Extraction of visual information to predict crowdfunding success

Simon J. Blanchard1 | Theodore J. Noseworthy2 | Ethan Pancer3 | Maxwell Poole3

1 McDonough School of Business, Georgetown University, Washington, DC, USA
2 Schulich School of Business, York University, North York, Ontario, Canada
3 Sobey School of Business, Saint Mary’s University, Halifax, Nova Scotia, Canada

Abstract
Researchers have increasingly turned to crowdfunding platforms to gain insights into entrepreneurial activity and dynamics. While previous studies have explored various factors influencing crowdfunding success, such as technology, communication, and marketing strategies, the role of visual elements that can be automatically extracted from images has received less attention. This is surprising, considering that crowdfunding platforms emphasize the importance of attention-grabbing and high-resolution images, and previous research has shown that image characteristics can significantly impact product evaluations. Indeed, a comprehensive review of empirical articles (n = 202) utilized Kickstarter data, focusing on the incorporation of visual information in their analyses. Our findings reveal that only 29.70% controlled for the number of images, and less than 12% considered any image details. In this manuscript, we contribute to the existing literature by emphasizing the significance of visual characteristics as essential variables in empirical investigations of crowdfunding success. We review the literature on image processing and its relevance to the business domain, highlighting two types of visual variables: visual counts (number of pictures and number of videos) and image details. Building upon previous work that discussed the role of color, composition, and figure–ground relationships, we introduce visual scene elements that have not yet been explored in crowdfunding, including the number of faces, the number of concepts depicted, and the ease of identifying those concepts. To demonstrate the predictive value of visual counts and image details, we analyze Kickstarter data using flexible machine learning models (Lasso, Ridge, Bayesian additive regression trees, and eXtreme Gradient Boosting). Our results highlight that visual count features are two of the top three predictors of success and highlight the ease at which researchers can incorporate some information about visual information. Our results also show that simple image detail features such as color matter a lot, and our proposed measures of visual scene elements can also be useful. By supplementing our article with R and Python codes that help authors extract image details (https://osf.io/ujnzp/), we hope to stimulate scholars in various disciplines to consider visual information data in their empirical research and enhance the impact of visual cues on crowdfunding success.

1 INTRODUCTION
Crowdfunding allows entrepreneurs to request funding from many investors simultaneously, usually in exchange for future products or equity stakes (Mollick, 2014; Mollick & Nanda, 2016). Recent legislative changes also give consumers and firms new ways to interact along these lines. Since 2016, when the US Securities and Exchange Commission allowed investors of various net worths to participate in crowdfunding, there has been a substantial increase in the amount of money raised on crowdfunding platforms; in 2020, it reached $500 million in the United States alone, according to the Crowdfunding Center.1 Dozens of different crowdfunding platforms support rewards-based investing (e.g., Kickstarter, Indiegogo) and equity-based investing (e.g., SeedInvest, StartEngine), and others specialize in individual help (e.g.,
GoFundMe) and non-profit funding (e.g., Mightycause). The number of international crowdfunding platforms also has grown, with platforms covering all regions of the world (e.g., JD Crowdfunding and Jingdong in China, Babyloan in France, Crowdprop in South Africa).

Given the growth of crowdfunding as an alternative funding source in recent years, researchers have increasingly turned to datasets from various crowdfunding platforms to gain insights into entrepreneurial activity and the dynamics of crowdfunding. Some operations management scholars have leveraged crowdfunding data to substantiate the role of new technologies in selecting and disseminating new ideas (Kornish & Hutchinson-Krupat, 2017), while others have sought to identify predictors of crowdfunding success (e.g., Z. Cui et al., 2019; Y. M. Wei et al., 2022), such as backers sought to identify predictors of crowdfunding success (e.g., Kuppuswamy & Bayus, 2018), referral timing (Burtsch et al., 2021), founder updates (Mejia et al., 2019), and frequency of communication (Xiao et al., 2021). Marketing scholars have typically focused on predictors of successful communication within project pages, such as information in videos (Li et al., 2019), emotions in text (Peng et al., 2022), and price advertising claims (Blaseg et al., 2020).

Recently, more attention is being focused on the role of visual elements in predicting crowdfunding success. This interest is likely due to three factors. First, crowdfunding platforms themselves provide guidance on how to develop project images that are most likely to attract backers. For instance, Kickstarter recommends that main project images be attention-grabbing, straightforward, and of high resolution, while Indiegogo stresses that image style and content are critical factors for success. This guidance indicates that platforms recognize the significant role played by images in a project’s success. Second, previous research has established that image characteristics such as brightness and contrast, as well as image compositions, can influence the persuasiveness of new product depictions. For example, studies by Noseworthy and Trudel (2011) and Pieters et al. (2010) have found that altering image characteristics can significantly impact product evaluations. Third, empirical researchers have increasingly been using vision application programming interfaces (API) and algorithms to augment their datasets and investigate the role of visual figures on company success (e.g., M. Zhang & Luo, 2023, in the context of Yelp; S. Zhang et al., 2022, in the context of AirBnB).

But to what extent do empirical researchers interested in crowdfunding utilize visual information in their analyses? As Kickstarter datasets are among the most frequently used sources of crowdfunding data due to their availability from sources like webrobots and Kaggle, and they serve as the foundation for our empirical study, we compiled a list of 399 empirical papers in operations, marketing, and management that utilized Kickstarter data. Our aim was to quantify the extent to which these articles incorporated visual information in their analyses, focusing on two types: visual counts (number of images and number of videos) and image details. After conducting a thorough manual review of the papers to include empirical analyses relevant to Kickstarter, we identified 202 articles for which we could assess whether variables included counts of images (either 0 or 1+ or a count), counts of videos (either 0 or 1+ or a count), and any image detail (e.g., colorfulness, contrast). The results are presented in Table 1.

Our findings indicate that approximately 46% of the empirical articles included a control variable to account for the presence of videos. However, only 29.70% of the articles controlled for the number of images, and less than 12% incorporated any image details in their analyses. Notably, we observed significant differences between subfields regarding the inclusion of visual characteristics. While management, marketing, and operations articles were roughly equally likely to have a variable capturing the presence of one (or more) videos, marketing articles were nearly twice as likely to include a variable for the count of images and were also more likely to consider image details. These highlight an opportunity to investigate potential important predictors insufficiently considered.

In the present article, our aim is to contribute to the existing literature on image processing and crowdfunding by emphasizing the significance of visual characteristics as essential variables in empirical investigations of the factors that contribute to crowdfunding success. To accomplish this, we undertake the following approach. First, we comprehensively review the literature on image processing and its relevance to the business domain. This review highlights the importance of two types of visual variables that should not be overlooked: visual counts and image details. Second, we expand upon the work of M. Zhang and Luo (2023) by incorporating the effect of scene elements that have not yet been explored in the context of crowdfunding. These scene elements include variables such as the number of faces present, the number of concepts depicted, and the ease of identifying those concepts. Finally, we present an analysis of Kickstarter data to demonstrate the predictive value of visual counts and image details. We employ nested variable sets across flexible models, including Lasso, Ridge, Bayesian additive regression trees (BART), and eXtreme Gradient Boosted Trees (XGBoost). We find consistent improvement when adding visual variables, particularly visual count variables, significantly enhancing both in-sample and out-of-sample root mean squared errors (RMSE) and mean absolute errors (MAE) metrics. Notably, these visual counts influence predictions nearly as much as the entire Linguistic Inquiry
and Word Count (LIWC) dictionary for text. We also find that BART and XGBoost, as implemented, outperform Ridge and Lasso, suggesting better accommodation of non-linearities using untransformed data. While image details slightly improve out-of-sample performance over baseline variables, their predictive value is less than that of visual count variables.

Our work makes two main contributions. First, we have observed that, despite the increasing empirical analyses of crowdfunding data, there is still a lack of attention given to image-related features. By emphasizing the importance and relative ease of controlling for visual counts and image details, we hope to encourage empirical researchers to include visual counts and image details in their studies of crowdfunding as they are important predictors of crowdfunding success. By supplementing our work with specific implementation and measures, we seek to stimulate crowdfunding, marketing, operations, and entrepreneurship scholars to expand their consideration of visual information data in their empirical research.

