken, Gordon's, and Moët in more depth (see Table 2). Results show that there are similarities between the brands (e.g., all brands had a topic about illustrations and text), but also clear differences. For example, both Moët and Gordon's seem to be portrayed mostly inside, during parties, or in a bar, whereas for Heineken words like outdoors and festival are more present. Table 2 further shows that Clarifai produces the most elaborate topics, with the most and also the greatest variety of words; therefore, Clarifai seems to be the most useful to give insights into how a brand is characterized and how this differs from other (competing) brands.

Table 2
Recognized topic models per brand for Heineken, Gordon's, and Moët.

Brand	Topic	Computer vision model	
Heineken	1	YOLOV2	–
		Google	Font, food, brand, cuisine, dish, advertising, product, text
		Clarifai	No person, business, desktop, illustration, text, symbol, design, food, sign, delicious, dinner, technology
	2	YOLOV2	Person, bottle
		Google	Fun, product
		Clarifai	People, adult, man, portrait, woman, indoors, one, music, wear, recreation, competition, festival, travel, outdoors
	3	YOLOV2	Bottle, diningtable, cup
		Google	Drink, beer, bottle, alcoholic beverage, liqueur, alcohol
		Clarifai	Drink, beer, no person, bar, glass, alcohol, bottle, cold, food, party, liquor
Gordon's	1	YOLOV2	Bottle, diningtable
		Google	Drink, liqueur, distilled beverage, alcoholic beverage, glass bottle, alcohol
		Clarifai	Drink, no person, glass, bar, alcohol, bottle, wine, food, liquor, party, container
	2	YOLOV2	Person
		Google	Fun, event, girl
		Clarifai	People, adult, woman, man, portrait, wear, festival, group, music, indoors, one, fun, recreation
	3	YOLOV2	Bottle, person
		Google	Font, product, brand, text, advertising
		Clarifai	No person, business, desktop, text, illustration, bill, symbol, sign vehicle, decoration, celebration, food, transportation system
Moët	1	YOLOV2	Person
		Google	Fun, girl, event
		Clarifai	People, woman, adult, man, portrait, music, wear, indoors, one, fashion, festival, group, party
	2	YOLOV2	Person
		Google	Advertising, product, poster
		Clarifai	Business, font, desktop, illustration, no person, text, design, brand, symbol, bill, vertical, people, retro, technology, art, paper
	3	YOLOV2	Bottle, dining table, wine glass
		Google	Drink, wine, alcoholic beverage, liqueur, bottle,
		Clarifai	No person, drink, wine, celebration, alcohol, glass, bottle, party, luxury, champagne, food

Discussion

With the increasing amount of visual brand-related UGC, there is a growing need for brands to monitor these images. At the same time, this increasing amount of visual brand-related UGC makes monitoring all this image content more complicated. Computer vision is a tempting solution, as it may provide brand owners with a cost-efficient way to constantly monitor visual brand-related UGC. Given the important role of visual brand-related UGC in brand management, it is crucial that the output of computer vision models can be trusted. Inaccurate computer vision output can seriously affect a brand's performance as it can lead to missing the opportunity to interfere with negative brand-related UGC or to understand a brand's target group. Furthermore, it is important to know whether these pre-trained models, which do not provide clear marketing variables as outcome, can be used for marketing purposes.

Despite these possible opportunities of computer vision, the current research is the first to discuss the application of computer vision in a marketing context and to examine the performance of three popularly used computer vision models for visual brand-related UGC. A series of qualitative and quantitative analyses showed that computer vision models can effectively detect objects that are present in visual brand-related UGC. Moreover, the study provides four concrete examples of how the computer vision output can be used for marketing purposes. The findings demonstrate that pre-trained computer vision models can be used to identify how brands are presented by consumers and how this relates to the representation of other brands in the same sector.

Out of the three tested models, Google Cloud Vision received the highest accuracy scores in the label evaluation procedure. However, it seems to be Clarifai that provides the most useful output labels in a marketing context. Whereas Google Cloud Vision performs best in the identification of objects that are present in the image, Clarifai makes the interpretation of the visual content easier by providing a greater variety of labels. For example, Clarifai gives more subjective labels (e.g., “sexy”) and labels referring to the overall scenery (e.g., “party”). Yet, when only looking at the objects that were recognized by both Google Cloud Vision and Clarifai, Google Cloud Vision still performed better in terms of accuracy. The lower accuracy scores are therefore not the result of the use of subjective labels. YOLOV2 was the least informative model. The 80 different output labels proved to be too limited to gain an accurate idea of the content of the images and resulted in great overlap between the label outputs for different brands.

