Journal of Advertising

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujoa20

Automatically Detecting Image–Text Mismatch on Instagram with Deep Learning

Yui Ha, Kunwoo Park, Su Jung Kim, Jungseock Joo & Meeyoung Cha

To cite this article: Yui Ha, Kunwoo Park, Su Jung Kim, Jungseock Joo & Meeyoung Cha (2021) Automatically Detecting Image–Text Mismatch on Instagram with Deep Learning, Journal of Advertising, 50:1, 52-62, DOI: 10.1080/00913367.2020.1843091

To link to this article: https://doi.org/10.1080/00913367.2020.1843091

JOURNAL OF ADVERTISING
2021, VOL. 50, NO. 1, 52–62
https://doi.org/10.1080/00913367.2020.1843091

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Automatically Detecting Image–Text Mismatch on Instagram with Deep Learning

Yui Ha[a]*, Kunwoo Park[b]*, Su Jung Kim[c]*, Jungseock Joo[b]* and Meeyoung Cha[d,e]*

[a]Korea Electric Power Corporation, Seoul, Republic of Korea; [b]University of California Los Angeles, Los Angeles, California, USA; [c]University of Southern California, Los Angeles, California, USA; [d]Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea; [e]Institute for Basic Science, Daejeon, Republic of Korea

ABSTRACT

Visual social media have emerged as an essential brand communication channel for advertisers and brands. The active use of hashtags has enabled advertisers to identify customers interested in their brands and better understand their consumers. However, some users post brand-incongruent content—for example, posts composed of brand-irrelevant images with brand-relevant hashtags. Such visual information mismatch can be problematic because it hinders other consumers' information search processes and advertisers' insight generation from consumer-initiated social media data. This study aims to characterize visually mismatched content in brand-related posts on Instagram and builds a visual information mismatch detection model using computer vision. We propose a machine-learning model based on three cues: image, text, and metadata. Our analysis shows the effectiveness of deep learning and the importance of combining text and image features for mismatch detection. We discuss the advantages of machine-learning methods as a novel research tool for advertising research and conclude with implications of our findings.

Instagram is one of the fastest growing photo- and video-sharing social media platforms and has attracted more than 1 billion monthly users worldwide. In the United States, there were approximately 107.2 million Instagram users by 2018, and this number is expected to grow to 120.3 million by 2023 (Nuñez 2020). With its increasing popularity, advertisers and brands have paid attention to Instagram's potential as a brand communication channel in social media. For the term *brand communication in social media*, we follow the definition of Alhabash, Mundel, and Hussain (2017) and refer to it as brand-related communication distributed via social media that enables Internet users to access, share, engage with, add to, and co-create. This definition includes both brand-generated posts (e.g., advertisements) and user-generated content (UGC) (Voorveld 2019), but we particularly focus on consumer-generated brand communication messages in this study because we are interested in examining how consumers use images and texts when they create brand-related posts in Instagram. Consumers create brand-related posts by using an image of a product with a hashtag indicating the brand (e.g., #apple, #chanel) as a form of brand engagement and loyalty (Phua, Jin, and Kim 2017). For consumers, this combined use of brand-related images and hashtags is a way to find brand-related information and connect with other consumers who have similar tastes (Sung, Kim, and Choi 2018). For advertisers and brands, consumer-generated social media data help to better

CONTACT Su Jung Kim sujung.kim@usc.edu; Jungseock Joo jjoo@comm.ucla.edu; Meeyoung Cha meeyoungcha@kaist.ac.kr

Yui Ha (PhD, Korea Advanced Institute of Science and Technology) is a senior researcher, Data Science Research Center, Korea Electric Power Corporation, Seoul, Republic of Korea.

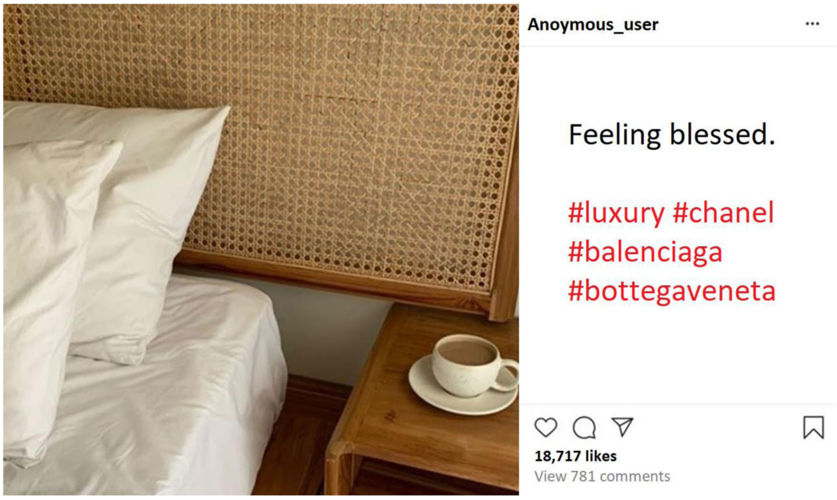
Kunwoo Park (PhD, Korea Advanced Institute of Science and Technology) is a postdoctoral researcher, Department of Communication, University of California Los Angeles, Los Angeles, California, USA.

Su Jung Kim (PhD, Northwestern University) is an assistant professor, Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, California, USA.

Jungseock Joo (PhD, University of California, Los Angeles) is an assistant professor, Department of Communication, University of California Los Angeles, Los Angeles, California, USA.

Meeyoung Cha (PhD, Korea Advanced Institute of Science and Technology) is a chief investigator, Institute for Basic Science, and an associate professor, School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea.

*Y.H. and K.P. contributed equally as first authors analyzing the data, and S.J.K. J.J, and M.C. contributed equally as co-corresponding authors.



Feeling blessed.

#luxury #chanel
#balenciaga
#bottegaveneta

Figure 1. An example of visual misinformation that uses brand-related hashtags with a brand-irrelevant image.

understand customers' needs—an element which was not accessible through traditional media (Berger et al. 2020).

