AS TASTY AS IT LOOKS: EXPLORING THE IMPACT OF IMAGE AESTHETICS ON FOOD DELIVERY SALES

Pengfei Ma
School of Management
Fudan University
670 Guoshun Road, Yangpu District, Shanghai 200433, China
19110690019@fudan.edu.cn

Cheng Zhang
School of Management
Fudan University
670 Guoshun Road, Yangpu District, Shanghai 200433, China
zhangche@fudan.edu.cn

Yifan Dou[1]
School of Management
Fudan University
670 Guoshun Road, Yangpu District, Shanghai 200433, China
yfdou@fudan.edu.cn

Hong Ling
School of Management
Fudan University
670 Guoshun Road, Yangpu District, Shanghai 200433, China
hling@fudan.edu.cn

ABSTRACT

Visual presentation is critical in online marketplaces where customers rely on images to evaluate products without physical interaction. This study examines the primary and contingent effects of image aesthetics on product sales by analyzing 5,074 dish images and 506,553 orders from 71 merchants on a leading food delivery platform. We classify signals into *promotional* and *reputational* types, with product images serving as promotional signals that help reduce customers' product uncertainty. Our hypothesis posits that additional information provided by the platform or merchants may either substitute or complement the value of these images, depending on the signal type (i.e., promotional or reputational). To address limitations associated with subjective ratings or low-level image features, we employ a deep learning algorithm to quantify image aesthetics more effectively. Our results show that image aesthetics positively influence sales. This effect is further strengthened by reputational signals (e.g., platform certification) but weakened by other promotional signals (e.g., advertisements). These findings highlight complementary and substitutive effects between different signal types, offering both theoretical insights and practical guidance for visual design strategies in e-commerce.

Keywords: Image aesthetics; Product uncertainty; Product sales; Moderating effect; Complementarity

1. Introduction

In online marketplaces, visual appeal has emerged as a critical differentiator, as customers make purchasing decisions based on limited information and without physical interaction. Among various presentation forms, image aesthetics are particularly influential in shaping customer perceptions and driving sales. Unlike traditional text-based

Cite: Ma, P., Zhang, C., Dou, Y., & Ling, H. (2025, May). As tasty as it looks: Exploring the impact of image aesthetics on food delivery sales. *Journal of Electronic Commerce Research*, *26*(2).
[1] Corresponding author.

information, visually engaging product images benefit from the well-documented picture superiority effect (Clark and Paivio, 1987; Unnava and Burnkrant, 1991), enabling consumers to process visual information more efficiently. Consequently, e-commerce merchants are increasingly investing in enhancing the aesthetic quality of product images to capture customer attention and positively influence purchasing behavior.

The importance of image aesthetics is especially pronounced on food delivery platforms, where product uncertainty is heightened. As experiential goods, food items differ from many other products in that their quality cannot be fully assessed until after purchase. A product image, like the one shown in Figure 1, engages customers' senses and stimulates their appetite, thereby influencing purchase intention. Further evidence suggests that images of food elicit stronger neural responses from customers than images of other products, underscoring the crucial role of food images in reducing product uncertainty (Versace et al., 2019). In this context, visually appealing food images function as primary signals of product quality, helping to reduce uncertainty and guide customer decision-making. The rapid growth of online food ordering, which generated $553.5 billion in revenue in 2023 and accounted for 15.4% of total e-commerce revenue, highlights the need for effective visual presentation strategies.[2] Thus, understanding the impact of image aesthetics on customer behavior is essential for both researchers and practitioners, especially in industries marked by high product uncertainty.

Figure 1. A food item example (Peking Duck) on a food delivery platform

Among the various dimensions of a customer's perception of product images, aesthetics has gained increasing importance across a wide range of products (Bloch et al., 2003). Aesthetic perception is deeply rooted in human nature, reflecting a fundamental aspect of sensibility that can heavily influence consumer decisions, often surpassing other visual elements. Aesthetics, defined as the study of human cognition and emotional response to the perception of beauty (Palmer et al., 2013), embodies the visual appeal that distinguishes products from competitors, enhances product recognition, and serves as a symbolic cue aiding consumers in their evaluation (Bloch et al., 2003). Prior research has shown that aesthetics influence consumer evaluations (Hagtvedt and Patrick, 2008), preferences (Schenkman and Jönsson, 2000), and loyalty (Cyr et al., 2006).

Despite its importance, the effect of image aesthetics on product sales remains inconclusive. While some studies have found a positive impact of image aesthetics on demand for properties (e.g., Zhang et al., 2022b), others have reported no significant effect (He et al., 2023; Zhang and Luo, 2023). A potential explanation for these mixed results lies in the perceived credibility of merchant-provided images. When merchants use images to promote their products, the incentive to boost sales may undermine the trustworthiness of these visuals, reducing their effectiveness in alleviating product uncertainty (Goh et al., 2013; Guan et al., 2023). These contrasting findings highlight the need to re-examine the role of image aesthetics, particularly in contexts where visual cues play a critical role in consumer decision-making. By addressing these gaps, this study aims to provide a deeper understanding of how image aesthetics influence product sales.

Furthermore, in the context of food delivery platforms, there exists an urgent need for a stable and standardized metric to quantify the visual appeal magnitude of food images. Current research has observed the surge in sales attributed to improved visual design in digital menus (Brewer, 2021; Le et al., 2023). Vermeir and Roose (2020) also assert that these visual design cues could affect customers' purchase intention through internal psychological processes including flavor expectation (Kpossa and Lick, 2020) or tastiness perception (Liu et al., 2022a). Practitioners in the

[2] https://www.statista.com/outlook/emo/ecommerce/worldwide#revenue

food delivery sector highlight the significance of high quality food photographs, which have been linked to a 15% increased sales (CLAID.AI, 2024; Pierce, 2022). However, some findings are based on surveys where customers either imagine the presence of food images or rate provided pictures, leading to arbitrary and inconsistent outcomes. Other studies rely on various pixel-based features such as color (Kpossa and Lick, 2020) or saturation (Liu et al., 2022a), which lack a holistic and unified evaluation of the visual appeal of food images. By employing computational method to assess the aesthetics of food images, we can quantify subjective perception, making the measurement more objective, stable and scalable.

Our study is also among the first to explore how other signals interact with image aesthetics to influence customer purchase decisions (Zhou et al., 2022). While a single signal can have a significant impact on consumers (Connelly et al., 2011), the presence of additional signals from merchants or third parties can alter the effectiveness of the focal signal, resulting in either substitutive or complementary effects (Bapna, 2019; Price and Dawar, 2002). Similar moderating effects have been observed in contexts such as equity investment (Bapna, 2019), product recommendations (Xu et al., 2020), and online healthcare markets (Zhou et al., 2022). Despite the growing body of empirical research on the combined effects of signals, there is a lack of theoretical development regarding the boundary conditions that distinguish substitutive signals from complementary ones. Establishing a clear and generalizable framework for these boundary conditions is crucial, particularly for extending the analysis to a wider range of signals, including image aesthetics in e-commerce.

In response, this paper categorizes signals into two distinct types: promotion signals and reputation signals, based on their source, operability, and credibility. Specifically, promotion signals are initiated by merchants to actively promote their offerings, using product descriptions, images, or advertisements to attract customers (Dimoka et al., 2012; Zhou et al., 2022). In contrast, reputation signals are generated from unbiased third parties, such as platform verifications or customer reviews, which provide an impartial assessment of the merchant's quality (Connelly et al., 2011; Wang et al., 2019). These categories reflect two aspects of information asymmetry: merchant-driven promotion versus third-party certification. We propose that signals within the same category (e.g., two promotional signals) tend to substitute for each other, reducing their combined impact. Conversely, signals across different categories (e.g., a promotional signal and a reputation signal) are likely to complement each other, enhancing their overall effect. Given that product images serve as promotion signals, we expect their impact to be moderated by additional platform information: promotional signals should weaken the effect of image aesthetics, while reputation signals should strengthen it. This framework offers a fresh perspective on the role of image aesthetics in e-commerce, highlighting its interplay with other signals in shaping customer perceptions and purchase decisions.

A key technical challenge in assessing image aesthetics lies in the limitations of existing methods for extracting visual features, which often result in potential measurement errors. Most prior studies focus on low-level visual features—such as illumination, shape, color, and texture—that are directly derived from pixel-level details (Sample et al., 2020). While these features capture basic visual properties (He et al., 2023; Wang et al., 2016; Zhang et al., 2022b), they fail to reflect the holistic perception of an image, particularly its aesthetic appeal, which involves a more comprehensive understanding. In contrast, high-level features relate to a semantic interpretation of the image, such as the conveyed sentiment or emotional response (Hou et al., 2023; Rao et al., 2019). However, existing approaches often lack the capability to extract these complex features effectively. Additionally, some studies rely on subjective ratings or perceptual measures of aesthetics (Hagtvedt and Patrick, 2008; Schenkman and Jönsson, 2000; Yamamoto and Lambert, 1994), which are prone to individual biases and limited scalability. As such, there is a pressing need for a more robust method for evaluating image aesthetics that can capture both low-level and high-level features, providing a holistic assessment aligned with human visual perception.

To address this challenge, we overcome the limitations of traditional methods by utilizing a deep learning model trained on human aesthetic perception data to extract high-level aesthetic features from images. The model predicts a distribution of aesthetic ratings based on human evaluations, providing a holistic assessment of the image's visual appeal. This approach ensures stability and scalability, allowing for application across diverse image datasets. We then analyze the impact of image aesthetics on product sales using a dataset of 5,074 dish images and 506,553 orders from 71 merchants on a food delivery platform. To address potential endogeneity concerns, we apply generalized propensity score matching (GPSM) and instrumental variable (IV) methods, ensuring the robustness of our findings.

In summary, we first employ a CNN to evaluate the dish image aesthetics as our key treatment variable. The impact of image aesthetics on product sales is significantly positive, exceeding 4.8%, and remains robust after addressing potential endogeneity through GPSM and IV, as well as measurement error using low-level image features. Additionally, we conduct moderation analysis based on the classification of two signal types. Our results show that the effect of image aesthetics is attenuated in the presence of promotion signals (e.g., advertisements and insurance), whereas it is strengthened by reputation signals (e.g., platform certification and new opening status).

Our analysis yields several interesting findings in the context of food delivery platforms. First, we propose and empirically validate the classification of signal categories as a boundary condition that determines whether the promotional signal of image aesthetics substitutes for other promotional signals or complements reputation signals in predicting product sales. Second, we advance the measurement of image aesthetics by introducing a novel deep learning approach, offering a more stable and scalable alternative to traditional methods. Finally, we provide new insights into the contingent effects of image aesthetics, highlighting its interaction with diverse types of information on e-commerce platforms.

The subsequent sections review the related research streams and propose our hypotheses, followed by an exposition of the algorithms for aesthetic measurement. We then present the data and perform regression analyses to explore the impact of aesthetics on product sales, along with other contingent factors that carry potential influences. The last section discusses the contribution to the literature and the implications to practitioners.

2. Literature Review

2.1. Signals Underlying Product Uncertainty

Information asymmetry widely exists in online markets where customers don't have access to enough amount of information to make a transaction (Akerlof, 1970). Signaling theory explains how different signals reduce the information asymmetry and enable a successful transaction (Connelly et al., 2011). Signals are observable pieces of information that deliver those inaccessible characteristics of a merchant or product relevant to the transaction (Connelly et al., 2011; Zhou et al., 2022). Thereby, various characteristics can be used to classify various signals. Current literature has employed categorization of description vs. performance (Dimoka et al., 2012), description vs. demonstration (Zhou et al., 2022) to clarify different roles of various signals.

Online customers face uncertainty from sellers and products simultaneously. The early literature on uncertainty in online markets typically adopts the seller's perspective to predict seller's potential opportunistic behaviors (Dimoka et al., 2012; Pavlou et al., 2007). Solutions are generally categorized as reputation and trust (Dellarocas, 2003; Pavlou et al., 2007). On the other hand, product uncertainty takes the perspective of the product, encompassing the buyer's challenge in evaluating the product and predicting its future performance (Dimoka et al., 2012). Typical product uncertainty includes description uncertainty, which refers to the integrity and accuracy of the product descriptions, and performance uncertainty, which refers to performance and persistence of the purchased product (Dimoka et al., 2012). Product fit uncertainty are also explored in latest research (Hong and Pavlou, 2014; Xu et al., 2020).

