Camera eats first: exploring food aesthetics portrayed on social media using deep learning

Alessandro Gambetti and Qiwei Han

Nova School of Business and Economics, Universidade Nova de Lisboa, Carcavelos, Portugal

Abstract

Purpose – The purpose of this paper is to explore and examine discrepancies of food aesthetics portrayed on social media across different types of restaurants using a large-scale data set of food images.

Design/methodology/approach – A neural food aesthetic assessment model using computer vision and deep learning techniques is proposed, applied and evaluated on the food images data set. In addition, a set of photographic attributes drawn from food services and cognitive science research, including color, composition and figure–ground relationship attributes is implemented and compared with aesthetic scores for each food image.

Findings – This study finds that restaurants with different rating levels, cuisine types and chain status have different aesthetic scores. Moreover, the authors study the difference in the aesthetic scores between two groups of image posters: customers and restaurant owners, showing that the latter group tends to post more aesthetically appealing food images about the restaurant on social media than the former.

Practical implications – Restaurant owners may consider performing more proactive social media marketing strategies by posting high-quality food images. Likewise, social media platforms should incentivize their users to share high-quality food images.

Originality/value – The main contribution of this paper is to provide a novel methodological framework to assess the aesthetics of food images. Instead of relying on a multitude of standard attributes stemming from food photography, this method yields a unique one-take-all score, which is more straightforward to understand and more accessible to correlate with other target variables.

Keywords Food aesthetics, Gastronomic experience, Social media, Computer vision, Deep learning

Paper type Research paper

1. Introduction

The growing prevalence of social media platforms has enabled customers to resolve uncertainty and communicate with businesses before making purchase decisions. For the restaurant industry, customers nowadays overwhelmingly use online social platforms such as Yelp or TripAdvisor for information-seeking and experience-sharing purposes (Hicks et al., 2012). Many tourism and hospitality research works acknowledge the strategic importance of social media from customer and provider perspectives (Chu et al., 2020; Leung et al., 2013). On the one hand, social media information about food and dining activities may affect customers’ restaurant choices (Liu et al., 2020). On the other hand, providers

The authors are very grateful to Mr Ben Collier who helped proof-read the paper, and to the editors and to the three peer-reviewers for their valuable comments that led to significant improvements of this paper. Moreover, the authors acknowledge financial support from Fundação para a Ciência e Tecnologia (UID/ECO/00124/2019) by LISBOA-01-0145-FEDER007722 and Social Sciences Data Lab PINFRA/22209/2016.
increasingly leverage social media as an advertising vehicle to communicate with existing and potential customers (Chu et al., 2020; DiPietro et al., 2012; Kwok and Yu, 2013), as well as to manage electronic word-of-mouth (Litvin et al., 2008; Pantelidis, 2010). For example, Kim et al. (2015) found positive effects of a restaurant’s activities on social media on firm value, after controlling for several firm characteristics. In particular, the proliferation of smartphones with high-quality cameras and affordable mobile data plans catalyzes a global phenomenon of “camera eats first,” i.e. taking and sharing food images on social media when customers are dining out in restaurants has become increasingly popular as a key component of their life experiences (Murphy, 2010). For example, researchers found that a substantial proportion of images shared on Instagram, an image-based social media platform, are related to food items (Holmberg et al., 2016; Hu et al., 2014).

Meanwhile, customers tend to select restaurants by searching for food images on social media generated by other customers (Oliveira and Casais, 2019). In a similar vein, restaurants also actively seek image-based social media marketing strategies, e.g. by posting food images on their social media accounts to encourage customer engagement (Lee et al., 2021a). Stepchenkova and Zhan (2013) explained such differences using the classic conceptualization of destination image that is divided into projected image of restaurant owners and perceived images of customers, respectively, which are inherently reflected in the content of images. Therefore, understanding how restaurants are portrayed on social media by both customers and restaurants is imperative not only for formulating adequate social media business strategies and operational actions, but also for designing effective management of customer gastronomic experiences (Sotiriadis, 2017).

Existing studies that use informational content from social media to understand online engagement between customers and restaurants mainly focus on customer reviews using textual analysis (He et al., 2013; Kwok and Yu, 2013; Lee et al., 2021b; Lu et al., 2013; Mariani et al., 2019; Pantelidis, 2010). For example, Pantelidis (2010) performed a content analysis of customer reviews on social media to characterize the “electronic gastronomic experience” phenomenon, which involves factors such as food, service and ambiance, among others. Meanwhile, images shared on social media as another important information source are being increasingly explored in hospitality research (Lo et al., 2011). However, one major challenge with studying images is that encoding images to generate meaningful representations typically requires a significant amount of manual effort and subjective evaluations from human coders, which may be prohibitively expensive for large-scale image content on social media (Hu et al., 2014).

Beyond the role of images in product presentation and consumer persuasion in tourism and hospitality research, there is a stream of literature about the relationship between food images and sensory gastronomic experience from the cognitive sciences showing that food images may affect human food perception and desire, known as food aesthetics (Spence et al., 2016). More specifically, food aesthetics refers to the sensory gastronomic experience from food presentation such as plating, decorating and styling (Schifferstein et al., 2020). In fact, in the culinary world, food aesthetics has long been acknowledged as the art of food presentation, that visual sensation of food represents as an important element of gastronomic experience as it is flavor (Deroy et al., 2014). As such, analyzing image content from social media would benefit from extracting perceptual features from food images and understanding how the visual appeal of food relates to the gastronomic experience and restaurant services. For example, Li and Xie (2020) examined the effect of image content on user engagement on social media and showed that high-quality images lead to higher engagement than low-quality ones.
In this paper, we aim to explore sensory gastronomic experiences portrayed on social media using a large-scale data set of food images collected from Yelp with two main research questions:

*RQ1.* How to characterize gastronomic experiences from the content of food images in terms of food aesthetics.

*RQ2.* How food aesthetics reflected in the food images would differ across several restaurant dimensions.

We propose a food aesthetics assessment model using computer vision and deep learning techniques to estimate aesthetic scores of all food images. More specifically, we adopt a customized multi-stage model that leverages pre-trained model parameters learned from a labeled general aesthetics data set (Murray et al., 2012) and fine-tune with a labeled food aesthetics data set (Sheng et al., 2018). The model is then applied to estimate the aesthetic scores of food images in a web-scraped data set containing food images posted on social media. Given the subjective nature of evaluating food aesthetics, we validate the model performance through two approaches. Firstly, we select a random sample of food images from our data set and evaluate them with an online survey conducted with human raters through Prolific, a crowdsourcing platform. We evaluate the degree of agreement between our model’s predicted aesthetics scores and human raters’ aesthetic preferences of food images. Secondly, we construct a set of photographic attributes for food images drawn from the food services and computer vision studies, including color, composition and figure–ground relationship attributes. Both approaches show largely consistent evaluations of food aesthetics as our model predictions.

Finally, we explore the discrepancies of restaurants in terms of average aesthetic scores across several dimensions, such as price levels, rating levels, chain status and cuisine types using one-way variance analysis (ANOVA) models. To understand the underlying mechanism that drives the differences in the aesthetics of food images on restaurants’ social media pages, we further evaluate the aesthetic appeal from customer-posted food images and restaurant-posted food images, respectively, using two-way ANOVA models. This would allow us to characterize the current image-posting activities by customers and restaurants on social media with different objectives.

The paper is organized as follows. Section 2 reviews the related literature about food aesthetics and food photography, as well as visual data analytics on social media. Section 3 explains the data collection steps to obtain a large collection of food images from Yelp. Section 4 proposes the food aesthetic assessment model using computer vision and deep learning techniques, and comparative photographic attributes. Section 5 presents the results from one-way and two-way ANOVA models, respectively. Section 6 discusses both theoretical and managerial implications, as well as major limitations of the study and future research.

### 2. Related literature

#### 2.1 Visual data analytics on social media

The growth of the tourism and hospitality research has been documented with the increasing availability of big data (Li et al., 2018; Mariani et al., 2018; Mariani, 2020; Mariani and Wamba, 2020). This stream of literature applies advanced analytical techniques to large quantities of a wide variety of data types, enabling a better understanding of tourism demand, tourist behavior, tourist satisfaction, etc. (Akter et al., 2021), including user-generated content (UGC) data, device data and transaction data (Li et al., 2018). The success
of tourism and hospitality research working with UGC data posted on social media may be attributed to its low cost and easy access (Mariani et al., 2019; Xiang et al., 2017) with two main types: online textual data (Mariani and Borghi, 2021a, 2021b) and online image data (Li et al., 2018). As such, tourism and hospitality studies leverage online (geotagged) image data collected from social media (Azevedo, 2021; Ghermandi et al., 2020; Vu et al., 2015) to perform statistical analysis, e.g. to count the number of images posted about tourism destinations. However, these studies typically do not fully account for the informational content of images, which may contain strong signals about tourism-related issues (Zhang and Luo, 2018).