Despite the benefits gained from analyzing brand-related posts, some users create posts containing brand-irrelevant images with brand-related hashtags, which may confuse other users, brands, and advertisers. For example, the post in Figure 1 is shared with hashtags referring to luxury brands #chanel and #balenciaga with pictures that portray the post creator's daily life. This image–hashtag mismatch can be problematic for brands, advertisers, and consumers. Such mismatch makes it difficult for brands and advertisers to accurately estimate the volume of conversation on their brands happening in Instagram and understand customer sentiments from such posts. Moreover, mismatched brand-related posts hinder consumers' information search processes for those who use brand-related hashtags to find brand information. Therefore, this issue of visual information mismatch, which we define as a type of incongruence where brand-relevant hashtags are used with brand-irrelevant images, should be addressed appropriately for advertisers and brands interested in using social media to reach, attract, and understand new and existing consumers.

To this end, we propose a unique method that applies computer vision to characterize and detect visual information mismatch in brand-related posts. Using 452,616 Instagram posts on fashion brands, we analyze the multimodal characteristics of visual information mismatch and develop a detection model. Computer vision is one of the artificial intelligence (AI) technologies especially useful in understanding consumer insights in digital advertising (Li 2019). AI aims to develop automated systems with intellectual capabilities such as perceiving, planning, reasoning, and acting to achieve a goal (Russell and Norvig 1995). In advertising, AI can be deployed to automatically analyze, generate, and personalize messages and understand consumer preferences and behaviors, affecting every facet of advertising from ad creation to consumer targeting to media planning to ad evaluation. With the growing interest in visual social media platforms, employing computer vision helps researchers examine large-scale multisensory data and understand their meanings and implications. To illustrate how computer vision and machine-learning techniques can be applied to visual social media, we pose three research questions:

> **RQ1:** How can we automatically classify visual information mismatch in brand-related posts using multimodal features extracted from images, texts, and metadata?
>
> **RQ2:** How does visual information mismatch in brand-related posts differ from visual information match regarding multimodal features extracted from images, texts, and metadata?
>
> **RQ3:** Does a brand-specific classification model for visual information mismatch detection outperform a general brand classification model for visual information mismatch detection?

Literature Review

The Multimodal Nature of Instagram and Image–Text Match As Contextual Congruence

What sets visual social media like Instagram apart from text-based social media is the posts' multimodal nature, especially the combined use of images and hashtags. Typically, users describe a subject, topic, or context of a situation with a hashtag, which consists of the symbol # and word that follows it (e.g., #fashion). Instagram users can search for a vast pool of visual content or interact with people who have common interests through hashtags (Daer, Hoffman, and Goodman 2014). While there is no restriction on selecting hashtags for a post, it is recommended to use appropriate hashtags in accordance with the community culture that follows the "folksonomy"—a user-driven classification of information (Ibba et al. 2015). From a brand communication perspective, the practices of exploration, networking, and hashtagging allow consumers to learn about unfamiliar brands (Sheldon and Bryant 2016), strengthen the connection with familiar brands (Pentina, Guilloux, and Micu 2018), and find a large number of images associated with user-defined hashtags.

Extant advertising research on Instagram has shown that a well-curated picture and hashtag(s) may lead to positive brand outcomes such as engagement (e.g., likes and comments), brand attitudes, and brand attachment (Kim and Phua 2020; Rietveld et al. 2020). These findings imply that a good alignment between a picture and hashtag(s) in a brand-related post is key to generating positive outcomes for brands and advertisers. This brings us to the concept of contextual congruence, which explains the degree of similarity between the advertised content and the editorial content. In visual social media, we argue that the fit between the content elements (i.e., brand-related images and texts) and the fit between a brand-related post and its surrounding elements form the basis of contextual congruence. Such coherence would lead to less intrusiveness and better acceptance of brand-related posts, as previous research on congruence suggests (King, Reid, and Macias 2004; Moore, Stammerjohan, and Coulter 2005).

In this study, we turn our attention to the opposite case (i.e., incongruence), where an image and hashtag(s) are poorly matched by the content creator, as illustrated in Figure 1. We consider the instance of misusing a brand-related hashtag with an image that does not include a product or logo of the brand mentioned in the hashtag as one specific type of contextual incongruence and refer to it as *visual information mismatch*. We expect that visual information mismatch in brand-related posts presumably will have negative impacts on consumers, brands, and advertisers. For consumers, visually mismatched posts become an obstacle for information searching and networking and, more broadly speaking, hurt the user-driven culture of hashtagging (i.e., folksonomy). For brands and advertisers, mismatched images and hashtags become "noise" in the data and provide little to no value when analyzing these social media posts to understand their consumers' thoughts and feelings.

Using Computer Vision to Detect Visual Information Mismatch

Given the potential negative consequences of visual information mismatch, it becomes critical to detect incongruent brand-related posts and adequately address them. We employ computer vision and machine learning tools to build a prediction model of visual information mismatch using variables extracted from images, hashtags, and metadata from Instagram. Previous research on detecting mismatched information includes predicting clickbait news or fake/deceptive consumer reviews. These studies have used machine learning to detect inconsistencies between a headline and body text in news articles (Yoon et al. 2019) or find distinctive characteristics of fake/deceptive reviews regarding content (e.g., sentiments) or reviewers (e.g., average review length) (Vidanagama, Silva, and Karunananda 2020).