Second, much recent research has focused on automated feature creation can be used to take advantage of multimedia formats such as video and images (Li et al., 2019; M. Zhang & Luo, 2023; S. Zhang et al., 2022). Most of the literature on automated feature extraction has been focused on visual processing characteristics (e.g., color, composition, and figure–ground) that are related more to execution than meaning. In addition to introducing several new such features (blur, number of segments, image quality/noise), we highlight the usefulness and ease of using APIs to automatically extract visual-scene elements such as the number of faces, the number of evoked concepts, and the ease of concept identification. Our research box contains the list of 399 articles we surveyed, the Python codes for our visual feature extractions, and the final dataset for analyses, and analysis codes in R are available at https://osf.io/uijnzp/. We also offer a detailed guide at https://bit.ly/uijnzp. By sharing these files, we aim to facilitate the understanding of the impact of visual cues on crowdfunding success and across disciplines.

2 A TAXONOMY OF VISUAL INFORMATION

A substantial body of literature suggests that specific image characteristics can have a significant impact on human judgment and decision-making processes across various contexts. Previous research on visual information has predominantly focused on two aspects: accounting for visual count variables, such as the number of pictures and videos, and examining the photographic properties of an image that are relevant to perception and interpretation. These properties include color features, composition, and figure–ground (M. Zhang & Luo, 2023; S. Zhang et al., 2022). In this section, we provide a brief summary of these image characteristics and explain why we anticipate their influence on the evaluation of new products in the context of crowdfunding. Additionally, we discuss the potential utility of considering visual scene elements in crowdfunding analysis.

2.1 Visual count variables: Number of pictures and number of videos

According to Kickstarter’s (2022) Creator Handbook, images and videos play a crucial role in engaging potential backers and conveying the essence of a project. They serve as the initial points of contact, capturing backers’ attention and making a favorable first impression. Scholarly research has recognized the significance of visual information in empirical studies (Blaseg et al., 2020; Li et al., 2019; Mejia et al., 2019). Mollick (2014) argues that incorporating videos indicates better preparation and higher project quality, which, in turn, leads to greater crowdfunding success (see also Kuppuswamy & Bayus, 2018; Tian et al., 2021; X. Wei et al., 2021). The number of images featured in projects also has a positive influence on backer contributions (X. Wei et al., 2021; Xiao et al., 2021). Project images offer a depth of communication that surpasses textual information, such as showcasing prototypes, and enhance perceptions of campaign informativeness, creator commitment, and credibility (Xiao et al., 2021). Mollick and Nanda (2016) suggest that crowds, as opposed to experts, prioritize visual content, making projects with more images and videos more appealing to the crowd. Moreover, the literature on information processing suggests that increasing the number of informational cues moderately enhances persuasion (e.g., Brown & Carpenter, 2000). Therefore, it is crucial to include variables that capture visual information, such as the number of pictures and videos. However, our findings indicate that the majority of projects do not incorporate visual counts in their analyses.

2.2 Image details

2.2.1 Color features, composition, and figure–ground relationship

In addition to considering the numbers of pictures and videos in crowdfunding projects, it is important to examine the characteristics of the images themselves as they can impact decision-making. The field of color psychology has provided scientific insights into how perceptions of color influence attitudes and choices, with color carrying meaning and influencing affect, cognition, and behavior in consumer contexts (Elliott & Maier, 2014). Practitioners in marketing, advertising, and graphic design have long recognized that color and composition play a role in influencing consumer behavior through factors such as atmospherics (Gorn et al., 2004), brand identity (Labrecque & Milne, 2013), and product evaluations (Miller & Kahn, 2005). Common applications of computer vision algorithms have made the automatic extraction of color features such as brightness (amount of light reflected), saturation (color intensity), colorfulness
Table 2: Image details.

| Color features                        |
|---------------------------------------|
| Brightness: the average value of the V channel in the HSV (hue, saturation, value) color space, normalized to [0,1] |
| Saturation: the average value of the S channel in the HSV color space, normalized to [0,1] |
| Colorfulness: a metric for how colorful the image is, computed as the sum of the standard deviation of the R-G (red-green) and B-Y (blue-yellow) color channels, normalized to [0,1] |
| Contrast: the standard deviation of the V channel in the HSV color space |
| Warm Hue: a metric for how much of the image’s color is in the range of red, orange, and yellow, computed as the proportion of pixels whose H value in the HSV color space is within the range |
| Clarity: a metric for how clear or sharp the image is, computed as the proportion of pixels whose V value in the HSV color space is greater than or equal to 0.7 |
| Blur: degree to which shapes of surfaces are difficult to perceive, calculated as the variance of Laplacian of the gray-scaled image |
| Image quality: predicted human judgments based on the pre-trained called LIBSVM by Mittal et al. (2012) |

Composition

Diagonal dominance: the average distance of a salient region’s points to the two diagonals of an image and taking the minimum of those distances, which is then subtracted from zero.

Rule of third: Using saliency maps from the StaticSaliencyFineGrained (OpenCV), adherence to the rule of third is calculated as the proximity of the centroid of the salient region and the nearest intersection in a 3 × 3 grid

Vertical physical balance: Proximity between the center of the image and the weighted center of the salient regions in the vertical axis based on saliency maps obtained from OpenCV

Horizontal physical balance: Proximity between the center of the image and the weighted center of the salient regions in the horizontal axis based on saliency maps obtained from OpenCV

Vertical color balance: Vertical color balance is a measure of how evenly the colors are distributed from top to bottom in an image, with values ranging from −1 (completely unbalanced) to 0 (perfectly balanced). This is computed using the superpixel pixel algorithm (Achanta et al., 2012)

Horizontal color balance: Horizontal color balance is a measure of how evenly the colors are distributed from left to right in an image, with values ranging from −1 (completely unbalanced) to 0 (perfectly balanced). This is also computed using the superpixel algorithm (Achanta et al., 2012)

Number of segments: Count of number of partitions (image segments) in an image, obtained using Otsu’s binarization to identify the threshold that minimizes weighted within-class variance on a grayscale image, and count of number of edge changes (Gonzalez & Woods, 2002)

Figure–ground relationship

Size difference: Using a GrabCut algorithm (Rother et al., 2004), the image is segmented in foreground and background. The size difference is the difference in area size

Color difference: The Euclidian distance between the mean color values of the foreground and background regions provides the color difference

Texture difference: Using Canny edge detection, texture difference is the absolute difference between the proportion of edge pixels between the foreground and background regions

Visual scene characteristics

Number of faces: the number of detectable front upright faces based on the pre-trained Viola–Jones algorithm of Pech-Pacheco et al. (2020)

Number of evoked concepts: image-level annotations identified as having 50% confidence from the Open Images Dataset using the Google Vision API (Krasin et al., 2017) and Microsoft Azure

Ease of concept identification: classification confidence of the most confidently identified annotation detected from the Open Images Dataset using the Google Vision API (Krasin et al., 2017) and Microsoft Azure

Note: Italics identified variables operationalized as in M. Zhang and Luo (2023). Underlines refer to variables that we added.

(chromaticity), contrast (variation in luminance), warmth hue (degree of red, orange, and yellow), and clarity (sharpness) relatively easy (see M. Zhang & Luo, 2023). Most color feature characteristics can be calculated directly from pixel values using pre-determined rules and transformations (see Table 2). Others, like image quality (absence of noise), is usually done by trained classifiers (Mittal et al., 2012).

Image composition refers to the arrangement of visual elements within an image. Similar to color features, there are various ways to automatically extract variables that characterize image composition. M. Zhang and Luo (2023) focused on several composition characteristics, including diagonal dominance (average distance of salient points to the diagonal), the use of the rule of thirds (placement of elements at intersecting lines dividing the image into equal ninths), physical balance (distribution of elements across horizontal and vertical lines), and color balance (even distribution of colors across the image). Other composition characteristics include blur (difficulty in perceiving shapes) and the number of image segments. Algorithms such as OpenCV (a popular, free, open-source computer vision algorithm) and the superpixel method developed by Anchanta et al. (2010) can be used to extract these composition variables by identifying salient regions and dividing the image into a grid.
Similarly, pre-trained algorithms can be employed to identify figure–ground relationships, which involve distinguishing the main object from its surroundings. Common figure–ground relationship variables include size difference (difference in area size between foreground and background), color difference (contrast between the foreground and background colors), and texture (variations in edge pixels between the foreground and background). Automatic extraction of figure–ground relationship variables typically involves using a classifier to separate the foreground from the background (e.g., GrabCut by Rother et al., 2004) before calculating the differences. Each of these image characteristics, including color features, composition, and figure–ground relationships, provides information about the quality and content of the images. They help identify blurry images or those of lower quality. However, other types of distortions can also impact image quality, such as noise, compression artifacts, transmission errors, and intensity shifts (Ponomarenko et al., 2009). Computer scientists have developed algorithms that approximate human judgments of overall image quality and noise perception, providing a more holistic assessment (Mittal et al., 2012).

