

Research

Brand visibility in packaging: a deep learning approach for logo detection, saliency-map prediction, and logo placement analysis

Alireza Hosseini¹ · Kiana Hooshanfar¹ · Pouria Omrani² · Reza Toosi³ · Ramin Toosi¹ · Zahra Ebrahimian¹ · Mohammad Ali Akhaee¹

Received: 1 February 2025 / Accepted: 24 April 2025

Published online: 24 May 2025

© The Author(s) 2025 [OPEN](#)

Abstract

The visibility of brand logos on packaging plays a crucial role in shaping consumer perception, directly influencing the product's success. Analyzing eye-tracking data across large groups of individuals is both costly and time-intensive. Therefore, there is a growing need to develop models that capture human visual attention behavior effectively. This paper introduces a framework that models attention in the human visual system to brand logos on packaging designs, to measure brand logo visibility and its impact on consumer perception. The proposed method consists of three main steps. The first step leverages YOLOv8 for logo detection across well-known datasets. The second step involves introducing a novel saliency prediction model tailored for the packaging context to model human visual attention. In the third step, by integrating logo detection with a saliency map generation, the framework provides a brand attention score. The effectiveness of the proposed method is assessed module by module, ensuring a thorough evaluation of each component. Comparing logo detection and saliency map prediction with SOTA models shows the superiority of the proposed methods. To investigate the robustness of the proposed brand attention score, we collected a dataset to examine previous psychophysical hypotheses related to brand visibility. The results show that the brand attention score is in line with all previous studies. Also, we introduced seven new hypotheses to check the impact of position, orientation, and other visual elements on brand attention. This research marks a stride in the intersection of cognitive psychology, computer vision, and marketing.

Article highlights

- Provides a framework to assess logo prominence on packaging and its impact on viewer focus.
- Introduces a new model to predict where people look in commercials and packaging images.
- Explores 12 consumer insights, including effects of logo position and orientation on visibility.

Keywords Brand attention · Neuro marketing · Logo detection · Saliency prediction

✉ Mohammad Ali Akhaee, akhaee@ut.ac.ir; Alireza Hosseini, arhosseini77@ut.ac.ir; Kiana Hooshanfar, k.hooshanfar@ut.ac.ir; Pouria Omrani, pouria.omrani@ieee.org; Reza Toosi, rtoosi81@gmail.com; Ramin Toosi, r.toosi@ut.ac.ir; Zahra Ebrahimian, z.ebrahimian@ut.ac.ir | ¹School of Electrical and Computer Engineering, University of Tehran, College of Engineering, Tehran, Iran. ²Faculty of Electrical Engineering, K. N. Toosi University of Technology, Tehran, Iran. ³Department of Computer Engineering, Faculty of Engineering, Golestan University, Gorgan, Iran.



1 Introduction

In today's dynamic business world, having a strong brand presence is crucial. The visibility of the brand is incredibly important for keeping up with consumer trends and staying competitive. Consumers often shape their perceptions of brands by considering factors such as visual attractiveness, functionality, and the social significance they convey, predominantly relying on visual cues [1]. For companies striving to establish and maintain a strong market presence, the packaging of their products, as an interface between the brand and the consumer, significantly influences the purchasing process [2].

The visual appeal of packaging, along with the prominent display of design elements, contributes to creating a lasting impression on the consumer and nurturing brand recognition. As consumers navigate the diverse market landscape, a well-designed package captures attention and effectively conveys the brand's values and identity, playing a key role in influencing the purchasing decision [3].

Several studies in marketing and consumer behavior have emphasized the role of effective packaging design in promoting brand recognition [4]. A well-designed packaging has been shown to significantly enhance brand awareness, purchase intent, and sales [5]. These investigations thoroughly explore various aspects of packaging design, conducting a detailed examination of elements such as packaging's shape, texture, and color [6–10]. Additionally, they explore the strategic considerations of precise positioning of design elements such as logos, aiming to uncover the subtle interactions between these factors and their impact on consumer perception and brand recognition.

Recognizing the impact of visual elements in packaging, particularly logos, on shaping brand recognition and recall is crucial [11, 12]. This visual aspect influences consumer responses, ultimately playing an important factor in determining the success of a product [13]. Logo, as a fundamental visual element, plays an essential role in packaging design, significantly influencing how consumers perceive and remember a brand [14]. A visually appealing package not only captures the consumer's attention but also enhances the visibility of the brand logo. On the flip side, weaknesses in design can hinder logo visibility, diminishing its potential impact on consumer awareness [15, 16].

Understanding the crucial influence of logo visibility on brand awareness highlights the importance of implementing effective methods to enhance logo visibility. Enhancing logo visibility is linked to understanding its strategic placement on the packaging. The positioning of a logo profoundly impacts its visibility, influencing its interaction with other design elements and resonance with consumers. Thus, it is imperative to focus on optimizing logo placement through strategic positioning on packaging. To this end, implementing advanced machine vision techniques to measure logo visibility becomes crucial to amplifying visibility. Identifying the logo's position within an image is the initial stage in assessing logo visibility [17]. The pursuit of brand visibility does not conclude with knowing the location of the logo; it extends to understanding the attention it commands within the consumer's visual field. This is where saliency prediction [18] emerges as a pivotal metric. Saliency prediction involves forecasting the perceptual prominence of the logo within the overall visual composition of packaging. Understanding the saliency prediction of the logo enables us to quantify its presence and visual impact, offering a detailed understanding of how much attention the brand attracts on the visual journey of consumers.

While there have been numerous studies on logo detection and saliency prediction, to the best of our knowledge, there is currently no specialized method for modeling human visual attention specifically for logos on packaging. The proposed method is positioned to provide a comprehensive framework for modeling human attention to brand logos in various packaging scenarios including automated logo detection and a novel saliency prediction algorithm. This approach is crafted to provide businesses with actionable insights aimed at optimizing logo visibility and creating engaging packaging designs that effectively connect with their target audience. It is composed of three key modules. The initial module of our design is brand logo detection, leveraging the cutting-edge YOLOv8, a state-of-the-art (SOTA) object detection model developed by Ultralytics [19]. This crucial step helps identify and precisely locate brand logos in visual content. Subsequently, the second module, utilizing a CNN-Transformer-based model generates saliency maps, a crucial element of our methodology. These maps highlight specific regions within the visuals that command the highest visual attention. These insights provide valuable information regarding viewer perception and cognitive responses. The third and concluding module efficiently integrates the outcomes of both logo detection and saliency map generation. This integration yields a score that quantifies the attention that the brand logo attracts within packaging or advertising visuals. Furthermore, it is noteworthy to mention that this approach has been validated against existing psychophysical studies related to brand logos in packaging. This validation underscores the capability of the model to simulate human visual attention on brand logos within packaging

and advertising imagery accurately. Consequently, this positions our model as a tool for investigating unexplored experiments regarding brand logos in packaging and advertising contexts. Through this approach, the proposed model provides a comprehensive analysis of brand visual attention, enabling businesses to make informed decisions to enhance their brand presence and impact. Our main contributions are as follows:

- An innovative framework that models human visual attention to brand logos on packaging.
- A new saliency prediction model, specifically designed for advertising images and packaging considering text maps, surpasses SOTA models in saliency prediction.
- Introduced a brand attention dataset explores 12 hypotheses from cognitive perspectives.
- Validated the effectiveness of the brand attention score through extensive comparisons with existing psychophysical studies.
- Introduced seven new hypotheses to understand the impact of logo position, orientation, and other design elements on brand visibility.

The rest of this work is organized into four sections. Section 2 delves into related work in the field, specifically focusing on optimizing logo placement through eye-tracking, brand logo detection, and saliency map prediction. Section 3 outlines the materials, methods, and modeling procedures employed in the research. Section 4 is dedicated to discussing the experiments conducted and the results obtained. Finally, Sect. 5 presents the main conclusions of the work, while proposing future directions and potential enhancements for the introduced architecture.

2 Related works

In this section, we will go through the domain of artificial intelligence (AI) and its applications in the field of marketing, specifically focusing on logo placement design in advertising images and packaging. We will also explore the techniques of brand logo detection and saliency map prediction, discussing their relevance to enhancing brand recognition and optimizing advertising effectiveness.