Although these mismatch detection mechanisms in clickbait news or fake/deceptive review detection can reach a high accuracy, they focus on text-based algorithms due to the nature of the data (e.g., headline, review content). Instagram posts are multimodal data that integrate images and corresponding hashtags, which require additional considerations for analyzing visual content. Unlike conventional content analysis relying on human coders manually annotating a handful of images, computer vision allows researchers to extract features from given images automatically. A feature is a measurable property of the analyzed image. For example, to automatically classify car images, one could select a few distinctive features of cars (e.g., wheels, windows, headlights) that are fed into a machine-learning model. Once a prediction model is trained on a reasonably sized data set using the extracted visual features, it can further process an unlimited amount of data without further annotations. Contrary to the manual coding approach, computer vision's scalability is more suitable and efficient for analyzing multimodal data from social media where millions of posts are generated daily (Nanne et al. 2020).

In advertising and marketing, more studies have begun to apply computer vision to analyze visual social media (Liu, Dzyabura, and Mizik 2020; Nanne et al. 2020; Tous et al. 2018; Vassey et al. 2020). Such studies use computer vision to detect or predict brand attributes from images or suggest better ways to curate social media content. The current study is one of the first efforts that applies computer vision to identify and predict visual information mismatch between images and hashtags in brand-related social media posts.

We propose a novel approach that quantifies image characteristics and uses multimodal features for visual information mismatch detection. We present a three-step analysis. First, we offer three categories of features from images, texts, and metadata. After manually annotating a sample of Instagram images, we apply computer vision and machine-learning approaches to extract the visual features from the entire set (research question 1). Second, based on the multimodal features, we develop a machine-learning classifier that detects visual information mismatch. By comparing evaluation performance across different feature combinations, we show the effectiveness and robustness of content-related features for mismatch detection (research question 2). Third, we develop a brand-specific detection model and compare it with a general detection model to test whether a brand-specific model outperforms the general model (research question 3).

Methods

Data Collection

We used data collected from Instagram in July 2017 during the time when Paris's haute couture fashion week took place. Because fashion brands and consumers actively post content during this period, focusing on this timeframe enabled us to examine data on diverse fashion brands. We searched for posts that mentioned hashtag(s) indicating any of the 64 world-renowned fashion brands, all of which have more than 50,000 followers on their official Instagram accounts. We selected the fashion industry because (1) following fashion was one of the primary motivations for following fashion brands on Instagram (Phua, Jin, and Kim 2017); (2) focusing on the single industry allows us to capture image–text mismatch more efficiently because Instagram posts about fashion brands have common visual characteristics, facilitating machines to quantify patterns from the data; and (3) the fashion industry has a clear hierarchy concerning brand value, represented by the prices of brand items. The hierarchy makes it easy to control for the effect of brand value by analyzing them within or across ranks in the hierarchy. Following the approach in Wang (2015), we categorized 64 brands into high-end (e.g., Loro Piana, Louis Vuitton), specialty retailers of private label apparel (e.g., Zara, Uniqlo), and fine jewelry and watches (e.g., Cartier, Tiffany).

Using the InstaLooter Python library, we collected Instagram posts from those 64 brands during the study period. To ensure we had enough time to observe user reactions, we collected engagement metrics (e.g., likes) on the posts for a week after the data collection period. We removed posts from users with more than 2 million followers because such posts created by microcelebrities may exhibit different patterns of content and engagement compared to those by average users. The data set included 452,616 posts by 217,474 users. Each post comprises content and metadata, such as username and the number of likes and comments.

Data Annotation

To determine whether a given post's image represents the target fashion brand indicated by a hashtag, we conducted a crowdsourcing task via CrowdFlower. While crowdsourcing overcomes the limitation of traditional surveys such as selection bias (de Winter et al. 2015), low quality of responses has been one of the major concerns (Sheehan 2018). For quality control, CrowdFlower has adopted a mechanism for filtering untrustworthy workers: Certain questions shown to workers as part of the questionnaire evaluate the validity of crowdsource workers; these are known as golden questions. Evaluating historical records of workers alongside responses to golden questions has allowed researchers to filter out untrustworthy workers' responses and achieve reliable outcomes in recent studies (e.g., Krauss et al. 2017).

Using the system's scores, we allowed only those individuals at level two or higher who had already completed more than 100 jobs and achieved an accuracy of more than 80% in their records to ensure the quality of the results. The annotators were paid for 3 to 5 cents for a task. Before the main annotation task, we trained annotators to label a picture with one of the three answers (*Yes*, *No*, *Not sure*) depending on whether the assigned hashtag (#fashion brand name) matched the image provided: To answer *Yes*, the image needed to contain the fashion items or logos of the tagged brand. To answer *No*, the image needed to not have a product or logo of the tagged brand. To answer *Not sure*, the annotator had to be unsure of the exact brand name of the product.

A total of 37,977 assessments were performed for 12,659 randomly selected posts by assigning three annotators per post. We considered only those posts that reached 100% agreement of the responses as the ground truth data; we excluded posts labeled *not sure*.

Table 1. A summary of descriptive statistics of labeled and unlabeled data.

	Labeled Data			Unlabeled Data		
Items	*M*	*SD*	Count	*M*	*SD*	Count
Number of likes per post	69.34	453.75		116.29	1119.05	
Number of comments per post	1.77	12.34		3.02	21.13	
Number of posts per person	1.26	1.15		2.08	6.34	
Number of posts			7,769			444,491
Number of users			6,159			213,838
Number of brands			63			73

Of the successfully labeled 7,769 posts, 3,509 posts were image–hashtag matched, while 4,260 posts were mismatched. We call the 7,769 labeled posts *labeled data* and the rest *unlabeled data* for this article. Table 1 summarizes the descriptive statistics of the two data sets.

Feature Extraction

Image Features

To quantify the visual characteristics of Instagram posts, we considered two types of image features—semantic and generic, which complement each other. Figure 2 demonstrates how semantic and generic features represent an image.