### 2.2.2 Visual scene characteristics

Even when considering the informational value of the number of pictures or videos in projects and their specific image characteristics such as color features, composition, and figure–ground relationships, we believe that there is further valuable information embedded within the images themselves. This additional information is likely to be relevant and can provide insights for predicting crowdfunding success. Specifically, we expand the taxonomy proposed in past research to add an additional category, visual scene characteristics, and we propose that capturing information about certain image attributes, such as the presence of faces, visual complexity, the number of identifiable concepts, and the ease of associating image content with existing schema, can contribute to our understanding of crowdfunding outcomes. To support our proposition, we draw upon the literature on visual scene processing, which sheds light on the significance of these image attributes.

#### Number of faces

The effect of facial images on consumer attention and behavior has been extensively studied in the advertising literature. Faces have been found to rapidly convey information and hold sociobiological significance (Dekowska et al., 2008; Hershler & Hochstein, 2005). For instance, Guido et al. (2019) discovered that in print advertising, the presence of facial stimuli can direct consumer attention to images and make communications more noticeable in a cluttered environment. Consequently, images that include faces are more likely to attract attention. This hypothesis has found some support in research showing that the number of faces featured in GoFundMe campaigns is associated with the number of backers and the amount of money raised (Rhue & Robert, 2018). However, it is important to note that reflexive attentional responses to facial stimuli might not necessarily translate into behavioral responses, such as sponsoring crowdfunding projects (Guido et al., 2019). X. Zhang et al. (2021) reported mixed findings regarding the impact of facial images on project funding, with results varying based on the type of project being featured. It appears that while facial images may increase attention, this alone may not be sufficient to drive crowdfunding success.

#### Number of evoked concepts

Scholarly work in imagery examines how consumers interact with aesthetic elements to frame category inferences, which often occurs independently of their examination of the surrounding contexts, though category inferences also may be contingent on how people comprehend the scenes and scenarios in which they first interact with novel objects (Buetti et al., 2016; Torralba & Olivia, 1999). For example, in a scenario in which a tech-savvy entrepreneur develops an innovative new home juicer, representations might manifest as distinct superordinate categories such as “kitchen appliance,” “human,” and “fruit.” Such categories offer only sparse information about what makes the juicer new or innovative. The categories offer structural regularity around a somewhat vague theme, but the theme is not random (Buetti et al., 2016; Torralba & Olivia, 1999); it relates some items in some meaningful ways to a target while excluding unrelated items (Buetti et al., 2016; Wolfe et al., 2003). A theme that incorporates the juicer may be a kitchen, but only if the objects in the scene are arranged to support the housing, preparation, and consumption of food. Processing these relationships is of paramount importance for planning and predicting suitable actions. We expect that multiple category cues influence consumers’ inferences about products, and the more distinctly that images evoke concepts (i.e., concepts that come to mind), the more they evoke functional and spatial relationships that help reduce ambiguity about what products are and do.

#### Ease of concept identification

By definition, “innovative products defy straightforward categorization” (Moreau et al., 2001, p. 496). As a result, it is not only the number of evoked categories that matters but also the ease with which the categories come to mind. Blanchard et al. (2012) find a strong association between the salience of information and the speed at which people recall it when making judgments; Moreau et al. (2001) find that the first knowledge structure that comes to mind is particularly important to a person’s judgment of new products. Accordingly, we expect that the ease with which the most salient concept is identified helps make the elements of visual scenes in images displayed on crowdfunding platforms less ambiguous.

### 3 Empirical setting and data

Kickstarter, a popular crowdfunding platform, has played a significant role in supporting projects, with over $5 billion in
pledges from 18 million backers and funding for more than 190,000 projects (Kickstarter, 2020). The platform operates under an all-or-nothing funding model, where backers are only charged if a project reaches its funding goal. Kickstarter has specific rules for technology projects, requiring them to showcase a non-rendered visual prototype to provide backers with a clear understanding of the project’s development stages. These prototypes are displayed as title images on the browsing page. Photorealistic renderings such as technical drawings, computer-assisted drawings models, or sketches are prohibited as they can potentially bias perceptions of the final product’s appearance. The Kickstarter project database is regularly scraped and made publicly available by weRobots (2017).

In our research, we focus on the role of visual elements in informing potential backers about the functionality of new products. To narrow down our analysis, we specifically examine projects listed in the “technology” section of Kickstarter, excluding offerings classified in other functional categories such as comics and illustrations, food, music, games, and publishing. Our dataset comprises 16,439 projects, which includes both successful (4187) and failed (12,252) projects for which we have complete data. Each project entry includes essential information such as project ID, funding goal, amount pledged, number of backers, whether it was featured as a staff pick, project launch date, and project deadline. Additionally, we collect the project’s title, a brief text description, and a link to the project’s title image visible in search results. To enrich our data, we augment the weRobots data by scraping full-text extractions of the project descriptions.

3.1 Baseline variables

As a baseline variable, we considered external factors that could influence the success of a project. One such factor is the potential feature of projects on Kickstarter’s highly visible sections called “staff picks” or “projects we love,” which have been associated with increased chances of success (Pisani, 2014). We also considered the geographic location of the project creator, as studies have shown that backers may have a higher willingness to pledge to projects based on the creator’s location (Mollick, 2014). To account for seasonality and temporal effects, we included two control variables related to the project’s launch date: the launch day of the year and the year of the project (ranging from 2009 to 2017). Additionally, we incorporated the duration of the project’s funding window, measured in the number of days the project could accept funding. While Kickstarter initially allowed projects to be open for 90 days, they now encourage shorter durations, typically 30 or 60 days. Several project content characteristics can influence project success. For example, projects with smaller funding goals are more likely to be backed and to achieve their objectives (Mollick, 2014). Accordingly, we controlled for funding goal size. Although many project creators were located outside the United States and used other currencies, the data scraper automatically converted the currencies using the exchange rate to USD at the time of scraping.

3.2 Text variables

To assess the impact of text descriptions on potential backers, we analyzed both the short text description (blurb) and the full-text description of the projects. For the full-text analysis, we utilized the LIWC dictionaries. Each dictionary (e.g., anger) contains words that belong to different psychological, emotional, and linguistic dimensions (e.g., hate, mad, angry, frustr*). To use LIWC on a text, the text is compared with each dictionary, and the percentage of words that match each category is calculated. We used LIWC-22 to create our 117 variables (see Boyd et al., 2022, for the list of dictionaries). The LIWC-22 has been extensively validated using diverse sources of publicly available texts, such as social media posts, reviews, articles, stories, and conversations from platforms like Facebook, Reddit, Twitter, Yelp, and more. Compared to earlier versions like LIWC-2015, the LIWC-22 offers additional variables related to determiners, cognition, states, motives, affect, and social behaviors, making it more comprehensive and suitable for our study. Boyd et al. (2022) found slightly higher correlations among self-reports, judges, and LIWC using the LIWC-22, further validating its effectiveness.

For the analysis of the blurb, we employed six variables. To account for text descriptions that emphasized the project’s innovativeness and superiority, we created two custom dictionaries containing the top 100 synonyms for the words “best” and “innovate.” These variables aimed to capture textual claims of being innovative and superior. Additionally, we included four text summary variables generated by LIWC: analytical language, clout/persuasiveness, authenticity, and emotional tone. These variables have been commonly used in automated text extraction in the crowdfunding literature and have an influence on text persuasiveness (Cohn et al., 2004; Kacewicz et al., 2013; Newman et al., 2003; Pennebaker et al., 2014).