Managerial and Scientific Implications

When choosing the right model, computer vision can provide marketers with new opportunities to monitor the performance of their brand. The current paper shows that both Google Cloud Vision and Clarifai can give insights into who uses a brand in what situations. Hence, computer vision can help marketers to monitor how a brand is portrayed by consumers and how this relates to the intended brand image (Fan & Gordon, 2014). For example, Moët is portrayed in relation to celebration and parties with women in the picture more than men. This confirms that Moët is considered a drink to celebrate. However, it also shows that this is apparently more common for women than for men. Moët can use computer vision output to 1) analyze how the brand is portrayed, and consequently, if this is in line with the intended brand image, 2) gain insights into who are their competitors (e.g., Gordon's is portrayed very similarly) and how do they differentiate from those competitors (e.g., Moët seems to be more popular amongst women), 3) develop new products based on recurrent moments of use, and 4) adjust the marketing strategy based on the target group, by focusing the content of their advertisements more on women or choosing outlets that will be seen by the target group, for example by advertising in a bar or party center.

In addition to the strengths and weaknesses of the individual computer vision models that were tested, results of this research also provide insights into the use of (pre-trained) computer vision in general to analyze visual content. With an average rated label accuracy of 5.01 (SD = 1.58) on a 7-point scale it seems that computer vision models are able to analyze visual brand-related UGC, even though the models were not trained on this specific type of visual content. Considering the rapid improvement of computer vision models, this accuracy is expected to go up, making the output of future (versions of) pre-trained computer vision models even more trustworthy. On top of that, the current paper showed that the output of pre-trained computer vision models can be used to gain important marketing insights that are hard to gather otherwise.

In combination with other metadata about the visual content, such as the number of people who viewed or responded to a post (e.g., by liking or commenting), computer vision can also help to predict downstream marketing outcomes, for example by revealing what objects in an image attract most attention amongst consumers. In the interpretation of computer vision output, research can be guided by theories from a variety of disciplines. Traditional advertising theories and insights from evolutionary psychology can guide researchers in the interpretation of why specific objects influence the attention towards image content. For instance, the presence of a person is known to attract attention towards stimuli from an evolutionary perspective (Wilson, 1975). Beneficial effects of this phenomenon have already been found in advertising (Droulers & Adil, 2015). Subsequently, theories on persuasion and eWOM can help analyze how this relates to downstream marketing outcomes (e.g., Gensler et al., 2013; Petty & Cacioppo, 1986).

This research focused on physical products since these consist of concrete objects that are easier to detect by computer vision than service products. Using computer vision to detect intangible services like a cloud service or social network seems less straightforward. Yet, the outcomes of computer vision models can also be used in the context of service brands. As opposed to physical products, service brands are not represented by one tangible object that can be detected by a computer vision model. Instead, computer vision can be used to identify objects that are often associated with service brands. For example, a specific travel agency might be associated with a beach and a glass of wine, which might indicate that people associate this brand with sunny destinations mostly whereas another travel agency is more often associated with a jungle or cultural highlights. Especially for these service brands, computer vision might also be helpful as a social media management tool. By automatically analyzing the content posted about a brand, a tourism company can quickly identify a group of sad-looking tourists in the pouring rain or an insurance company can identify which one of their customers is having trouble with a broken car on the side of the road.

There are also implications for the scientific community. Content analysis is a common technique in marketing (e.g., De Vries et al., 2012). The current research shows that certain content analysis tasks can be automated by using computer vision. As a result, it will be less time consuming to analyze marketing content. This makes it easier for researchers to apply content analysis to a larger database, leading to greater generalizability of the results.

The results described in this paper are only examples of what is possible with computer vision. To show these possibilities, the current research used images that were already tagged with the brand name. However, in a similar way it is possible to identify not-tagged content that is relevant for brands. For example, by searching an Instagram feed for images that are related to drinks, a brand such as Moët can see how present they are in the market place and how they are represented as compared to other drinks to discover what makes them unique. Furthermore, the output can be analyzed by using topic modeling (or t-SNE plotting) to identify brands that are most similar to the brand at stake, which can hence be considered the brand's biggest competitors. This information can in turn be used to adjust the marketing strategy to maximize the effectiveness of a campaign.