2.1 Optimizing logo placement with eye-tracking

Neuromarketing, an increasingly influential field of study, uniquely utilizes neuroscience knowledge to directly assess product packaging, eliminating the need to depend on consumers' self-reported preferences [20]. By incorporating advanced methodologies like neuroimaging and physiological measurements, neuromarketing employs a more direct and objective approach to assessing consumer responses. This represents a notable shift away from traditional survey-based approaches. A key methodology in neuromarketing is eye tracking [21], providing a detailed examination of visual attention patterns. By studying where and how consumers focus their gaze, researchers gain valuable insights into elements that capture attention and drive perception, uncovering processes beyond conscious awareness [22, 23].

Specific parameters govern visual behavior, with fixations playing a central role in this context. Fixations, characterized by eye movement, represent moments when the visual system actively acquires information [24]. Numerous studies exploring eye movements, the attention mechanism, and consumer behavior have consistently emphasized the importance of analyzing fixations based on their frequency and duration [25]. By understanding the patterns and characteristics of fixations, researchers gain insights into how individuals allocate their visual attention and engage with stimuli. This knowledge proves particularly valuable in fields such as neuromarketing, where assessing consumer responses relies on a detailed understanding of visual attention dynamics.

Employing eye-tracking techniques, previous researches underscore the critical role of packaging design, investigating the influence of specific attributes like color, shape, and labeling on consumer perceptions of the product [26].

Strategic positioning of packaging design components is central to practical marketing efforts [27]. Inadequate placement may cause crucial design elements to go unnoticed, impacting product evaluation [15, 16]. An important study reveals a consumer preference for high-power brands when the brand logo is positioned on the upper side of the packaging, contrasting with diminished appeal when placed on the lower side [6, 7]. The effectiveness of capturing participants' attention by placing packaging content at the top is emphasized by Rebollar et al. [8]. Building on existing research, Piqueras-Fiszman et al. [9] explored the impact of packaging shape and images on consumer attention, with a focus on the logo. Their findings showed that squared-shaped packaging significantly heightened attention toward the logo.

Additionally, the study demonstrated the substantial influence of incorporating images on capturing consumer attention. This highlights the complex balance required in packaging design to ensure that attention is not only captured but also sustained, emphasizing the need for strategic placement and thoughtful integration of visual elements to prevent essential components from being marginalized.

2.2 Brand logo detection

Logo detection, a subfield of object detection, has witnessed substantial advancements over the years. In its initial stages, logo detection heavily relied on manually crafted visual attributes, including the Scale-Invariant Feature Transform (SIFT) and the Histogram of Oriented Gradients (HOG), combined with traditional classification models like Support Vector Machines (SVM) [28–30]. However, these approaches faced notable constraints. They were time-consuming because of their region-selective search method using sliding windows. They also struggled to handle different types of logos and were not very efficient at adapting to new situations [17]. In recent years, deep learning has emerged as the prevailing paradigm for logo detection. These approaches can be categorized into different strategies, including Region-based Convolutional Neural Network (R-CNN) models and YOLO-based models. R-CNN models [31], Fast R-CNN [32], and Faster R-CNN [33] have made noteworthy contributions to the field of logo detection. Hoi et al. [34] introduced the Deep Logo-DRCN scheme, which investigated various techniques within the field of deep region-based convolutional networks (DRCN) for improved logo detection. Similarly, Oliveira et al. [35] proposed an automatic graphic logo detection system based on Fast R-CNN, known for its robustness under unconstrained imaging conditions. Their approach involved utilizing transfer learning and data augmentation to train a CNN model, enabling multiple detection of potential regions containing objects. Additionally, Li et al. [36] developed Faster R-CNN for logo detection, incorporating transfer learning, data augmentation, and clustering to optimize hyper-parameters and anchor precision in the Region Proposal Network (RPN), resulting in a significant improvement in detection accuracy.

Feature Pyramid Networks (FPN) are crucial in addressing the multi-scale problem in object detection [37]. FPN notably enhances small object detection without escalating computational demands. Recent works have employed FPN to improve logo detection. Meng et al. [38] proposed OSF-Logo, incorporating the Regulated Deformable Convolution (RDC) module in a specific layer of FPN. This integration allows adaptive adjustments of convolution kernel positions, facilitating geometric adaptations to logos. In addition, Jin et al. [39] developed Brand Net, utilizing FPN to extract multi-scale features for logo recognition. To enhance small object detection in the context of logo recognition, FPN have also been integrated into Detection Transformers (DETR) [40]. Velazquez et al. [41] integrated FPN into DETR, enhancing small object detection. Nevertheless, this approach results in an increased computational load during backward propagation. More recently, Hou et al. [42] proposed the Multi-Scale Feature Decoupling Network (MFDNet) to distinguish between multiple logo categories. MFDNet incorporates a Balanced Feature Pyramid (BFP) for merging multi-scale features and a Feature Offset Module (FOM) with an anchor region proposal network for the optimal selection of logo features.

Driven primarily by the compelling demand for speed and real-time object detection applications, You Only Look Once (YOLO) was developed [43]. YOLO models, known as single-stage detectors, have played a central role in revolutionizing object detection for their ability to achieve both accuracy and speed. Early versions of YOLO, such as YOLOv2 [44] and YOLOv4 [45], set new benchmarks in the field. More recent iterations, including YOLOv7 [46] and YOLOv8 [19], represent the current SOTA in object detection. YOLO models are widely employed, particularly in the domain of logo detection. Palecek et al. [47] presented Scaled YOLOv4, outperforming traditional two-stage models such as Faster R-CNN in both speed and accuracy. It achieved a relative improvement of up to 46%, running up to twice as fast. Notably, logo detectors utilizing YOLOv7 and YOLOv8 remain unexplored, presenting an opportunity for potential improvements in balancing accuracy and speed, potentially reaching the SOTA in logo detection.

2.3 Saliency map prediction

Saliency prediction in computer vision involves the identification and anticipation of the most significant or salient regions within an image or video frame, likely to capture human attention. This process holds practical utility in various applications. CNNs are commonly used for saliency prediction tasks. Kroner et al. [48] introduced an encoder-decoder framework that incorporates several convolutional layers, each set at various dilation rates, to effectively grasp features on multiple scales. Jia et al. [49] used deep CNN models to extract more useful visual features for saliency prediction. TempSal [50] enables sequential saliency map generation through a temporal information-based model, astutely exploiting human temporal attention patterns. The incorporation of transfer learning principles amplifies the potential of CNN

models in the domain of saliency prediction [51, 52]. The fusion of RNN with CNN represents a hybrid approach in the field of both image and video saliency prediction, as introduced by Droste et al. [53].

Researchers have been inspired by the achievements of attention in natural language processing (NLP) and have started applying these models to computer vision tasks such as saliency prediction. Cao et al. [54] proposed a saliency prediction method named VGG-SSM. Their pipeline consists of three parts: feature extraction, multi-level integration, and a self-attention module. They demonstrated that refining global information from deep layers through a self-attention mechanism, in coordination with fine details in distant portions of a feature map, yields a comprehensive data enhancement process. Additionally, Lou et al. [55] developed a transformer-based method with both DenseNet and ResNet backbones.

The works mentioned earlier were created for general use, while numerous other works have been suggested specifically for advertising purposes. L  v  que et al. [56] collected an eye-tracking database of video advertising and evaluated their analysis with SOTA deep learning-based saliency models. Liang et al. [57] compiled an eye-tracking dataset comprising 1000 advertising images. Subsequently, they introduced a method that incorporates text features within advertising images, which considers the interaction between text region and pictorial region. Kou et al. [58] proposed confidence scores fusion for saliency prediction in advertising images, which is helpful to improve the robustness and performance. Another study, conducted by Jiang et al. [59], introduces the concept of salient Swin-Transformers. In this work, the researchers initially curated a dataset of e-commerce images for saliency prediction tasks. Subsequently, they proposed a novel multi-task learning framework that demonstrated SOTA performance in e-commerce scenarios.