From the perspective of product uncertainty, our study further makes a categorization for product-related signals, namely promotion signals vs. reputation signals. Promotion signals pertain to features related to the merchants' incentives to promote their products. These signals are proactively provided and operated by merchants, indicating a relatively lower credibility (Dimoka et al., 2012; Zhou et al., 2022). On the other hand, Reputation signals pertain to features recording a current status of the product. These signals are provided and recorded by third parties, including platforms or customers. Merchants need to adhere to specific guidelines and operate formally in order to have these signals approved. Without interest conflict with the transaction, signals from third parties are of high credibility (Connelly et al., 2011; Price and Dawar, 2002; Zhou et al., 2022). To make a comparison, promotion and reputation signals are two fundamental informational mechanisms for resolving product uncertainty: signaling for description from merchants and signaling addressing performance from third-parties (e.g., Dimoka et al., 2012; Mavlanova et al., 2012; Setia et al., 2020). Signaling for description is delivering signals proactively manipulated or modified by merchants to convey concealed or limited quality information (Lu and Chen, 2021), with pricing and promotions being a typical form of signaling information (Liu et al., 2022b). Studies have also revealed various pricing-relevant signaling factors, such as advertising (Kirmani, 1990), money-back guarantees (Lee et al., 2005), and insurance (Zhang et al., 2022a). Meanwhile, signaling addressing performance relates to signals aimed at reducing imperfect market information and showing products' performance or current status, beyond the control of merchants. Factors such as trust (Pavlou et al., 2007) and reputation (Hong and Pavlou, 2017) operate within this mechanism.

To date, studies on the joint effects of diverse categories of signals remain limited (Zhou et al., 2022). Significant joint effects have been found in equity investment (Bapna, 2019), product launches (Price and Dawar, 2002) and online healthcare markets (Zhou et al., 2022). Scant literature theorizes the boundary condition that separates two signals as substitutes for each other from two signals as complements of each other (Zhou et al., 2022). Xu et al. (2020) employ the perspective of product uncertainty to illustrate that recommendation sources can be either complementary or substitutable, postulating that factors within the same category—whether description, performance uncertainty—may substitute for one another, while those across categories may exhibit complementarity. Zhou et al. (2022) posits the relative credibility of two signals as boundary condition, which a high credibility signal may substitute for low credibility signal, while two signals both with high credibility may complement each other.

It is an essential issue regarding making the best use of different information combinations. Proper combinations can maximize the effect of information on consumers' response and, consequently, on product sales, while improper ones can be detrimental to merchants' uncertainty reduction efforts such as promotion. We focus on a broader range of information signals to explore the interactive effect of image aesthetics combined with other displayed information through an explanation of different signal categories.

2.2. Image Features

In the domain of computer vision, image features are typically categorized as low-level and high-level features. Low-level features include specific elements that can be derived using statistical or pattern recognition methods to analyze pixel values, such as color and composition (Zhang et al., 2022b). Algorithmic outputs of local regions of an image are also included in low-level features (Wang et al., 2018). Aggregating these low-level features can yield the overall semantic meaning of the image, referred to as high-level features (Dhar et al., 2011; Rao et al., 2019). Generally, high-level features contain more holistic information and reflect the human's mental perception better, such as sentiment (Hou et al., 2023) and overall image quality (Zhang et al., 2022b).

Aesthetics is defined as a reflection in human minds and emotions in relation to the perception of beauty (Palmer et al., 2013). Regarding the role of aesthetics in product sales, existing literature typically presents two perspectives on its influence. One perspective emphasizes the artistic aspect and proposes the content-based spillover effect and content-independent spillover effect, both of which enhance customers' evaluations of the product (Hagtvedt and Patrick, 2008). Another perspective analyzes aesthetics in broader contexts, affirming the mental arousal and positive emotions elicited by aesthetics (Guan et al., 2023). Empirical evidence also demonstrates a significant association between aesthetics and low-level image features, including color and composition (Zhang et al., 2022b), indicating the compelling presentation and effective communication of product details embedded in image aesthetics (Guan et al., 2023; Jiang and Benbasat, 2007; Liao et al., 2016). Table 1 below summarizes the related literature. We only include the ones investigating high-level image features related to aesthetics. Moreover, causal inference methods regarding the features unstudied in our paper are noted as n.a. in Table 1.

As summarized in Table 1, Maier & Dost (2018) and Sohn (2017) use manual validation or subjects' recall to get the aesthetic features. Human-judged aesthetic features may not reach a consensus among consumers. Subjective and arbitrary reality also impedes the generalization of the conclusion. Literature has begun to utilize computational methods such as convolutional neural network (CNN, Lecun et al., 1998) to get the aesthetic features. For example, Zhang et al. (2022b) utilize CNN to classify the property image quality, confirming the relationship between image quality and property demand. Li et al. (2022) utilize CNN to get image sentiment, examining the U-shaped relationship between image sentiment and perceived usefulness of reviews. Instead of just investigating one aspect of aesthetic features, Guan et al. (2023) and Wang & Ding (2022) directly derive image aesthetics by CNN and make causal inference through lab experiments. Guan et al. (2023) test the outcome of post-purchase rating. Wang & Ding (2022) check the moderating role of image aesthetics on monetary referral. Shifting the focus to the effect of image aesthetics on product sales, our study uses CNN-derived image aesthetics to illustrate direct aesthetic impacts in food delivery platforms.

In the context of food delivery platforms, the aesthetic quality of food images plays a pivotal role in influencing consumer perception and purchasing behavior. On the one hand, considering the experiential nature of food products and the attractiveness of food images, the effect of image aesthetics of the displayed food can be more pronounced than any other products on e-commerce platforms such as books or electronic devices (Versace et al., 2019). Visual cues serve as critical determinants in shaping consumers' perception of food quality within the context of online food sales (Choi et al., 2024). The substantial effect of food images can be confirmed by brain activities (Killgore and Yurgelun-Todd, 2006; Toepel et al., 2009). On the other hand, existing literature on images of food delivery platforms primarily relies on subjective surveys (Brewer, 2021; Cai and Chi, 2021) or low-level image features (Kpossa and Lick, 2020; Liu et al., 2022a) to examine their influence on product sales. This approach can introduce arbitrariness, subjectivity and pixel-level limitation into the results. In contrast, CNN-derived image aesthetics offer a holistic perception of food images through computational methods, thereby rendering our findings more objective, stable and scalable. Moreover, our paper represents an early foray to introduce generalized propensity score matching (GPSM) and the instrument variable (IV) approach to analyze the causal effect of image aesthetics by addressing the endogeneity concerns of self-collected empirical data.

Table 1: Literature review on image aesthetics.

Literature	High-level Image features	Method	Dependent Variable	Empirical Data	Causal Inference	Findings
Zhang et al. (2022b)	Image quality	CNN	Property demand	Airbnb property listings	n.a.[3]	Property image quality positively affects Airbnb property booking rate. The image quality can be explained by several low-level image features.
Guan et al., (2023)	Image aesthetics	CNN	Post-purchase rating	Amazon product and review-related data	Laboratory experiment	Aesthetic images negatively affect the post-purchase rating.
Wang and Ding (2022)	Image aesthetics	CNN	Product sales	Pinduoduo's marketing platform	Laboratory experiment	Image aesthetics positively moderate the relationship between monetary rewards and product sales in referral programs.
Li et al. (2022)	Image sentiment	CNN	Perceived review usefulness	Yelp restaurant and review data	n.a.	A U-shaped relationship exists between review photo sentiment and review usefulness. Photo sentiment affects review enjoyment positively and linearly.
Maier and Dost (2018)	Fitness of contextual background	Manual Validation	Liking Purchase Intention	n.a.	Experiment	Fitting contextual background can increase imagery fluency and mental imagery, and in turn, product liking and purchase intention.
Sohn (2017)	Visual complexity Visual congruence	Subject's recall	Satisfaction	Survey	n.a.	Perceived visual complexity and visual congruence affect fluency perceptions, which consequently affects satisfaction.
This study	Image aesthetics	CNN	Product sales	Food delivery platform data	GPSM, IV	Image aesthetics can positively affect product sales. The effect of image aesthetics can be substitutive or complementary combining with other displayed information.

[3] n.a. is the abbreviation for not applicable.

3. Hypotheses Development

The literature has documented that textual and visual product descriptions serve as merchant-provided information channels and convey signals to potential customers (Wells et al., 2011). Within this context, the aesthetic features examined in this study are categorized as promotional signals (Dimoka et al. 2012). Consequently, merchants dedicate efforts to creating visually-appealing images to enhance customer loyalty and encourage purchase intentions (Cyr et al., 2006; Yoo et al., 2024). However, prior studies present mixed findings on the effects of image aesthetics. Some studies suggest that the impact of image aesthetics may be limited, particularly when compared with the influence of image content (He et al., 2023; Zhang and Luo, 2023) or product reviews (Goh et al., 2013; Guan et al., 2023). Conversely, other studies affirm the positive influence of aesthetics, highlighting aspects such as perceived quality (Zhang et al., 2022b), emotional response (Hou et al., 2023), and specific aesthetic elements like color and composition (Goswami et al., 2011; Zhang et al., 2022b).

This inconsistency motivates us to examine the main effect of image aesthetics as a starting point. Intuitively, merchant-provided product images are expected to play a crucial role in consumers' purchase decisions by signaling product quality; typically, these images are exposed to consumers before they scroll down to read reviews. In some cases, there can be a lack of reviews due to a cold start (Liang et al., 2024) or the product's position in the long tail (Aguiar and Waldfogel, 2018), and product images become the primary source of product information. In such instances, an aesthetically appealing image can evoke a positive emotional response from consumers and enhance their purchase intentions. The vividness of the image also helps highlight product details, which in turn fosters a more favorable attitude toward the product (Coyle and Thorson, 2001; Kumar and Tan, 2015). To sum up, we hypothesize the positive effect of image aesthetics on product sales.

H1: image aesthetics is positively related to product sales.

Apart from the promotional signal conveyed by image aesthetics, the signaling literature suggests that other types of signals (e.g., reputation signals) may communicate different product characteristics (Connelly et al., 2011). Signals from distinct categories can reduce different dimensions of information asymmetry, indicating potential synergistic effects (Xu et al., 2020). In other words, alternative promotional signals (i.e., non-aesthetic signals) are expected to substitute for the effect of image aesthetics, while reputation signals are likely to complement it.

In this study, we consider two representative signals: display advertisements as a promotional signal (similar in nature to image aesthetics) and merchant certifications as a reputation signal (different in nature from image aesthetics). In the e-commerce context, display advertisements are commonly viewed as promotional signals that capture consumers' attention (Dimoka et al., 2012; Kirmani, 1990), operating in a manner like image aesthetics. As such, their effects in reducing product uncertainty may overlap or even conflict. For instance, frequent exposure to advertisements from the same merchant might reduce customers' focus on the aesthetic appeal of the product image. Conversely, a visually striking product image could attract attention on its own, even without the merchant investing in prominent display positions. Based on this, we hypothesize a substitutive relationship between image aesthetics and advertisements.

H2: Advertisements negatively moderate the relationship between image aesthetics and product sales.

Lastly, merchant certification is issued by platforms that conduct inspections to verify compliance with specific standards. These certifications aggregate insights from multiple sources, including platform assessments and consumer feedback, serving as a third-party guarantee to support consumer purchase decisions (Dimoka et al., 2012). By acting as merchant-side information, certifications help bridge the information asymmetry between merchants and consumers, thereby reducing product uncertainty and enhancing customers' purchase intentions.

Given their distinct informational roles—each compensating for the limitations of the other (Connelly et al., 2011; Mavlanova et al., 2012)—image aesthetics and certifications may exhibit a complementary effect. For example, the positive impact of image aesthetics may be amplified by the increased reputation conferred by platform certification. When the merchant is certified, customers are more inclined to trust both the merchant and the displayed image, thereby strengthening the effect of image aesthetics. While image aesthetics may not directly enhance the merchant's reputation, they can still serve as a supportive signal of trustworthiness through their vividness and implied attention to product quality. Accordingly, we propose:

H3: Platform certification of merchants positively moderates the relationship between image aesthetics and product sales.