An emerging research technique that incorporates visual content analysis of large-scale UGC image data has been empowered by the rapid adoption of modern data science techniques such as computer vision and deep learning (Mariani and Baggio, 2022). Computer vision is a computational approach to approximate human visions to extract, analyze and understand visual cues. More specifically, computer vision techniques intend to automate the function of the human visual system, such as detecting and recognizing objects and classifying images (Forsyth and Ponce, 2011). Moreover, the recent development of deep learning can further boost the performance of computer vision tasks achieving state-of-the-art results. As an emerging area of machine learning, deep learning is based on artificial neural networks, in which neurons are organized into the architecture of multiple interconnected layers (LeCun et al., 2015).

We provide a list of tourism and hospitality research papers using visual data analytics on social media in Table 1, reporting research gaps, image data and the proposed solutions, respectively. Given its great potential in extracting the complex structural representation of images at large scale, instead of labor-intensive manual encodings that work with hundreds of images (Stepchenkova and Zhan, 2013), these studies leverage computer vision and deep learning techniques to assess cultural ecosystems service (Richards and Tunçer, 2018), recommend advertising images for tourism destinations (Deng and Li, 2018), recognize the entities appearing in images (Ren et al., 2021; Wang et al., 2020), investigate tourists’ behavior and perception (Zhang et al., 2019a, 2021b,c) and explore tourists’ photos reflecting regional characteristics (Cho et al., 2022). However, none of these studies provides the image quality assessment, which is important in visual data analytics on social media.

Moreover, computer vision and deep learning techniques have also been applied to extract meaningful visual cues from food images. Zhou et al. (2019) review the technical articles that use deep learning in the food domain for food recognition, calorie estimation and quality detection tasks. Zhang and Luo (2018) extract composition attributes of food images from Yelp to predict restaurant survival. Zhang et al. (2020) use a combination of food image attributes and deep learning generated representations to capture visual cultural bias in food classification. However, studies using food images in hospitality research are still limited. Yang et al. (2017) explore the effect of two types of presentation formats (textual and imagery) involving online restaurant reviews; however, the number of food images was accounted for manually, action that might not easily scale up to characterize visual content from large-scale food image data sets.

**2.2 Food aesthetics and the science of food presentation**

There has been a growing body of evidence from the cognitive neurosciences that supports the phrase “we eat first with our eyes” (Delwiche, 2012), as visual stimuli may modify the perception of taste, smell and flavor of food. van der Laan et al. (2011) comprehensively reviewed neuroimaging studies involving hundreds of participants, revealing how a small number of human brain regions are activated in response to food images. Essentially, the
<table>
<thead>
<tr>
<th>SN</th>
<th>Reference</th>
<th>Research gap</th>
<th>Data</th>
<th>Content extraction methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stepchenkova and Zhan (2013)</td>
<td>Comparative content analysis of destination images</td>
<td>500 Flickr travel images in Peru</td>
<td>Manual encoding to identify destination attributes</td>
</tr>
<tr>
<td>2</td>
<td>Bossard et al. (2014)</td>
<td>Dishes and ingredients recognition from food images</td>
<td>Food101 data set, including 101,000 images</td>
<td>Machine learning model to mine discriminative visual components</td>
</tr>
<tr>
<td>3</td>
<td>Latorre et al. (2014)</td>
<td>Use image-focused social media to study tourism consumption</td>
<td>400,000 Flickr and Instagram images geotagged in Zaragoza, Spain</td>
<td>Statistical analysis of image geotags</td>
</tr>
<tr>
<td>4</td>
<td>Amato et al. (2017)</td>
<td>Identify food trends and popularity on social media</td>
<td>Twitter food images</td>
<td>Deep learning technique to classify food categories</td>
</tr>
<tr>
<td>5</td>
<td>Richards and Tunçer (2018)</td>
<td>Assess cultural ecosystems service</td>
<td>20,000 landscape images from Singapore</td>
<td>Computer vision and clustering algorithms to group images</td>
</tr>
<tr>
<td>6</td>
<td>Deng and Li (2018)</td>
<td>Select appropriate photos for destination promotion</td>
<td>YFCC 100M data set including 20,974 Flicker images from New York City</td>
<td>Machine learning model to predict emotion from image content</td>
</tr>
<tr>
<td>7</td>
<td>Ma et al. (2018)</td>
<td>Predict hotel review helpfulness</td>
<td>68,896 Yelp and Tripadvisor images of US cities</td>
<td>Transfer learning model to extract image features</td>
</tr>
<tr>
<td>8</td>
<td>Giglio et al. (2019)</td>
<td>Identify tourism attractiveness</td>
<td>30,951 online destination photos of Australia</td>
<td>Transfer learning model for photo classification</td>
</tr>
<tr>
<td>9</td>
<td>Zhang et al. (2019a)</td>
<td>Explore tourists' cognition</td>
<td>35,336 Flickr tourists' photos in Beijing</td>
<td>Deep learning model to identify types of scenes</td>
</tr>
<tr>
<td>10</td>
<td>Wang et al. (2020)</td>
<td>Identify image classification categories</td>
<td>26,392 Flicker images from six Italian cities</td>
<td>Machine learning model for photo category classification</td>
</tr>
<tr>
<td>11</td>
<td>Ren et al. (2021)</td>
<td>Evaluate differences in contents of photos by hotel managers and travelers</td>
<td>53,813 Tripadvisor hotel images in Macau</td>
<td>Deep learning model for entity recognition</td>
</tr>
<tr>
<td>12</td>
<td>Yu and Egger (2021)</td>
<td>Explore interplay between the effects of color and user engagement</td>
<td>4,757 Instagram images with top destination attributes</td>
<td>Deep learning model for detecting dominant colors of images</td>
</tr>
<tr>
<td>13</td>
<td>Zhang et al. (2021b)</td>
<td>Identify tourists' perceptions of urban space</td>
<td>14,841 Flickr images depicting urban space</td>
<td>Deep learning models for scene recognition and semantic segmentation</td>
</tr>
<tr>
<td>14</td>
<td>Zhang et al. (2021c)</td>
<td>Identify tourists' perceptions of urban space</td>
<td>YFCC 100M data set including 2,227 tourist photos in Beijing</td>
<td>Deep learning model for landscape element recognition</td>
</tr>
<tr>
<td>15</td>
<td>Cho et al. (2022)</td>
<td>Explore tourists' photos reflecting regional characteristics</td>
<td>168,216 Flicker destination images in Seoul</td>
<td>Deep learning model for tourist's photo classification</td>
</tr>
</tbody>
</table>

Table 1. Summary of visual data analytics on social media in tourism and hospitality research
visual appeal has a significant influence on the overall pleasure that food elicits (Spence et al., 2016). For example, Jansen et al. (2010) found that simply showing appealing food images may evoke hunger and affect people's consumption choices.

Numerous studies also aim to understand what components of visual cues cognitively influence people's perceptual food tastes, such as color (Chatterjee, 2004; Spence et al., 2010), shape (Deroy and Valentin, 2011; Piqueras-Fiszman et al., 2012) and orientation (Michel et al., 2015). Zellner et al. (2011) further argue that visual complexity, such as the arrangement of food on the plate, could influence customers' dining experience (Spence et al., 2014). For example, Velasco et al. (2016) showed that people prefer balanced over unbalanced visual compositions when plating food. Overall, people tend to respond more positively to food images with high aesthetic level (Peng and Jemmott, 2018).

3. Data

The data set used in this study was obtained from the Yelp open data set repository (https://www.yelp.com/dataset). This data set included more than 160,000 businesses operating in various sectors in ten metropolitan North American areas, such as Montreal, Calgary, Toronto, Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison and Cleveland. Without the loss of generality, we decided to focus on restaurants conducting business in the Las Vegas metropolitan area because of its high touristic activities. According to Statista, Las Vegas was the most visited metropolitan area in the data set with 42.52 million tourists overall in 2019 (Lock, 2021b), about 6 million of them were international tourists (Lock, 2021a).