Semantic Image Features

We annotated 1,000 sampled images with tags such as *smiling face* or *body part* using the grounded theory approach to capture semantic information from images. After several rounds of merging the tags, we identified 10 semantic categories while allowing duplicates: selfie (18%), body snap (23%), marketing (22%), product only (20%), nonfashion (17%), face (26.8%), logo (53.5%), brand logo (16.2%), smile (7.9%), and outdoor (10.4%). Using these semantic categories, we asked crowdworkers to annotate 3,169 images sampled from the 7,769 posts (i.e., labeled data). Here, we proposed to use a transfer learning approach based on the convolutional neural network (CNN) to automatically annotate semantic categories of the entire data set using only the semantic labels of the 3,169 images. In particular, we fine-tuned ResNet-50, which had been pretrained on the ImageNet data set, which is composed of more than 10 million images (Russakovsky et al. 2015). This transfer learning approach enabled us to learn specific patterns for a downstream task (i.e., semantic tag classification) based on the pretrained networks on a huge set of images (i.e., ImageNet) and thus generalizes well for unseen data during training (Goodfellow, Bengio, and Courville 2016).

Generic Image Features

To capture image characteristics in addition to the predefined semantic categories, we employed a separate CNN model that takes an image as input and outputs a total of 2,048 features. In particular, we adopted AlexNet (Krizhevsky, Sutskever, and Hinton 2012), which was also pretrained on the ImageNet data set. These generic features were seen as a low-dimensional representation of the raw images.

Text Features

To capture the characteristics of hashtags, we measured term frequencies that are inversely weighted by document frequency (term frequency–inverse document frequency [TF-IDF]) based on the following criteria: (1) top-100 most popular hashtags on Instagram and (2) top-100 world-renowned fashion brand names. Compared to frequency, TF-IDF could capture the relative importance of a hashtag occurrence considering its general popularity (Rajaraman and Ullman 2011).

Metadata Features

We also used metadata to quantify the characteristics of Instagram posts other than the content elements. We computed the number of followings, followers, and posts created by each user. We also measured the number of received likes and comments as a proxy for popularity.

Analysis

To answer research question 1, we developed a machine-learning classifier that detects visual information mismatch based on the multimodal features extracted from images, texts, and metadata. By comparing various evaluation metrics, we determined which features or combinations of features yielded the highest effectiveness and robustness levels. To answer research question 2, we inferred the mismatch label on a more extensive set of data using the machine-learning model trained for research question 1 and compared characteristics of general brand posts with an image–hashtag match with those with an image–hashtag mismatch. To answer research question 3, we compared the prediction models

Semantic Image Features (10 binary)				
Selfie	Body snap	Marketing	**Product-only**	Nonfashion
0.000	0.001	0.002	**0.999**	0.000
Face	Logo	**Brand logo**	Smile	Outdoor
0.000	0.783	**0.891**	0.000	0.000
Generic Image Features (2,048 dimensions)				
0	1	...	2046	2047
0.002	0.013	...	0.289	0.224

Figure 2. An example on how an image is represented by semantic and generic features.

for general brand-related posts and brand-specific posts and examined whether a brand-specific classification model boosts the performance of detecting image–hashtag mismatching compared to the general model.

Results

Developing a Visual Information Mismatch Detection Model for Brand-Related Posts

Research question 1 asked how we can automatically classify visual information mismatch in brand-related posts. We tested the effectiveness of the proposed computer vision-based features for visual information mismatch detection, where the image of a post does not correspond to the brand-related hashtag(s) in the same post. To evaluate classification performance, we split the 7,769 labeled posts into training and test sets using the ratio of 80:20. We compared the performance of three classification algorithms: logistic regression, nonlinear support vector machine (SVM), and random forest. We conducted fivefold cross-validation using a grid search to optimize each model's performance by varying prediction thresholds. As a result, the decision threshold of 0.53 was chosen.

Figure 3 shows the classification performance in terms of accuracy, precision, F1 score, and the area under the curve of the receiver operating characteristic (AUC ROC). Accuracy indicates how many instances are predicted correctly compared to the entire set. Precision indicates how many instances are correct among those a model predicts as a positive class. F1 score is the harmonic mean of precision and recall (i.e., how many instances are retrieved among true labels). AUC ROC is an aggregate measure of performance by varying prediction thresholds. Among the four metrics, F1 score and AUC ROC are generally considered standard metrics for evaluating classification performance for their robustness to a biased distribution.

We focused on three main observations from Figure 3. First, among the three algorithms, random forest showed the best performance across all feature combinations. Hence, we use the random forest classifier to report the results for the rest of this article. Second, the image feature alone marked the highest accuracy of 0.837 among the three independent feature sets compared to the text feature (0.830) and the metadata feature (0.622). Third, the combination of image and text features showed the highest classification performance with an accuracy of 0.864, a precision of 0.934, an AUC ROC of 0.938, and an F1 score of 0.853. Adding the metadata features to the model did not lead to a performance change, suggesting that features from images and hashtags together play a significant role in mismatch detection.

Finding the Differences between Mismatched and Matched Posts

Research question 2 asked how visual information mismatch in brand-related posts differs from matched posts in multimodal features. We compared each of the features using the *t* test for continuous variables and the chi-square test for binary variables. Figure 4 shows the features exhibiting a significant difference. For clarity, we presented only the variables for which the effect size is greater than 0.5 and the *p* value is lower than 0.05. We found significant differences in four features—nonfashion, logo, product only, and hashtag length—where the first three are visual semantic features. Most of the text features turned out to be not as important as visual features except for hashtag length. The results suggest that visually mismatched posts are more likely to be composed of nonfashion images and lengthy hashtags.