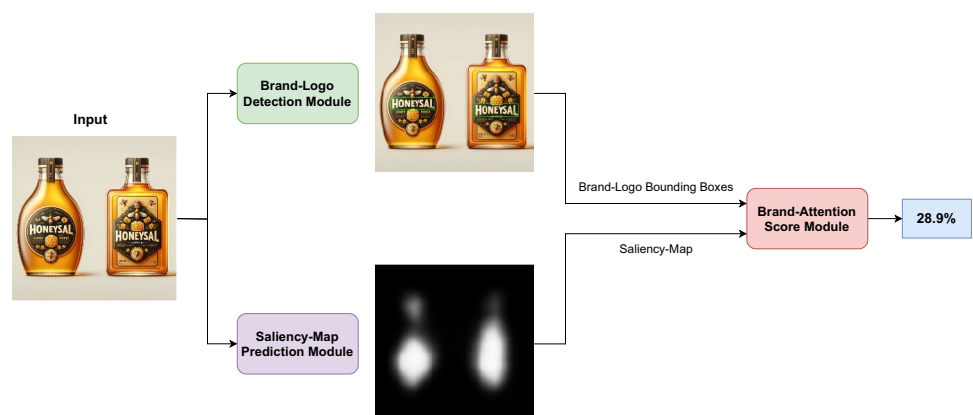
3 Proposed method

The primary aim of our research is to design a system for a comprehensive evaluation of the visual prominence of brand logos within the context of packaging or advertising images. To achieve this objective, the proposed methodology encompasses three closely related main steps as illustrated in Fig. 1. The first module is brand logo detection and is supported by the SOTA object detection model, YOLOv8. This module identifies and locates brand logos within the imagery, forming the foundational basis for subsequent analysis. Then, the second module focuses on generating saliency maps, a critical aspect of our approach. The saliency maps illuminate the regions within the image that command the highest degree of visual attention, providing valuable insights into viewer perception and cognition. The final module consolidates the outcomes of the brand logo detection and saliency map generation modules. This approach gives a score that measures how much attention the brand logo gets in the packaging or advertising image. This combination of techniques offers valuable insights for businesses aiming to optimize the visual prominence of their brand logos in marketing materials.

3.1 Brand logo detection

In the initial stage of the proposed method, our focus lies on brand logo detection. For this task, we employ the YOLOv8 model, specifically trained for logo detection purposes. When presented with an input image I with spatial dimensions $H \times W$ and C color channels, our Logo YOLOv8 model processes this image. The output of this model consists of a 1D

Fig. 1 Overview of the proposed brand-attention method



list of bounding boxes, denoted as B , where each bounding box (b) is represented as a tuple containing the coordinates $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$

$$B = \text{LOGO_YOLOv8}(I) = [b_1, b_2, \dots, b_n] \quad (1)$$

The number of logo boxes detected in the image is represented by n . This detection is the first fundamental step in our brand attention system.

3.2 Saliency map prediction

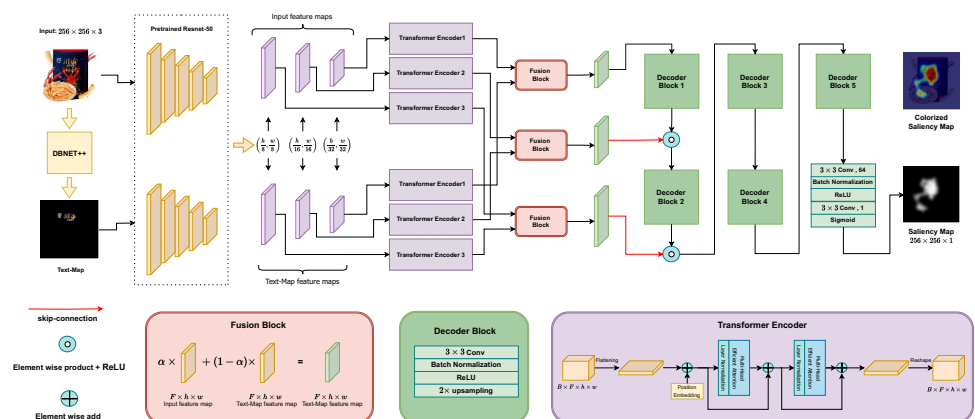
Our primary objective in the second stage is to generate saliency maps for images, with a specific focus on advertising and packaging designs. We introduce a novel saliency map prediction model tailored to address the unique requirements of both advertisements and packaging images (Fig. 2). This model is inspired by the TranSalNet network [55], with major improvements made to boost its efficiency and performance. One component of the proposed method involves incorporating the influence of text into the saliency map. Previous studies have shown that text is just as important as other visual elements in packaging and advertising. These studies found that text is instrumental in capturing people's attention, and they used eye-tracking data to confirm this [57, 59].

In the proposed model, we initiate the process by detecting text within the image. To achieve this, the text detection model proposed by Lia et al. [60] is employed, which outputs a text map. The text map and the original image are subsequently processed through a CNN decoder, resulting in multiple feature maps. To efficiently capture and process information from feature maps, we apply transformation through Transformer encoders. This enables the model to consider complex relationships and dependencies within visual content. To ensure a seamless integration of these elements, we introduce a pivotal component of the model: the *Fusion Block*. This block is strategically designed to merge the feature maps derived from both the text map and the original image. By doing so, it enables the simultaneous utilization of visual and text-map features, thereby enhancing the overall interpretative capabilities of the proposed model. After the fusion block, we use a CNN decoder, which is supported by skip connections coming from the encoder section. This integrated process ensures the restoration of long-range context-enhanced feature maps obtained from the fusion block. These enhanced feature maps serve as the foundation for constructing the final saliency map, capturing the regions of the image that attract the most visual attention. Figure 2 illustrates the proposed saliency model, providing a visual representation of its architecture and the various components that comprise our refined saliency map prediction system. As depicted, the model comprises five principal components, each of which will be explained in more detail in subsequent sections of this paper.

3.2.1 Text detector

We employ the cutting-edge DBNet++ network [60], which has emerged as a front-runner in the domain of text detection, consistently achieving SOTA accuracy across a spectrum of five scene text detection benchmarks. These benchmarks cover a diverse range of challenges, from handling horizontal and multi-oriented text to curved text, demonstrating the versatility and performance of DBNet++. The DBNet++ operates on images with spatial dimensions of $H \times W$ and C

Fig. 2 The block diagram of the proposed saliency model



channels, allowing it to accurately identify text regions within these images. By deploying this innovative network, we can precisely extract and isolate text from non-textual information, ultimately generating text maps. Given an input image $I \in \mathbb{R}^{H \times W \times C}$, the DBNet++ detects text regions denoted as R . As a consequence, a text map, denoted as $t_{\text{map}} \in \mathbb{R}^{H \times W \times C}$, is generated as follows:

$$t_{\text{map}}(x, y, c) = \begin{cases} I(x, y, c) & \text{if } (x, y, c) \in R \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

3.2.2 CNN encoder

A CNN encoder is designed as our feature extractor. The primary objective of this CNN encoder is to extract essential features from both the image and the text-map while ensuring that the spatial information is distinctly preserved. To achieve this, three sets of convolutional layers are used, each designed to capture features at different spatial scales. Specifically, we extract feature maps with spatial dimensions of $(w/8, h/8)$, $(w/16, h/16)$, and $(w/32, h/32)$. For the image and text-map image feature extraction, the ResNet-50 architecture is used [61]. This backbone is efficient, using fewer parameters than deeper architectures. It balances depth and performance, effectively extracting detailed features for saliency prediction [55].

3.2.3 Transformer encoder

After the initial CNN Encoder stage, which focuses on enhancing long-range and contextual information within our data, we designed three distinct transformer encoders to efficiently capture and process this enriched information. In the proposed pipeline, transformer encoders are integrated to handle the unique characteristics of both original images and text-maps. Specifically, three sets of multi-scale feature maps, denoted as i_1 , i_2 , and i_3 , are derived from the image data. These sets have spatial dimensions of $(w/32, h/32)$, $(w/16, h/16)$, and $(w/8, h/8)$, respectively. Each set is then fed into its respective transformer encoder. To adapt the input size of the transformer encoder and reduce computational complexity, we employ 1×1 convolution layers (conv1×1) with a stride of one. These convolution layers are applied to the input tensors, including i_1 , i_2 , and i_3 , to decrease their channel dimensions while preserving spatial dimensions. The conv1×1 operation specifically reduces the dimensions of i_1 , i_2 , and i_3 from 2048, 1024, and 512 to 768, 768, and 512, respectively. This dimension reduction streamlines the data for subsequent processing within the transformer encoder, aligning it with the required input dimensions and optimizing computational efficiency. Likewise, the textual components of the data, denoted as t_1 , t_2 , and t_3 , undergo dimension reduction through conv1×1 layers employing the same filter size and stride. This process ensures their alignment with the reduced dimensions of the visual components.

To facilitate position awareness and optimize the transformer encoders for effective processing of spatial information within these feature maps, we integrate position embeddings (PE) [62] into the input before feeding it into the transformer encoders. Each transformer encoder in the proposed model consists of two identical layers featuring Multi-Head Efficient Attention (MEA) [63] and multi-layer perceptron (MLP) blocks. Notably, the model's design deviates from Transalnet regarding the number of heads and layers in each transformer encoder. Specifically, transformer encoders employ one efficient attention head and a 2-layer MLP. These tailored configurations are designed to meet the specific requirements of our model, ensuring the efficient processing of the enriched feature maps. Additionally, the MLP block in each transformer encoder consists of two layers with a GELU activation function. Layer normalization (LNORM) and residual connections are applied before and after each block, ensuring stable and effective feature processing.