3.3 Visual data

Among the many ways to operationalize our image detail variables as measures, we opted to rely on measures that are both common in practice and readily accessible from common packages such as OpenCV, the Python Imaging Library, or SciPy, with many variables adapted from M. Zhang and Luo (2023). Table 2 lists the measures.

For visual scene variables, we proposed three variables: the number of faces, the number of evoked concepts, and the ease of concept identification. For the number of faces, we use the number of detectable front upright faces based on the pre-trained Viola–Jones algorithm of Pech-Pacheco et al. (2020), which has been validated elsewhere. For the number of evoked concepts and ease of classification, there are...
numerous data sources and pre-trained models for multilabel annotation to identify the concepts that humans who evaluate images are likely to evoke. We first chose to identify elements in images automatically, using Google’s Cloud Vision API, which annotates input images with discrete labels that are then confidence-scored. When annotating images, the algorithm first relies on Google’s Open Images dataset (Krasin et al., 2017), which contains 478,000 crowdsourced images annotated by humans with more than 59 million image-level labels across 19,000 categories. Using TensorFlow (Google’s internal machine learning framework), a deep convolutional neural network (CNN) assigns multiple labels to the evaluated images. We therefore counted the number of labels extracted by the Google Vision API for which classification confidence was at least 50%. We used maximum classification confidence across all identified concepts to measure ease of concept identification. Although there are alternative measures (e.g., average, median), capturing the ease of concept identification with the highest confidence label is consistent with past categorization research, which indicates that the most salient category exemplar presents distinct advantages for both recall (Blanchard et al., 2012) and inferences (Moreau et al., 2001).

We argued the number of annotations provided by commercial APIs is a variable that is relevant to visual scenes because the annotations are likely to reflect concepts that are relevant to product usage. It would be surprising to find entrepreneurs in our Kickstarter data using depictions of concepts that are difficult to identify or have no relation to their innovative products, particularly in settings in which the entrepreneurs are attempting to communicate about their products to potential backers. Even if we were to assume that our arbitrarily chosen classifier could perfectly recover the ground truth labels that would be assessed by humans (which it does not), such visual elements may be unrelated to a product’s usage scenario. Consider SHRU (see Figure 1), an intelligent cat toy offered on Kickstarter as “not another cat toy, it’s a cat companion, design to be your cat’s new best friend.”

The toy looks like a small ball, the cat has whiskers and fur, and the image contains a caption. We seek to provide empirical evidence that our annotations capture aspects of the visual scene relevant to how the product is intended to be used. To do so, we randomly selected a subsample of 1603 projects in our datasets and subjected them to a validation task using Amazon Mechanical Turk (mTurk) workers on Amazon. Specifically, we created a custom classification task in the mTurk API, whereby mTurk workers independently evaluated the labels identified by the Google Vision API for each project. Specifically, participants viewed the project title, product image, and “blurb” (short description). They then read:

A computer program identified the following labels as part of the above image. Some of the labels relate to a broader scene that depicts how the product is meant to be used, and others have nothing to do with the scene being conveyed in the image. Which of the following labels are not related to how the product is meant to be used? Unselect the labels that do not directly relate to what you believe is the gist of the scene being conveyed in the image. [see instructions].

Participants also answered two questions to help remove ambiguity about whether the annotations were relevant to the product: “To what extent do all the labels relate to one another to describe a product usage scenario?” (1, “not at all related to a scenario”; 9, “completely related to a scenario”), and then a nominal “yes” or “no” question about whether the image conveyed how someone would use this product.

The [see instructions] button displayed a layer that provided a sample task with labels accompanied by a training task to illustrate how to select relevant labels. Specifically, it showed the project title, image, and blurb for the SHRU as an example (see the Appendix). We then provided participants with further clarification related to relevance labels:

Here, we have a cat toy. Of all the labels, not all directly relate to how the toy is meant to be used. For instance, cat, ball, play, small to medium-sized cats, kitten would all be obviously relevant to how the designer intended the product to be used (e.g., a ball that cats or kittens can play with). However, carnivore and whiskers (although both about a cat) do not relate to how the product is meant to be used. Likewise, photo caption isn’t helpful in explaining the use of the object. We would expect you to unselect photo caption, carnivore, and whiskers.

The mTurk task was structured such that each project had five workers perform the evaluation. The task was completed using 403 unique workers who evaluated 1603 images...
for total compensation of $1141.80. We classified any label as validated if at least three of the five workers who performed the task agreed that the label was related to the product usage. We assumed the label validation rate would be rather high because it would be surprising to find the concepts unrelated, particularly regarding the most salient object in the scene (i.e., most easily identifiable), and because they appeared in a setting in which entrepreneurs were attempting to communicate a product to potential backers.

As expected, on average, most labels (84%) captured by the Google Vision API were explicitly retained as related to the product once validated by the mTurk workers, with an average reduction of 1.97 labels. The number of original labels had a 0.938 ($p < 0.001$) correlation with the number of validated labels, which reflects the idea that a rational entrepreneur would not contrive an unrelated scene when accurate communication is critical. The average decrease in primary classification confidence was 1.31%, which correlated with the unvalidated primary classification confidence of 0.891 ($p < 0.001$). These label validation results substantiate our assumption that most of the distinct labels cataloged by the visual API consist of objects that relate spatially or categorically to what we described conceptually as the product.

4 | PREDICTIVE VALUE OF VISUAL INFORMATION

To this point, we have documented that visual information is likely to play a role in crowdfunding success, and we have presented a series of concepts: visual counts and image details. In this section, we use our empirical dataset from Kickstarter to illustrate how such information can be automatically extracted and used within machine learning models to show the importance of visual information.

4.1 | Forecasting framework

To assess the importance of visual information contained in images, we compare five sets of variables: (1) baseline (six project characteristics only); (2) baseline + textual details (from blurb and main text, 123 extracted variables); (3) baseline + visual counts (number of images and videos, two variables); (4) baseline + image details; and (5) all variables. We fit the same models for the five sets of variables.

4.2 | Comparative models

Our outcome is (logged) total dollars raised $y_i$ and predictors $x_i = \{x_{i1}^P, x_{i2}^P, ..., x_{iN}^P\}$ for $N$ crowdfunding projects. Our statements regarding visual information are agnostic to the kind of classification model used. We needed to choose a learning model that can accommodate unexpected nonlinearities and variable selection to make sure that we truly capture the predictive value of visual information.

4.2.1 | Regularized regressions (Ridge and Lasso)

First, we considered regularized regression given large predictor spaces and potential for overfitting. Regularized regression methods, such as Ridge and Lasso, are popular techniques to conduct regressions while avoiding overfitting and improving model performance on out-of-sample data. To do so, they add a regularization term to the regression objective function. In both methods, the objective is to minimize the sum of squared errors between the predicted outcome and the actual outcome, subject to a constraint that is a function of the predictor variables.

Ridge regression adds a penalty term to the squared error loss function, which is proportional to the sum of the squared values of the regression coefficients. The loss function minimized in Ridge is given by:

$$\text{Loss}_{ridge} = \sum_{i=1}^{N} L(y_i, \hat{y}_i) + \lambda \sum_{j=1}^{P} \beta_j^2,$$

(1)

where $\lambda$ is the tuning parameter that controls the amount of shrinkage applied to the coefficients, and $L(y_i, \hat{y}_i)$ is the squared error loss. The shrinkage parameter $\lambda$ penalizes the sum of squared coefficients in the model and shrinks them toward zero. The larger the value of $\lambda$, the more the coefficients are shrunk toward zero, resulting in a more biased but less variable model.

Lasso regression, instead of using a shrinkage parameter affect coefficients, imposes a penalty to the objective function of a linear regression based on the $L_1$ norm of the vector of coefficients (i.e., sum of absolute value of the coefficients), which is weighted against the residual sum of squares between the predicted values and observed values using a regularization parameter $\lambda$:

$$\text{Loss}_{lasso} = \sum_{i=1}^{N} L(y_i, \hat{y}_i) + \lambda \sum_{j=1}^{P} |\beta_j|.$$

(2)

This has the advantage of providing a simpler and more interpretable model, by selecting only the most relevant predictor variables and encouraging some to be exactly zero.