Next, we filtered the data set as follows. Firstly, we removed restaurants from the Yelp data set that had closed, because they might otherwise lead the analysis to be biased toward restaurants with poor performance. Secondly, we excluded restaurants with very few food images, and retained only those with at least 10 food images posted by either the business owner or customers. In addition, as food images associated with each restaurant had not been updated since the official release of the data set, we web-scraped all the food images of these restaurants posted on Yelp. In total, we collected 50,018 food images for 577 restaurants.

Furthermore, we also obtained the poster of each image, as we believe that business owners and customers may have different incentives when posting food images. Lastly, we also included several restaurant characteristics, including:

- **Price level**, indicating the average cost of a meal, divided into four categories (“$”: under $10, “$$”: $10–$30, “$$$: $31–$60 and “$$$$”: over $60);
- **Stars rating**, indicating the average star rating given by customers, distributed from 1 to 5;
- **Cuisine styles**, indicating the regional origin of the cuisines styles, divided into Italian, Mexican, Chinese, Japanese, American-New and American-Trad; and
- **Chain status**, indicating whether the restaurant belongs to a chain (e.g. McDonald’s, Burger King, Starbucks, etc.).

To check the restaurant’s chain status, following Zhang and Luo (2018), we counted the number of restaurant names in the data set, and assigned those appearing more than five times as belonging to a restaurant chain [1].
4. Methodology

4.1 Food aesthetics assessment model

In this section, we introduce a novel approach to characterize food aesthetics from food images. This process aims to estimate an aesthetic score for each food image in terms of its aesthetic quality. Pragmatically, having one single reference metric may significantly reduce complexity in interpreting and understanding why an image is more visually pleasing.

The food aesthetic assessment model architecture is illustrated in Figure 1. Overall, the model consisted of three stages:

1. We implemented a transfer learning approach to adapt a pre-trained neural image assessment model (NIMA) based on a labeled image data set of generic aesthetics (Talebi and Milanfar, 2018).
2. We fine-tuned the model by training on a labeled data set of food images (Sheng et al., 2018).
3. We used the trained model to estimate the aesthetic scores of food images in our data set.
4.1.1 Neural image assessment. NIMA model is a combination of a convolutional neural network (CNN) and a normalized earth moving distance (EMD) loss function (Levina and Bickel, 2001) that predicts the aesthetic quality of an image based on a probability distribution (Talebi and Milanfar, 2018). Firstly proposed by Fukushima (1980) and further developed by LeCun et al. (1989), the CNN model is a class of neural network architectures that is proven successful in addressing many computer vision tasks, such as image classification, recognition and segmentation. Given an image, the CNN model aims to automatically identify the underlying intricate structure through convolution operations to extract its visual content.

Containing about 250,000 generic images, the AVA data set was used to train the NIMA model (Murray et al., 2012). Each image in the data set was rated from 1 (low aesthetics) to 10 (high aesthetics) by an average of 200 amateur photographers competing for the DPChallenge (https://www.dpchallenge.com/), which is a contest that rewards the best team that has submitted the best image. Each image corresponds to a distribution of votes that was converted into a probability distribution across the ten rating classes. Essentially, AVA images have the mean ratings of about 5.5 and standard deviations of 1.4. Thus, the aesthetic quality of an image can be summarized by its mean score across all the ratings. Following Talebi and Milanfar (2018), we chose the MobileNet as the backbone model (Howard et al., 2017), which is mainly designed for mobile computer vision tasks because of its computational efficiency. In particular, we modified the architecture of the MobileNet replacing the last layer of the network with a fully connected layer of ten neurons (one for each rating class), all activated by a softmax activation function normalizing the output of the network to a probability distribution. The idea is that each neuron should output the probability of its related rating class.

\[
\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j} e^{x_j}}
\]  

(1)

In the training pipeline, input images were rescaled proportionally such that the shortest side between height and width had size 256 pixels, and then a patch of size 224 × 224 pixels was randomly cropped to prevent statistical overfitting. Target distributions, expressing the aggregation of human ratings, would be represented as an empirical probability mass function \( p = \{p_{s_1}, p_{s_2}, \ldots, p_{s_{10}}\} \) for each ordinal \( s_i \) class, and \( N = 10 \) classes in total. As a result, the aesthetic score \( \mu = \sum_{i=1}^{N} p_{s_i} \cdot s_i \) is the weighted average of each ordinal class value multiplied by its related probability. The ground-truth probability distributions attributable to the customers’ rating activity were learned minimizing the EMD loss. To be exhaustive, the EMD loss can be defined as the smallest cost to shift from one cumulative probability distribution (CDF) to another.

\[
\text{EMD}(p, \hat{p}) = \left( \frac{1}{N} \sum_{k=1}^{N} \left| |CDF_p(k) - CDF_{\hat{p}}(k)| \right|^r \right)^{1/r}
\]

(2)

In equation (2), \( p \) refers to the true probability distribution, \( \hat{p} \) refers to the predicted probability distribution, and \( r \) is a power coefficient, usually set equal to 2 for easier optimization.
Directly training on the AVA data set, which is over 30 gigabytes, may have incurred high computational costs. In fact, we deployed transfer learning to facilitate our task. Transfer learning is the process of keeping information obtained while resolving one task and applying it to a different but related task (Zhuang et al., 2021). Practically, we downloaded a pre-trained NIMA model based on the MobileNet architecture and fine-tuned the model only on food images (Majumdar et al., 2018). Pre-trained implementations are a core of transfer learning. In particular, pre-trained models are saved networks that were previously trained on a large data set, in this case on a large-scale image aesthetic classification task.

4.1.2 Fine tuning using food data. We fine-tuned the NIMA model such that, during inference, the model receives and scores a food image (input), returning an aesthetic score in a 0 to 1 bounded continuous range (output).

Firstly, we downloaded the Gourmet Photography Dataset (GPD), which is a large-scale data set for aesthetic assessment of food photographs (Sheng et al., 2018). GPD images were retrieved from various social media websites with diverse food classes (e.g. Pizza, Sushi and Tacos) and geo-information (e.g. French, Italian and Japanese). Moreover, complementary images were also retrieved from many food categorization data sets to enrich data complexity. For instance, pictures from the Food101 data set were included (Bossard et al., 2014). Then, image labeling was performed using Amazon’s Mechanical Turk. A group of 25 workers was asked to classify images in two binary classes: “Beautiful Aesthetic” and “Ugly Aesthetic.” Finally, eight additional expert photographers with adequate aesthetic perceptions were hired to validate the annotations from the MTurks. For each image, they were asked to agree or disagree with the dominant annotated label. In this regard, images that more than four of the experts agreed on the label were kept in the data set. As a result, the final curated and validated data set contains 12,000 human-annotated food images in two binary “Beautiful Aesthetic” or “Ugly Aesthetic” balanced classes.

Secondly, we fine-tuned the NIMA model on the GPD data set. To keep the aesthetic score (output) in a continuous range as in NIMA, we minimally altered the architecture of the pre-trained model. Specifically, we added a fully connected layer of two neurons (one for each class) after the softmax layer of the mother model; consequently, the second-to-last layer activation function was replaced by a ReLU activation function, mathematically \( \text{ReLU}(x) = \max(0,x) \), and, as usual, the final layer was activated by a softmax activation function. We then split the data into 80% training and 20% validation and followed the same training pipeline as pointed out in Subsection 4.1.1 with the following specifications: we divided the data into batches of 64 examples per batch, and we trained on four epochs. By definition, an epoch is an entire pass through the whole data set. We discovered by trial-and-error that four passes were the optimal number of epochs in this setting, because they maximized the following metrics without overfitting the model. At testing, the model achieved an 89.52% accuracy score, and sufficiently balanced recall and precision scores of 91.86% and 87.89%, respectively.