The introduced methodology distinguishes itself through the adoption of efficient attention, as proposed by Shen et al. [63], diverging from the conventional self-attention mechanism. Traditional self-attention is mathematically represented as

$$s(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

In this formula, Q , K , and V are the query, key, and value vectors, while d_k is the embedding dimension. However, this approach is limited by its $O(N^2)$ computational complexity, which presents major challenges when processing high-resolution images.

Efficient attention, on the other hand, optimizes this process by normalizing the keys and queries before their interaction. Represented as,

$$E(Q, K, V) = \rho_q(Q)(\rho_k(K)^T V) \quad (4)$$

where ρ_q and ρ_k are normalization functions. This approach addresses the redundancy in the context matrix generation of standard self-attention. It reduces the computational complexity to $O(d^2n)$, with a memory complexity of $O(dn + d^2)$, assuming $d_v = d$ and $d_k = d/2$. Here, d represents the embedding dimension. This model's efficient attention mechanism prioritizes a comprehensive understanding of the input feature, avoiding the computation of pairwise similarities. By treating keys as attention maps k_j^T and focusing on semantic information rather than positional similarities, it achieves a significant computational efficiency improvement without sacrificing representation richness. The diagram depicting the efficient attention mechanism discussed above is presented in Fig. 3 [63].

It can be summarized that for a given sample input m consisting of t_1 to t_3 (representing textual content) and i_1 to i_3 (representing image-based features), the transformer encoder process can be mathematically described as follows:

$$z_0 = \text{conv}_{1 \times 1}(m) \oplus \text{PE} \quad (5)$$

$$z'_l = \text{MEA}(\text{LNorm}(z_{l-1}) \oplus z_{l-1}) \quad (6)$$

$$z_l = \text{MLP}(\text{LNorm}(z'_l) \oplus z'_l) \quad (7)$$

where z_l represents the output feature maps of the l -th layer in the transformer encoder. The feature maps that go through transformer encoders 1, 2, and 3 are contextually enhanced and are referred to as i_j^* for $j = 1$ to 3 for image and t_j^* for $j = 1$ to 3 for text map image.

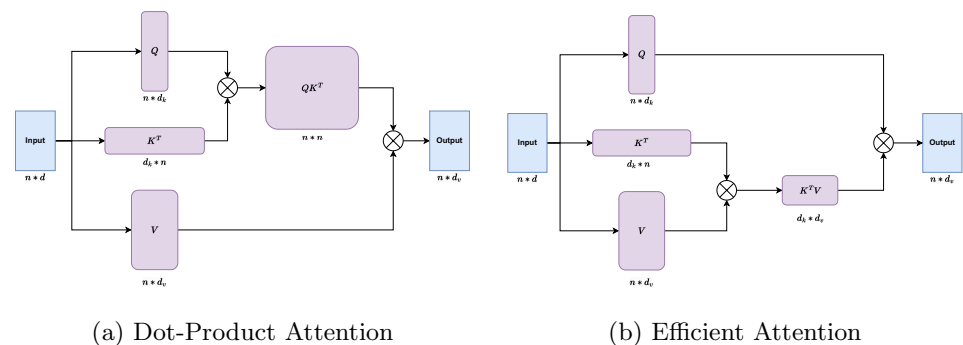
3.2.4 Fusion block

After generating enhanced visual features for the image and text map image, it is imperative to merge these features effectively. The proposed fusion process involves assigning weights to the visual and textual modalities. We introduce weighting factors, denoted as α , which determine the influence of visual and textual data, respectively.

$$I_{f_j}^* = \sigma(\alpha) \cdot i_j^* + (1 - \sigma(\alpha)) \cdot t_j^* \quad (8)$$

In this equation, $I_{f_j}^*$ represents the final feature representation after the fusion process. The selection of the α parameter is of paramount importance since it governs the equilibrium between the visual and textual modalities. In the proposed model, we treat α as a learnable parameter, enabling the model to determine the optimal value for this factor. This dynamic approach allows the model to adapt and effectively combine visual and textual information based on the unique demands of the task at hand, thereby enhancing the overall performance and versatility of the model. To ensure that α remains within the valid range $[0, 1]$ after optimization, a sigmoid function is applied. The sigmoid function, denoted as $\sigma(\cdot)$, maps real-valued inputs to the interval $[0, 1]$, making it an ideal choice for constraining the α parameter.

Fig. 3 Architecture of dot-product and efficient attention [63]



3.2.5 CNN decoder

The CNN decoder plays a key role in integrating and restoring long-range context-enhanced feature maps obtained from the fusion block. Its primary objective is to reconstruct the saliency maps while restoring the original image resolution. The proposed CNN decoder is designed to facilitate efficient and effective pixel-level classification, enabling the prediction of saliency maps. Within the network, several key operations are performed to enhance the model's performance. After each 3×3 convolution operation ($\text{Conv}_{3 \times 3}$), the batch normalization (BNorm) is applied to promote convergence. Besides, the activation function ReLU is used in all blocks, with Sigmoid employed in the final block. After initial down-sampling of the input image to a 32-scale by the encoder network, a pivotal process in the CNN Decoder involves a 2-scale up-sampling. This method uses nearest-neighbor interpolation and happens in the first five decoding stages. It creates the saliency map that has the same size as the original input image.

To improve the feature map's long-range and multi-scale context during the decoding process, the up-sampled feature map is fused with the output from the fusion blocks, denoted as $i_{f_j}^*$ for $j = 1$ to 3. This fusion is acquired through the corresponding skip-connection, using an element-wise product operation, and ensures that the model benefits from comprehensive contextual information at different scales. The operations within each CNN decoder block can be represented as follows.

$$O_i = \begin{cases} i_{f_1}^*, & i = 1 \\ \text{ReLU}\left(2X_{\text{Upsample}}(O_{i-1}) \cdot i_{f_i}^*\right), & i = 2, 3 \\ 2X_{\text{Upsample}}(O_{i-1}), & i = 4, 5, 6 \end{cases} \quad (9)$$

$$O_i^* = \text{ReLU}\left(\text{BNorm}(\text{Conv}_{3 \times 3}(O_i))\right), \quad \text{for } i : 1 \text{ to } 6 \quad (10)$$

$$S = \text{Sigmoid}(\text{Conv}_{3 \times 3}(O_6^*)) \quad (11)$$

where O_i refers to the output of the i -th decoding stage before the convolution operation, O_i^* refers to the output after applying the convolution, batch normalization, and ReLU operations, and S represents the final saliency map predicted by the proposed model.

3.2.6 Loss function and evaluation metrics

Drawing inspiration from established conventions in the domain of saliency map prediction models and referencing other saliency prediction frameworks [53, 55, 64], our model employs a composite loss function. This function combines three metrics: Kullback-Leibler divergence (KL), Linear Correlation Coefficient (CC) and Mean Squared Error (MSE) loss.

Let g^s represent the ground truth of the saliency map, g^f denote the ground truth of the saliency fixation map, and S denote the network's predicted saliency map. The overarching loss function is defined as:

$$\text{Loss} = \lambda_1 \cdot \text{KL}(g^s, S) + \lambda_2 \cdot \text{CC}(g^s, S) + \lambda_3 \cdot \text{MSELoss}(g^s, S) \quad (12)$$

where each component is elucidated as follows:

- *KL divergence*: A standard measure of dissimilarity between probability distributions, is expressed as:

$$\text{KL}(g^s, S) = \sum_{i=1}^n g_i^s \log \left(\epsilon + \frac{g_i^s}{S_i + \epsilon} \right) \quad (13)$$

Here, ϵ serves as a regularization constant, set to 2.2×10^{-16} as used in previous studies [55].

- *CC*: CC is defined as the ratio of the covariance between g^s and S to the product of their standard deviations, signifying similarity. The formula is presented as:

$$CC(g^s, S) = \frac{\text{cov}(g^s, S)}{\sigma(g^s) \cdot \sigma(S)} \quad (14)$$

Here, $\sigma(\cdot)$ designates the standard deviation, and $\text{cov}(\cdot)$ stands for the covariance.