4.2.2 | Bayesian additive trees

Bayesian additive regression trees (Chipman et al., 2010) have been shown to provide a flexible approach to make predictions while avoiding strong parametric assumptions (Hill et al., 2020), particularly when cross-validation is used to choose hyperparameters (Dorie et al., 2019). Assuming that $(y_i, x_i)$ are independent and identically distributed (i.i.d.), a
single tree in a BART models the expected value of the continuous outcome as:

\[ E(Y|x) = g(x; T, M), \tag{3} \]

where \( g(x; T, M) \) is the function that follows the rules described by regression tree \( T \), and \( M \) is the set of terminal node predictions for each node. A single tree is built via recursive partitioning, which divides up the \( p \)-dimensional space of the predictors into mutually exclusive areas. Each area is as homogeneous as possible (i.e., contains observations of similar expected value). In addition to building different algorithms to estimate single trees, researchers have worked on ensemble methods that produce forecasts based on multiple decision trees. This includes boosted trees (e.g., AdaBoost) and bootstrap aggregated trees (e.g., random forests). Such ensembles typically outperform single trees because they can reduce overfitting (Hastie et al., 2008).

One notable advancement in ensemble methods for decision trees has been the development of Bayesian additive regression trees (Chipman et al., 2010). Still assuming that \((y_i, x_i)\) are ind, Bayesian additive regression trees model predict the outcome as follows:

\[ E(Y|x) = \Phi \left[ \sum_{j=1}^{m} g(x; T_j, M_j) \right], \tag{4} \]

where \( g(x; T_j, M_j) \) is still the function that allows one to follow the tree-specific regression rules described by the tree \( T_j \), and \( M_j \) is the set of terminal node predictions (i.e., expected outcome values) for each node. For estimation, Chipman et al. (2010) proposed a Bayesian backfitting MCMC algorithm that incorporates a regularization prior to avoiding having a single tree too influential and can outperform many existing machine learning methods on prediction tasks. It has also been used as a predictive benchmark in marketing (Schoenmueller et al., 2022).

4.2.3 | XGBoost

Another ensemble approach, XGBoost (Chen & Guestrin, 2016), is popular in marketing (e.g., Rafieian & Yogananarasi, 2021) and operations management (Schaer et al., 2022). Similar to BART, XGBoost is an ensemble method that combines multiple decision trees to make predictions. However, the approach is different. XGBoost seeks to find the optimal set of decision trees by minimizing the following loss function:

\[ \text{Loss}_{XGBoost} = \sum_{i=1}^{N} L(y_i, \hat{y}_i) + \sum_{j=1}^{m} \Omega(T_j), \tag{5} \]

where the first term represents the data loss, and the second term \( \Omega(T_j) \) is a regularization term to penalize large values in the terminal node predictions and trees. XGBoost uses gradient boosting to build each tree by iteratively correcting the errors made by previous trees. From an initial set of parameters, it calculates the negative gradient of the loss function as pseudo-residuals. The tree is then fitted to these pseudo-residuals, and this process continues with a weighting factor determined by a learning rate parameter. XGBoost also includes techniques like shrinkage, column subsampling, and early stopping to improve performance and prevent overfitting.

BART and XGBoost differ in several ways. On the one hand, XGBoost incorporates several regularization techniques and frequently achieves the highest performance while being scalable for large datasets and high dimensionality. On the other hand, BART provides several advantages regarding interpretability: Its Bayesian approach allows one to quantify uncertainty, impute missing values and intervals using posterior distributions. This can be particularly helpful when trying to assess the relative importance of subsets of features or to do significance tests. It is also considered to be simpler to implement because packages require fewer tuning parameters.

4.3 | Predictive results

We divided the data into an in-sample training step and an out-of-sample testing set to train and evaluate our models. We used the training set (80%) to train the model and the testing set (20%) to evaluate it. For regularized regressions, we first scaled the variables prior to using cross-validation on mean-square errors within the training data to select the tuning parameter \( \lambda \) for each of Ridge and Lasso. For BART, we used 10-fold cross-validation within the training set to choose the hyperparameters in our classifier (e.g., number of trees \( m \), shrinkage parameter \( k \)) as implemented in BART-Machine (Kapelner & Bleich, 2013). For XGBoost, we also used 10-fold cross-validation as it has many parameters (e.g., max depth, subsample, column sample by tree, number of rounds, and early stopping rounds). The decision of using cross-validation across all models was made to minimize our use of researcher degrees of freedom (i.e., our choice of hyperparameters being responsible for model results). Table 3 shows the results for in-sample RMSE (root mean squared errors) and MAE (mean absolute errors). We note that the RMSE and MAE for Ridge and Lasso are based on final descaled predictions so that the values are comparable to that of BART and XGBoost and so that they are more easily interpretable. We also note that for missing values for some image characteristics that could not be computed, we used the KNNimpute function in the R library Caret for Lasso, Ridge, and XGBoost, whereas we used the recommended procedure by Kapelner and Bleich (2013; see Kapelner & Bleich, 2015).

Some results are consistent across all four models. Compared to a model trained on only the baseline variables, the addition of variables (whether they are counts of visuals, text, or image details) improves both in-sample and
### Table 3

Kickstarter: Comparisons of prediction across subsets of variables.

| Model   | Variable sets       | In-sample Root mean squared errors (RMSE) | Mean absolute errors (MAE) | Out-of-sample RMSE | MAE   |
|---------|---------------------|------------------------------------------|----------------------------|---------------------|-------|
| Ridge   | 1) Base             | 3.6232                                   | 2.8988                     | 3.4593              | 2.9123 |
|         | 2) Base + Visual counts | 3.3868                               | 2.3885                     | 2.8455              | 2.3722 |
|         | 3) Base + Text details | 3.2502                               | 2.1387                     | 2.7277              | 2.2185 |
|         | 4) Base + Image details | 3.5279                               | 2.7246                     | 3.2850              | 2.7300 |
|         | 5) All              | 3.1482                                   | 2.0118                     | 2.5426              | 2.0734 |
| Lasso   | 1) Base             | 3.4899                                   | 2.8982                     | 3.4597              | 2.9137 |
|         | 2) Base + Visual counts | 3.0079                               | 2.3851                     | 2.8423              | 2.3685 |
|         | 3) Base + Text details | 2.8435                               | 2.1151                     | 2.7308              | 2.2108 |
|         | 4) Base + Image details | 3.3583                               | 2.7246                     | 3.2913              | 2.7343 |
|         | 5) All              | 2.7010                                   | 1.9957                     | 2.5408              | 2.0663 |
| BART    | 1) Base             | 3.1243                                   | 2.5784                     | 3.1893              | 2.6316 |
|         | 2) Base + Visual counts | 2.4555                               | 1.9638                     | 2.4874              | 1.9946 |
|         | 3) Base + Text details | 2.3648                               | 1.8891                     | 2.5316              | 2.0395 |
|         | 4) Base + Image details | 2.9235                               | 2.4034                     | 3.0468              | 2.4972 |
|         | 5) All              | 2.2371                                   | 1.7817                     | 2.3542              | 1.8838 |
| XGBoost | 1) Base             | 3.1335                                   | 2.5902                     | 3.2050              | 2.6494 |
|         | 2) Base + Visual counts | 2.4494                               | 1.9605                     | 2.4849              | 1.9938 |
|         | 3) Base + Text details | 2.3421                               | 1.8756                     | 2.5461              | 2.0523 |
|         | 4) Base + Image details | 2.9422                               | 2.4185                     | 3.1074              | 2.5454 |
|         | 5) All              | 2.2039                                   | 1.7551                     | 2.3521              | 1.8864 |

We also find some differences between the models. It is notable that BART and XGBoost fit better than Ridge or Lasso across all variable sets, in-sample and out-of-sample. Their ability to improve predictive accuracy is most notable when few variables are used, but the improvements are also notable when the entire set of variables is used for out-of-sample RMSE (XGBoost = 2.35; BART = 2.35; Ridge = 2.54; Lasso = 2.54) and MAE, which is not optimized directly by any of the models (BART = 1.88; XGBoost = 1.88; Ridge = 2.07; Lasso = 2.07).