4.1.3 Calibration of aesthetic scores using temperature scaling of probabilities. Afterwards, we calibrated the probabilities of the model for better probability interpretations. In general, calibrated probabilities help a model match the true likelihood of events. Moreover, the training data can potentially be biased in some hidden fashion, and, in this case, a calibrated probability may significantly fix such an undesirable issue. In addition, neural networks tend to output overconfident probabilities (Guo et al., 2017). Practically, we performed probability calibration with temperature scaling (Jaynes, 1957; LeCun et al., 2015). Temperature scaling is the closest expansion to the Platt scaling (Platt, 1999). In fact, with the GPD test data, we learned a new parameter \( T > 0 \) that would scale...
the logits of the neural network before activation. By definition, logits are the not activated output of a model. Mathematically, for \( k \) classes and a logit vector of predictions \( z_i \), we computed the calibrated probabilities \( \hat{q}_i \) by applying the softmax function \( \sigma_{SM} \) to the logits vector divided by the learned temperature \( T \). In equation (3) we show the mathematical formula.

\[
\hat{q}_i = \max_k \sigma_{SM}(z_i/T)^k
\]

For completeness, \( T \) was optimized while recursively minimizing the cross-entropy loss function between the ground-truth labels and the temperature scaled probabilities computed from the test data. The optimal \( T \) converged to 1.5369. To be rigorous, with the described temperature scaling approach, the class prediction for a generic vector \( \hat{q}_i \) remains unaltered. Then, it is possible to conclude that the accuracy of the model remains unchanged. While, on the other hand, probabilities are more accurate, and reflect more the true aesthetic score of an image.

For each Yelp web-scraped image, we then extracted the two logit values at the last layer, scaled them by the learned temperature value \( T \), and activated them via a softmax activation function to interpret them as probabilities. The two probability values, which sum to 1, express the probability of belonging to either the positive class or the negative class. As the two values express the same meaning, but in two opposite directions, we extracted as aesthetic score the probability of an image to pertain to the positive class, i.e. the “Beautiful Image” class. Lastly, for better interpretation, all the image aesthetic scores were multiplied by 10 to be transformed into the 0–10 range. Figures 2 and 3 illustrate the examples of the top and bottom images for aesthetic scores, respectively.

4.1.4 Aesthetic assessment model performance validation. Validating deep learning models is an essential step in model development. Ideally, a model should be fundamentally able to approximate human performance with a certain degree of confidence. In order to test whether our model performs accurately, we opted for an out-of-bag evaluation survey directly sent to members of the general public. To do so, we randomly picked 20 image pairs such that their aesthetic scores were at least 1.50 apart in each pair, which is about half a standard deviation of the aesthetic scores. Finally, we redacted the survey such that each respondent could select the most visually appealing image in the pair. Moreover, we added a third option granting the possibility to skip the question in case of any doubt of the respondent. We did so with the intent of not forcing a surveyee to express a judgment. Furthermore, we added two attention questions to make the results more robust. Intentionally, we swapped the meaning of two randomly picked questions, asking which image was the ugliest in the pair. Valid attention answers would show the opposite image selection compared to their relative positive ones.

We then sent the survey to 100 people through the Prolific platform, a crowdsourcing platform for running digital research surveys. We chose Prolific because it allows for the effective sampling strategy across the USA to assign the survey to a representative sample. Moreover, we guaranteed that all the 100 workers were paid a fair amount given their contribution to this study. After deleting the answers from inattentive respondents, we scrutinized 66 valid submissions. Out of 20 image pairs, 16 showed agreement between the hired workers’ perceptions and the aesthetic scores, meaning that our model is reliably able to approximate human performance with 80% confidence.
4.2 Photographic attributes

In the restaurant industry, a food image on social media may trigger a potential customer’s different emotions relative to its perception. Previous studies survey many image attributes from the photographic literature that help compare and classify images (Datta et al., 2006; Wang et al., 2013). Hence, we refer to such photographic attributes by defining them and conjecturing how each may be relevant for the purpose of this study. We considered investigating 15 visual attributes and divided them into three major components: color, composition and figure–ground relationship detailed as below (Zhang and Luo, 2018; Zhang et al., 2021a).

4.2.1 Color. Based on previous research, Gorn et al. (1997) recognized two components of image visual appeal: from boredom to excitement, and from tension to relaxation. For each component, the latter is preferred over the former. To study both dimensions, all the images were converted into their relative HSV representation (Levkowitz and Herman, 1993), which is an acronym standing for Hue, Saturation (Chroma) and Value (Brightness). Hue can be defined as the perception of the color portion, ranging from 0° to
360°. Specifically, perceived colors are red, yellow, blue and green, or a combination. Saturation represents the purity and intensity of a color, ranging from 0 to 255. The closer the value of each pixel to 0 saturation, the closer it is to grey on a grey-scale. Finally, the value characterizes the brightness of color; it ranges from 0 to 255, with 0 being completely black and 255 being the brightest:

- **Attribute 1: Brightness.** As pointed out earlier, brightness refers to the overall lightness of an image. Ideally, people would prefer to view brighter images as they convey more pleasure and enjoyment (Valdez and Mehrabian, 1994). As a further consequence, a brighter image makes the visual object displayed in the image easier to spot. As the objects of this study are meal dishes, we conjecture that a higher brightness may overall lead to better image quality. Technically, for each image, we computed the brightness score by averaging the value dimension of each pixel. In Figure 4, we provide an example of a sufficiently bright image and its relative counterexample.
**Attribute 2: Contrast.** Contrast may be considered as a complementary metric to brightness. In fact, it expresses how the illumination is evenly spread across each pixel (Valdez and Mehrabian, 1994). The lower the contrast, the smoother an image is in showing all the details; thus, we hypothesize that the lower the contrast of an image, the higher the aesthetic quality. In computational terms, for each image, we computed the contrast score as the standard deviation of the value dimension. In Figure 4, we provide an example of an image with a sufficiently high contrast value and its relative counterexample.

**Attribute 3: Saturation.** Saturation may be associated with the color intensity and purity of an image. Generally, a higher saturation value is correlated to a more
intense color perception (Gorn et al., 1997; Valdez and Mehrabian, 1994), arousing more joyful feelings. In this regard, we predict that higher saturation is a proxy for higher image quality. In practice, for each image, the saturation score was computed by averaging the saturation dimension of each pixel. Figure 4 shows an example of a correctly saturated image and its relative counterexample.

- **Attribute 4: Clarity.** Clarity represents the concentration of HSV colors in a given image (Levkowitz and Herman, 1993). In fact, clear images display well-defined objects and details in the photographic space, leading to a fluid information transmission to the observer. We suppose that the higher the clarity, the more appealing an image is. Technically, for each image, the clarity score was computed as the proportion of normalized value pixels that exceed the threshold of 0.7. In Figure 4, we provide an example of a clear image and its relative counterexample.

- **Attribute 5: Warm.** Generally, warm hues arouse joyful feelings to an observer. Examples of warm colors span from red to yellow, with all the intermediate shades included. Parallely, cool hues (e.g. green, blue) arouse more relaxation (Gorn et al., 1997; Valdez and Mehrabian, 1994). Computationally, for each image, we calculated the warm level score as the proportion of warm hues in the hue channel lower than 60 or greater than 220. In Figure 4, we provide an example of a warm image and its relative counterexample.

- **Attribute 6: Colorfulness.** We referred to Hasler and Suesstrunk (2003) to calculate how colorful each image is. In contrast to all the previous attributes, the standard RGB color model was adopted. The RGB framework is a three-channel color in which the combination of red, green and blue are mixed. Theoretically, it is possible to define image colorfulness as the departure from a grey-scale image, i.e. a grey-scale image would have 0 colorfulness. For this work, we conjecture that more colorful images lead to better image quality. In practical terms, colorfulness $C$ was computed as follows:

\[
rg = R - G \\
yb = 0.5(R + G) + B
\]

in which $R$, $G$ and $B$ correspond to the red, green and blue channels. Next, we computed the standard deviation and mean of both $rb$ and $yb$:

\[
\sigma_{rgb} = \sqrt{\sigma_{rb}^2 + \sigma_{yb}^2} \\
\mu_{rgb} = \sqrt{\mu_{rb}^2 + \mu_{yb}^2}
\]

Finally, colorfulness $C$ is defined as the combination of mean and standard deviation such that:

\[
C = \sigma_{rgb} + 0.3 \cdot \mu_{rgb}
\]

In Figure 4, we provide an example of a colorful image and its relative counterexample.

4.2.2 Image composition. Image composition is the study of how visual objects are displayed within an image. According to Freeman (2007), how elements are placed leads the observer to promptly recognize the focal point of an image. For this work, we identified the following six attributes.
• **Attribute 7: Diagonal dominance.** According to Grill and Scanlon (1990), for an image to be diagonally dominant, the most salient object should be placed in proximity to the interception of the two diagonals. In this regard, the overall perception leads to a sense of openness. Practically, for each image, we computed the diagonal dominance score by, firstly, identifying the most salient region of an image; secondly, calculating the Manhattan distance between the center of the most salient region and each of the two diagonals. Then, the diagonal dominance score is the minimum value across the two previously calculated values. In our hypothesis, diagonally dominant images should carry a higher aesthetic value. In Figure 5, we provide an example of a diagonally dominant image and its relative counterexample.