To draw a bigger picture of how visually mismatched posts differ from matched posts, we inferred the missing 452,616 posts' labels using the random

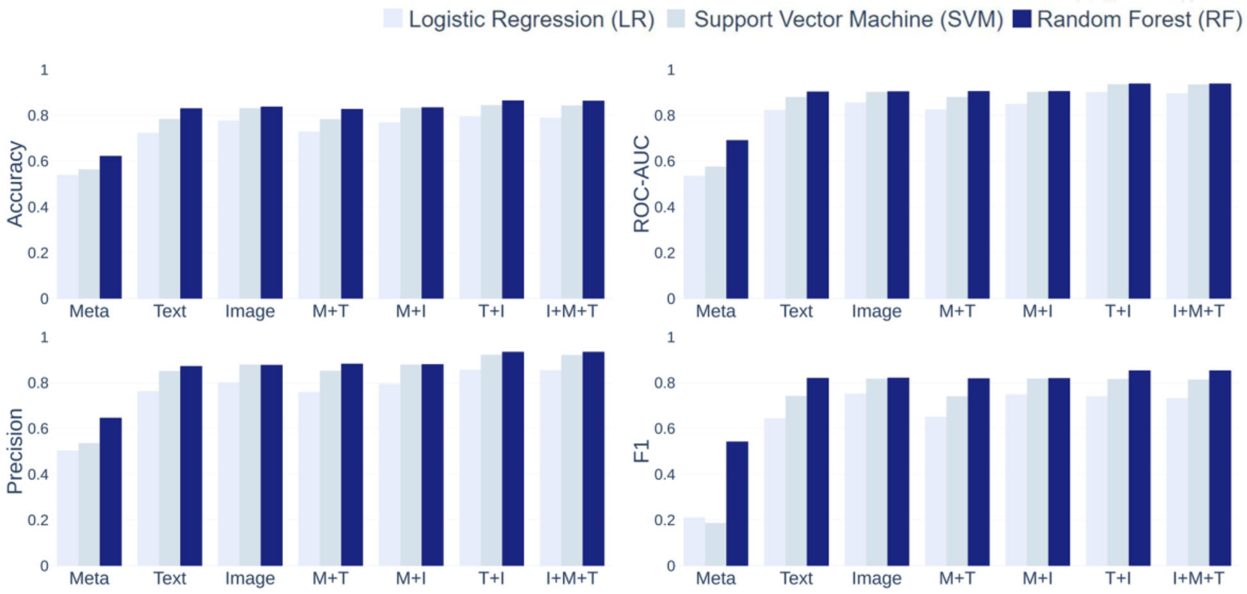


Figure 3. Evaluation results of the three classification models with an 80:20 random split.

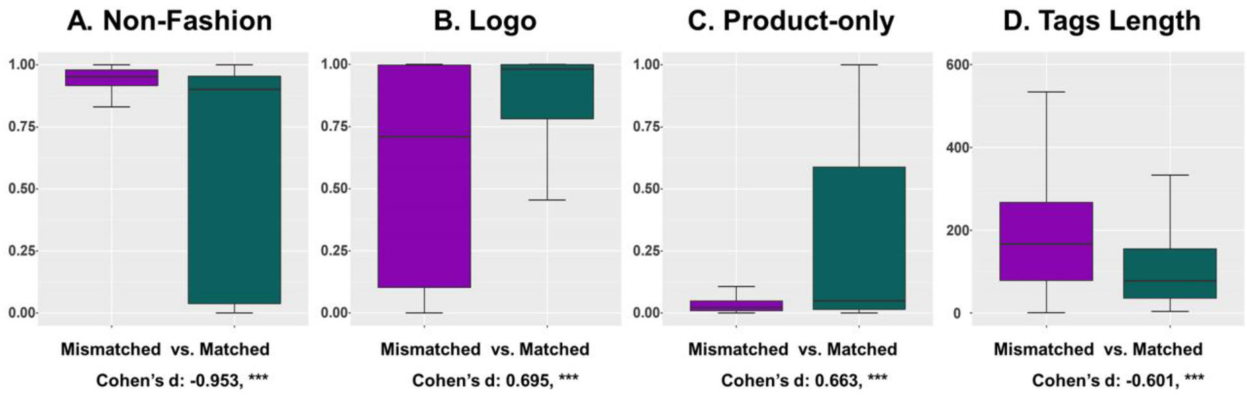


Figure 4. Comparison between mismatched and matched posts (results with Cohen's *d* greater than 0.5 and *p* value lower than 0.05 is shown).

forest trained on the labeled data set. Our model classified 30,342 (11%) posts as mismatches and 247,085 (89%) as matched posts. We then conducted statistical tests to identify any features that show a significant difference between the two groups. Again, the visual feature on the nonfashion feature showed the largest difference with Cohen's *d* of 2.137. These observations suggest that visual information mismatch in brand-related posts can be identified by their content characteristics, especially visual properties.

Improving Classification Performance by Building a Brand-Specific Classification Model

Research question 3 asked whether brand-specific classification models would perform better in mismatch detection. Our data set comprised a random sample of Instagram posts of the most popular fashion brands with price range varying significantly. We expected a detection model using a specific brand's data set to identify visual information mismatch better because items from a particular brand may share common characteristics distinct from those of other brands (Fionda and Moore 2009). To investigate how much the model can improve from being trained for a specific brand compared to being trained on the entire set of heterogeneous brands, we trained a random forest classifier separately using the data set of each of the following luxury brands: Cartier, Hermès, and Chanel. We focused on these brands for their brand awareness in general[1] and popularity on Instagram.[2]

The brand-specific model employed the labeled data for each brand (specifically, 677, 529, and 689 labeled posts for Cartier, Hermès, and Chanel, respectively) to have train and test sets using the 80:20 split. On the other hand, the general model

Table 2. Classification results for the three selected brands from the brand-specific and general models.

	Brand-Specific Model		General Model	
Features	Accuracy	AUC ROC	Accuracy	AUC ROC
Cartier (I + M + T)	0.822	0.815	0.664	0.770
Chanel (I + M + T)	0.804	0.880	0.551	0.689
Hermès (I + M + T)	0.755	0.849	0.685	0.727

Note. AUC ROC = area under the curve of the receiver operating characteristic; I = image; M = meta; T = text.

used 6,213 labeled records containing 63 brands for training and then was tested on each of the three brands' test data set. Thus, we compared brand-specific and general models' performance using the same test set, but each model was trained differently.