Overall, the results show that the addition of visual information, particularly visual count variables, improves the performance of all three models on in-sample and out-of-sample metrics. Additionally, the BART and XGBoost model outperforms the Ridge and Lasso models across all variable sets, likely indicative of their ability to better accommodate non-linearities using the raw data (i.e., our implementations of Lasso and Ridge did not include two-way interactions). Across models, we also find that although the image details variables reduce out-of-sample performance beyond the base model, their predictive value is notably lower than the two visual count variables. Visual count variables seem to help predict even better than all text details, when BART and XGBoost are given. Given the similar performance of BART to XGBoost and the usefulness of being able to sample the posterior for uncertainty quantification as discussed below, the subsequent descriptive analyses will focus on the BART model results.

#### 4.4 Importance of visual counts and image details

We have argued that visual elements are important predictors and should be incorporated when predicting crowdfunding
success; visual information provides significant predictive power in both in-sample and out-of-sample. Although these findings support that adding visual count variables and image details significantly improves prediction, there is no evidence that removing image variables would significantly affect fit. One way to obtain such evidence is to retrain a BART model without any specific covariate and compare the pseudo-$R^2$ of the model with that of BART models retrained using “null” samples where the predictors are removed (see Bleich et al., 2014). Using 100 resamples where we remove all image details (but keep visual counts), we find that all of the resamples see a reduction in pseudo-$R^2$. If we remove only the new image details we proposed (visual scene elements, with the addition of blur, quality/noise, and number of segments), we find that 85% of resamples saw a reduction in pseudo-$R^2$. This suggests that not only are image details important beyond visual count variables but that the proposed predictors can add important information for prediction.

Next, we use two strategies from interpretable machine learning (e.g., Molnar, 2020; Molnar et al., 2018) to provide confidence in the importance of the visual attributes: feature importance and partial dependence plots.

### 4.4.1 Feature importance

According to Molnar (2020), feature importance in machine learning refers to the increase in the prediction error of the model that occurs when we break the relationship between the feature and the true outcome. There exist many ways to calculate variable importance scores. In regression, one often compares standardized regression coefficients as differences in the size of unstandardized coefficients that could be due to feature scaling. In BART models, variable importance is commonly assessed via inclusion proportions in the post-burn-in MCMC iterations (Bleich et al., 2014). Here, we use the proportion of times each variable is chosen in a tree among all the posterior draws of the sum-of-tree model. In Figure 2, for the sake of illustration, we present the Top-40 highest feature importance features along with the confidence intervals (95%). In Table 4, we present some summary statistics coded by variable type.

Consistent with the results in Section 3, we find that visual count variables are important. Whereas the average variable has a 0.63% probability of being included in the final tree, the visual count variables (number of pictures: 2.31%, number of videos: 1.94%) are higher than most. Moreover, several image detail measures are most likely to be included. All color image detail variables have a significant probability of inclusion, with brightness (1.36%, SE = 0.27%) and contrast–brightness (1.36%, SE = 0.27%). For figure–ground, the size difference is likely to be included (1.49%, SE = 1.32%); however; texture and color differences are not. For composition, horizontal color balance (0.83%, SD = 0.69%) and number of segments (0.38%, SE = 0.37%) are most important.

Focusing on the variables that are novel in our research (i.e., beyond M. Zhang & Luo, 2023), we find that the variables that are novel to our inquiry for color (blur: 0.46%, SE = 0.43%; quality–noise: 1.28%, SE = 0.54%) and composition (number of segments) are significant. For our visual scene variables (all novel), the most important is the number of annotations provided by Google (1.32%, SD = 0.66%), but the maximum classification confidence is not (0.27%, SE = 0.29%). It also seems that the importance of the number of annotations from Google Vision is greater than that provided by Microsoft Azure (0.09%, SE = 0.02%) and that the maximum classification confidence provided by Azure (0.09%, SE = 0.02%) has a non-zero probability of inclusion. Finally, we note that the number of upright faces has a small but significant probability of inclusion (0.09%, SE = 0.05%).

Looking at whether confidence intervals exclude zero across features allows us to look at which type of features are more likely to have non-zero inclusion. We find that the confidence interval excludes zero for all visual count and base variables other than the day of the year, and that nearly two-thirds of the textual variables from the blur and text have a significant probability of being included in the model. Of the image details, all color variables (brightness, contrast, blur, quality/noise, colorfulness) are significant, and they are, on average, the most important. Only one of the figure–ground variables had a non-zero probability of inclusion (size difference: 1.49%, SE = 0.14%). The visual scene variables all had a significant probability of inclusion other than the maximum classification confidence provided by Google Vision.

Although these findings support that adding visual count variables and image details improves prediction, there is no evidence that removing image variables would significantly affect fit. One way to obtain such evidence is to retrain a BART model without any specific covariate and compare the pseudo-$R^2$ of the model with that of BART models retrained using “null” samples where the predictors are removed (see Bleich et al., 2014). Using 100 resamples where we remove all image details (but keep visual counts), we find that all of the resamples see a reduction in pseudo-$R^2$. If we remove only the new image details we proposed (visual scene elements, with the addition of blur, quality/noise, and number of segments), we find that 85% of resamples saw a reduction in pseudo-$R^2$.
a reduction in pseudo-$R^2$. This suggests that not only are image details important beyond visual count variables but that the proposed predictors can add important information for prediction.

### 4.4.2 Exploring functional forms—Partial dependence plots

All the prior analyses have focused on illustrating whether the features are useful for prediction and have been agnostic to how any features associate with the outcome. Yet, one of the most useful aspects of tree-based models is that they can flexibly accommodate non-linear relationships between covariates and outcomes. One way to explore the function form of the relationship between any single feature and the outcome is to use partial dependence functions (PDFs; Friedman, 2001). PDFs provide the expected value of the response variable as a function of a particular predictor variable while averaging out the effects of all other predictor variables in the model. For any given predictor variable, one first fixes the values of all other predictor variables and then computes the expected value of the response variable over a range of values for the chosen predictor variable. This process is repeated for different reference values of the other predictor variables, and the resulting PDFs are averaged to obtain an overall estimate of the relationship between the response variable and the chosen predictor variable. Plotting the results of this process through partial dependence plots can be useful to identify regions of the predictor variable space where the response variable is particularly sensitive to changes in the predictor variable or to visualize the shape of the relationship between the response variable and the predictor variable. To the extent that interpretable features are used in the model, this approach provides interpretability to otherwise opaque models (cf., Molnar, 2020).

In Figure 3, we present plots for several of the proposed visual variables with the highest proportion inclusion. Nevertheless, we notice an intuitive pattern for the relationship between variables and log dollars raised. For example, there is a significant increase in log dollars raised going from zero to one video, but less evidence going from one to more, which is content with the recommendation that authors performing empirical analyses of Kickstarter data should at least include a control for whether the project has at least one video. Intuitively, we find that consideration of image quality (i.e., absence of noise) is fairly linearly associated with dollars raised but blur is not. For images, the number of pictures is generally positively associated with dollars raised, and there are slowly diminishing marginal returns going beyond five. Moreover, the number of concepts seems to peak at five, consistent without our intuition that the elements contained in the image might help disambiguate what the product is and what it does without introducing unnecessary complexity. The reduction in dollars raised associated with having too many associations may be related to the findings on the number of image segments (a composition feature) that more annotations also likely capture more image segments.

### 4.5 Robustness: Predictive value of number of pictures - The role of due to post-failure project editing

We have found that visual count variables (number of images and number of videos) are important predictors of success (log dollars raised). This association is unlikely to be causal
as numerous factors can simultaneously predict the number of visuals and odds of success (stage of prototype development). We cannot rule out that possibility. However, it is important to note that projects can be edited once the campaign is finished. Anecdotally, we have noticed that some projects have removed images and edited the text post failure—presumably in an attempt to defend against copycats and to redirect potential customers outside of the platform. 6

If that is true, then the association we find between the number of images and dollars may be due to unsuccessful projects removing their images. To investigate whether that is a likely explanation, we created an interface to evaluate blinded project descriptions to see if the text of failed projects (including canceled and suspended) with less than 1000 characters in the project description had been edited post failure. This interface did not highlight anything else about the project other than its description, so the code was blind to the number of images and videos for the project. Out of 3046 projects manually inspected, 146, or 4.79%, had clearly modified their listing after the fact.