• **Attribute 8: Rule of thirds.** According to Krages (2005), any image can be split into nine equal regions with two horizontal and vertical lines for each dimension. The Rule of Thirds asserts that the dominant element should be located in proximity of the intersection of those imaginary lines (Meech, 2004). In this study, we hypothesize that the more an image follows the Rule of Thirds, the more aesthetically pleasing it is. Technically, for each image, the Rule of Thirds score was computed as the minimum distance among the center of the salient region and each of the four intersection points. In Figure 5, we provide an example of an image correctly following this standard and its relative counterexample.

• **Attributes 9 and 10: Horizontal and vertical physical visual balance.** In the photography literature, visual balance is related to the distribution of items (Krages, 2005). Hypothetically, by half-splitting an image either vertically or horizontally, the more that the elements displayed are mirrored into each other from both sides, the more that the image is physically balanced. Symmetric images would be the most accurate example of perfectly balanced images. Linking to the aesthetics concept, there are several studies stressing that humans consider visually balanced images as more aesthetically appealing (Arnheim, 1974; Bornstein et al., 1981). In computational terms, for both vertical and horizontal physical visual balance, we divided each image into ten segments using the SLIC Superpixel algorithm (Achanta et al., 2012), and computed the weighted center of the image. To perform the last operation, we weighed the center of each segment by its relative saliency score. Finally, following Wang et al. (2013), the horizontal physical visual balance was extracted by calculating the distance between the weighted center and the vertical line halving the image. On the other hand, the same approach was applied for the vertical visual balance, considering the horizontal line halving the image instead. In Figure 5 we provide examples of visually balanced images and their relative counterexamples.

• **Attributes 11 and 12: Horizontal and vertical color visual balance.** The same idea from Attributes 9 and 10 was followed to compute color visual balance (Krages, 2005). For the horizontal visual balance, we vertically divided each image into two equal portions. Then, for each pixel on the right-hand side, we took its symmetric counterpart on the left-hand side, and computed the pair’s Euclidean cross-pixels distance. The horizontal color visual balance score is the average of all such pair distances. The same approach was adopted to compute the vertical visual balance, but splitting the image horizontally instead. Perceptually, for all the visual balance attributes, we argue that more balanced images may entail a higher aesthetic value.
In Figure 5, we provide examples of visually balanced images and their relative counterexamples.

4.2.3 Figure–ground relationship. The figure–ground relationship component evaluates an image into two constituents: foreground and background. In practical terms, the figure, which corresponds to the foreground, is the most salient region of an image, a region in which primary elements should catch the eye of an observer. On the other hand, the background refers to the ground, space in which secondary elements should be portrayed. Inherently, research in advertising suggests that images in which figure and ground are clearly separated tend to obtain more attention from viewers (Larsen et al., 2004; Schloss and
Palmer, 2011). To explore this, we further refer to three visual attributes, namely, the size difference, the color difference and the texture difference. For each attribute, we deployed the Grabcut algorithm to separate each image foreground from its related background (Rother et al., 2004). More specifically, Grabcut extracts a black (ground) mask and a white (figure) mask of the same image size.

- **Attribute 13: Size difference.** The size difference attribute measures the difference in the number of pixels between the figure and the ground. Thus, it is possible to measure the proportion of primary and secondary elements displayed. Following Wang et al. (2013), for each picture, we calculated the size difference score by subtracting the number of pixels of the figure from the number of pixels of the ground. Finally, we standardized the value obtained by dividing it by the total number of pixels. In Figure 6, we provide an example of an image with a substantial size difference and its relative counterexample.

- **Attribute 14: Color difference.** The color difference attribute measures the difference in the color between the foreground and the background (Larsen et al., 2004; Schloss and Palmer, 2011). For each picture, we computed the color difference score by calculating the Euclidian distance between the figure and ground RGB vectors. We hypothesize that a higher score correlates to a more aesthetically pleasing image, as the viewer can clearly spot the primary elements displayed. In Figure 6, we provide an example of a colorfully different image and its relative counterexample.

- **Attribute 15: Texture difference.** The texture difference attribute detects the density of the edges of both foreground and background (Larsen et al., 2004; Schloss and Palmer, 2011). For each picture, we deployed the Canny edge detection algorithm to detect the edges in both dimensions (Canny, 1986). Then, the texture difference score was computed as the absolute value of the difference between the foreground edge density and background edge density. In Figure 6, we provide an example of a sufficiently texture-different image and its relative counterexample.

![Figure 6.](image-url)

Set of examples for figure–ground attributes
4.3 Correlation between the aesthetic score and the photographic attributes

We scored each Yelp web-scraped image with the attributes mentioned above. We then conducted a Pearson correlation analysis to understand how the aesthetic scores correlate with such photographic attribute values for the whole set of scraped images. Specifically, we decomposed the aesthetic score into two components:

(1) aesthetic score from images posted by customers; and
(2) aesthetic score from images posted by restaurant owners.

Correlation results are summarized in Table 2. More specifically, for both customer- and owner-posted images, about the majority of the photographic attributes are positively correlated with the aesthetic scores, largely indicating a coherence among them. Furthermore, we also observed that some attributes such as warm and saturation are not correlated or even negatively correlated with the aesthetic scores. We believe that the aesthetic scores may contain composite aesthetic quality of food images that does not have to align with all the photographic attributes.

4.4 Methodological contribution

To summarize, we make two main methodological contributions that can effectively extract both high-level and low-level visual contents from large-scale online food images using computational approaches. On one hand, we propose a deep learning based image quality assessment model that provides a high-level evaluation framework of food aesthetics. It may directly assign aesthetic scores to food images that reflect the image quality without further processing. However, existing hospitality literature adopt deep learning methods to perform basics tasks such as image classification, scene and entity recognition (Cho et al., 2022; Zhang et al., 2021b,c). On the other hand, we use computer vision methods to obtain a large set of low-level visual features that are indicative of aesthetic qualities from the computer science field (Datta et al., 2006; Deng et al., 2017; Machajdik and Hanbury, 2010; Wang et al., 2017).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Customer p-value</th>
<th>Owner p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brightness</td>
<td>0.16***</td>
<td>0.20***</td>
</tr>
<tr>
<td>Saturation</td>
<td>0.00</td>
<td>-0.05***</td>
</tr>
<tr>
<td>Contrast</td>
<td>0.07***</td>
<td>0.10***</td>
</tr>
<tr>
<td>Clarity</td>
<td>0.21***</td>
<td>0.25***</td>
</tr>
<tr>
<td>Warm</td>
<td>-0.04***</td>
<td>-0.06***</td>
</tr>
<tr>
<td>Colorfulness</td>
<td>0.07***</td>
<td>0.039***</td>
</tr>
<tr>
<td>Diagonal dominance</td>
<td>-0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Rule of thirds</td>
<td>0.01</td>
<td>0.06***</td>
</tr>
<tr>
<td>Horizontal PVB</td>
<td>0.04***</td>
<td>-0.06***</td>
</tr>
<tr>
<td>Vertical PVB</td>
<td>0.01**</td>
<td>-0.03</td>
</tr>
<tr>
<td>Horizontal CVB</td>
<td>0.03***</td>
<td>0.02</td>
</tr>
<tr>
<td>Vertical CVB</td>
<td>0.05***</td>
<td>0.04**</td>
</tr>
<tr>
<td>Size difference</td>
<td>-0.04***</td>
<td>-0.02</td>
</tr>
<tr>
<td>Color difference</td>
<td>0.05***</td>
<td>0.04**</td>
</tr>
<tr>
<td>Texture difference</td>
<td>0.02***</td>
<td>-0.03*</td>
</tr>
</tbody>
</table>

Table 2. Correlations between customer and owner aesthetic scores with each photographic attribute and p-value

Notes: Both photographic attributes and the aesthetic score are aggregated by the mean at restaurant level. ***p < 0.01; **p < 0.05; *p < 0.1
We apply these feature extraction methods to characterize the concepts derived from food photography, and thus allow comparing visual features with aesthetic scores.