Limitations and Future Research

Despite the demonstrated possibilities, this research also had some limitations. The current study only looks at the general models of Google Cloud Vision and Clarifai whereas both systems have additional modules available for the analysis of specific content (e.g., wedding). We did not include those modules to allow for a better comparison to the results of YOLOV2 (as YOLOV2 did not have additional modules), but also because the possibilities of Google Cloud Vision and Clarifai differ. For example, Google Cloud provides an emotion analysis of the detected faces, whilst Clarifai provides more labels regarding demographics of people in the image. Future research can also take these specific modules into account to evaluate the accuracy of the different computer vision models more thoroughly.

Moreover, while this research focused on the possibility to use computer vision to analyze visual brand-related UGC, it did not yet make a connection to common marketing outcomes, such as brand performance or engagement (e.g., amount of likes). This study serves as a stepping stone for future research to further examine the possibilities of computer vision in the field of marketing. For example, by using computer vision to predict which images are going to be most popular or by relating the objects present in the image to people's brand attitude or purchase intention. As a result, computer vision can not only be used to monitor visual brand-related UGC but also serve as a guide to produce effective MGC for marketers and as a multi-functional tool in marketing research for scholars.

The current research focused on the computer vision model's ability to recognize objects in an image. This is the most promising application of computer vision because object detection is already far developed (Szegedy et al., 2017). However, much marketing research focuses on a more diverse range of aspects in visual brand-related UGC than the objects that are present in an image. For example more implicit messages, such as the credibility of visual brand-related UGC (Lin, Spence, & Lachlan, 2016). Such subjective analyses are difficult to perform automatically and still require manual (re) coding of the output.

Declaration of Competing Interest

None.

Acknowledgements

We are grateful to Emiel Krahmer for thinking along with the design of the label evaluation. We would also like to thank our fellow researchers at the Interactive Marketing Research Conference 2018, for the useful feedback they provided during the High Density Session.

References

Advertising Standards Authority UK (2019), “Recognizing ads: Social media,” Retrieved from https://www.asa.org.uk/advice-online/recognising-ads-social-media.html (last accessed August 25th, 2019).

Aker, Cemal and Sinan Kalkan (2017), “Using deep networks for drone detection,” *Paper presented at IEEA VVS, Lecce, Italy.* Retrieved from: arXiv:1706.05726v1.

Bakhshi, Saeideh and Eric Gilbert (2015), “Red, purple, and pink: The colors of diffusion on pinterest,” *PLoS One*, 10, 2, 1–20.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker (2015), “Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software*, 67, 1–48.

Bianco, Simone, Marco Buzzelli, Davide Mazzini, and Raimondo Schettini (2017), “Deep learning for logo recognition,” *Neurocomputing*, 245, 23–30.

Chari, Simos, George Christodoulides, Caterina Presi, Jil Wenhold, and John P. Casaletto (2016), “Consumer trust in user-generated brand recommendations on facebook,” *Psychology and Marketing*, 33, 12, 1071–81.

Chau, Patrick Y.K., Grace Au, and Kar Y. Tam (2000), “Impact of information presentation modes on online shopping: An empirical evaluation of a broadband interactive shopping service,” *Journal of Organizational Computing and Electronic Commerce*, 10, 1, 1–22.

Chen, Tao and Mark Dredze (2018), “Vaccine images on twitter: Analysis of what images are shared,” *Journal of Medical Internet Research*, 20, 4e130.

Chen, Tianqi and Carlos Guestrin (2016), *Xgboost: A scalable tree boosting system*. San Francisco, California: Paper Presented at KDD. Retrieved from: https://dl.acm.org/citation.cfm?id=2939785.

Cheng, Gong, Junwei Han, and Xiaoqiang Lu (2017), “Remote sensing image scene classification: Benchmark and state of the art,” *Proceedings of the IEEE*, 105, 10, 1865–83.

Chu, Shu-Chuan and Yoojung Kim (2011), “Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites,” *International Journal of Advertising*, 30, 1, 47–75.

Clarifai (2019a), https://clarifai.com/ (Last accessed February 18th, 2019).

——— (2019b), https://clarifai.com/models/general-image-recognition-model-aaa03c23b3724a16a56b629203edc62c (Last accessed February 18th, 2019).