Comparing the projects (modified vs. not) on project characteristics, we find a few differences. First, failed projects modified after the fact had fewer pictures on average (0.27 vs. 1.14, \( p < 0.01 \)) and videos (0.13 vs. 0.41, \( p < 0.01 \)). This suggests that, indeed, failed projects who edited their listing also scrubbed their pictures and videos. Second, we found that these edited projects were also marginally more likely to have been labeled staff picks by Kickstarter (3% vs. 1%, \( p = 0.06 \)) and had raised more money (log dollars of 5.05 vs. 2.45) despite having failed by Kickstarter standards and not received funding. This would suggest that the importance of the number of pictures and the number of videos depends on whether it is indicative of pre-raise video, at least to an extent.

---

**FIGURE 3** Partial dependence plots for added image characteristics (y-axis is the partial effect). Note: Plotted at quintiles. [Color figure can be viewed at wileyonlinelibrary.com]
4.6 | Summary

We investigated the importance of visual count variables (number of pictures and videos) and image details. The results suggest that visual information has considerable predictive value. Specifically, although summarizing visual information into counts (the number of pictures and videos) is most helpful, the title image of the project contains important additional information for predicting success. Easily extracted image details pertaining to color, composition, and figure–ground are also importantly, and we have provided evidence that novel automatically extracted visual scene features such as the number of upright faces, the number of concepts and classification confidence are helpful.

5 | DISCUSSION AND CONCLUSION

The fields of operations management, marketing, and management have a rich history of using technological tools to enhance and expedite processes related to new idea development. Within this context, there has been a focus on identifying novel predictors of crowdfunding success and utilizing pre-launch processes and tools to aid entrepreneurs in selecting the most promising ideas. In line with this trend, we explore the untapped potential of automated extraction of visual information as a valuable data source for studying crowdfunding in the domains of marketing, management, and operations management.

To begin, we conducted a comprehensive review of the literature to identify existing visual information features that can be extracted from project listings, including visual counts such as the number of pictures and videos, as well as image details encompassing color, composition, and figure–ground relationships. Building upon these established features, we proposed three additional variables (blur, quality/noise, and number of segments) that we believed to be promising in predicting crowdfunding success. Furthermore, drawing insights from the literature on visual scene dynamics, we advocated for the inclusion of visual scene elements, such as the number of faces, the number of evoked concepts, and classification confidence. In our empirical analysis, focused on Kickstarter projects, we employed nested comparisons of baseline, text, and visual features across various predictive models (Lasso, Ridge, BART, and XGBoost) to demonstrate the incremental predictive value of visual features as a whole. Additionally, using techniques from interpretable machine learning and our trained BART model, we highlighted the significance of individual features through proportion inclusions and covariate importance tests. We also utilized partial dependence plots to illustrate potential non-linear relationships. Through our research, we provide several valuable contributions to both researchers and practitioners.

5.1 | Research implications

Although empirical research on crowdfunding generally has considered visual information images by counting the number of pictures and videos attached to the project page (if at all; see Table 1) or focused on manually coded visual features such as founders’ race (Younkin & Kuppuswamy, 2018), we build on recent literature in marketing, which has shown the usefulness of automatically extracting visual attributes can help predict firm success (M. Zhang & Luo, 2023; S. Zhang et al., 2022), to show that automatically extractable visual features can predict crowdfunding success. In doing so, we expand not only the list of documented important visual attributes across established categories of color, composition, and figure–ground by adding blur, number of segments, and image quality (noise) but also provide evidence of the usefulness of adding visual scene elements such as the evoked concepts and faces. Although we found that visual scene elements do incrementally predict crowdfunding success, we did find that image details contributed less to predictive accuracy than visual count variables. One possibility is that we did not use the best available APIs to identify our constructs. Another is that we only focused on the title image, and extracting image details from all visuals would be more beneficial. We encourage future research to look into how to summarize the information image details across multiple images.

Although the study of visual scene dynamics is not a novel area of research, the study of how products are visually processed typically has been assessed through individual human judgments and based on experimental data (e.g., Noseworthy & Trudel, 2011; Veryzer & Hutchinson, 1998). This individual judgment process is resource-intensive and time-sensitive, and it necessitates a trade-off of ecological validity for internal validity. Therefore, and noting the impracticality of using human judges to classify tens of thousands of project images according to whether the visual elements are relevant to a crowdfunding product, we advocate for the use of validated commercial APIs to extract the number of annotations and classification confidence as proxies for the number of concepts and ease of concept identification. We provided preliminary evidence that visual scene elements can inform crowdfunding success and validated this with a separate task of human judgments of relevancy.

5.2 | Practical implications

For consumers, it is difficult to predict crowdfunding success because many unobservable variables (e.g., characteristics of the product creator, product complexity) simultaneously influence product creators’ strategies and outcomes. Rather than developing a statistical model or machine learning algorithm that would outperform the prediction of crowdfunding success—a task many authors have attempted (e.g., Etter et al., 2013; Greenberg et al., 2013)—we used our knowledge
of visual information processing and visual scene dynamics to identify potentially important predictors. We hope our efforts motivate researchers who are interested in developing predictive models to carry out the simple task of augmenting their datasets with easily extracted visual counts and image details. Tools made available to non-technical investors (e.g., Sidekick.epfl.ch, Kicktraq) generally focus on textual and baseline characteristics, and we hope that our findings about the importance of visual counts and image details can help them make scalable products that can better predict success.

For project creators, product designers and marketers are responsible for both the selection and composition of the focal image and various other factors that could also influence funding outcomes. Whereas Kickstarter and platforms tend to focus on textual and baseline characteristics, and we hope that our findings about the importance of visual counts and image details can help them make scalable products that can better predict success.

For project creators, the advice given is general. For example, Kickstarter recommends keeping things simple, and this aligns well with our observation that too many image segments and evoked concepts (annotations) are associated with lower log dollars raised. Likewise, it mentions using high-resolution assets that correlate with our finding that image quality (i.e., lack of noise) is important. However, a pre-trained model that natively handles non-linearities and interactions, such as ours, can help companies select alternative pictures to feature as the splash image. From a set of baseline characteristics and baseline variables that are project-specific, one can feed alternative pictures and see how the odds of success are impacted. We note that although such predictive exercises are helpful, they would still be unlikely to be adequate substitutes for A/B testing that could be implemented on product pages.

For platforms like Kickstarter, the success of campaigns directly impacts their revenue as Kickstarter only distributes funds and earns its commission if the project exceeds its goals. It then becomes essential for such platforms to identify projects that are likely to succeed and find ways to understand funder dynamics as a function of the marketing tools available (Kim et al., 2020). Platforms have different tools at their disposal, some more indirect (e.g., altering the duration window) and some more direct (e.g., highlighting projects, altering fees). One of the most commonly discussed levers at the disposal of the platform is the labeling and highlighting of projects as “staff picks” or “projects we love,” which have been linked to increased odds of success (Pisani, 2014). Yet, platforms are not able to highlight every single project, and projects only generate funds for the platform if the project exceeds the stated goals. One possibility is that the platform strategically allocates staff pick labels not to projects that are clearly going to exceed their funding thresholds but instead focus on giving staff picks to projects that are likely to be completed at levels of funding near the campaign’s goal and where the staff pick label can push them across the finish line. Although the staff pick feature is associated with increased (log) dollars when looked at descriptively (M = 5.20 without; M = 10.53 with; p < 0.01) and through BART covariate importance tests (p < 0.01), it is unclear whether the difference in dollars raised is due to an effect of highlighting the product via staff picks or whether it is due to selection of projects that are, at least, highly unlikely to fail.

### 5.3 | Summary

In this article, we have built on a recent body of work in visual processing and marketing that has studied how interpretable image details and attributes can be automatically extracted to predict firm-level outcomes (M. Zhang & Luo, 2023; S. Zhang et al., 2022). One could wonder whether such efforts are necessary when modern machine learning algorithms have the capability to automatically learn patterns and feature representations from data. In our view, such feature engineering remains crucial for achieving accurate predictions and eventual theory development. First, we have highlighted that visual count information (number of pictures, number of videos), a feature that is not extracted from image details, is the most important variable in predicting the outcome. Second, relying solely on non-interpretable models (e.g., CNNs) for prediction can fail short in practice unless one has an immense amount of data as automatic feature extraction can introduce unnecessary complexity, diminish performance, and make interpretation more challenging. Third, past research in marketing has demonstrated the significant impact of feature engineering on seemingly visual data. For instance, Blanchard et al. (2020) showed that feature engineering enabled a simple Bayesian model to outperform a CNN that did not incorporate researcher-extracted features. This highlights the potential value of carefully engineered features even in visual domains. Ultimately, we aim to emphasize the importance of including relevant features in predictive models. By showcasing the predictive power of these feature sets, we hope to not only encourage researchers interested in empirical analyses to consider and incorporate them into their own analyses but also hope to stimulate research that highlights novel data sources such as social media (R. Cui et al., 2018), tags (Nam & Kannan, 2014), and textual features (Toubia et al., 2021) and facilitate the generation of important features.