The objective of this loss function is to minimize the KL and MSELoss while concurrently maximizing the value of CC. This dynamic balance is achieved through the fine-tuning of the coefficients λ_i , where i ranges from 1 to 3. By employing the Optuna framework [65], we have systematically determined the values for these coefficients to achieve optimal training. Based on the achieved experiments, these coefficients have been chosen to optimize the model's performance, with a specific focus on reducing KL while concurrently enhancing CC, aligning closely with the intended outcome of the proposed model.

In our comprehensive evaluation framework, we use three additional metrics-Similarity (SIM), Normalized Scan-path Saliency (NSS) and Area under ROC Curve(AUC)-to provide an assessment of the model's performance. While these metrics are not directly embedded within the training loss function, they play an important role in the evaluation phase.

- *SIM*: SIM gauges the linear relationship between the elements of g^s and S , where the minimum value at each position is summed to calculate the coefficient:

$$\text{SIM}(g^s, S) = \sum_{i=1}^n \min(g_i^s, S_i) \quad (15)$$

- *NSS*: NSS measures the similarity between the predicted S and g^f by comparing the fixations with the saliency map values:

$$\text{NSS}(g^f, S) = \frac{1}{\sum_i (g_i^f)} \sum_i \left(\frac{S_i - \mu(S)}{\sigma(S)} \right) g_i^f \quad (16)$$

where, $\sigma(\cdot)$ designates the standard deviation, and $\mu(\cdot)$, $\text{cov}(\cdot)$ stands for the mean and covariance, respectively.

3.3 Brand-attention score

After localizing brand bounding boxes (B) and generating the saliency maps for both packaging and advertising images, we can quantitatively assess the prominence of the brand within an image. The intuition involves converting the saliency map image into a list of pixel probabilities, ensuring that the cumulative probability sums to 1. Subsequently, we calculate the sum of probabilities associated with pixels contained within the image region.

Algorithm 1 Brand-attention score calculation

```

Data: B, S
Result: Brand-Attention Score
S[S < Threshold] = 0
SNorm = S / sum(S)
if B is None then
  | return 0
end
else
  Brand-Attention Score = 0
  for b in B do
     $x_{min}, y_{min}, x_{max}, y_{max} = b$ 
    for y in range( $y_{min}, y_{max} + 1$ ) do
      for x in range( $x_{min}, x_{max} + 1$ ) do
        | Brand-Attention Score += SNorm[x, y]
      end
    end
  end
  return Brand-Attention Score
end

```

The pseudo-code for calculating the brand attention score is presented in Algorithm 1. This pseudo-code outlines the procedure for computing the brand attention score based on the provided saliency map and bounding boxes. It involves removing saliency map values below a threshold, normalizing the remaining values to probabilities, and then calculating the score by summing the normalized values within the specified bounding box regions. Using the saliency map and this algorithm, we can obtain an attention score for every object or text (not only the brand logo) for which bounding boxes are provided or selected by users.

4 Experiments and results

In this section, we go through the datasets, training setup, and result analysis for both logo detection and saliency prediction. Moreover, the outcomes underscore the enhanced efficacy of the proposed technique compared to leading-edge methods across diverse evaluation metrics. The following part introduces the brand attention module and the proposed dataset. The brand attention module is then validated based on earlier hypotheses, with results thoroughly analyzed using feedback from human participants. The section concludes by proposing and discussing new hypotheses regarding brand visibility in packaging. The computational tasks described in this section were executed using the PyTorch framework on a workstation equipped with an Intel Core i-9 CPU and an NVIDIA GeForce RTX3090 GPU.

4.1 Logo detection

4.1.1 Datasets

Recent advances in computer vision have led to the development of specialized datasets tailored to logo detection. In particular, the growing demand for robust logo recognition in packaging applications has motivated our selection of datasets that capture the variations in packaging design and branding [17]. Our research focuses on two logo detection datasets: FoodLogoDet-1500 [42] and LogoDet-3 K [66], selected for their unique attributes that make them well-suited for the complexities of logo detection in product packaging. While these well-known datasets might initially seem limited in terms of packaging and brand design diversity, a closer examination reveals significant variation. As shown in Fig. 4, LogoDet-3 K

Table 1 Summary of selected logo detection datasets

Dataset	#Images	#Objects	#Logos
FoodLogoDet-1500	99,768	145,400	1500
LogoDet-3K	158,652	194,261	3000

Fig. 4 Sample images from FoodLogoDet-1500, LogoDet-3 K, and SalECI datasets



(a) FoodLogoDet-1500 (b) LogoDet-3K



(c) SalECI

comprises nine super categories with numerous sub-categories, while FoodLogoDet-1500 includes 63 sub-categories covering a wide range of packaging types. A summary of the selected datasets is provided in Table 1, and Fig. 4 presents sample images from these datasets.

4.1.2 Training setup

The dataset used for logo detection contains numerous classes, which are not essential for our specific case. Therefore, all classes have been aggregated into one for logo detection. Due to the inability of the proposed model to converge on large-scale datasets, a two-stage fine-tuning process has been implemented. In the first stage, the small version of the YOLOv8 model is fine-tuned, initially pre-trained on the COCO dataset, over the FoodLogoDet-1500 dataset. This initial fine-tuning serves as a crucial step to help the model adapt to the characteristics of the data and mitigate convergence issues. The fine-tuning process in this stage is carried out using the Adam optimizer across 100 epochs, with a

Table 2 Metrics on models fine-tuned over Foodlogo-det-1500

Method	mAP_{50}	mAP_{50-95}	Precision	Recall
Faster RCNN [33]	0.821	0.595	0.778	0.753
DETR [39]	0.849	0.640	0.806	0.781
MFDNet [42]	0.879	0.635	0.836	0.811
YOLOv7 [46]	0.932	0.698	0.90	0.866
YOLOv8 [19]	0.936	0.704	0.904	0.879

Values in bold denote the best-performing result for each metric

Table 3 Metrics on models pretrained on FoodLogo and fine-tuned over FoodLogoDet-1500+LogoDet3k dataset

Method	mAP_{50}	mAP_{50-95}	Precision	Recall
Faster RCNN [33]	0.80	0.57	0.76	0.74
DETR [39]	0.85	0.63	0.80	0.78
MFDNet [42]	0.87	0.62	0.82	0.80
YOLOv7 [46]	0.88	0.61	0.84	0.81
YOLOv8 [19]	0.94	0.71	0.91	0.88

Values in bold denote the best-performing result for each metric

batch size set to 32, and involves specifying a learning rate of 10^{-2} and a momentum of 0.9. During the second stage, we continued the fine-tuning process on both the FoodLogoDet-1500 and the larger LogoDet-3k datasets. This approach ensures that the model further adapts to a broader range of data patterns. The second-stage fine-tuning is conducted for 50 epochs with a batch size of 64, using the same hyperparameters as in the first stage. This two-stage fine-tuning strategy has proven effective in addressing the model convergence challenge. The entire process takes approximately 60 hours to complete.

4.1.3 Method comparison

We compared the proposed logo detection model with several SOTA methods. In addition to YOLOv7 [46] and MFDNet [42], we also evaluated Faster RCNN [33] and DETR [39] to provide a comprehensive performance comparison. Results are shown in Table 2 and Table 3. As can be observed, YOLOv8 significantly outperforms the other methods across various metrics, such as mAP_{50} , mAP_{50-95} , precision, and recall, in both stages of evaluation.

4.2 Saliency map prediction

4.2.1 Dataset

In the domain of saliency map prediction tasks, various general-purpose datasets, including SALICON [67], CAT2000 [68], MIT1003 [69], and MIT300 [70] have been established. However, this paper uniquely centers its focus on commercial and advertisement images. To address this specific focus, we leverage the Saliency E-commerce Images (SalECI) dataset introduced by Jiang et al. [59]. The SalECI dataset comprises 257,302 fixations obtained through eye-tracking experiments involving 25 subjects. The dataset comprises 972 e-commerce images, each paired with corresponding fixation maps and text boundaries. This dataset acts as an important tool for exploring saliency within the realm of commercial and advertising stimuli.