De Vries, Lisette, Sonja Gensler, and Peter S.H. Leeflang (2012), “Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing,” *Journal of Interactive Marketing*, 26, 83–91.

Droulers, Oliver and Safaa Adil (2015), “Could face presence in print ads influence memorization?” *The Journal of Applied Business Research*, 31, 1403–8.

Erkan, Ismail and Chris Evans (2016), “The influence of ewom in social media on consumers' purchase intentions: An extended approach to information adoption,” *Computers in Human Behavior*, 61, 47–55.

Fan, Weiguo and Michael D. Gordon (2014), “The power of social media analytics,” *Communications of the ACM*, 57, 6, 74–81.

Ferwerda, Bruce and Marko Tkalcic (2018), “You are what you post: What the content of instagram pictures tells about user's personality,” *Paper presented at HUMANIZE, Tokyo, Japan.* Retrieved from: http://www.divaportal.org/smash/get/diva2:1183753/FULLTEXT01.pdf.

Figure Eight (2019), https://www.figure-eight.com/ (Last accessed February 18th, 2019).

Gensler, Sonja, Franziska Völckner, Yuping Liu-Thompkins, and Caroline Wiertz (2013), “Managing brands in the social media environment,” *Journal of Interactive Marketing*, 27, 242–56.

Google Cloud (2019), "Vision," Retrieved from https://cloud.google.com/vision/ (Last accessed February 18th, 2019).

Guo, Yandong, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao (2016), "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *Computer vision – ECCV 2016*. Bastian Leibe, Jiri Matas, Nicu Sebe, Max Welling, editors. Amsterdam, The Netherlands: Springer, Cham, 87–102.

Hennig-Thurau, Thorsten, Charles F. Hofacker, and Björn Bloching (2013), "Marketing the pinball way: Understanding how social media change the generation of value for consumers and companies," *Journal of Interactive Marketing*, 27, 237–41.

Hernández-Méndez, Janet and Francisco Muñoz-Leiva (2015), "What type of online advertising is most effective for etourism 2.0? An eye tracking study based on the characteristics of tourists," *Computers in Human Behavior*, 50, 618–25.

Hollenbeck, Candice R. and Andrew M. Kaikati (2012), "Consumers' use of brands to reflect their actual and ideal selves on facebook," *International Journal of Research in Marketing*, 29, 395–405.

Hosseini, Hossein, Baicen Xiao, and Radha Poovendran (2017), "Google's cloud vision api is not robust to noise," *Paper presented at ICMLA, Cancun, Mexico.* Retrieved from: arXiv:1704.05051v2.

Ismagilova, Elvira, Emma Slade, and Michael Williams (2016), "Persuasiveness of ewom communications: Literature review and suggestions for future research," in *Social Media: The Good, the Bad, and the Ugly Lecture Notes in Computer Science*. Y.K. Dwivedi, et al, editors. Swansea, UK: Springer Cham, 354–9.

Jaakonmäki, Roope, Oliver Müller, and Jan Vom Brocke (2017), "The impact of content, context, and creator on user engagement in social media marketing," *Paper presented at: HICCS 2017, Waikoloa, Hawaii.* Retrieved from: http://hdl.handle.net/10125/41289.

Kaplan, Andreas M. and Michael Haenlein (2010), "Users of the world, unite! The challenges and opportunities of social media," *Business Horizons*, 53, 59–68.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015), "Deep learning," *Nature*, 521, 7553, 436–44.

Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnic (2014), "Microsoft coco: Common objects in context," in *Computer Vision – ECCV 2014*. David Fleet, Tomas Pajdla, Bernt Schiele, Tinne Tuytelaars, editors. Zurich, Switzerland: Springer Cham, 740–55.

Lin, Xialing, Patric R. Spence, and Kenneth A. Lachlan (2016), "Social media and credibility indicators: The effect of influence cues," *Computers in Human Behavior*, 63, 264–71.

Liu, Liu, Daria Dzyabura, and Natalie Mizik (2017), "Visual listening," *Extracting brand image portrayed on social media.* Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2978805. (Last accessed February 2nd, 2019).

Marr, David (1982), *Vision: A computational investigation into the human representation and processing of visual information*. New York, New York: Freeman.

Mazloom, Masoud, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn Van Dolen (2016), "Multimodel popularity prediction of brand-related social media posts," *Paper presented at MM'16, Amsterdam, The Netherlands.* Retrieved from: https://dl.acm.org/citation.cfm?id=2967210.