4. Data and Method

We partnered with a major food delivery platform in China to gather data from a university area with over 50,000 student residents within a 3-kilometer radius.[4] The student population, drawn from various regions across the country,

[4] The company name and data details cannot be fully disclosed due to a confidential agreement.

offers a diverse and representative sample, capturing a wide range of aesthetic preferences in online food ordering. The dataset comprises 506,553 orders, encompassing 5,074 distinct dishes from 71 different restaurants, collected over a one-month period in 2021.

Descriptive statistics for the restaurant-level data are presented in Table 2. For each successful order, the order price must meet a minimum threshold, referred to as the *Start delivery threshold*. The total order price comprises the discounted price (if any promotional conditions are met), the delivery *Fee*, and any *Extra fee*. Promotional discounts are applied when the order price or dish combination qualifies for active promotions. The average *Start delivery threshold* is 14.5 (standard deviation: 4.3). The delivery fee and extra fee are capped at 5 and 1, respectively. On average, restaurants offer 6 promotional activities (standard deviation: 1.9).

We include two dummy variables, *Crowdsource* and *City-wide*, to indicate the delivery mode of the restaurant. The platform offers three typical delivery service options. The basic mode is guaranteed delivery (where both *Crowdsource* and *City-wide* equal 0). When *Crowdsource* equals 1, the restaurant utilizes an outsourced delivery service. When *City-wide* equals 1, the delivery is guaranteed with personnel covering the entire city area.[5] In our sample, 86.52% of the restaurants use outsourced delivery, while only 9.97% opt for city-wide guaranteed delivery.

The variable *Rating* reflects the customer's post-purchase review, ranging from 3 to 5, with an average score of 4.7. The *Advertisement* dummy indicates whether the restaurant advertises on the platform, with a mean value of 0.2 (standard deviation: 0.4). The *Certification* dummy denotes whether the restaurant holds a platform-issued certification, with an average value of 0.4. Likewise, the *Food insurance* dummy and *New* dummy indicate whether the restaurant has purchased food safety insurance and whether it is newly listed on the platform, respectively. Restaurants opened within the last 30 days are classified as new.

Beyond the firm-provided data, a key challenge in this study is to develop a reliable measurement of image aesthetics for food dish images. Given the absence of pre-existing models or datasets specifically designed for evaluating the aesthetic quality of food imagery, we adapt a two-stage convolutional neural network (CNN) model inspired by Talebi and Milanfar (2018). This approach involves an initial pretraining phase followed by a finetuning phase to extract high-level features and assess the aesthetic appeal of food images.

As depicted in Figure 2, the model is first pretrained on a diverse dataset of general images, not restricted to food, to maximize generalizability in image feature extraction. Following this, the model is finetuned on a curated set of food dish images to enhance its accuracy in assessing aesthetic quality specific to the context of our study. This two-stage strategy achieves a balance between broad applicability and specialized precision: the extensive variety of general images in the pretraining phase ensures robust feature learning, while the targeted finetuning on food images refines the aesthetic evaluation. Such a hybrid training methodology aligns with recent advancements in deep learning, including generative pre-trained transformers (GPT; Brown et al., 2020), which leverage both generalization and domain-specific adaptation.

In the first pretraining stage, the model is trained on the ImageNet dataset (Deng et al., 2009) for general image classification. The output layer in this stage consists of 1,000 neurons, corresponding to 1,000 distinct categories (e.g., 'cat', 'dog'), enabling the model to capture a broad range of image content. This pretraining phase aims to enhance the model's performance in subsequent tasks by minimizing the cross-entropy loss function (Mao et al., 2023), thereby improving the model's capability to understand general image features.

[5] Guaranteed delivery signifies that the platform's employed riders are tasked with the delivery of ordered food directly to the customers. Each delivery is backed by a guarantee from the platform. For example, customer A orders a pizza through the platform. The platform assigns a dedicated rider to pick up the pizza from the restaurant and ensure it is delivered to customer A within the promised time. If there is an issue with the delivery (e.g., delay or loss), the platform offers compensation or arranges a redelivery. Crowdsource describes a process wherein the platform lists the delivery task on a crowdsourcing system. Freelance riders in proximity to the pick-up location can then opt to accept the task and take charge of the delivery process. However, there are instances when no riders may accept the delivery task, leading to a failed delivery attempt. For example, customer B orders dumplings through the platform. The platform posts the delivery task on a crowdsourcing system, and nearby freelance riders see the task and accept it at their discretion. A rider accepting the task will pick up the dumplings and deliver them to customer B. If no riders are willing to accept the task, the delivery attempt may fail. City-wide delivery service is a premium offering by the platform, often selected by established brand restaurants. In this mode, dedicated riders are assigned to ensure delivery across the city, irrespective of the distance between the restaurant and the customer's location. For example, a well-known chain coffee shop opts for the city-wide delivery service. Customer C, who lives on the other side of the city, 15 km away from the coffee shop, places an order. The platform assigns a dedicated rider to pick up the order from the coffee shop and deliver it to customer C, regardless of the distance.

Table 2: Descriptive Statistics of the restaurants (*N*=71)

Name	Explanation (Unit)	mean	s.d.[6]	min	max
Start delivery threshold	Price threshold for delivery (yuan[7])	14.5	4.3	0.0	20.0
Promotion num	Num of promotion (Each)	6.0	1.9	2.0	10.0
Fee	Delivery fee (yuan)	0.2	0.7	0.0	5.0
Extra fee	Package fee (yuan)	0.1	0.3	0.0	1.0
Rating	Review rating (points)	4.7	0.3	3.0	5.0
Dummy: *Crowdsource*	Crowdsourced delivery	0.9	0.3	0.0	1.0
Dummy: *City-wide*	City-wide delivery	0.1	0.3	0.0	1.0
Dummy: *Advertisement*	Advertising	0.2	0.4	0.0	1.0
Dummy: *Certification*	platform certification	0.4	0.5	0.0	1.0
Dummy: *Food insurance*	Food safety insurance	0.8	0.4	0.0	1.0
Dummy: *New Opening*	New restaurant	0.2	0.4	0.0	1.0

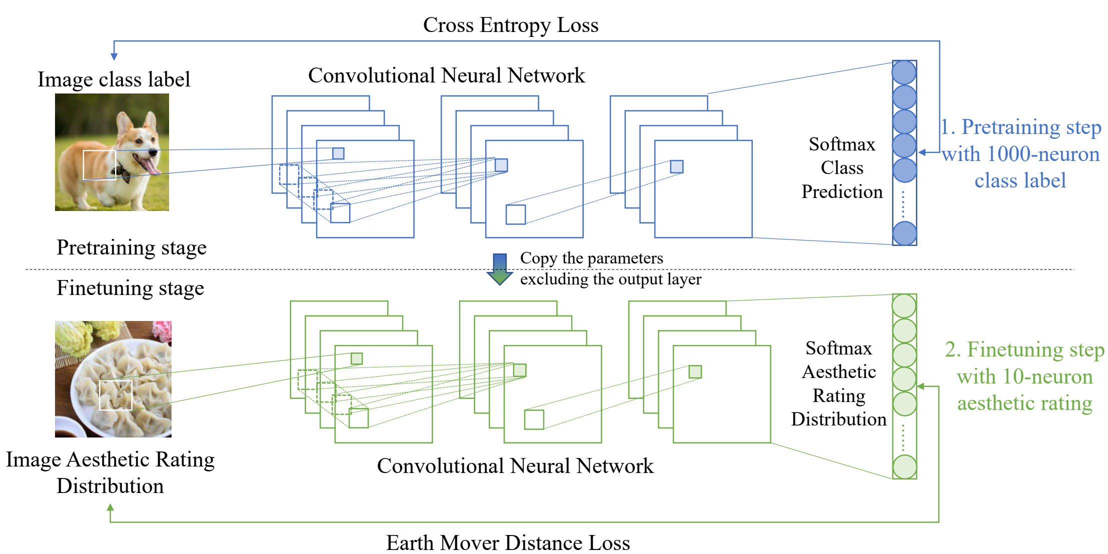

Figure 2. Model setting illustration

[6] S.d. is the abbreviation for standard deviation.
[7] Yuan is the currency unit in China.

To adapt the model for human aesthetic perception, we further finetune it on the Aesthetics Visual Assessment (AVA) dataset (Murray et al., 2012) during the second stage. The AVA dataset contains a large collection of photos rated on a 10-point aesthetic scale by users of a digital photo-sharing platform. Each photo is accompanied by a discrete rating distribution based on user evaluations. To tailor the model to the context of food imagery, we perform an additional finetuning step using a subsample of food images from the AVA dataset, enhancing the model's relevance for assessing the aesthetic appeal of food dishes.

In this second stage, the output layer comprises 10 neurons, reflecting the 10-point scale of aesthetic ratings. The loss function is designed to learn the cumulative aesthetic distribution using the Earth Mover's Distance (Rubner et al., 1998), which penalizes larger deviations more heavily. Consequently, the model outputs a 10-item aesthetic distribution. For subsequent analysis, we compute the mean value of this distribution to represent the aesthetic score of each input image.

We also integrate VGG (Simonyan and Zisserman, 2015), ResNet (He et al., 2016), and MobileNet (Howard et al., 2017) as the convolutional neural network (CNN) components in our architecture. The performance metrics for each model are summarized in Table 3. The networks are trained over 13 epochs. During the initial 5 epochs, only the parameters of the modified output layer are updated, following the efficient finetuning strategy validated by Ericsson et al. (2021). In the subsequent 8 epochs, all model parameters are optimized.

Table 3: The comparison between three networks

CNN components	# of parameters	Training Time	EMD Loss (test set)	Time (Prediction of 1000 images)
VGG-16	14.7M	6h35m	0.0723	17.008s
ResNet-101	42.7M	8h24m	**0.0701**	17.182s
MobileNet	**3.2M**	**5h29m**	0.0710	**12.306s**

We conduct the experiments on a Linux server equipped with an NVIDIA RTX A4000 GPU (16GB) and TensorFlow version 2.5.0. Among the three models, MobileNet exhibits medium loss values while significantly reducing both the number of parameters and the training/prediction time. Although all three CNN architectures demonstrate comparable performance, we select MobileNet due to its simplicity and superior learning efficiency.

To validate the model's accuracy, we recruited 201 participants (101 female and 100 male) to assess the aesthetic value of 20 food dish images. Each participant was asked to rate 10 images on a scale from 1 to 10, yielding an average of 100 ratings per image. To reduce potential bias from fatigue or habituation, the order of image presentation was randomized. The results indicate a strong positive correlation between human ratings and model predictions, with a correlation coefficient of 0.75 ($p < 0.05$).

Figure 3 illustrates representative examples of images with minimum, mean, and maximum aesthetic values. Images with the minimum aesthetic value are characterized by poor photographic quality, lacking professional composition or beautification. Notably, one example does not even depict food. The second row displays images with a mean aesthetic value, where the photographer has employed visible enhancements, such as adding decorative elements or removing the background to emphasize the food in the foreground. These modifications modestly increase the image's attractiveness. The third-row features images with the maximum aesthetic value, where the photographer has applied visual effects and carefully composed the scene to highlight the food. Overall, higher aesthetic values correspond to enhanced visual presentation, which may stimulate customers' appetite and increase their engagement with the product. In summary, the aesthetic value metric effectively captures both the photographer's efforts in image beautification and the customers' holistic aesthetic perception.

Figure 3. Image examples at the min/mean/max aesthetic value

Table 4 presents the descriptive statistics of the dish-level data, including aesthetic values. The variable *Aesthetics*, derived from the neural network's predictions, ranges from a minimum of 3.4 to a maximum of 6.3. The mean operation on the output distribution narrows the range of aesthetic values. *Price* has a mean of 19.1 (standard deviation: 19.2), while the *Discount Price* averages 14.1 (standard deviation: 13.8). In some cases, these values may be zero, as certain dishes are classified as accessories to main courses and are not available for individual purchase. *Dishsale,* representing the sales volume of a dish over a one-month period, has a mean of 99.8 and a standard deviation of 342.0. In the following analysis, *Dishsale* may be log-transformed to address skewness. The variable *Discount* is calculated using the formula $1 - {}^{Discount\ Price}/_{price}$, capturing the promotional intensity of a dish. The average discount level is 0.2, with a standard deviation of 0.3. Furthermore, we categorize each dish as either a meal or a drink based on its name. *Meal* suggests that the dish is a meal, with a mean value of 0.7 and a standard deviation of 0.5.