Table 2 shows the results of evaluating the brand-specific and general models based on accuracy and AUC ROC. The brand-specific model showed a better performance than the general model across all three brands. While the latter showed an average accuracy of 60%, the former achieved an average accuracy of over 80%.

Table 3 demonstrates how the two models predicted the TRUE-labeled instances, which correspond to visual information mismatch. The horizontal axis presents the brand-specific model's decisions, and the vertical axis indicates the predictions of the general model. The predictions were made on 135, 138, and 106 posts for Cartier, Chanel, and Hermès. Results show that the two models made a different prediction for 58 (41%), 54 (39.1%), and 86 (81.2%) posts for the three brands, respectively. Moreover, the brand-specific model was two to five times more accurate than the general model. These results suggest that classification model performance can be increased by separately training a machine-learning model using brand-specific labeled data, which offers a practical application for fashion brands aiming to detect and filter visually mismatched posts.

Discussion

Visual social media have several advantages as a brand communication channel: They attract potential and existing customers and encourage them to connect with brands and other consumers. They also provide a pool of images of products or services embedded in everyday settings. Advertisers and brands can better understand their consumers by analyzing brand-related images and texts, brand community networks, and brand engagement metrics in brand-generated content. Despite these advantages, visual information mismatch in brand-related posts creates contextual incongruence, which violates the consumer-driven culture of searching for brand information and connecting with other consumers. It also challenges advertisers' efforts to use social media data to generate consumer insights.

To mitigate these concerns, we proposed a novel method to automatically detect visual information mismatch by combining features extracted from images, texts, and metadata using computer vision and machine learning. Our model showed that the combination of image and text features achieved the highest accuracy. To highlight the importance of the combined usage, we present four examples of Instagram posts with different prediction scores from image and text classifiers in Figure 5.[3] While the two scores were similar in the first and fourth posts, the image- and text-based classifiers made different predictions in the second and third examples. The second post had a high image-classifier score due to the image irrelevant to fashion, and the third post had a high text-classifier score due to the length of hashtags. This illustration reveals the problem of using a single feature when detecting mismatch in multimodal data and emphasizes the importance of using combined multimodal features for better model performance in detecting mismatch for visual social media content.

We also found that a brand-specific detection model performs better than a general model trained on the entire data set comprising diverse fashion brands. The split-and-train approach improved up to five times for the three luxury brands compared to the general detection model. This finding provides implications for advertisers who want to identify and filter out visual information mismatch in the brand's social media account.