Table 4 Comparing the saliency prediction accuracy for the proposed and nine other SOTA methods over SalECI. #Param indicates the number of parameters in the model

Method	#Param	CC \uparrow	KL \downarrow	AUC \uparrow	NSS \uparrow	SIM \uparrow
Contextual Encoder-Decoder (CEC) [48]	20M	0.459 \pm 0.136	1.1346 \pm 0.23	0.76 \pm 0.066	0.925 \pm 0.268	0.373 \pm 0.06
DeepGazellE [52]	104M	0.561 \pm 0.124	0.995 \pm 0.215	0.842 \pm 0.055	1.327 \pm 0.318	0.399 \pm 0.065
UNISAL [53]	4M	0.6 \pm 0.15	0.768 \pm 0.262	0.845 \pm 0.056	1.574 \pm 0.522	0.514 \pm 0.094
EML-Net [49]	47M	0.510 \pm 0.16	1.227 \pm 0.903	0.807 \pm 0.062	1.232 \pm 0.407	0.536 \pm 0.103
VGGSSAM [54]	42M	0.691 \pm 0.126	0.682 \pm 0.259	0.815 \pm 0.048	1.324 \pm 0.362	0.58 \pm 0.091
Transalnet [55]	72M	0.717 \pm 0.061	0.873 \pm 0.079	0.824 \pm 0.054	1.723 \pm 0.203	0.534 \pm 0.043
VGGSSM [54]	43M	0.728 \pm 0.121	0.599 \pm 0.237	0.829 \pm 0.043	1.396 \pm 0.359	0.611 \pm 0.089
Temp-SAL [50]	242M	0.719 \pm 0.065	0.712 \pm 0.126	0.813 \pm 0.077	1.768 \pm 0.182	0.629 \pm 0.048
SSwin transformer [59]	–	0.687 \pm 0.175	0.652 \pm 0.478	0.868 \pm 0.072	1.701 \pm 0.497	0.606 \pm 0.101
Ours	66M	0.75\pm0.050	0.578\pm0.117	0.892\pm0.033	1.89\pm0.204	0.645\pm0.040

Values in bold denote the best-performing result for each metric

Fig. 5 Comparison of the saliency maps of different models over SALECI

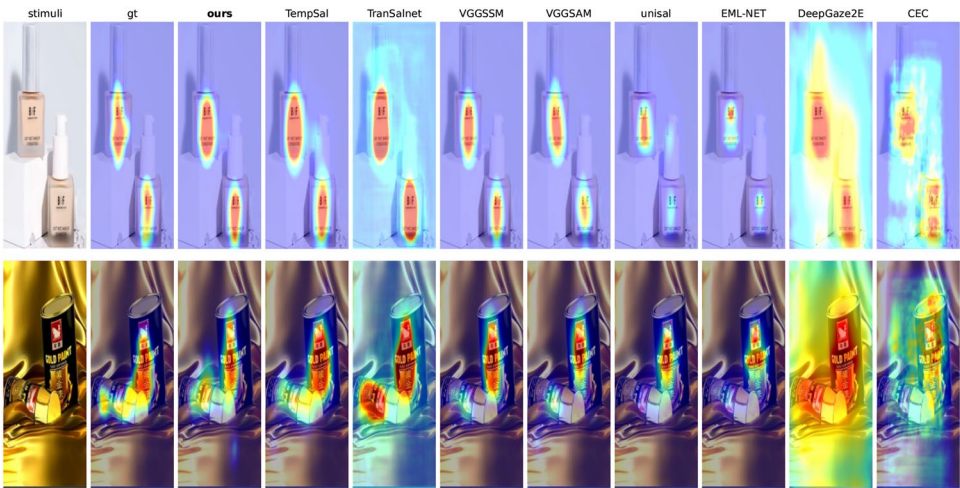


Table 5 Efficiency and complexity comparison across various saliency prediction SOTA methods

Method	FLOPs (G)	Model-params (M)	Model-size (MB)	Memory-usage (MB)	Inference-time (MS)
Temp-SAL [50]	35.81	116.21	443.30	679.67	24.04
DeepGazellE [52]	43.775	104.05	396.91	7167.92	536.50
Transalnet [55]	25.47	71.98	274.59	377.34	14.368
Ours	24.30	66.21	252.56	333.63	13.036

Bold values represent the best result

4.2.2 Training setup

The proposed method was trained over the SaleCI dataset [59] using a step learning rate scheduler with a step size of 4 and a gamma value of 0.1. The initial learning rate was set to 5×10^{-4} , and weight decay was applied at a rate of 10^{-4} . The Adam optimizer is used for training. Additionally, the optimal values for the loss function weighting coefficients, λ_i , were determined to be $\lambda_1 = 10$, $\lambda_2 = -3$, $\lambda_3 = 5$. The initial starting value for α was set at 0.5, and it dynamically adjusts to 0.659 during the training process.

4.2.3 Method comparison

Comparisons with SOTA methods: We compare our approach with ten SOTA saliency prediction models using the SaleCI dataset (see Table 4). Our method outperforms the competing techniques across all evaluation metrics, including CC, KL, NSS, and SIM, while maintaining a competitive number of parameters. Notably, our model achieves superior performance compared to methods such as Transalnet [55], SSwin Transformer [59], and Temp-SAL [50], thereby establishing a new SOTA for commercial saliency prediction.

Qualitative results: Figure 5 illustrates the visual quality of the predicted saliency maps. Our model’s predictions are notably closer to the ground truth when compared to other leading methods like Temp-SAL [50] and SSwin Transformer [59], further validating the quantitative improvements demonstrated in Table 4.

Efficiency analysis: Our proposed model achieves higher efficiency than existing SOTA methods by integrating optimized attention mechanisms with fewer heads and layers, reducing computational complexity while maintaining superior performance. Despite incorporating a text detector and fusion block, our model maintains a lower parameter count than Transalnet and significantly reduces FLOPs, memory usage, and model size. As shown in Table 5, our model achieves the lowest FLOPs, smallest model size, reduced memory usage, and fastest inference time, making it highly efficient and well-suited for real-world applications. These results emphasize the effectiveness of our architectural improvements in enhancing saliency prediction for commercial images.

Table 6 Impact of the pre-trained feature extractor

Training strategy	CC ↑	KL ↓	NSS ↑	SIM ↑
Without pre-trained weights	0.648	0.772	1.551	0.549
Pre-trained (frozen)	0.738	0.5814	1.840	0.627
Pre-trained (fine-tuned)	0.750	0.578	1.890	0.645

Values in bold denote the best-performing result for each metric

Table 7 Effect of using text feature map

Method	CC ↑	KL ↓	NSS ↑	SIM ↑
Ours (no text feature map)	0.721	0.696	1.860	0.624
Ours	0.750	0.578	1.890	0.645

Values in bold denote the best-performing result for each metric

Table 8 Effect of α in the fusion block

α Value	CC ↑	KL ↓	NSS ↑	SIM ↑
0.5	0.713	0.636	1.760	0.608
0.6	0.720	0.620	1.810	0.617
0.65	0.726	0.612	1.809	0.614
Learnable	0.750	0.578	1.890	0.645

Values in bold denote the best-performing result for each metric

Table 9 Effect of different loss term combinations

KL	CC	MSE	CC ↑	KL ↓	NSS ↑	SIM ↑
✓			0.709	0.620	1.736	0.598
	✓		0.720	1.358	1.845	0.601
✓	✓		0.725	0.632	1.810	0.628
✓		✓	0.725	0.614	1.795	0.617
	✓	✓	0.710	0.726	1.770	0.570
✓	✓	✓	0.750	0.578	1.890	0.645

Values in bold denote the best-performing result for each metric

4.2.4 Ablation study

Effect of pre-trained feature extractor: We conducted an ablation study to examine the effect of using a pre-trained feature extractor. Three training strategies were compared: training from scratch (without pre-trained weights), using pre-trained weights with frozen parameters, and fine-tuning a pre-trained feature extractor during training. As shown in Table 6, the model performs best when fine-tuning is applied, while freezing still offers improvements over training from scratch.

Effect of using text feature map: We carried out an experiment comparing the model without text features against our full model with text feature maps. Table 7 shows that incorporating the text feature map improves performance across all metrics.

Effect of trainable fusion weight (α): The fusion weight α balances visual and text features. Fixed α values may not capture the varying importance of these modalities. Allowing α to be learnable lets the model adaptively adjust this balance, leading to improved performance. Table 8 shows that the learnable α outperforms fixed values.

Effect of different terms in loss function: We evaluate various loss combinations using CC, KL, and MSE to study their impact on model performance. Our experiments show that the KL term is essential for saliency prediction, CC further enhances performance, and including MSE improves generalization. Table 9 summarizes the results.

4.3 Brand attention

In this section, we evaluate the effectiveness of the proposed brand attention module by comparing it with the observations in psychophysical studies. To test the model, we have designed a dataset where each group of images is the same in every way, apart from one particular logo feature it is examining. Wrapping up this section, we introduce some new hypotheses in this field that have not been explored yet.