Table 4: Summary statistics of dishes (*N*=5,074)

VarName	Explanation (Unit)	mean	s.d.	min	max
Aesthetics	Image aesthetics (points)	5.1	0.5	3.4	6.3
Price	Price (yuan)	19.1	19.2	0.0	200.0
Discount Price	Discount price (yuan)	14.1	13.8	0.0	168.0
Dishsale	Sales of a dish (orders)	99.8	342.0	0.0	14,994.0
Discount	Discount rate (calculated)	0.2	0.3	0.0	1.0
Dummy: *Meal*	Food type: Meal or Drink	0.7	0.5	0.0	1.0

5. Results

5.1. The Impact of Image Aesthetics on Product Sales

To estimate the effect of aesthetics, we conduct our regression analysis with the following model.

$$ln\_dishsale_{i,k} = \alpha_0 + \alpha_1 Aesthetics_{i,k} + \alpha_2 \boldsymbol{controls_{i,k}} + \alpha_3 restaurant_k + u_{i,k}, \quad (1)$$

where k indicates the index of the restaurant and $restaurant_k$ is the dummy of this restaurant. α_3 represents a restaurant-specific intercept, capturing the fixed effect of the restaurant. Due to the similarity in dish flavors within the same restaurant menu, incorporating restaurant fixed effects can also help control for the type of flavors in the dishes. i indicates the index of the dish. $\boldsymbol{controls_{i,k}}$ represents the control variables related with the specific dish *i* and restaurant *k*. Here we incorporate both dish-level and restaurant-level variables as controls in our model. Since these product attributes can clearly influence consumers' perceptions, and consequently, product sales. It is essential to eliminate potential biases from other attributes to accurately assess the impact of image aesthetics on product sales. Specifically, the dish-level control variables address price-related factors, such as dish price and discounts, as well as food type-related issues. The restaurant-level control variables contain fee-related factors, such as delivery and packaging fees, service-related factors, including delivery mode and food safety insurance, and operation status factors, like rating and platform certification. However, due to the issue of collinearity, the restaurant fixed effects and restaurant-level control variables cannot coexist in the same model. We incorporate dish-level control variables only when restaurant fixed effects are present in the model. $Aesthetics_{i,k}$ denotes the image aesthetic value of dish *i*, while α_1 represents the effect of image aesthetics on product sales. The dependent variable, $ln\_dishsale_{i,k}$, is the logarithmic transformation of product sales, $Dishsale_{i,k}$, defined as $ln\_dishsale_{i,k} = \ln(Dishsale_{i,k} + 1)$.

Table 5. Regression Result of the main effect of *Aesthetics*

	(1) *ln_dishsale*	(2) *ln_dishsale*	(3) *ln_dishsale*
Aesthetics	0.621***	0.276***	0.145***
	(0.051)	(0.051)	(0.050)
Price		-0.029***	-0.027***
		(0.001)	(0.001)
Discount		0.235**	0.007
		(0.096)	(0.107)
Controls	NO	YES	YES
Restaurant FE	NO	NO	YES
N	5,074	5,074	5,074
R-squared	0.026	0.285	0.502

ln_dishsale is the log form of *Dishsale*, transformed by *ln_dishsale*=ln(*Dishsale*+1).
Standard errors in parentheses are robust to heteroskedasticity. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 reports the regression coefficients for the *Aesthetics* variable, alongside the coefficients for key dish-level covariates (i.e., price and discount). Column (1) presents the results from the model excluding control variables. Column (2) incorporates control variables, while Column (3) further includes restaurant fixed effects to account for unobserved heterogeneity at the restaurant level.

The coefficients for *Aesthetics* in Table 5 are all significant at the 1% level, indicating a strong positive effect of image aesthetics on product sales. The negative coefficients for *Price* and the positive coefficients for *Discount* align with established market practices. As control variables and restaurant fixed effects are included, the coefficients for *Aesthetics* decrease from 0.621 to 0.145, suggesting that part of the positive impact of image aesthetics is explained by other control variables and restaurant-specific factors. Nevertheless, the significant coefficient of 0.145 implies a 14.5% increase in sales attributable to image aesthetics.

Given the mean value of *Dishsale* at 99.8 orders, a one-point increase in the aesthetic value of a product image corresponds to an additional 14.5 dish orders within a restaurant. When considering the aesthetic effect across different restaurants, this impact translates into a 27.6% increase in sales, equating to 27.5 additional dish orders. Thus, the evidence supports **H1.**

5.2. Robustness Tests

To strengthen the causal interpretation of our findings, we extend the analysis beyond Table 5 to address potential endogeneity issues. One possible source of endogeneity arises from selection bias by restaurants. Specifically, restaurants may enhance the aesthetics of their product images based on expected sales, making image aesthetics an endogenous variable influenced by the restaurant's promotional intentions. Additionally, omitted variable bias presents another concern, as unobserved factors such as cooking style or service quality may simultaneously affect both image aesthetics and product sales. Lastly, the CNN-based measurement of image aesthetics requires validation against alternative metrics to ensure robustness.

In this section, we conduct three robustness checks at both the model and measurement levels to validate our results. First, to address the endogeneity issue arising from anticipatory selection bias by restaurants, we employ generalized propensity score matching (GPSM) based on observable variables. Second, to mitigate the endogeneity induced by omitted variables, we use a fixed-effects estimation combined with an instrumental variable approach, re-evaluating the effect of aesthetics while accounting for unobservable factors. Third, drawing on the image measurement framework from Zhang et al. (2022b), we apply an alternative aesthetic measure based on 12 interpretable low-level features to assess potential measurement error in our algorithm.

5.2.1. Generalized Propensity Score Matching

We employ generalized propensity score matching (GPSM) to address selection bias related to the continuous treatment variable, image aesthetics. The GPSM method, developed by Hirano and Imbens (2004), extends the conventional propensity score matching (PSM) framework (Rosenbaum and Rubin, 1983) from binary to continuous treatments. Under the same conditional independence assumption, GPSM estimates the treatment effect based on the predicted level of the independent variable, rather than the observed levels, to mitigate endogeneity concerns. GPSM has been widely used to evaluate continuous treatment effects, such as in studies examining IT usage (Atasoy et al., 2016) or the degree of environmental adaptation (Baráth and Fertő, 2024).

The GPSM analysis follows three main steps. In the first step, we estimate the conditional density of the treatment variable given the covariates, expressed as $r(t, \boldsymbol{x}) = f_{T|\boldsymbol{X}}(t|\boldsymbol{x})$, where T represents the treatment level (image aesthetics in our study), and $\boldsymbol{X}$ denotes the observable covariates associated with selection bias, such as factors influencing a restaurant's choice of aesthetic values. The function r is the generalized propensity score. In the second step, we estimate the conditional expectation of the outcome variable (*ln_dishsale* in our context) as a function of both the treatment level and the generalized propensity score. The specification typically includes higher-order terms and an interaction term between the treatment and the generalized propensity score. We model this as $\beta(t, r) = \mathrm{E}[Y|T = t, R = r] = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 r + \alpha_4 r^2 + \alpha_5 t * r$, where we use a quadratic approximation in our analysis. In the final step, we average the conditional expectation of the outcome over the generalized propensity score to derive the dose-response function, $\mu(t) = E[\beta(t, r(t, \boldsymbol{x})]$. This function allows us to estimate the expected outcome at any given treatment level. The average treatment effect at a specific treatment level can also be obtained by taking the derivative of the dose-response function.

Using the program developed by Bia and Mattei (2008), we estimated the dose-response function to examine the effect of image aesthetics on product sales. Control variables and restaurant fixed effects are retained in the estimation. To assess the balancing property after incorporating the generalized propensity score, we divided the *Aesthetics* variable into five levels and tested whether the adjusted mean of each covariate at a given level was statistically different from the adjusted means at other levels. The results indicate that the means of the covariates do not differ significantly across the five levels of aesthetics at the 1% significance level. This outcome confirms that the balancing property is satisfied, suggesting that GPSM effectively adjusts for selection bias induced by observable covariates.

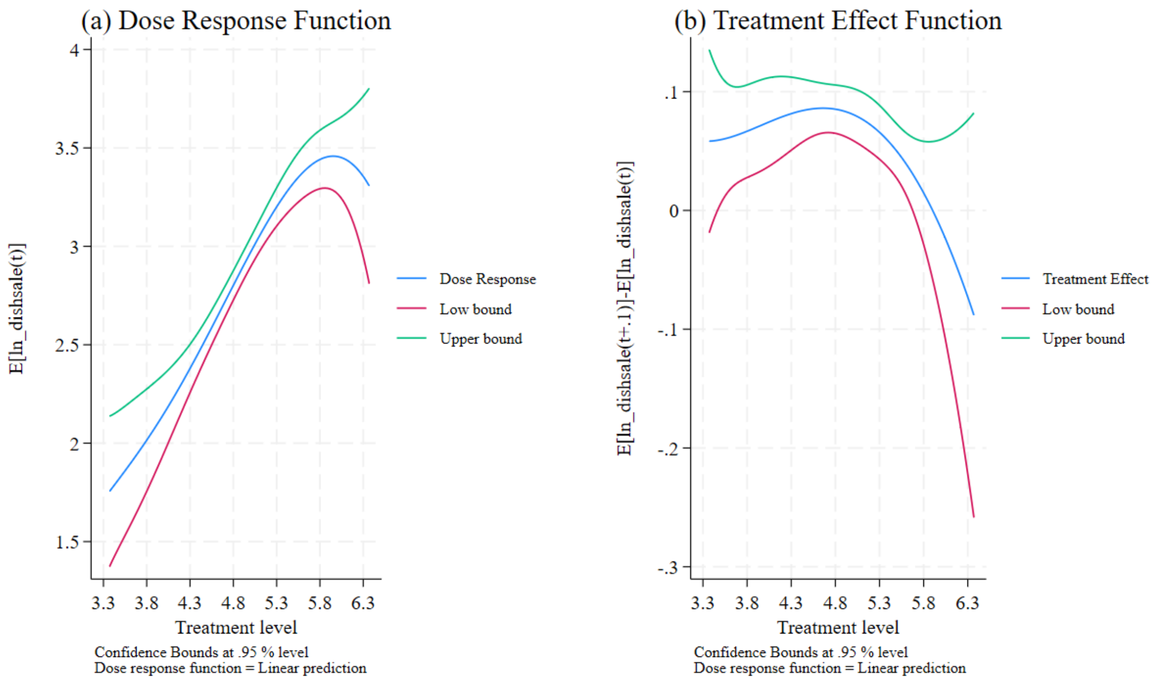

Figure 4. Generalized Propensity Score Matching Results

Figure 4 illustrates the dose-response function and the corresponding treatment effect function from GPSM. The dose-response function in Figure 4(a) depicts the outcome distribution across different levels of the treatment variable, *Aesthetics*. The outcome variable, *ln_dishsale*, which represents the growth rate of dish sales, exhibits a monotonically increasing trend as the aesthetic value increases, except at the extreme ends of the aesthetics range, where the confidence intervals are wider. The treatment effect function in Figure 4(b) represents the marginal effect of aesthetics on the dose-response function, corresponding to the derivative of the dose-response function with respect to *Aesthetics*. The confidence intervals at aesthetic levels below 3.48 and above approximately 5.66 include zero, indicating that the treatment effect is not statistically significant in these ranges. However, the range between 3.48 and 5.66 demonstrates a significant positive treatment effect, with values consistently above zero, confirming the positive impact of aesthetics at moderate levels. The average treatment effect is 0.048 across all the treatment level, which is lower than 0.145 in the previous results. The maximum effect can reach 0.086 at the treatment level of 4.66. This observation may be attributed to the relatively larger sample size in the middle range of aesthetics. In summary, our findings support **H1**, confirming the positive effect of image aesthetics on product sales, even after accounting for selection bias due to observable variables using GPSM.

5.2.2. Fixed-Effects Estimation with Instrumental Variable

Our second robustness check employs an instrumental-variable approach to mitigate endogeneity concerns arising from unobserved factors and to further validate our results. Additionally, to address merchant heterogeneity, we continue to include restaurant-specific dummies and utilize a merchant-level fixed-effects model with instrumental variables.