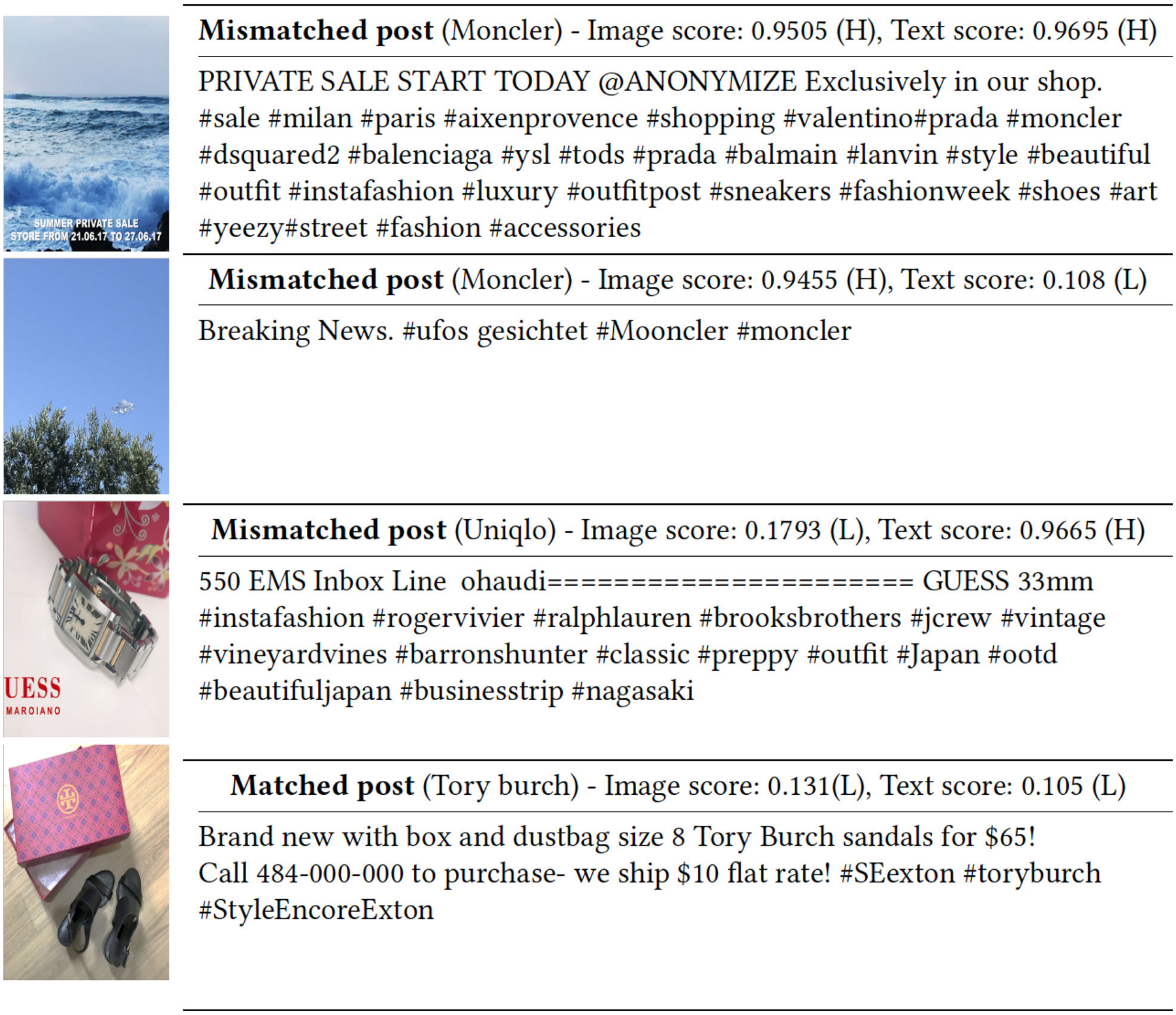
Research Implications

In addition to this research's methodological contributions, our approach to detecting visual information mismatch can help expand research on contextual congruence and brand identity. First, a previous study on contextual congruence looked at the similarity between ad content and editorial content (Moore, Stammerjohan, and Coulter 2005). With more consumers generating brand-related content, how brand-related UGC fits into brand-generated content or aligns with the flow of the visual social media feed is a new research stream where the concept of contextual congruence can be further developed. Our proposed method can help researchers understand which multimodal features become the basis of contextual

Table 3. Classification results for brand-specific and general models based on labeled data for each brand.

		Brand-Specific Model					
		Cartier		Chanel		Hermès	
		False	True	False	True	False	True
General model	False	7 posts (5.2%)	41 posts (30.4%)	17 posts (12.3%)	45 posts (32.6%)	3 posts (2.8%)	63 posts (59.4%)
	True	17 posts (12.6%)	70 posts (51.8%)	9 posts (6.5%)	67 posts (48.6%)	23 posts (21.8%)	17 posts (16%)

Note. The successful classification rates of all brand-specific models (upper-right corner of each quadrant box) are two times higher than those of the general models (lower-left corner of each quadrant box).



Mismatched post (Moncler) - Image score: 0.9505 (H), Text score: 0.9695 (H)

PRIVATE SALE START TODAY @ANONYMIZE Exclusively in our shop. #sale #milan #paris #aixenprovence #shopping #valentino#prada #moncler #dsquared2 #balenciaga #ysl #tods #prada #balmain #lanvin #style #beautiful #outfit #instafashion #luxury #outfitpost #sneakers #fashionweek #shoes #art #yeezy#street #fashion #accessories

Mismatched post (Moncler) - Image score: 0.9455 (H), Text score: 0.108 (L)

Breaking News. #ufos gesichtet #Mooncler #moncler

Mismatched post (Uniqlo) - Image score: 0.1793 (L), Text score: 0.9665 (H)

550 EMS Inbox Line ohaudi===================== GUESS 33mm #instafashion #rogervivier #ralphlauren #brooksbrothers #jcrew #vintage #vineyardvines #barronshunter #classic #preppy #outfit #Japan #ootd #beautifuljapan #businesstrip #nagasaki

Matched post (Tory burch) - Image score: 0.131(L), Text score: 0.105 (L)

Brand new with box and dustbag size 8 Tory Burch sandals for $65! Call 484-000-000 to purchase- we ship $10 flat rate! #SEexton #toryburch #StyleEncoreExton

Figure 5. Examples of mismatch prediction using two classifiers (H = high; L = low).

congruence and which of them lead to higher brand engagement and favorable brand attitudes. We can also focus on how visual identities that are implicitly expressed (e.g., color, shape, other objects that implicitly signify certain brands) can play a role in studying visual information match or mismatch. Our current analytic approach considers such implicit cues as generic image features. Still, future research can incorporate them as semantic features, especially in developing brand-specific mismatch detection models.

Limitations and Suggestions for Future Research

This study bears several limitations, which become the basis for future research. First, the target domain was limited to fashion to control the variation in content patterns. Future research can explore how our method can be applied to other brand categories that are also popular on Instagram (e.g., entertainment). Second, while the computer vision approach can overcome manual labeling limitations, it still requires substantial labels to train a machine-learning model. An AI-driven system that automatically identifies potential categories based on unsupervised learning can benefit future studies to reduce the cost. Third, this study examined whether an image contains a brand's logo or items for deciding visual information mismatch. Still, an image could be seen as relevant to a brand in a broader sense when it implicitly connects to brand identity (without showing a logo). Future studies could investigate different kinds of mismatch, including brand-relevant images shared with brand-irrelevant hashtags. Besides overcoming these limitations, future research needs to examine the characteristics of social media users who create visual information mismatch, specifically their age, gender, income, and brand engagement behaviors (Turunen 2015). This will help to identify which consumer groups are more likely to create visually mismatched content. From a methodological standpoint, the next step is to develop mismatch detection for video content. As a sequence of images and audio represents video clips, computer vision and speech recognition technology can help analyze video content on social media.

Conclusion

This study proposed a novel approach that adopts machine learning and computer vision to detect visual information mismatch. As more and more interactions between brands and consumers occur in social media, having the right tool to detect noise in the data will help advertisers analyze multimodal data to generate consumer insights with higher accuracy and efficiency. As recent articles in the *Journal of Advertising* called for more advertising research that uses real social media data (Voorveld 2019) with the help of AI (Li 2019), this study demonstrates one of the ways in which advertising scholars and practitioners can apply computer vision to visual social media data as an analytic tool for brand communication.

Notes

1. https://luxe.digital/business/digital-luxury-ranking/most-popular-luxury-brands/.
2. Each brand-indicative hashtag has been hit more than 50 million times.
3. We replaced all images used in the original posts with those taken by one of the authors due to copyright protection.

Funding

This study was funded by the Institute for Basic Science (IBS-R029-C2); the National Research Foundation of Korea (NRF-2017R1E1A1A01076400); and the National Science Foundation (SBE-SMA #1831848).

ORCID

Kunwoo Park http://orcid.org/0000-0003-2913-9711
Su Jung Kim http://orcid.org/0000-0003-2025-4019
Jungseock Joo http://orcid.org/0000-0002-4707-8919
Meeyoung Cha http://orcid.org/0000-0003-4085-9648

References

Alhabash, S., J. Mundel, and S. A. Hussain. 2017. Social media advertising: Unraveling the mystery box. In *Digital advertising: Theory and research*, eds. Shelly Rogers and Esther Thorson, 285–99. New York, NY: Routledge.