4.3.1 Dataset

While aiming to validate various hypotheses concerning logo placement and packaging design, we have created a dataset comprising 650 images. This collection is a systematically designed platform for testing various ideas connected to packaging design and how people perceive brands. To ensure a rich and varied base for our study, 95% of the images in this dataset are sourced from the Internet templates, complemented by those generated by DALL-E, an advanced AI image generation tool. Each image has been carefully modified to align with specific research questions, with alterations ranging from subtle logo repositioning to more substantial design transformations. Our dataset is organized into 12 hypotheses, each examining different aspects of design and branding. For each hypothesis, we analyze 18 ± 3 images per hypothesis. In each hypothesis image set, all logo characteristics are fixed, except the one under experiment. This setup provides us with an in-depth insight into the influence of packaging design and logo placement on brand perception.

4.3.2 Previous hypotheses analysis

We evaluate the effectiveness of our brand attention model by comparing its output to data from human observers who have studied logo attention. This comparison involves aligning the model's predictions with findings from psychophysical studies, ensuring its accuracy in predicting how humans notice logos for real-world applications. The following subsequent items provide a summary of the studies that form the basis of this comparative analysis, showcasing their relevance in the context of brand logo attention:

1. *Study 1:* Piqueras-Fizman et al. [9] examined the influence of packaging shape and the presence of an image on attention to different elements, like the logo. It was found that a squared shape, as opposed to a rounded one, drew more attention to the logo. They have also demonstrated that the photo element on product packaging was highly influential in drawing consumer interest rather than text. Additional studies have also suggested that geometric and pictorial cues can guide visual attention, although the extent of these effects may vary depending on context [1, 2]. Backing these ideas, our initial results, as displayed in Table 10, suggest that packaging with a squared shape indeed garnered more attention to the logo compared to rounded shapes. Notably, the obtained results generally align with existing research, indicating that while packaging shape and imagery seem to influence consumer attention, these effects should be interpreted cautiously given potential contextual variations [71]. Figure 6 showcases a sample of images created for testing this study.
2. *Study 2:* The findings from studies proposed by Dong et al. [7] and Riaz et al. [6] underscore a noteworthy connection between logo placement and consumer purchase intention. The research indicates that high-power brands tend to

Fig. 6 Sample images illustrating the influence of packaging shape (left) and the presence of an image on directing attention to different elements, such as the logo (right)





Horizontal-Vertical Packaging Orientation		Text vs Image	
Square	Round	Text	Image
			
Score: 28.51	Score: 27.33	Score: 27.9	Score: 27.2

Fig. 7 Testing images to demonstrate how logo position impacts brand attention. Top-to-bottom logo positioning (top) and all-around logo positioning (bottom)

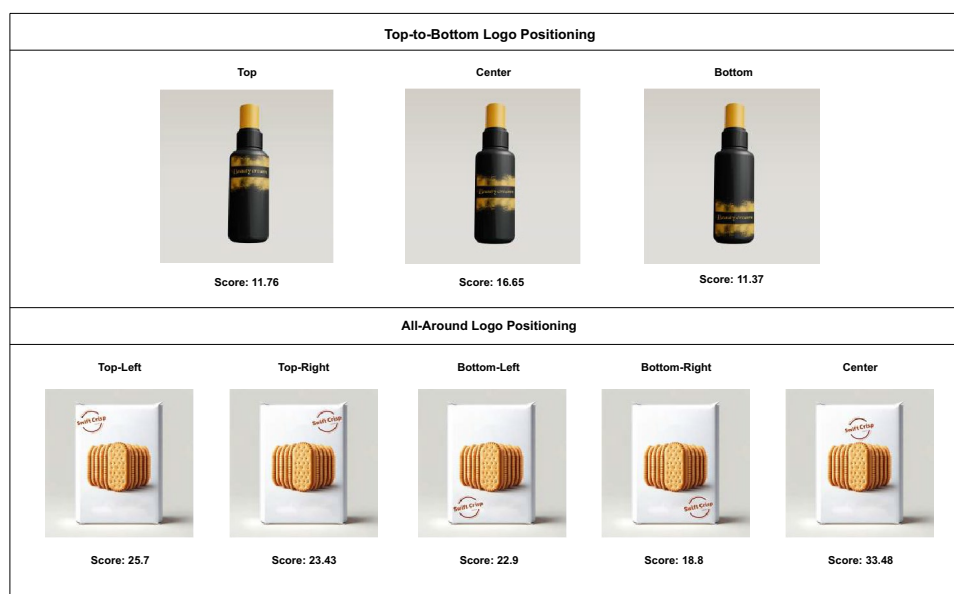


Table 10 Comparing the impact of top-to-bottom logo positioning, text vs. image, and square-round packaging orientation hypotheses on brand attention score

Hypothesis	Position	Mean	SE
Top-to-bottom logo positioning	Down	28.89	5.19
	UP	34.05	5.64
Text vs image	Image	31.71	4.61
	Text	37.23	4.62
Square-round packaging orientation	Round	25.82	4.86
	Square	27.02	4.01

Values in bold denote the best-performing condition

benefit from having their logos placed at the top of packaging, while low-power brands may be favored when logos are positioned lower. This effect appears to be influenced by cultural reading patterns and the natural hierarchy of visual cues [71]. The study explores strategic logo placement using the concept of power metaphors, suggesting that top-of-packaging placement may enhance perceived brand power. It should be noted that these relationships are context-dependent and that variations in experimental conditions warrant a cautious interpretation of the results in practical settings. The results of our proposed model, as detailed in Table 10, generally support these observations, reinforcing their potential relevance for marketing and brand strategy. Figure 7 illustrates visual examples developed for this study.

4.3.3 Proposed hypotheses

Similarly, as outlined in the preceding section, the proposed brand attention method serves as a robust foundation for exploring brand marketing and visual analytics. Beyond the ongoing studies, we have introduced several new hypotheses that investigate aspects not extensively covered in the existing literature. These hypotheses represent unexplored territories as we strive for a comprehensive understanding of brand perception and consumer behavior. This exploration guides future psychophysical tasks, providing a framework for new investigations in the field.

- *Positioning of brand logos:* Previous studies have mostly examined logo placement at either the top or bottom of packaging [6, 7]. However, a research gap exists regarding the effects of central and off-center placements (e.g., upper-left, upper-right, bottom-left, and bottom-right) on brand attention. Moreover, while many studies have documented that higher placements generally capture more attention due to reading patterns and visual hierarchy [72], few have systematically explored a broader range of placements in a controlled experimental setting. In our experiments, each

placement condition was tested to ensure statistical robustness. The proposed model predicts that positioning the brand logo at the center of the packaging significantly enhances brand attention compared to other positions, as outlined in Table 11. Furthermore, our findings suggest that upper placements tend to attract more attention than lower placements, with the upper-left corner outperforming the upper-right corner-potentially due to the left-to-right scanning habit in Western reading cultures [72]. Similarly, among the lower positions, the bottom-left appears more effective than the bottom-right.