As mentioned above, the cooking status of the dish could be an omitted variable and potentially causes endogeneity issues. Additionally, dish ingredients may influence both the aesthetic quality of the food image and product sales, which are not fully accounted for in our control variables. To address these concerns and identify the causal effect, we use the average aesthetic quality of other dishes within the same restaurant as an instrumental variable for the aesthetic value of the focal dish image. Typically, restaurant operators curate dish images together, resulting in similar visual styles across images from the same restaurant, which suggests a strong correlation in their aesthetic qualities. Hence, we select the average aesthetics of the other dishes within the same restaurant (referred to as *AAOD*) as our final instrumental variable.

In our sample, the correlation between the aesthetic quality of the focal dish image and *AAOD* is 0.602, supporting the relevance of this instrument. At the same time, it is unlikely that the aesthetics of other dish images directly influence the sales of the focal dish, as consumers typically only view the image of the dish they intend to purchase. In our data, the correlation between *Dishsale* and *AAOD* is 0.096, which supports the exogeneity of this instrument.

To further ensure that the instrument variable satisfies the exclusion restriction, we flexibly control for dish- and restaurant-level factors, ensuring that the aesthetics of other dishes impact the sales of the focal dish only through its own image. We also include restaurant-level fixed effects to control for systematic biases arising from the same restaurant context. Given the similarity of food types within a single restaurant, this approach helps us better assess the impact of image aesthetics on similar dishes. However, restaurant-level control variables and restaurant fixed effects will not coexist in a model due to collinearity. We employ a two-stage least squares (2SLS) method for model estimation. Accordingly, we employ the following 2SLS fixed-effects model to estimate the causal effect of image aesthetics on dish sales.

$$Aesthetics_{i,k} = \beta_0 + \beta_1 AAOD_{i,k} + \beta_2 \boldsymbol{controls_{i,k}} + \beta_3 restaurant_k + e_{i,k}, \tag{2}$$

$$ln\_dishsale_{i,k} = \gamma_0 + \gamma_1 \widehat{Aesthetics}_{i,k} + \gamma_2 \boldsymbol{controls_{i,k}} + \gamma_3 restaurant_k + u_{i,k}, \tag{3}$$

where k indicates the index of the restaurant and $restaurant_k$ indicates the dummy of this restaurant. β_3 and γ_3 represent a restaurant-specific intercept that captures heterogeneities across restaurants. i indicates the index of the dish. $\boldsymbol{controls_{i,k}}$ represents the control variables related to the specific dish *i* and restaurant *k*, as described in the main model, Equation (1). $Aesthetics_{i,k}$ represents the image aesthetic value of dish *i*, while $AAOD_{i,k}$ represents the average aesthetics of the other dishes in restaurant *k* except for dish *i*.

Equation (2) represents the first step of 2SLS to address the endogeneity issue, where the image aesthetics of the focal dish is regressed against the average image aesthetic value of other dishes in the same restaurant. Equation (3) is the second step to get γ_1 as the effect of image aesthetics on dish sales. The estimation results are in Table 6.

In column (1) of Table 6, we exclude all control variables and restaurant fixed effects. In column (2), we add the control variables, and column (3) incorporates both control variables and restaurant fixed effects. The first-stage results, presented in Panel B, show a highly significant coefficient for *AAOD* at the 1% significance level. The KP *F*-statistics across all three columns exceed 1,860, confirming that *AAOD* is a strong instrumental variable for the aesthetic quality of the focal dish image.

Panel A presents the second-stage results. Across all specifications—whether control variables or restaurant fixed effects are included—Aesthetics consistently shows a highly significant effect on *ln_dishsale*. The coefficient decreases from 1.557 to 0.229 when control variables and restaurant fixed effects are included, highlighting the strong explanatory power of restaurant-level factors. Nonetheless, *Aesthetics* maintains a positive impact of 22.9% on dish sales relative to other dishes within the same restaurant. Specifically, a one-point increase in *Aesthetics* leads to 22.9 additional orders. In summary, the 2SLS results with restaurant fixed effects align with our main findings, with the estimated coefficient exhibiting a slight increase from the previous result of 0.145. The effect of image aesthetics on dish sales remains consistently positive, providing strong support for **H1**.

To address potential concerns regarding the validity of our instrumental variable, we follow Clarke and Matta (2018) in assessing the robustness of the 2SLS results under partial violations of the exclusion restriction. The exclusion restriction requires that the instrumental variable (IV) be uncorrelated with unobserved error terms. In our study, the IV, *AAOD*, exactly identifies the endogenous variable, *Aesthetics*, which complicates the verification of its validity. To mitigate this issue, we adopt the approach proposed by Conley et al. (2012), which relaxes the exclusion restriction by allowing for some degree of correlation between the IV and the error term. This framework enables practical inference under such conditions. Following Clarke and Matta (2018), we implement this analysis in Stata with the union of confidence intervals (UCI) approach.

Figure 5 presents the results, showing the estimated coefficient of *Aesthetics* across different levels of the correlation parameter, δ.[8] Across the range of δ from -5 to 5, the estimated coefficient of *Aesthetics* remains positive, with upper and lower bounds ranging from 0.026 to 0.434. This demonstrates the robustness of our 2SLS results. Compared to the point estimate of 0.229 (the coefficient at $\delta = 0$), the confidence interval under the UCI approach

[8] According to Clarke and Matta (2018), instrumental variable estimation is modeled as

$$Y = X\beta + Z\gamma + \epsilon,$$
$$X = Z\Pi + V.$$

Here Z denotes the instrumental variable and X represents the potentially endogenous variables. Classical IV estimation imposes the prior that $\gamma = 0$. In our analysis, we relax this restriction by allowing γ to vary within a flexible range, where δ parameterizes the range of γ. For further details, please refer to the original paper.

widens but remains consistently positive, further supporting the reliability of our findings. Thus, the impact of image aesthetics is reaffirmed, providing additional evidence for H1.

Table 6: Robustness check with instrumental variable and restaurant fixed effect

	(1) *ln_dishsale*	(2) *ln_dishsale*	(3) *ln_dishsale*
	Without controls Without Restaurant FE	With controls Without Restaurant FE	With controls With Restaurant FE
Panel A: 2SLS estimates			
Aesthetics	1.557*** (0.090)	0.639*** (0.106)	0.229*** (0.056)
Controls	NO	YES	YES
Restaurant FE	NO	NO	YES
Observations	5,074	5,074	5,074
Panel B: First stage			
AAOD	0.977*** (0.018)	0.949*** (0.022)	-53.295*** (0.857)
R-squared	0.362	0.369	0.839
KP *F*-statistics	2,868.35	1,860.48	3,865.14

ln_dishsale is the log form of *Dishsale*, transformed by *ln_dishsale*=ln(*Dishsale*+1).
Standard errors in parentheses are robust to heteroskedasticity. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

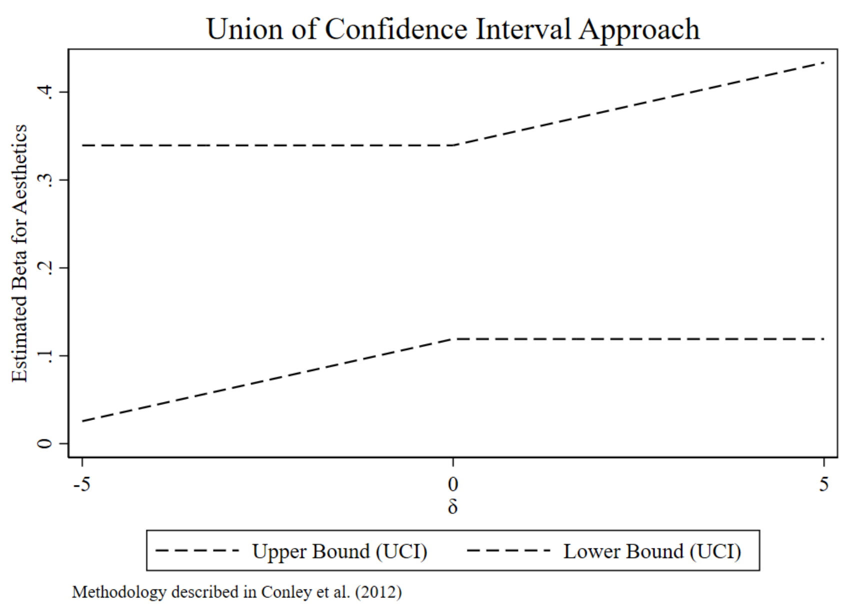

Figure 5. The estimated results of *Aesthetics* with UCI approach

5.2.3. Aesthetic Measurement with Low-level Image Features

To mitigate potential measurement errors from the CNN algorithms, we conduct an additional robustness check using an alternative measure of image aesthetics. Specifically, we adopt the approach from Zhang et al. (2022b), which utilizes 12 interpretable low-level image features to assess the impact of aesthetic quality on the demand for Airbnb properties. Following this method, we extract the same 12 low-level image features for our dataset. We then regress dish sales on these features to evaluate the effectiveness of image aesthetics from a low-level feature perspective and to further corroborate our previous findings.

We follow the same definitions and algorithms as described in Zhang et al. (2022b), with the features detailed in Table 7. Composition captures the arrangement of visual elements, including *Diagonal dominance*, *Rule of thirds*, *Visual balance intensity*, and *Visual balance color*. Color reflects the pixel details and vividness of the image, which can influence viewers' emotional responses; this includes *Warm hue*, *Saturation*, *Brightness*, *Brightness contrast*, and *Image clarity*. The figure-ground relationship measures the distinctions between foreground elements and the background scene in terms of size, color, and texture differences. While these features assess aesthetic perception from a different perspective than our primary analysis, they nonetheless represent key aspects of visual aesthetics.

Therefore, we hypothesize a significant association between these features and sales. However, given the number of variables and the complex, unknown relationships between individual features and overall aesthetics, we refrain from making causal inferences between low-level image features and sales as in our primary analysis. Instead, we offer these results as a robustness check to address potential measurement errors in our aesthetics metric and advise readers to interpret them with this context in mind.

Due to the transfer of the algorithm from Airbnb property images to food images, some features are not applicable to certain samples. We removed the samples without feature values, leaving 3,988 samples. Then, a regression analysis is conducted to analyze the relationship between *Dishsale* and these low-level image features. Column (1) in Table 8 reports the result of the coefficients of these low-level image features. We control for the issues relating to the dish and the restaurant fixed effects. In the category of composition, *Visual balance intensity* demonstrates a noteworthy value of 0.628 at a statistically significant level of 1%, which is the most salient factor among others. *Visual balance color* also shows a coefficient of -0.058, which is statistically significant at the 5% level. *Diagonal dominance* and *Rule of thirds* have no significant coefficient in the regression. In the category of color, both *Warm hue* and *Saturation* have a significant impact on *ln_dishsale*. *Warm hue* has a coefficient of 0.607 and *Saturation* is 1.194 at a significance level of 1%. The *Brightness*, *Contrast of brightness* and *Clarity* show no significance in our sample. Lastly, in the category of figure-ground relationship, *Texture difference* exhibits a negatively significant coefficient of -0.393 at the 5% level, while *Size difference* shows a coefficient of -0.543, significant at the 10% level. *Color difference* does not have a statistically significant impact on *ln_dishsale*.

Table 7: low level image aesthetic features in (Zhang et al., 2022b)

Classification	Feature
Composition	*Diagonal dominance*
	Rule of thirds
	Visual balance intensity
	Visual balance color
Color	*Warm hue*
	Saturation
	Brightness
	Contrast of brightness
	Image clarity
Figure-ground Relationship	*Size difference*
	Color difference
	Texture difference

It is important to highlight the contextual differences between property images (e.g., on Airbnb) and food images on delivery platforms, which may influence the significance of aesthetic features. Property images often convey multi-layered information about interior design, including elements such as space, furniture, and lighting. In contrast, typical food images on delivery platforms often focus solely on the dish itself without additional objects in the scene. In some cases, merchants even omit the plate to display the food directly. Consequently, compositional techniques commonly applied to property images, such as *Diagonal dominance* or *Rule of thirds*, which is less relevant for food images. However, *Warm hue* and *Saturation*, which represent the vividness of the whole image, still exhibit statistically positive impacts.