### ACKNOWLEDGMENTS

The authors would like to thank Zhengrong Gu for technical assistance and seminar participants at UC San Diego, Fudan University, McGill University, Shanghai Tech, and the Theory and Practice in Marketing Conference for helpful feedback. Ethan Pancer is supported in part by funding from the Social Sciences and Humanities Research Council of Canada (Grant SSHRC 435-2018-1319).

### ORCID

Simon J. Blanchard ORCID https://orcid.org/0000-0002-1803-3543

### ENDNOTES

1. https://www.thecrowdfundingcenter.com/.
2. https://www.kickstarter.com/help/images.
References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence, 34*(11), 2274–2282.

Blanchard, S. J., Aloise, D., & DeSarbo, W. S. (2012). The heterogeneous p-median problem for categorization based clustering. *Psychometrika, 77*(4), 741–762. https://doi.org/10.1007/s11336-012-9283-3

Blanchard, S. J., Dyachenko, T. L., & Kettle, K. L. (2020). Locational choices: Modeling consumer preferences for proximity to others in reserved seating venues. *Journal of Marketing Research, 57*(5), 878–899. https://doi.org/10.1177/0022243720941525

Blaseg, D., Schulze, C., & Skiera, B. (2020). Consumer protection on Kickstarter. *Marketing Science, 39*(1), 211–233. https://doi.org/10.1287/mksc.2019.1203

Bleich, A., Kaplner, A., George, E. I., & Jensen, S. T. (2014). Variable selection for BART: An application to gene regulations. *The Annals of Applied Statistics, 8*(3), 1750–1781. https://doi.org/10.1214/14-AOAS755.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. University of Texas at Austin. https://www.liwc.austin.edu/static/documents/LIWC-22-Manual%20%20%20Development%20%20%20Psychometrics.pdf

Brown, C. L., & Carpenter, G. S. (2000). Why is the trivial important? A reasons-based account for the effects of trivial attributes on choice. *Journal of Consumer Research, 26*(4), 372–385. https://doi.org/10.1086/309569

Buetti, S., Cronin, D. A., Madison, A. M., Wang, Z., & Lleras, A. (2016). Towards a better understanding of parallel visual processing in human vision: Evidence for exhaustive analysis of visual information. *Journal of Experimental Psychology: General, 145*(6), 672–707. https://doi.org/10.1037/xge0000163

Burtt, G., Gupta, D., & Martin, P. (2021). Referral timing and fundraising success in crowdfunding. *Manufacturing & Service Operations Management, 23*(3), 676–694.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA (pp. 785–794).

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics, 4*(1), 266–298. https://doi.org/10.1214/09-AOAS285

Cohn, M. A., Mehli, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*, 687–693. https://doi.org/10.1111/j.0956-7976.2004.00741.x

Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2018). The operational value of social media information. *Production and Operations Management, 27*(10), 1749–1769. https://doi.org/10.1111/poms.12707

Cui, Z., Kumar, P. M., & Goncalves, D. (2019). Scoring vs. ranking: An experimental study of idea evaluation processes. *Production and Operations Management, 28*(1), 176–188. https://doi.org/10.1111/poms.12910

Dekowska, M., Kuniecki, M., & Jaśkowski, P. (2008). Facing facts: Neuronal mechanisms of face perception. *Acta Neurobiologiae Experimentalis, 68*(2), 229–252.

Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science, 34*(1), 43–68. https://doi.org/10.1214/18-STS667

Elliot, A. J., & Maier, M. A. (2014). Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology, 65*(1), 95–120. https://doi.org/10.1146/annurev-psych-010213-115035

Etter, V., Grossgausler, M., & Thiran, P. (2013). Launch hard or go home! Predicting the success of Kickstarter campaigns. *Proceedings of the first ACM conference on Online Social Networks (COSN’13), Boston, MA (pp. 177–182).

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232. https://doi.org/10.1214/aoas132

Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing*. Prentice-Hall.

Gorn, G. J., Chattopadhyay, A., Sengupta, J., & Tripathi, S. (2004). Waiting for the web: How screen color affects time perception. *Journal of Marketing Research, 41*(2), 215–225. https://doi.org/10.1509/jmkr.41.2.215.28668

Greenberg, M. D., Parbo, D., Harirahan, K., & Gerber, E. (2013). Crowdfunding support tools: Predicting success & failure. *CHI’13 Extended Abstracts on Human Factors in Computing Systems, Paris, France (pp. 1815–1820).

Guido, G., Pichieri, M., Pino, G., & Natarajan, R. (2019). Effects of face images and face pareidolia on consumers’ responses to print advertising: An empirical investigation. *Journal of Advertising Research, 59*(2), 219–231. https://doi.org/10.2501/JAR-2018-030

Hastie, T., Tibshirani, R., & Friedman, J. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics, 9*(3), 432–441.

Hill, J., Linero, A., & Murray, J. (2020). Bayesian additive regression trees: A review and look forward. *Annual Review of Statistics and Its Application, 7*(3), 251–278. https://doi.org/10.1146/annurev-statistics-031219-041110

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research, 45*(13), 1707–1724. https://doi.org/10.1016/j.visres.2004.12.021

Kaplner, A., & Bleich, J. (2013). bartMachine: Machine learning with Bayesian additive regression trees arXiv preprint arXiv:1312.2171. https://arxiv.org/abs/1312.2171

Kaplner, A., & Bleich, J. (2015). Prediction with missing data via Bayesian additive regression trees. *Canadian Journal of Statistics, 43*(2), 224–239. https://doi.org/10.1002/cjs.11248

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2013). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology, 32*(3), 125–143.

Kickstarter. (2020). Stats. https://www.kickstarter.com/help/stats

Kickstarter. (2022). *Creator handbook: Telling your story*. http://bit.ly/3XzZni

Kim, C., Kannan, P. K., Trusov, M., & Oradanini, A. (2020). Modeling dynamics in crowdfunding. *Marketing Science, 39*(2), 339–365. https://doi.org/10.1287/mksc.2019.1209

Kornish, L. J., & Hutchison-Krupat, J. (2017). Research on idea generation and selection: Implications for management of technology. *Production and Operations Management, 26*(4), 633–651. https://doi.org/10.1111/poms.12664

Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Kamali, S., Malli, M., Pont-Tuset, J., Veit, A., Belongie, S., Gomes, V., Gupta, A., Sun, C., Chechik, G., Cai, D., … Murphy, K. (2017). OpenImages: A public dataset for large-scale multi-label and multi-class image classification. *International Journal of Computer Vision, 128*(7), 1956–1981.

Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. In D. Cumming & L. Hornuf (Eds.), *The economics of crowdfunding* (pp. 151–182). Palgrave Macmillan, Cham.
SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Blanchard, S. J., Noseworthy, T. J., Pancer, E., & Poole, M. (2023). Extraction of visual information to predict crowdfunding success. Production and Operations Management, 32, 4172–4189. https://doi.org/10.1111/poms.14083

APPENDIX A
mTurk image classification validation
Below, you will find a product, a product description, and an image.

A computer program identified the following labels as part of the above image. Some may be relevant descriptions of the image but have nothing to do with the product. Which of the following labels are about the product? Unselect the labels that are not about the product in the description. [see instructions].

For example, in the SHRU: we have a cat toy. Of all the labels, not all are directly relevant to the product. For instance, cat, ball, small to medium-sized cats, kitten would all be obviously relevant to the product. Play is an important part of the product concept, but carnivore and whiskers (although both about a cat) may not have much to do with the product. Likewise, photo caption is not helpful. We would expect you to unselect photo caption, carnivore, and whiskers.

[begin instructions menu]

In these tasks, we want you to help you identify which labels are related to the product in the image.