5. Results
In this section of the work, we discuss the results of a descriptive research to show the performance of the food aesthetics deep learning model correlating with several business variables. The goal is to provide sufficient tangible evidence about the capabilities of deep learning in the gastronomic sector. To do so, we performed many ANOVA among the food aesthetic scores and the business variables introduced in Section 3. Specifically, we considered the variation of the aesthetic scores among different heterogeneous groups such as the stars rating, the price level, whether the restaurant is a chain, and different restaurants serving different cuisine styles. Not less importantly, we reviewed how the image poster might affect such aesthetic score variations. We structure the workflow of this section as follows: firstly, before fitting any model, we document how data were pre-processed; secondly, we discuss and report the modeling results.

5.1 Data pre-processing
We averaged the aesthetic scores at a restaurant level. For example, for a given restaurant, given images A, B and C with aesthetic scores of 5, 7 and 9, respectively, the average aesthetic score of such restaurant would be 7. From now on, we will use the terms “average aesthetic score” and “aesthetic score” interchangeably. In addition, for each restaurant, we decomposed the aesthetic score as we did in Subsection 4.3:

- aesthetic score from customer-posted images; and
- aesthetic score from owner-posted images.

As for the business variables, we categorized the stars rating and the price level variables as follows. For the former, the average rating is distributed from 1 to 5. To categorize it, we selected as thresholds 2.5 and 4, and restaurants rated lower (or equal) and higher (or equal) were classified as low-rated and high-rated, respectively. In the middle of the two thresholds, we identified mid-rated restaurants. As for the price level variable, according to the Yelp price subdivision mentioned in Section 3, we specified restaurant price levels as “$” being classified as cheap, “$$” being classified as affordable, and “$$$” and “$$$$” together being classified as expensive. In Table 3, we provide the summary statistics of the variables in the data set.

5.2 Food aesthetics across different restaurant business variables
We performed several one-way ANOVA models to explore the variation of the food aesthetic scores across different types of restaurant variables. For each one-way ANOVA model discussed later, the null hypothesis states that all the sub-group distributions are statistically equal. Whereas, the alternative hypothesis states that at least one of the distributions is statistically different. In case the alternative hypothesis was true, we also performed Tukey-HSD tests to compare the means across all the possible group pair combinations. Here, the null hypothesis states that the two means are statistically equal, whereas the alternative hypothesis states that they are statistically different.

As far as the stars rating, we rejected the ANOVA null hypothesis, meaning that differently rated restaurants have different aesthetic score distributions ($p$-value 0.000). We observed that low-rated restaurants (2.18) had a higher average aesthetic score with respect
to the mid-rated (1.56) and high-rated (1.94) ones. Aesthetic scores are reported in brackets. As for the Tukey-HSD test, we tested the average aesthetic score of high-rated versus low-rated restaurants, high-rated versus mid-rated restaurants and low-rated versus mid-rated restaurants. For the first pair, we could not reject the null hypothesis (p-value 0.315). Thus, our first descriptive result (I) states that high-rated restaurants do not necessarily display more beautiful images than the low-rated ones, and their scores are statistically equivalent. On the other hand, we rejected the null hypothesis for the other pairs, meaning that the differences of the aesthetic scores among high-rated and low-rated restaurants compared to the mid-rated ones are statistically different.

Next, we applied the same analysis method across differently priced restaurants. Explicitly, we tested the distribution of the aesthetic scores across cheap (1.96), affordable (1.73) and expensive (2.15) restaurants. However, we could not reject the ANOVA null hypothesis (p-value 0.110). As for the consequences of this result, there was no need to perform Tukey-HSD tests. Thus, we summarize our second descriptive result (II) as follows: there is no significant aesthetic score variation across differently priced restaurants, meaning that the price charged does not influence the photographers to take, for example, more (or less) beautiful images.

Table 3. Summary statistics of the variables in the data set

<table>
<thead>
<tr>
<th>Name</th>
<th>Categories</th>
<th>Count</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average aesthetic score</td>
<td>–</td>
<td>577</td>
<td>1.85</td>
<td>1.46</td>
</tr>
<tr>
<td>Brightness</td>
<td>–</td>
<td>577</td>
<td>143.41</td>
<td>13.98</td>
</tr>
<tr>
<td>Contrast</td>
<td>–</td>
<td>577</td>
<td>53.54</td>
<td>5.61</td>
</tr>
<tr>
<td>Clarity</td>
<td>–</td>
<td>577</td>
<td>0.37</td>
<td>0.08</td>
</tr>
<tr>
<td>Warm</td>
<td>–</td>
<td>577</td>
<td>0.86</td>
<td>0.06</td>
</tr>
<tr>
<td>Colorfulness</td>
<td>–</td>
<td>577</td>
<td>115.86</td>
<td>14.96</td>
</tr>
<tr>
<td>Diagonal dominance</td>
<td>–</td>
<td>577</td>
<td>-49.05</td>
<td>7.96</td>
</tr>
<tr>
<td>Rule of thirds</td>
<td>–</td>
<td>577</td>
<td>-83.18</td>
<td>8.64</td>
</tr>
<tr>
<td>Horizontal PVB</td>
<td>–</td>
<td>577</td>
<td>-7.87</td>
<td>1.49</td>
</tr>
<tr>
<td>Vertical PVB</td>
<td>–</td>
<td>577</td>
<td>-5.51</td>
<td>0.89</td>
</tr>
<tr>
<td>Horizontal CVB</td>
<td>–</td>
<td>577</td>
<td>-0.58</td>
<td>0.04</td>
</tr>
<tr>
<td>Vertical CVB</td>
<td>–</td>
<td>577</td>
<td>-0.57</td>
<td>0.04</td>
</tr>
<tr>
<td>Size difference</td>
<td>–</td>
<td>577</td>
<td>-0.18</td>
<td>0.1</td>
</tr>
<tr>
<td>Color difference</td>
<td>–</td>
<td>577</td>
<td>78.51</td>
<td>12.65</td>
</tr>
<tr>
<td>Texture difference</td>
<td>–</td>
<td>577</td>
<td>8.72</td>
<td>8.24</td>
</tr>
</tbody>
</table>
As far as the cuisine types, we tested the variation of the aesthetic scores across restaurants serving Italian (1.6), Mexican (1.68), American-New (2.01), American-Trad (2.07), Chinese (1.91) and Japanese cuisines (1.42). We rejected the ANOVA null hypothesis at the 0.95 confidence level (p-value 0.035). However, for the Tukey-HSD test, only the pair Mexican versus American-Trad is statistically significant at the 0.95 confidence level out of 15 pairs tested (p-value 0.044), leading us to define the third descriptive result (III): the cuisine type is not a factor that image-takers take into consideration for more (or less) beautiful images.

Finally, we analyzed the variation of the aesthetic scores across chain (2.30) versus non-chain (1.62) restaurants. As a result, we rejected the ANOVA null hypothesis (p-value 0.000). Thus, we can formulate our fourth (IV) descriptive result stating that chain restaurants tend to have higher aesthetic scores than the non-chain ones. We may conceptualize that, even if the aesthetic scores of both categories tend to be clustered in the low end of the range, chain restaurants may experience a cross-sharing of images to display in the platform across the multiple restaurants in the region, and, more importantly, they may hire professional photographers to take more aesthetically appealing images of the dishes they sell. Consequently, the business owner may upload such images on the platform and use them as a promotional tool to attract new customers. This last hypothesis leads us to study the aesthetic score divergence across images posted by the business owners (3,670 images) or by the customers (46,348 images). In fact, the averages of these two groups of picture amount to 6.20 and 1.21, respectively, speculating that business owners tend to upload more beautiful images. Hence, we tested the variation of the distribution of the aesthetic scores of these two groups. Inherently, we rejected the ANOVA null hypothesis (p-value 0.000). Thus, we can conclude that owner-posted images statistically outscore customer-posted images.

5.3 Role of image posters
Again, it may be hypothesized that business owners have more positive incentives to post more appealing images, as their intentions may be to attract customers by displaying the best images on the platform. To do so, restaurateurs may hire professionals so that a superior standard of quality is guaranteed. On the contrary, the same may not be valid for customers, who might not have any incentive to do so. Moreover, most random customers may not own professional photographic equipment, and this disadvantage lowers the quality of their images.