Table 8: Regression coefficients of low-level aesthetic features on *Dishsale*

		(1) *ln_dishsale*	(2) *Aesthetics*
Composition	*Diagonal dominance*	0.028 (0.024)	-0.033*** (0.006)
	Visual balance intensity	0.628*** (0.190)	0.044 (0.068)
	Visual balance color	-0.058** (0.028)	0.008 (0.009)
	Rule of thirds	-0.055 (0.038)	-0.017* (0.010)
Color	*Warm hue*	0.607*** (0.159)	0.044 (0.040)
	Saturation	1.194*** (0.170)	0.262*** (0.038)
	Brightness	0.312 (0.523)	0.093 (0.167)
	Contrast of brightness	0.634 (0.426)	1.026*** (0.142)
	Clarity	0.153 (0.346)	0.232** (0.113)
Figure-ground Relationship	*Size difference*	-0.543* (0.292)	-0.509*** (0.098)
	Color difference	-0.019 (0.024)	-0.027*** (0.008)
	Texture difference	-0.393** (0.196)	-0.107* (0.056)
	Controls	YES	NO
	Restaurant FE	YES	NO
	N	3,988	3,988
	R-squared	0.509	0.066

ln_dishsale is the log form of *Dishsale*, transformed by *ln_dishsale*=ln(*Dishsale*+1).
Standard errors in parentheses are robust to heteroskedasticity. **p* < 0.1, ***p* < 0.05, ****p* < 0.01

Furthermore, to investigate whether the 12 image features can account for the image aesthetics in our study, we conducted a regression analysis to assess the extent to which these features predict our *Aesthetics*. Column (2) in Table 8 reports the results. In the category of composition, *Diagonal dominance* exhibits a notable coefficient of -0.033 at a significance level of 1%. *Rule of thirds* has a coefficient of -0.017 at 10% significance level, which means weak explanation power to *Aesthetics*. In the category of color, *Saturation*, *Contrast of brightness* and *Clarity* show significant values of 0.262, 1.026 and 0.232, respectively. A much larger value than any other features emphasize the importance of *Contrast of brightness* in predicting our *Aesthetics*. In the category of figure-ground relationship, all three differences of size, color, texture show significant relevance with *Aesthetics*. In summary, *Aesthetics* in our study predominantly captures elements related to color and figure-ground relationships, while also incorporating aspects of *Diagonal dominance*. This metric serves as a comprehensive measure, integrating the 12 low-level aesthetic features identified in prior research, thus providing a holistic assessment of visual appeal.

In conclusion, our findings demonstrate that certain image features associated with aesthetics are significantly correlated with product sales. The robustness of the relationship between image aesthetics and sales is evident across various measurements, confirming the consistent positive impact of visual appeal on consumer purchasing behavior.

5.3. The Moderating Effects of Advertisement and Certification

Following **H2** and **H3**, we perform a moderation analysis by examining the interaction effects between the key variables.

$$ln\_dishsale_{i,k} = \eta_0 + \eta_1 Aesthetics_{i,k} \times \mathrm{Moderator_k} + \eta_2 Aesthetics_{i,k} + \eta_3 \boldsymbol{controls_{i,k}} + u_{i,k}. \quad (4)$$

The notation of the equation (4) remains the same as before. $\mathrm{Moderator_k}$ represents the moderator variable of the k restaurant in our analysis. We note that since the key moderator variables are measured at the restaurant level, restaurant fixed effects are no longer included in the interaction analysis due to collinearity. However, we retain the dish and restaurant level control variables to enhance the explanatory power of our findings. In summary, the interaction analysis is conducted across dishes from different restaurants.

Table 9 presents the coefficients of the interaction terms between *Aesthetics* and *Advertisement*, as well as between *Aesthetics* and *Certification*, using a stepwise analysis approach. In Column (1), the interaction term between *Advertisement* and *Aesthetics* is negative and significant (coefficient = -0.373, p < 0.01). This result suggests that the positive effect of image aesthetics on product sales diminishes as the merchant increases promotional efforts through advertising. In other words, when consumers are exposed to stronger promotional signals, they are more likely to notice the product due to advertising, which in turn reduces the influence of image aesthetics. This outcome is consistent with the notion of substitutive effects between advertisement and image aesthetics, as both are categorized as promotional signals. Hence, the findings support **H2**.

In contrast, the coefficient of the interaction term between *Aesthetics* and *Certification* is 0.252 ($p < 0.05$), as shown in Column (2) of Table 9, indicating a statistically significant positive relationship. This result suggests that platform certification enhances the positive effect of image aesthetics on product sales. This finding can be explained by the differing categorization of the two signals: while image aesthetics serves as a promotional signal, certification functions as a fair information signal provided by the platform. The presence of certification increases the perceived trustworthiness of the merchant, which in turn amplifies the credibility of the aesthetic appeal conveyed by product images. Consequently, customers are more likely to respond positively to image aesthetics when the merchant is certified, thereby strengthening its impact on sales. Thus, the evidence supports **H3**.

Column (3) of Table 9 explores whether the substitutive and complementary effects can coexist within a single model. The results indicate that the interaction term between *Aesthetics* and *Advertisement* remains significantly negative (coefficient = -0.453, $p < 0.01$), while the interaction term between *Aesthetics* and *Certification* continues to show a significantly positive coefficient of 0.333 ($p < 0.01$). These findings suggest that image aesthetics can be simultaneously enhanced by platform certification and attenuated by the merchant's advertising efforts. In other words, while platform certification bolsters the credibility and impact of image aesthetics, increased advertising may detract from its effectiveness as a promotional signal. Figure 6 visualizes these moderating effects. *Advertisement* inverts the positive relationship between *ln_dishsale* and *Aesthetics*, while *Certification* steepens the curve. This graphical evidence confirms that advertisement negatively moderates the effect of image aesthetics on product sales, whereas platform certification positively moderates it, as hypothesized.

Moreover, we also conduct a 2SLS analysis for the interaction effects, as shown in Column (4) of Table 9. The instrumental variable setting is the same as Table 6 and only the second step results are presented for simplicity. However, our goal is not to estimate the causal effects of the interaction term but to provide additional evidence that accounts for endogeneity, thereby corroborating our moderating findings. Although the significant level of the interaction term between *Aesthetics* and *Certification* declines to 10%, the signs of the coefficients for both interaction terms remain unchanged, further verifying the consistency of our results.

Table 9: Interaction test for *Advertisement* and *Certification*

	(1)*ln_dishsale*	(2) *ln_dishsale*	(3) *ln_dishsale*	(4) *ln_dishsale* (2SLS)
Aesthetics X *Advertisement*	-0.373*** (0.129)	-	-0.453*** (0.133)	-2.091*** (0.241)
Aesthetics X *Certification*	-	0.252** (0.108)	0.333*** (0.112)	0.414* (0.224)
Aesthetics	0.344*** (0.055)	0.194*** (0.062)	0.249*** (0.064)	0.873*** (0.134)
Controls	YES	YES	YES	YES
N	5,074	5,074	5,074	5,074
R-squared	0.286	0.285	0.287	0.255

ln_dishsale is the log form of *Dishsale*, transformed by *ln_dishsale*=ln(*Dishsale*+1).
Standard errors in parentheses are robust to heteroskedasticity. *p < 0.1; **p < 0.05; ***p < 0.01

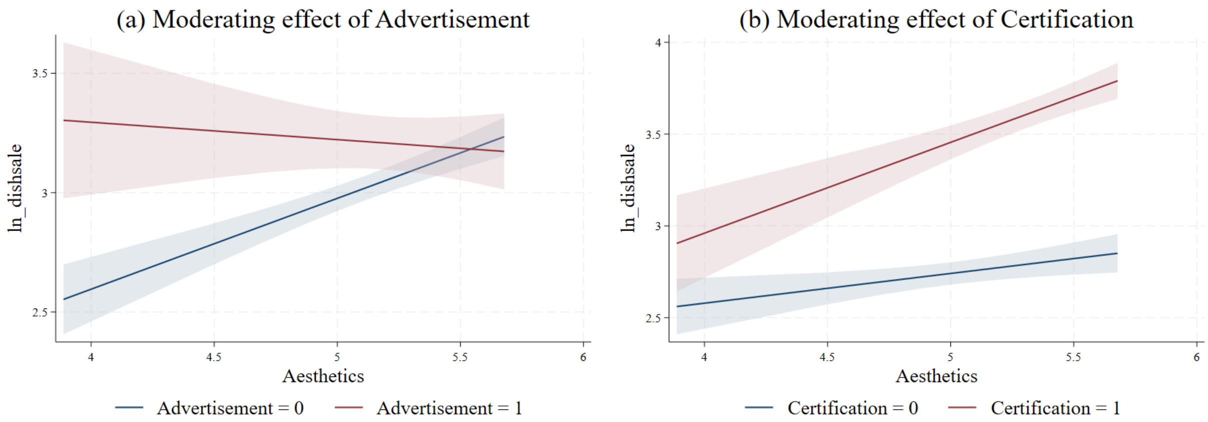

Figure 6. plots of the moderating effects for *Advertisement* and *Certification* with confidence intervals

5.4. Robustness Check on Signal Categorization

We conduct a robustness check by considering alternative sources of signals, specifically using merchant-provided insurance details (Zhang et al., 2022a) and announcements of new merchant openings (Aguiar and Waldfogel, 2018). These variables serve as proxies for promotional signals and reputation signals, respectively, to examine whether the previously observed interaction effects persist within and across different signaling mechanisms.

On e-commerce platforms, product insurance serves a dual role: it provides post-purchase protection for customers and acts as a pre-purchase promotional signal for merchants. By offering insurance, merchants aim to attract more customers and differentiate their products from competitors, effectively signaling a commitment to product quality (Zhang et al., 2022a). From a signaling perspective, insurance functions similarly to advertising, as it conveys a message of reliability and reduces customers' perceived risks.

Given this overlap in signaling intent, both insurance and image aesthetics fall into the category of promotional signals. When strong insurance coverage is offered, it can substitute for the reassurance typically provided by appealing product images. Customers may feel less need to scrutinize the product's visual presentation, knowing that the insurance offers a safeguard against potential dissatisfaction. Consequently, the reliance on image aesthetics is reduced, leading to a diminished effect. Thus, we hypothesize a substitutive relationship between insurance and image aesthetics, as customers shift their attention away from the visual appeal toward the explicit quality assurance provided by the insurance.

New merchant openings provide another distinct type of information on e-commerce platforms. In our sample, any restaurant that has been in operation for less than one month is designated as "new" by the platform. Such a "new" label can generate interest among customers who are eager to explore fresh options, thereby offering potential benefits to the merchant (Aguiar and Waldfogel, 2018). However, the label also signals a lack of reviews and ratings (Liang et al., 2024). When combined with image aesthetics, the "new" label may exhibit a complementary effect due to its role as a reputation signal. For newly opened restaurants lacking customer feedback, visual presentation through high-quality images becomes an important cue for customers to assess the product. The label may attract customers willing to try new offerings, prompting them to pay closer attention to the product's visual appeal. As a result, the positive impact of image aesthetics is amplified in this context, leading to a complementary relationship between the "new opening" label and image aesthetics.

Table 10 presents the regression results with the same model specifications as in the previous analysis. Column (1) shows the interaction effect between *Aesthetics* and *Insurance*, with a coefficient of -0.257, significant at the 5% level. The main effects of *Aesthetics* (coefficient = 0.501) is also verified in the regression. These findings support the hypothesis that the presence of insurance diminishes the effect of image aesthetics on product sales. Since both insurance and image aesthetics are categorized as promotional signals aimed at reducing product uncertainty, they exhibit a substitutive relationship. The overlap in their signaling function leads to a reduced marginal impact of aesthetics when insurance is present. This substitutive effect between factors within the same category of promotional signals aligns with the results observed in **H2**.