McDonald, Malcom and Ian Dunbar (2012), *Market segmentation: How to do it and how to profit from it*. West Sussex, UK: John Wiley & Sons Ltd.

Muntinga, Daniël G., Marjolein Moorman, and Edith G. Smit (2011), "Introducing cobras: Exploring motivations for brand-related social media use," *International Journal of Advertising*, 30, 1, 13–46.

Netbase (2018), "Netbase report reveals most loved brands of 2018," Retrieved from https://www.netbase.com/press-release/netbase-report-reveals-the-most-loved-brands-of-2018/ (Last accessed August 25th, 2019).

Papernot, Nicolas, Patrick McDaniel, Somesh Jha, Matt Fredrikson, Z. Berkay Celik, and Ananthram Swami (2015), "The limitations of deep learning in adversarial settings," *Paper Presented at IEEE EuroS&P, Saarbrucken, Germany.* Retrieved from: arXiv:1511.07528v1.

Petty, Richard E. and John T. Cacioppo (1986), "The elaboration likelihood model of persuasion," *Advances in Experimental Social Psychology.*, 19, 123–205.

Pew Research Center (2018a), "Social media fact sheet," Retrieved from http://www.pewinternet.org/fact-sheet/social-media/ (Last accessed August 25th, 2019).

——— (2018b), "Teens, social media, and technology," Retrieved from http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/ (Last accessed August 25th, 2019).

R Core Team (2019), *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org.

Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi (2016), "You only look once: Unified real-time object detection," *Paper Presented at IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada.* Retrieved from https://www.cvfoundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html.

——— and Ali Farhadi (2017), "Yolo9000: Better, faster, stronger," *Paper Presented at IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii.* Retrieved from: http://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html.

Serrano, Shanna and Tauheed Ramjaun (2018), "Exploring visual brand-related electronic word-of-mouth on instagram," *Journal of Promotional Communications*, 6, 1, 72–88.

Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi (2017), "Inception-v4, inception-resnet and the impact of residual connections on learning," *Paper Presented at the AAAI-17, San Francisco, California.* Retrieved from: https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/viewPaper/14806.

Szeliski, Richard (2011), *Computer vision: Algorithms and applications*. London, UK: Springer-Verlag.

The Mathworks, Inc (2018), *Matlab and statistics toolbox release 2018b*. Mississippi: Natick. Retrieved from https://nl.mathworks.com/products/matlab.html?s_tid=hp_products_matlab.

Thompson, Laure and David Mimno (2018), "Authorless topic models: Biasing models away from structure," *Paper Presented at COLING, Santa Fe, New Mexico.* Retrieved from: https://par.nsf.gov/biblio/10092208.

Tous, Ruben, Mauro Gomez, Jonatan Poveda, Leonel Cruz, Otto Wust, Mouna Makni, and Eduard Ayguadé (2018), "Automated curation of brand-related social media images with deep learning," *Multimedia Tools and Applications*, 77, 20, 27123–42.

Tran, Kenneth, Xiaodong He, Lei Zhang, Jian Sun, Cornelia Carapcea, Chris Thrasher, … Chris Sienkiewicz (2016), "Rich image captioning in the wild," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 49–56. Retrieved from https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w12/html/Tran_Rich_Image_Captioning_CVPR_2016_paper.html.

Van der Maaten, Laurens and Geoffrey Hinton (2008), "Visualizing data using t-sne," *Journal of Machine Learning Research*, 9, 2579–605.

Van Noord, Nanne, Ella Hendriks, and Eric Postma (2015), "Towards discovery of the artist's style: Learning to recognize artists by their artworks," *IEEE Signal Processing Magazine*, 32, 4, 46–54.

Vilnai-Yavetz, Iris and Sigal Tifferet (2015), "A picture is worth a thousand words: Segmenting consumers by facebook profile images," *Journal of Interactive Marketing*, 32, 53–69.

Wilson, Edward O. (1975), *Sociobiology*. Cambridge, Massachusetts: Harvard University Press.

Word It Out (2019). https://worditout.com/word-cloud/create (Last Accessed August 25th, 2019).

Zeiler, Matthew D. and Rob Fergus (2013), "Visualizing and understanding convolutional networks," in *Computer vision – ECCV 2014*. David Fleet, Tomas Pajdla, Bernt Schiele, Tinne Tuytelaars, editors. Zurich, Switzerland: Springer Cham, 818–33.