To deepen the analysis, we performed several two-way ANOVA tests to explore the interaction on the aesthetic scores of the poster variable (business owners or customers) with the stars rating, the price level, the chain status and the cuisine types. For the two-way ANOVA models, the null hypothesis states that the interactions are statistically equal, whereas the alternative hypothesis states that they are statistically different. Results are summarized in Table 4.

Furthermore, as shown in Table 5, we performed several Tukey-HSD post hoc tests just for the significant two-way ANOVA models to check the variation of the aesthetic scores across each possible pair combination. Methodologically, all Tukey-HSD null hypothesis state that all the sub-group pair combinations are statistically equal as in Subsection 5.2. However, the aesthetic scores regarding each pair are decoupled from the image posters. For example, for the stars rating, the high-rated class is divided into high-rated/customers and high-rated/owners. The same example logic may be extended to all the other groups and sub-groups.

5.3.1 Stars rating. From the Tukey-HSD tests, we found that all the pair combinations were statistically significant at the 0.05 significant level, except for the pair low-rated/
customers and mid-rated/customers. In particular, considering customer-posted images, high-rated restaurants display a higher aesthetic score versus the low-rated (+0.88) and mid-rated (+0.82) ones. The statistically significant mean differences across the categories are reported in brackets. However, the opposite is true concerning owner-posted images. In fact, low-rated restaurants show a higher aesthetic score with respect to mid-rated (+1.36) and high-rated (+2.50) restaurants. Last but not least, it is important to remark on the superiority of owner-posted images when comparing poster–heterogeneous pairs. Numerically, low-rated, mid-rated and high-rated owner-posted related images achieve aesthetic scores of 7.80, 6.44 and 5.30, respectively; whereas, customer-posted related images achieve scores of 0.80, 0.86 and 1.68, respectively.

5.3.2 Price level. We found that nine pair combinations out of 15 were statistically significant at the 0.01 significance level, while the other six combinations were not at any other significance level. Concerning owner-posted related images, neither expensive (6.49) nor affordable (6.06) nor cheap (6.31) restaurants were statistically different. The same outcome was valid for customer-posted related images, but with lower aesthetic scores of 1.58, 1.48 and 0.91, respectively. However, for each price level, we again found the higher appeal of owner-posted related images versus customer-posted related images. In this regard, concerning expensive, affordable and cheap restaurants, owner-posted related images outscore customer-posted related images by +4.91, +4.58 and +5.40, respectively.

5.3.3 Chain restaurants. For chain and non-chain restaurants, we found that owner-posted related images score significantly higher than customer-posted related images (+6.68 and +3.73 statistically significant mean differences, respectively). In addition, we also observed that all the other combinations were statistically significant. In particular,
concerning owner-posted related images, they score higher for chain restaurants compared to the non-chain ones (±2.34 statistically significant mean difference); whereas, the opposite is true for the customer-posted related images (−0.62 statistically significant mean difference).

6. Discussion and conclusions
6.1 Conclusions
Nowadays, it is not uncommon for both customers and restaurants to post food images on social media, although they may have different objectives. On the one hand, posting food

<table>
<thead>
<tr>
<th>Stars-Poster Group 1</th>
<th>Tukey-HSD group tests</th>
<th>Stars-Poster Group 2</th>
<th>Mean diff</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-rated/Customers (1.68)</td>
<td>High-rated/Owners (5.30)</td>
<td>3.62***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>High-rated/Customers (1.68)</td>
<td>Low-rated/Customers (0.80)</td>
<td>−0.88**</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>High-rated/Customers (1.68)</td>
<td>Low-rated/Owners (7.80)</td>
<td>6.12***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>High-rated/Customers (1.68)</td>
<td>Mid-rated/Customers (0.86)</td>
<td>−0.82**</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>High-rated/Customers (1.68)</td>
<td>Mid-rated/Owners (6.44)</td>
<td>4.76***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>High-rated/Owners (5.30)</td>
<td>Low-rated/Customers (0.80)</td>
<td>−4.51***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>High-rated/Owners (5.30)</td>
<td>Low-rated/Owners (7.80)</td>
<td>2.50***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>High-rated/Owners (5.30)</td>
<td>Mid-rated/Customers (0.86)</td>
<td>−4.44***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>High-rated/Owners (5.30)</td>
<td>Mid-rated/Owners (6.44)</td>
<td>1.14***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Low-rated/Customers (0.80)</td>
<td>Low-rated/Owners (7.80)</td>
<td>7.00***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Low-rated/Customers (0.80)</td>
<td>Mid-rated/Customers (0.86)</td>
<td>0.07</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Low-rated/Customers (0.80)</td>
<td>Mid-rated/Owners (7.80)</td>
<td>5.64***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Low-rated/Owners (7.80)</td>
<td>Mid-rated/Customers (0.86)</td>
<td>−6.94***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Low-rated/Owners (7.80)</td>
<td>Mid-rated/Owners (6.44)</td>
<td>−1.36***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Mid-rated/Customers (0.86)</td>
<td>Mid-rated/Owners (6.44)</td>
<td>5.58***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Price-Poster Group 1</td>
<td>Price-Poster Group 2</td>
<td>Mean diff</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>Affordable/Customers (1.48)</td>
<td>Affordable/Owners (6.06)</td>
<td>4.58***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Affordable/Customers (1.48)</td>
<td>Cheap/Customers (0.91)</td>
<td>−0.57</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td>Affordable/Customers (1.48)</td>
<td>Cheap/Owners (6.31)</td>
<td>4.84***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Affordable/Customers (1.48)</td>
<td>Expensive/Customers (1.58)</td>
<td>0.10</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Affordable/Customers (1.48)</td>
<td>Expensive/Owners (6.49)</td>
<td>5.01***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Affordable/Owners (6.06)</td>
<td>Cheap/Customers (0.91)</td>
<td>−5.15***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Affordable/Owners (6.06)</td>
<td>Cheap/Owners (6.31)</td>
<td>0.26</td>
<td>0.895</td>
<td></td>
</tr>
<tr>
<td>Affordable/Owners (6.06)</td>
<td>Expensive/Customers (1.58)</td>
<td>−4.48***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Affordable/Owners (6.06)</td>
<td>Expensive/Owners (6.49)</td>
<td>0.43</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Cheap/Customers (0.91)</td>
<td>Cheap/Owners (6.31)</td>
<td>5.40***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Cheap/Customers (0.91)</td>
<td>Expensive/Customers (1.58)</td>
<td>0.67</td>
<td>0.756</td>
<td></td>
</tr>
<tr>
<td>Cheap/Customers (0.91)</td>
<td>Expensive/Owners (6.49)</td>
<td>5.58***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Cheap/Owners (6.31)</td>
<td>Expensive/Customers (1.58)</td>
<td>−4.74***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Cheap/Owners (6.31)</td>
<td>Expensive/Owners (6.49)</td>
<td>0.18</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Expensive/Customers (1.58)</td>
<td>Expensive/Owners (6.49)</td>
<td>4.91***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Chain-Poster Group 1</td>
<td>Chain-Poster Group 2</td>
<td>Mean diff</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>Chain/Customer (0.86)</td>
<td>Chain/Owners (7.54)</td>
<td>6.68***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Chain/Customer (0.86)</td>
<td>Non-chain/Customer (1.48)</td>
<td>6.02**</td>
<td>0.022</td>
<td></td>
</tr>
<tr>
<td>Chain/Customer (0.86)</td>
<td>Non-chain/Owners (5.20)</td>
<td>4.35***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Chain/Owners (7.54)</td>
<td>Non-chain/Customer (1.48)</td>
<td>−6.06***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Chain/Owners (7.54)</td>
<td>Non-chain/Owners (5.20)</td>
<td>−2.33***</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Non-chain/Customer (1.48)</td>
<td>Non-chain/Owners (5.20)</td>
<td>3.73***</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Summary of all the pairs Tukey-HSD tests

Notes: In brackets the average aesthetic score for each combination; ***p < 0.01, **p < 0.05, * p < 0.1

Exploring food aesthetics portrayed on social media
images on social media tends to increase customers’ enjoyment of the dining experience (Diehl et al., 2016; Zhu et al., 2019). On the other hand, restaurants are eager to share impressive images on social media as a low-cost digital marketing channel to gain exposure and generate interest for potential customers (Oliveira and Casais, 2019). Essentially, this paper presents an analytical approach to examine the current practices of food image posting activities on social media platforms, with a special focus on the visual content of those images. In particular, we characterize the aesthetics of food images for each restaurant and identify the discrepancies between restaurants across several dimensions. We find that restaurants with different rating levels, cuisine types and chain status have different aesthetic scores. More specifically, results show that chain restaurants, restaurants with lower ratings and restaurants serving American cuisine tend to receive higher aesthetic scores.