Table 10: Interactive effect with *Insurance* and *New opening*

	(1) *ln_dishsale*	(2) *ln_dishsale*	(3) *ln_dishsale*	(4) *ln_dishsale* (2SLS)
Aesthetics X *Insurance*	-0.257** (0.118)	-	-0.260** (0.120)	-0.455* (0.232)
Aesthetics X *New Opening*	-	0.525*** (0.139)	0.527*** (0.142)	1.108*** (0.277)
Aesthetics	0.501*** (0.103)	0.228*** (0.055)	0.425*** (0.104)	0.902*** (0.196)
Controls	YES	YES	YES	YES
N	5,074	5,074	5,074	5,074
R-squared	0.277	0.278	0.279	0.266

ln_dishsale is the log form of *Dishsale*, transformed by *ln_dishsale*=ln(*Dishsale*+1).
Standard errors in parentheses are robust to heteroskedasticity. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Column (2) of Table 10 reports the results for the interaction between *Aesthetics* and *New Opening*. The interaction term exhibits a significantly positive coefficient of 0.525. The main effects of *Aesthetics* (coefficient = 0.228) and *New Opening* (not reported in the table) are also included in the regression. These findings indicate that the "new opening" label enhances the effect of image aesthetics on product sales, confirming a complementary relationship between image aesthetics and new opening. The positive interaction suggests that the presence of the new opening label increases the importance of visual appeal in driving customer engagement, as the label compensates for the lack of prior reviews by amplifying the credibility of the aesthetic signal. This result also supports the notion of complementary effects between signals from different categories—namely, promotional signals (image aesthetics) and reputation signals (new opening). Thus, the evidence is consistent with **H3**.

Column (3) of Table 10 presents a model that simultaneously includes the interaction terms of *Aesthetics* X *Insurance*, and *Aesthetics* X *New Opening*. The interaction term for *Aesthetics* X *Insurance* has a coefficient of -0.260, significant at the 5% level. In contrast, the interaction term for *Aesthetics* X *New Opening* yields a positive coefficient of 0.527, significant at the 1% level. These results once again demonstrate the coexistence of substitutive and complementary effects. The negative interaction with insurance confirms a substitutive relationship, where insurance and image aesthetics both serve as promotional signals, reducing product uncertainty in similar ways. Conversely, the positive interaction with new opening highlights a complementary relationship, as the new opening label enhances the credibility of image aesthetics by acting as a reputation signal. Similarly, we visualize the interaction effects in Figure 7. As shown in Figure 7a, the moderation effect of *Insurance* does not invert the slope as *Advertisement*, but flattens the curve. In contrast, *New Opening* steepens the curve and amplifies the effect of image aesthetics on product sales. The graphical evidence reaffirms these interaction effects.

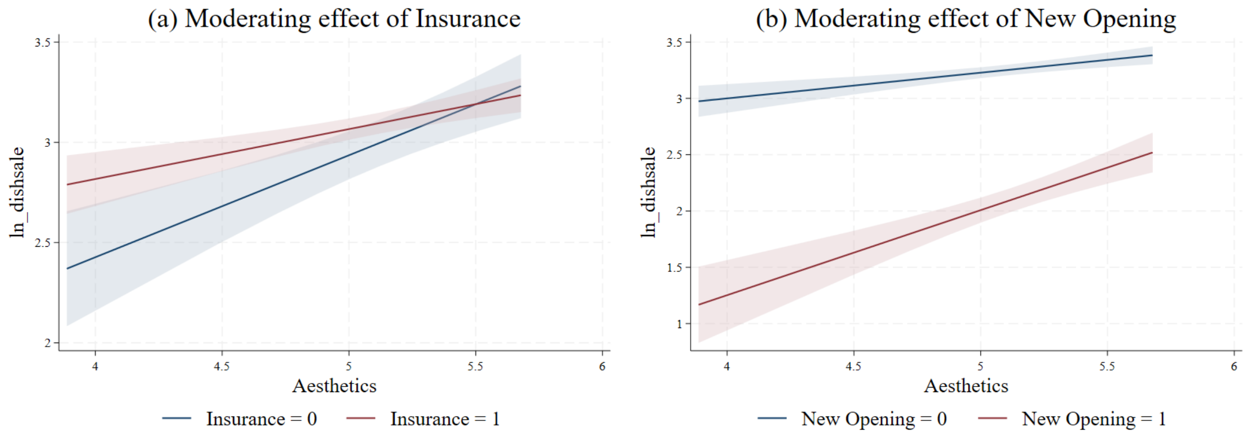

Figure 7. plots of the moderating effects for *Insurance* and *New Opening* with confidence intervals

In column (4) of Table 10, we perform a 2SLS analysis to address potential endogeneity concerns. The significance level of the interaction term for *Insurance* drops to 10%. Nonetheless, the signs of the two interaction terms remain consistent, suggesting that the substitutional and complementary effects continue to coexist as previously observed. Together, these findings provide robust evidence for the nuanced interplay between signals from the same and different categories, supporting the hypotheses outlined in **H2** and **H3**.

6. Discussion

This study examines the role of image aesthetics in influencing product sales within the framework of signaling theory. Our findings show that image aesthetics can significantly boost sales by over 4.8%, aligning with the positive effects reported in prior research (e.g., Goswami et al., 2011; Zhang et al., 2022b). Unlike mixed findings reported in earlier studies, we find a consistent positive impact in the context of food delivery, where food images help reduce product uncertainty, akin to visuals for experience goods like Airbnb listings.

Moreover, our analysis reveals the moderating effects on image aesthetics. It acts as a promotional signal, showing a substitutive relationship with advertising and insurance, where overlapping functions diminish its impact. In contrast, reputation signals such as platform certification and new merchant status complement image aesthetics, enhancing its effectiveness and leading to higher sales. This interplay mirrors findings on diverse recommendation signals (Xu et al., 2020) and credibility indicators (Zhou et al., 2022), while also supporting the view of product uncertainty reduction (Dimoka et al., 2012).

Our paper makes two key contributions to the theoretical understanding of the use of visual information in e-commerce, especially in the context of food delivery platforms. First, we advance the framework of signal categorization by identifying boundary conditions that distinguish between substitutive and complementary effects of signals. Drawing on signaling theory and the perspective of product uncertainty, we classify signals based on their source, operability, and credibility into two categories: promotional signals and reputation signals. Signals within the same category, such as image aesthetics and insurance, act as substitutes due to their overlapping promotional functions and lower credibility. In contrast, signals across categories, such as image aesthetics and platform certification, exhibit complementary effects because they address different aspects of uncertainty. This categorization can be extended to a broader range of signals in e-commerce, highlighting consistent patterns of substitutive and complementary effects.

Second, our study is among the first to apply signaling theory to study image aesthetics. We investigate the main effect of image aesthetics using both GPSM and instrumental variable approaches. Together, these analyses offer valuable insights into the role of image aesthetics in online transactions. Additionally, our CNN-based model provides a scalable and reliable method for measuring image aesthetics, enabling high-level feature extraction and quantification of aesthetic perception.

In addition, our findings offer valuable practical implications. Despite that the food delivery industry reached a market size of \$353.3 billion in the U.S. in 2024,[9] small vendors on food delivery platforms still often struggle with

[9] https://www.statista.com/topics/3294/online-food-delivery-services-in-the-us/#topicOverview

digital menu organization, causing lower sales performance compared to large franchise restaurants. Our study provides strong evidence that leveraging image aesthetics can help boost product sales. Unlike franchise restaurants that employ marketing professionals to carefully craft their visual presentations, small vendors can benefit from aesthetically enhanced food images to attract customers. Moreover, our model offers a hand-on calculator of image aesthetics, enabling merchants to refine their image effectively.

Additionally, our findings on the interactive effects of signals provide practical guidance for optimizing the information presented to customers. While small vendors can implement various strategies to enhance their menu display, each approach incurs operational costs, such as platform advertisements or customer insurance. Given budget constraints, maximizing profit is a key objective for these restaurants. Our study highlights the importance of choosing the right mix of informational signals. Using the signal categorization framework, we show that promotional efforts like advertisements or insurance may substitute for the benefits of enhancing image aesthetics, leading to inefficient use of resources. Instead, we recommend small vendors focus on obtaining platform certification, which complements the effort put into image beautification without incurring additional costs. By strategically arranging information, merchants can boost product sales and improve profitability more effectively.

This research has limitations. First, there is an issue with the interpretability of the aesthetic assessment model. While other studies focus on low-level image features that are objective and easily explained, our approach captures consumer perceptions, resulting in a trade-off with reduced interpretability. Second, the study lacks time-series data. Analyzing changes in image aesthetics over time could help mitigate concerns about omitted variable bias and provide stronger causal evidence. Given the cross-sectional nature of our data, we relied on GPSM and instrumental variable methods. Future research using panel data could apply techniques like difference-in-differences to strengthen causal inferences.

Acknowledgement

This work was supported by National Natural Science Foundation of China (72033003, 72271058, 72342012), and the Major Project of National Social Science Fund of China (21&ZD119). We also gratefully acknowledge the helpful comments of the Associate Editor and two anonymous reviewers.

REFERENCES

Aguiar, L. and Waldfogel, J. (2018). Quality Predictability and the Welfare Benefits from New Products: Evidence from the Digitization of Recorded Music. *Journal of Political Economy*, *126*(2), 451–902.

Akerlof, G. A. (1970). The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500.

Atasoy, H., Banker, R. D. and Pavlou, P. A. (2016). On the Longitudinal Effects of IT Use on Firm-Level Employment. *Information Systems Research*, *27*(1), 6–26.

Bapna, S. (2019). Complementarity of Signals in Early-Stage Equity Investment Decisions: Evidence from a Randomized Field Experiment. *Management Science*, *65*(2), 933–952.

Baráth, L. and Fertő, I. (2024). The relationship between the ecologisation of farms and total factor productivity: A continuous treatment analysis. *Journal of Agricultural Economics*, *75*(1), 404–424.

Bia, M. and Mattei, A. (2008). A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score. *The Stata Journal*, *8*(3), 354–373.

Bloch, P. H., Brunel, F. F. and Arnold, T. J. (2003). Individual Differences in the Centrality of Visual Product Aesthetics: Concept and Measurement. *Journal of Consumer Research*, *29*(4), 551–565.

Brewer, P. (2021). The effect of online restaurant menus on consumers’ purchase intentions during the COVID-19 pandemic. *International Journal of Hospitality Management*, *94*(2021), 102777.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv.

Cai, R. and Chi, C. G.-Q. (2021). Pictures vs. reality: Roles of disconfirmation magnitude, disconfirmation sensitivity, and branding. *International Journal of Hospitality Management*, *98*(2021), 103040.

Choi, E., Eom, H. and Moon, J. (2024). Understanding Product Attributes Associated with Food Sales: Sensory Attributes for Online Versus Offline Options of a Multi-Channel Retailer. *Journal of Electronic Commerce Research*, *25*(1).

CLAID.AI. (2024). *How Food Photos Boost Deliveries by 15%: The Impact of Visual Content on Food Tech*. https://claid.ai/blog/article/impact-of-visual-content-on-food-tech/

Clark, J. M. and Paivio, A. (1987). A Dual Coding Perspective on Encoding Processes. In M. A. McDaniel and M. Pressley (Eds.), *Imagery and Related Mnemonic Processes: Theories, Individual Differences, and Applications* (5–33. pp.). Springer.

Clarke, D. and Matta, B. (2018). Practical Considerations for Questionable IVs. *The Stata Journal: Promoting Communications on Statistics and Stata*, *18*(3), 663–691.

Conley, T. G., Hansen, C. B. and Rossi, P. E. (2012). Plausibly Exogenous. *The Review of Economics and Statistics*, *94*(1), 260–272.

Connelly, B. L., Certo, S. T., Ireland, R. D. and Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, *37*(1), 39–67.

Coyle, J. R. and Thorson, E. (2001). The Effects of Progressive Levels of Interactivity and Vividness in Web Marketing Sites. *Journal of Advertising*, *30*(3), 65–77.

Cyr, D., Head, M. and Ivanov, A. (2006). Design aesthetics leading to m-loyalty in mobile commerce. *Information & Management*, *43*(8), 950–963.

Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, *49*(10), 1407–1424.

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, and Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Dhar, S., Ordonez, V. and Berg, T. L. (2011). High level describable attributes for predicting aesthetics and interestingness. *2011 IEEE Conference on Computer Vision and Pattern Recognition*, 1657–1664.

Dimoka, A., Hong, Y. and Pavlou, P. A. (2012). On Product Uncertainty in Online Markets: Theory and Evidence. *MIS Quarterly*, *36*(2), 395–426.

Ericsson, L., Gouk, H. and Hospedales, T. M. (2021). How Well Do Self-Supervised Models Transfer? *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5414–5423.

Goh, K.-Y., Heng, C.-S. and Lin, Z. (2013). Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. *Information Systems Research*, *24*(1), 88–107.