Berger, J., A. Humphreys, S. Ludwig, W. W. Moe, O. Netzer, and D. A. Schweidel. 2020. Uniting the tribes: Using text for marketing insight. *Journal of Marketing* 84 (1):1–25.

Daer, A. R., R. Hoffman, and S. Goodman. 2014. Rhetorical functions of hashtag forms across social media applications. In Proceedings of the 32nd ACM International Conference on The Design of Communication CD-ROM, Colorado Springs, CO: Association for Computing Machinery, Article 16.

de Winter, J. C. F., M. Kyriakidis, D. Dodou, and R. Happee. 2015. Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. *Procedia Manufacturing* 3:2518–25.

Fionda, A. M., and C. M. Moore. 2009. The anatomy of the luxury fashion brand. *Journal of Brand Management* 16 (5–6):347–63.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep learning*. Cambridge, MA: MIT Press.

Ibba, S., M. Orrù, F. E. Pani, and S. Porru. 2015. Hashtag of Instagram: From folksonomy to complex network. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 279–84. Lisbon, Portugal: SCITEPRESS - Science and Technology Publications, LDA.

Kim, T., and J. Phua. 2020. Effects of brand name versus empowerment advertising campaign hashtags in branded Instagram posts of luxury versus mass-market brands. *Journal of Interactive Advertising* 20 (2):95–16.

King, K. W., L. N. Reid, and W. Macias. 2004. Selecting media for national advertising revisited: Criteria of importance to large-company advertising managers. *Journal of Current Issues and Research in Advertising* 26 (1):59–67.

Krauss, M. J., R. A. Grucza, L. J. Bierut, and P. A. Cavazos-Rehg. 2017. "Get drunk. Smoke weed. Have fun.": A content analysis of tweets about marijuana and alcohol. *American Journal of Health Promotion: AJHP* 31 (3): 200–8. doi:10.4278/ajhp.150205-QUAL-708

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 1097–1105. New York: ACM.

Li, H. 2019. Special section introduction: Artificial intelligence and advertising. *Journal of Advertising* 48 (4): 333–37.

Liu, L., D. Dzyabura, and N. Mizik. 2020. Visual listening: Extracting brand image portrayed on social media. *Marketing Science* 39 (4): 1226.

Moore, R. S., C. A. Stammerjohan, and R. A. Coulter. 2005. Banner advertiser - Web site context congruity and color effects on attention and attitudes. *Journal of Advertising* 34 (2):71–84. doi:10.1080/00913367.2005.10639189

Nanne, A. J., M. L. Antheunis, C. G. van der Lee, E. O. Postma, S. Wubben, and G. van Noort. 2020. The use of computer vision to analyze brand-related user generated image content. *Journal of Interactive Marketing* 50:156–67.

Nuñez, M. 2020. Instagram is reaching upper limit of U.S. user growth, so expect even more ways to Insta-shop. eMarketer, January 6.

Pentina, I., V. Guilloux, and A. C. Micu. 2018. Exploring social media engagement behaviors in the context of luxury brands. *Journal of Advertising* 47 (1):55–69.

Phua, J., S. V. Jin, and J. Kim. 2017. Gratifications of using Facebook, Twitter, Instagram, or Snapchat to follow brands: The moderating effect of social comparison, trust, tie strength, and network homophily on brand identification, brand engagement, brand commitment, and membership intention. *Telematics and Informatics* 34 (1):412–24.

Rajaraman, A., and J. D. Ullman. 2011. *Mining of massive datasets*, New York, NY: Cambridge University Press.

Rietveld, R., W. van Dolen, M. Mazloom, and M. Worring. 2020. What you feel, is what you like: Influence of message appeals on customer engagement on Instagram. *Journal of Interactive Marketing* 49:20–53.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* 115 (3):211–52.

Russell, S. J., and P. Norvig. 1995. *Artificial intelligence: A modern approach*. Englewood Cliffs, NJ: Prentice Hall.

Sheehan, K. B. 2018. Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs* 85 (1):140–56. doi:10.1080/03637751.2017.1342043

Sheldon, P., and K. Bryant. 2016. Instagram: Motives for its use and relationship to narcissism and contextual age. *Computers in Human Behavior* 58:89–97.

Sung, Y., E. Kim, and S. M. Choi. 2018. #Me and brands: Understanding brand-selfie posters on social media. *International Journal of Advertising* 37 (1):14–28.

Tous, R., M. Gomez, J. Poveda, L. Cruz, O. Wust, M. Makni, and E. Ayguadé. 2018. Automated curation of brand-related social media images with deep learning. *Multimedia Tools and Applications* 77 (20):27123–42.

Turunen, L. L. M. 2015. Challenging the hierarchical categorization of luxury fasion brands. *Nordic Journal of Business* 64 (2):119–38.

Vassey, J., C. Metayer, C. J. Kennedy, and T. P. Whitehead. 2020. #Vape: Measuring e-cigarette influence on Instagram with deep learning and text analysis. *Frontiers in Communication* 4:75.

Vidanagama, D. U., T. P. Silva, and A. S. Karunananda. 2020. Deceptive consumer review detection: a survey. *Artificial Intelligence Review* 53 (2):1323–52.

Voorveld, H. A. M. 2019. Brand communication in social media: A research agenda. *Journal of Advertising* 48 (1): 14–26.

Wang, T. 2015. *The value of luxury brand names in the fashion industry*. Claremont, CA: Claremont McKenna College.

Yoon, S., K. Park, J. Shin, H. Lim, S. Won, M. Cha, and K. Jung. 2019. Detecting incongruity between news headline and body text via a deep hierarchical encoder. *Proceedings of the AAAI Conference on Artificial Intelligence* 33:791–800.