To explain these findings, we further evaluate the aesthetic appeal from customer-posted food images and restaurant owner-posted food images, respectively. Overall, we show that owner-posted food images have higher aesthetic scores than customer-posted food images across all restaurant types, implying that restaurant owners have intentionally posted high-quality food images on social media for marketing purposes. In particular, we perform post hoc comparison analyses to scrutinize further the differences in food images between the groups of posters and restaurants. Results show that restaurants with higher ratings and non-chain status tend to receive food images with higher aesthetic scores from the customers, indicating that user-posted food images reflect their perceptual restaurant quality (Atwal et al., 2019). By contrast, owners of lower-rated restaurants and chain restaurants tend to post food images with higher aesthetic scores, while no significant differences are identified between restaurants with different price levels (Holmberg et al., 2016). This indicates that owners of lower-quality restaurants are likely to more proactively perform social media marketing by posting high-quality professional food images (Lepkowska-White, 2017).

6.2 Theoretical and methodological implications
This paper offers several theoretical and methodological contributions in hospitality literature about understanding the social media activities of the restaurant industry. The first contribution is to provide a novel methodological framework to assess the aesthetics in food images. By leveraging advanced computer vision and deep learning techniques, we are able to extract meaningful informational content from food images that pertain to the perception of food, which is a long-acknowledged phenomenon as shown in the practices of food photography. Instead of relying on a multitude of standard attributes stemming from the photographic literature, our method leverages a unique one-take-all score, which is more straightforward to understand and more accessible to correlate with other target variables such as restaurants’ page metrics (number of visitors, likes, etc.). We provide an open-source Python package that includes a pre-trained model for generating aesthetic scores and photographic attributes for food images at: https://bit.ly/3zbqLJY.

The second contribution is to bridge the studies on food presentation from cognitive science literature to describe the sensory gastronomic experience by demonstrating its usefulness in characterizing the aesthetic quality of real-world food images. Given that the proposed method is applied to large quantities of food images with strong heterogeneity from different types of customers and restaurants, it seems to robustly capture the inherent aesthetics that indicate the gastronomic experience from cognitive science. This would allow researchers to knowingly employ the method to evaluate the quality of food images in a wide range of settings.
This paper also contributes to the expanding exploration of underlying objectives of image-posting on social media by different stakeholders. In particular, we identify interesting discrepancies between customer- and owner-posted food images, which may reflect the role of projected and perceived images of tourism destinations, respectively (Andreu et al., 2000; Marine-Roig and Ferrer-Rosell, 2018). As such, this paper extends the understanding about projected and perceived gastronomic experiences that highlight the role of image posters in tourism and hospitality research (Ren et al., 2021; Stepchenkova and Zhan, 2013; Zhang et al., 2021c).

6.3 Practical implications
Armed with the findings from this paper, we present several important practical implications from the viewpoint of restaurants and social media platforms. Firstly, as we show that restaurant owners tend to post food images with higher aesthetic quality than their customer base, such an outcome may be caused by an asymmetry of incentives between them. For example, restaurant owners may consider posting more visually appealing images as their social media marketing strategies are to promote engagement with customers and influence their choices (Oliveira and Casais, 2019; Spence et al., 2016). On the contrary, customers may not be naturally interested in the quality of their posted images because they do not receive any reward for doing so. To overcome this issue, restaurant owners may choose to incentivize their customers to write not only informative reviews but also share visually appealing food images of their dining experiences on social media (Li and Xie, 2020). For example, restaurants may consider organizing food photography challenges on social media to allow their customers to become participants of the contest or collaborate with influencers to post high-quality food images and allow their followers to imitate.

Secondly, we also note significant discrepancies between owners of different types of restaurants. We demonstrate that chain restaurants indeed have coordinated social media marketing strategies by posting professionally shot branded food images as would be found in an advertisement (Holmberg et al., 2016). However, high-rated, non-chain restaurants seem to be lagging behind, and as such, should devise more proactive image-based social media marketing strategies. For example, Stapinski (2013) recorded high-end chefs across the world intended to ban food photography from their restaurants back in 2013 in order not to disrupt the ambiance. But now, most restaurants would like to be featured as the “most instagrammable restaurants” on various channels. For example, during the COVID-19 pandemic, even Michelin-star restaurants started to share food images from their home kitchens on social media to keep engaging with their customers (MICHELIN Guide, 2020).

Thirdly, social media platforms may consider designing incentive mechanisms for encouraging users to post food images with high aesthetics. For example, Yelp confers “elite” badges to its users who are recognized to share well-written reviews and high-quality images to receive benefits. Hence, social media platforms may similarly benefit from learning aesthetics from food images to filter and display high-quality images in prominent spots.

Lastly, the opportunities for using increasing social media activities to monitor and manage gastronomic experiences require the restaurant industry to get empowered with innovative data-driven strategies (Akter et al., 2021; Mariani and Wamba, 2020), enabling big data analytics capabilities to understand the gastronomic experiences derived from emerging data sources and to promote the customer engagement.
6.4 Limitations and future research

A few limitations arise from this study, with the majority of them stemming from the data used both to answer the research question and to train the deep learning aesthetic model. Hence, in this section, we will concisely discuss each. Firstly, to avoid sampling bias, we selected the Yelp official data set as our primary data, and we restricted such data to the city of Las Vegas only. However, we lack the knowledge of whether the results can be generalized to other areas. This limitation exists primarily because of time constraints and computational costs. Specifically, web-scraping for gathering complete information regarding each restaurant is computationally costly, such that extending the research question to a set of other areas would incur high extra costs. Secondly, as we do not have the ground truth labels of food images used in this paper, we mainly rely on transfer learning and labeled data from other researchers to build and train the aesthetic model. This pipeline enables us to train the model by leveraging existing knowledge about aesthetics to speed up the training process; however, such a pipeline also poses some constraints. Transfer learning may often suffer the problem of “negative transfers,” meaning that there may be some dissimilarities between the initial (pre-trained model) and the target problems. This may lead our model’s predicted aesthetic scores to be potentially biased toward the other labeled data sets.

However, to the best of our knowledge, there are no systematic standards among the scientific community to address the problem of such dissimilarity. The alternative option that we considered was to build our data set of images and aesthetic labels by using a crowd-sourced service. Such an alternative would have granted us more flexibility in collecting the data and building the model. However, such a procedure would have been time and computationally costly. Moreover, we would need to address several potential biases arising from the labeling process in the data. Finally, when analyzing the results, we do not aim to provide a causal interpretation of the aesthetic scores to affect restaurants across different dimensions. This limited us in formulating sufficiently strong conclusions, thus making the results of this research descriptive.

For future research, this work may be naturally extended to the following avenues. Firstly, given that other factors such as the restaurant ambiance may also influence the customers’ gastronomic experience, ambience aesthetic models can be built by similarly training and evaluating ambiance images of restaurants posted on social media. Consequently, we would be able to characterize the ambiance aesthetics and examine whether there are any discrepancies between different types of restaurants and image posters. Secondly, we would also study how customer reviews interact with the associated images on social media. For example, key metrics such as the number of “likes” or “upvotes” received by customer reviews can be correlated with the informativeness of reviews using textual analysis and the aesthetics of embedded images. Lastly, methodological contributions can be further reinforced, e.g. by developing a deep learning model to beautify food images. Then, we may perform an experimental study to understand whether improved aesthetics of food images would also lead to more engagement on social media.

Note

1. Yelp uses a naming convention of “name–location–number” to uniquely identify each restaurant. For example, pizza-hut-las-vegas-8 would indicate the 8th Pizza Hut restaurant in Las Vegas. Therefore, it would be possible to easily identify chain status of each restaurant by checking whether the maximum number of the restaurant name exceeds 5.
References

Exploring food aesthetics portrayed on social media


Piqueras-Fiszman, B., Alcaide, J., Roura, E. and Spence, C. (2012), “Is it the plate or is it the food? Assessing the influence of the color (black or white) and shape of the plate on the perception of the food placed on it”, Food Quality and Preference, Vol. 24 No. 1, pp. 205-208.


Corresponding author
Qiwei Han can be contacted at: qiwei.han@novasbe.pt