Goswami, A., Chittar, N. and Sung, C. H. (2011). A study on the impact of product images on user clicks for online shopping. *Proceedings of the 20th International Conference Companion on World Wide Web*, 45–46.

Guan, Y., Tan, Y., Wei, Q. and Chen, G. (2023). When Images Backfire: The Effect of Customer-Generated Images on Product Rating Dynamics. *Information Systems Research*, *34*(4), 1641–1663.

Hagtvedt, H. and Patrick, V. M. (2008). Art Infusion: The Influence of Visual Art on the Perception and Evaluation of Consumer Products. *Journal of Marketing Research*, *45*(3), 379–389.

He, J., Li, B. and Wang, X. (2023). Image Features and Demand in the Sharing Economy: A Study of Airbnb. *International Journal of Research in Marketing*, *40*(4), 760–780.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Hirano, K. and Imbens, G. W. (2004). The Propensity Score with Continuous Treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (73–84. pp.). John Wiley & Sons, Ltd.

Hong, Y. and Pavlou, P. A. (2014). Product Fit Uncertainty in Online Markets: Nature, Effects, and Antecedents. *Information Systems Research*, *25*(2), 328–344.

Hong, Y. and Pavlou, P. A. (2017). On Buyer Selection of Service Providers in Online Outsourcing Platforms for IT Services. *Information Systems Research*, *28*(3), 547–562.

Hou, J.-R., Zhang, J. and Zhang, K. (2023). Pictures That Are Worth a Thousand Donations: How Emotions in Project Images Drive the Success of Online Charity Fundraising Campaigns? An Image Design Perspective. *MIS Quarterly*, *47*(2), 535–584.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications* (arXiv:1704.04861). arXiv.

Jiang, Z. and Benbasat, I. (2007). Research Note—Investigating the Influence of the Functional Mechanisms of Online Product Presentations. *Information Systems Research*, *18*(4), 454–470.

Killgore, W. D. S. and Yurgelun-Todd, D. A. (2006). Affect modulates appetite-related brain activity to images of food. *International Journal of Eating Disorders*, *39*(5), 357–363.

Kirmani, A. (1990). The Effect of Perceived Advertising Costs on Brand Perceptions. *Journal of Consumer Research*, *17*(2), 160–171.

Kisielius, J. and Sternthal, B. (1984). Detecting and Explaining Vividness Effects in Attitudinal Judgments. *Journal of Marketing Research*, *21*(1), 54–64.

Kpossa, M. R. and Lick, E. (2020). Visual merchandising of pastries in foodscapes: The influence of plate colours on consumers' flavour expectations and perceptions. *Journal of Retailing and Consumer Services*, *52*(2020), 101684.

Kumar, A. and Tan, Y. (2015). The Demand Effects of Joint Product Advertising in Online Videos. *Management Science*, *61*(8), 1921–1937.

Le, T. T., Tuyet, N. B. T., Anh, T. L., Kim, N. D. T., Thai, N. T. T. and Lan, A. N. (2023). The effects of online restaurant menus on consumer purchase intention: evidence from an emerging economy. *British Food Journal*, *125*(7), 2663–2679.

Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lee, B.-C., Ang, L. and Dubelaar, C. (2005). Lemons on the Web: A signalling approach to the problem of trust in Internet commerce. *Journal of Economic Psychology*, *26*(5), 607–623.

Li, H., Ji, H., Liu, H., Cai, D. and Gao, H. (2022). Is a picture worth a thousand words? Understanding the role of review photo sentiment and text-photo sentiment disparity using deep learning algorithms. *Tourism Management*, *92*, 104559.

Liang, C., Hong, Y. and Gu, B. (2024). Monitoring and the Cold Start Problem in Digital Platforms: Theory and Evidence from Online Labor Markets. *Information Systems Research*, *Advance online publication*.

Liao, C., To, P.-L., Wong, Y.-C., Palvia, P. and Kakhki, M. D. (2016). The Impact of Presentation Mode and Product Type on Online Impulse Buying Decisions. *Journal of Electronic Commerce Research*, *17*(2), 153–168.

Liu, S., Wu, L. L., Yu, X. and Huang, H. (2022a). Marketing online food images via color saturation: A sensory imagery perspective. *Journal of Business Research*, *151*(2022), 366–378.

Liu, X., Wu, X., Shi, W., Tong, W., Parkside, F. K. and Ye, Q. (2022b). The Impacts of Electronic Word-of-Mouth on High- Involvement Product Sales: Moderating Effects of Price, Brand Origin, and Number of Customers. *Journal of Electronic Commerce Research*, *23*(3), 177–189.

Lu, B. and Chen, Z. (2021). Live streaming commerce and consumers' purchase intention: An uncertainty reduction perspective. *Information & Management*, *58*(7), 103509.

Maier, E. and Dost, F. (2018). Fluent contextual image backgrounds enhance mental imagery and evaluations of experience products. *Journal of Retailing and Consumer Services*, *45*(2018), 207–220.

Mao, A., Mohri, M. and Zhong, Y. (2023). Cross-Entropy Loss Functions: Theoretical Analysis and Applications. *Proceedings of the 40th International Conference on Machine Learning*, 23803–23828.

Mavlanova, T., Benbunan-Fich, R. and Koufaris, M. (2012). Signaling theory and information asymmetry in online commerce. *Information & Management*, *49*(5), 240–247.

Murray, N., Marchesotti, L. and Perronnin, F. (2012). AVA: A large-scale database for aesthetic visual analysis. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2408–2415.

Palmer, S. E., Schloss, K. B. and Sammartino, J. (2013). Visual Aesthetics and Human Preference. *Annual Review of Psychology*, *2013*(64), 77–107.

Pavlou, P. A., Liang, H. and Xue, Y. (2007). Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective. *MIS Quarterly*, *31*(1), 105–136.

Pierce, M. (2022). *High-quality food photos can increase orders on restaurant delivery apps by 35% | Snappr For Enterprise Blog*. https://www.snappr.com/enterprise-blog/high-quality-food-photos-can-increase-orders-on-restaurant-delivery-apps-by-35, https://www.snappr.com/enterprise-blog/high-quality-food-photos-can-increase-orders-on-restaurant-delivery-apps-by-35

Price, L. J. and Dawar, N. (2002). The joint effects of brands and warranties in signaling new product quality. *Journal of Economic Psychology*, *23*(2), 165–190.

Rao, T., Li, X., Zhang, H. and Xu, M. (2019). Multi-level region-based Convolutional Neural Network for image emotion classification. *Neurocomputing*, *333*(2019), 429–439.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, *70*(1), 41–55.

Rubner, Y., Tomasi, C. and Guibas, L. J. (1998). A metric for distributions with applications to image databases. *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 59–66.

Sample, K. L., Hagtvedt, H. and Brasel, S. A. (2020). Components of visual perception in marketing contexts: a conceptual framework and review. *Journal of the Academy of Marketing Science*, *48*(3), 405–421.

Schenkman, B. N. and Jönsson, F. U. (2000). Aesthetics and preferences of web pages. *Behaviour & Information Technology*, *19*(5), 367–377.

Setia, P., Bayus, B. L. and Rajagopalan, B. (2020). The Takeoff of Open Source Software: A Signaling Perspective Based on Community Activities. *MIS Quarterly*, *44*(3), 1439–1458.

Simonyan, K. and Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv.

Sohn, S. (2017). Consumer processing of mobile online stores: Sources and effects of processing fluency. *Journal of Retailing and Consumer Services*, *36*(2017), 137–147.

Talebi, H. and Milanfar, P. (2018). NIMA: Neural Image Assessment. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, *27*(8), 3998–4011.

Toepel, U., Knebel, J.-F. and Hudry, J. (2009). The brain tracks the energetic value in food images. *NeuroImage*, *44*(2009), 967–974.

Unnava, H. R. and Burnkrant, R. E. (1991). An Imagery-Processing View of the Role of Pictures in Print Advertisements. *Journal of Marketing Research*, *May, 1991*(28), 226–231.

Vermeir, I. and Roose, G. (2020). Visual Design Cues Impacting Food Choice: A Review and Future Research Agenda. *Foods*, *2020*(9), 1495.

Versace, F., Frank, D. W., Stevens, E. M., Deweese, M. M., Guindani, M. and Schembre, S. M. (2019). The reality of “food porn”: Larger brain responses to food-related cues than to erotic images predict cue-induced eating. *Psychophysiology*, *2019*(56:e13309).

Wang, M., Li, X. and Chau, P. Y. K. (2016). The Impact of Photo Aesthetics on Online Consumer Shopping Behavior: An Image- Processing-Enabled Empirical Study. *37th International Conference on Information Systems*. 37th International Conference on Information Systems, Dublin.

Wang, Q., Li, B. and Singh, P. V. (2018). Copycats vs. Original Mobile Apps: A Machine Learning Copycat-Detection Method and Empirical Analysis. *Information Systems Research*, *29*(2), 273–291.

Wang, X. and Ding, Y. (2022). The impact of monetary rewards on product sales in referral programs: The role of product image aesthetics. *Journal of Business Research*, *145*, 828–842.

Wang, X., Guo, J., Wu, Y. and Liu, N. (2019). Emotion as signal of product quality: Its effect on purchase decision based on online customer reviews. *Internet Research*, *30*(2), 463–485.

Wells, J. D., Valacich, J. S. and Hess, T. J. (2011). What Signal Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions. *MIS Quarterly*, *35*(2), 373–396.

Xu, J. (David), Benbasat, I. and Cenfetelli, R. T. (2020). The Relative Effect of the Convergence of Product Recommendations from Various Online Sources. *Journal of Management Information Systems*, *37*(3), 788–819.

Yamamoto, M. and Lambert, D. R. (1994). The Impact of Product Aesthetics on the Evaluation of Industrial Products. *Journal of Product Innovation Management*, *11*(4), 309–324.

Yoo, S. H., Ozer, M. and Jingjun, X. (2024). Relative effects of the different bundles of web-design features on intentions to purchase experience products online. *Decision Support Systems*, *179*(2024), 114171.

Zhang, C., Yu, M. and Chen, J. (2022a). Signaling Quality with Return Insurance: Theory and Empirical Evidence. *Management Science*, *68*(8), 5847–5867.

Zhang, M. and Luo, L. (2023). Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp. *Management Science*, *69*(1), 25–50.

Zhang, S., Lee, D., Singh, P. V. and Srinivasan, K. (2022b). What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features. *Management Science*, *68*(8), 5644–5666.

Zhou, J., Kishore, R., Amo, L. and Ye, C. (2022). Description and Demonstration Signals as Complements and Substitutes in an Online Market for Mental Health Care. *MIS Quarterly*, *46*(4), 2055–2084.

APPENDIXES

Appendix A. Subgroup analysis of the heterogeneous effects of aesthetics

We further examine the heterogeneous effects of aesthetics on product sales through subgroup analysis. The analysis considers four moderating variables: *Advertisement*, *Certification*, *Insurance*, and *New opening*. The detailed results are presented below.

Table A1. Heterogeneous analysis on *Advertisement*

	Advertisement=0	*Advertisement*=1
Restaurant num	60	11
Dish num	4,100	974
Aesthetics Mean	5.068	5.144
Aesthetics Coefficient	0.363***(0.055)	0.223*(0.126)

Table A2. Heterogeneous analysis on *Certification*

	Certification=0	*Certification*=1
Restaurant num	46	25
Dish num	3,066	2,008
Aesthetics Mean	4.965	5.262
Aesthetics Coefficient	0.223***(0.062)	0.502***(0.092)

Table A3. Heterogeneous analysis on *Insurance*

	Insurance=0	*Insurance*=1
Restaurant num	16	55
Dish num	1,234	3,840
Aesthetics Mean	5.140	5.064
Aesthetics Coefficient	0.528***(0.107)	0.175***(0.058)

Table A4. Heterogeneous analysis on *New Opening*

	New Opening=0	*New Opening*=1
Restaurant num	55	16
Dish num	4,267	807
Aesthetics Mean	5.077	5.110
Aesthetics Coefficient	0.142**(0.056)	0.816***(0.112)

The results demonstrate significantly positive effects across all subgroups. Specifically, the subgroup of new restaurants shows the largest effect of image aesthetics, with a coefficient of 0.816 significant at the 1% level, demonstrating the substantial impact of new openings. The platform certification strengthens the effect of image aesthetics, while both advertising and providing insurance weaken it.