- **Bold distinction in packaging:** Many packaging designs incorporate bold text or objects, yet the impact of these elements on brand logo attention has been under-explored. To investigate this, we conducted experiments in which we selectively bolded or emphasized non-logo textual elements and graphical objects on the packaging, while keeping the logo constant. The proposed model predicts that when non-logo elements are visually enhanced (e.g., through bolding), they can act as competing focal points, potentially reducing the relative visual attention directed toward the brand logo. This finding aligns with theories of selective visual attention, which suggest that salient distractors can divert attention from a primary target [73, 74]. Moreover, research in design studies shows that the interplay of contrasting visual elements-such as bold versus regular typography-affects the overall perceptual hierarchy and may compromise brand identity consistency if not balanced properly [71, 75]. Table 11 details the experimental outcomes, demonstrating a measurable reduction in the brand attention score when non-logo elements are emphasized.
- **Presence of person in packaging:** It is well known that human faces instinctively capture visual attention [76]. However, their influence on brand attention within packaging contexts has received limited investigation. In our experiments, including a person or face in the packaging led to a measurable decrease in attention toward the brand logo, as shown in Table 11. This result aligns with previous observations that faces, due to their strong attentional pull, can divert gaze from other visual elements [77].
- **Multi packaging:** The influence of presenting multiple packages of a brand in a single image has received limited study, particularly regarding its impact on brand logo visibility and attention. In our experiments, each condition was tested using images, with the multi-packaging condition displaying between 2 to 4 packages of the same brand, compared to images with a single package. The proposed model predicts that images containing multiple packages are more effective at absorbing brand attention than single-package images, likely due to the increased opportunity for the brand logo to be detected in varied spatial configurations [1, 71]. These findings suggest that repetition may enhance visual attention; however, this effect should be interpreted cautiously, as factors such as shelf arrangement, ambient lighting, and overall brand identity may moderate the outcome [78].
- **Multi objects in packaging:** The effect of featuring multiple objects in packaging design (e.g., presenting a single orange versus 2 to 4 oranges) on brand logo attention remains underexplored. The proposed model predicts that packaging designs with multiple objects divert attention from the brand logo, as shown in Table 11. Our experimental results, comparing images with a single object against those with 2 to 4 objects, indicate that additional objects increase visual clutter and diminish the logo's prominence [73, 74]. The achieved results imply that simpler packaging featuring only one object is more effective in maintaining higher brand logo attention.
- **Horizontal-vertical packaging orientation:** The proposed model predicts that horizontally oriented packaging enhances brand logo attention more effectively than vertically oriented packaging. This may be because a horizontal layout offers a broader, more balanced visual field that can emphasize the logo's scale and prominence [75, 79]. Experimental results in Table 11 indicate a clear preference for horizontal packaging designs. These findings are consistent with established visual processing principles and design theories.
- **Horizontal-vertical brand logo orientation:** We examine the influence of logo orientation on attention while keeping other packaging elements constant. The proposed model indicates that vertical logos capture more attention, possibly due to the additional processing required for vertical text [80]. This outcome is consistent with research suggesting that deviations from canonical orientations enhance visual salience by engaging additional attentional mechanisms [81]. Table 11 presents the corresponding increase in brand attention for vertical logos.
- **Brand logo color:** The influence of logo color on the capture of consumer attention has been a topic of interest in research on marketing and color psychology [10, 82]. However, comprehensive studies comparing a wide range of colors in the context of brand logos remain limited. In our study, we examined eight colors—red, blue, green, yellow, orange, purple, black, and white—to assess their impact on brand attention. Our experimental results (see Table 12) indicate that, under controlled conditions, logos rendered in red tend to attract higher attention scores. Nevertheless, this outcome should be interpreted with caution; while red's associations with alertness and prominence [82] may enhance salience, design literature emphasizes that brand colors are chosen for long-term identity consistency and contextual relevance [75, 83].

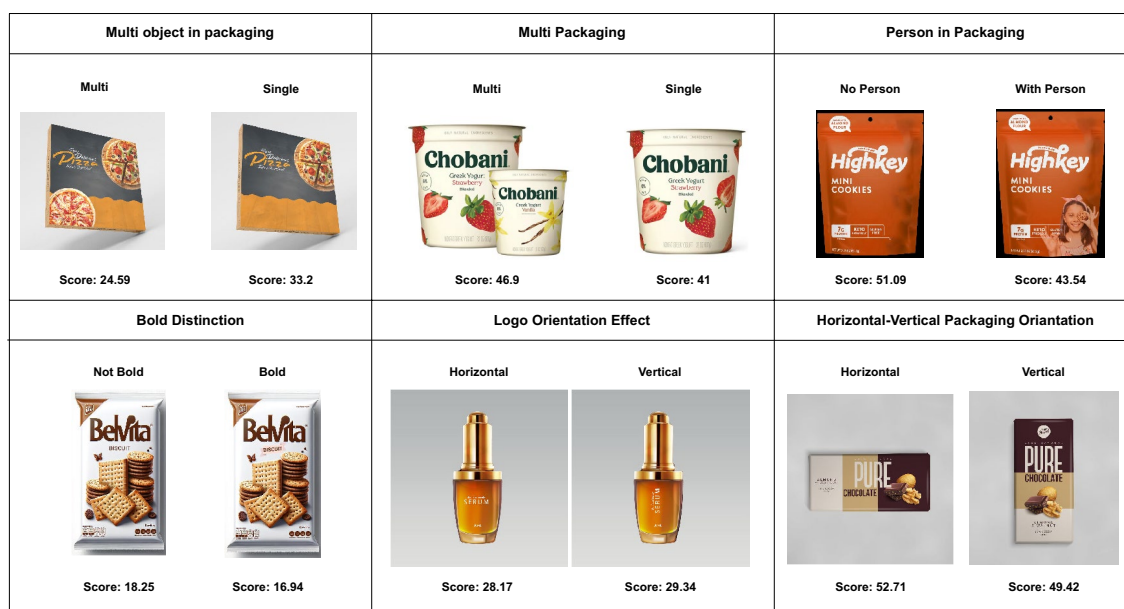


Fig. 8 Sample images for assessing the proposed hypotheses: multi objects (top left), multiple packaging (top center), presence of a person (top right), bold distinction (bottom left), horizontal-vertical brand logo orientation (bottom center), and horizontal-vertical packaging orientation (bottom right)

Table 11 Comparing the impact of top-to-bottom logo positioning, all-around logo positioning, bold distinction, horizontal-vertical brand logo orientation, horizontal-vertical packaging orientation, presence of person, multi-object and multi packaging on brand attention score

Hypothesis	Position	Mean	SE
Top-to-bottom logo positioning	Down	28.89	5.19
	UP	34.05	5.64
All-around logo positioning	Center	40.02	7.06
	Down-Right	15.05	3.05
	Down-Left	18.8	3.41
	UP-Right	16.51	3.05
	UP-Left	20.24	3.34
	Center	24.92	4.12
Bold distinction	Boldness	19.98	2.27
	Not Bold	21.1	2.35
Horizontal-vertical brand logo orientation	Horizontal	29.91	4.05
	Vertical	34.54	4.8
Horizontal-vertical packaging orientation	Vertical	27.92	4.92
	Horizontal	36.92	5.59
Person in packaging	With Person	32.26	5.76
	No Person	36	6.16
Multi object in packaging	Multi	32.5	5
	One	40.95	5.29
Multi packaging	Single	31.64	4.16
	Multi	39.52	4.73

Values in bold denote the best-performing condition

- **Packaging color:** Packaging color has a substantial impact on brand visual attention, yet there exists limited research comparing various colors systematically. In our experiments, we examined how different packaging colors affect the visibility of the brand logo. The proposed model predicts that packaging color significantly influences brand attention, with less intense, warmer, and simpler colors enhancing logo visibility. To ensure a robust evaluation, the packaging was modified exactly once for each color condition-red, blue, green, yellow, orange, purple, black, and white-thereby

Fig. 9 Visualization of the brand-logo color and packaging color influence on brand attention



isolating the impact of each color on consumer perception. The achieved results, as shown in Table 12, indicate that white, due to its neutral nature, allows the brand logo to stand out more effectively. This finding aligns with previous studies suggesting that neutral backgrounds can enhance logo salience by providing high contrast [10].

Figures 7, 8 and 9 present samples from the brand attention dataset, serving as empirical evidence for the evaluation of our proposed hypotheses.

5 Conclusion and discussion

The importance of logos within packaging emerges as an influential visual cue, profoundly shaping consumer perception and promoting brand recognition. This paper introduces a module specifically designed to model human attention to brand logos in packaging. The module comprises three main components: fine-tuned YOLOv8 logo detection, a novel CNN-Transformer-based saliency map prediction model that outperforms existing methods in predicting visual attention,

Table 12 Comparing the impact of packaging color and brand logo color on brand attention score

Hypothesis	Position	Mean	SE
Packaging color	Black	36.82	5.65
	Brown	37.85	5.62
	Orange	37.46	5.46
	Yellow	37.45	5.61
	Green	36.38	5.55
	Blue	37.51	5.66
	Red	38.23	5.73
	White	40.84	5.89
	Brand logo color	White	4.33
		Brown	4.4
		Orange	4.63
		Yellow	4.66
		Green	4.93
		Blue	4.87
		Black	4.7
		Red	4.79

Values in bold denote the best-performing condition

and a derived brand attention score. To validate our approach, we compared its predictions against established psycho-physical studies, demonstrating that the proposed method aligns well with known trends in brand attention. Our study contributes to bridging a research gap by verifying established hypotheses while introducing seven new ones—such as the impact of multi-packaging, multiple objects, and color variations on brand attention. These contributions advance the literature by offering a quantifiable measure of brand salience that integrates both traditional design theories and computational methods. By utilizing the capabilities of this module, it becomes possible to simulate human visual attention to brand logos under controlled conditions, thereby opening new opportunities for testing unexplored hypotheses in branding. For example, our model suggests that positioning the brand logo at the center or upper left of the packaging increases its visibility, while it predicts that a red logo and white packaging can enhance the brand attention score under the tested conditions. While the practical utility of the proposed module is highlighted for designers in advertising and packaging, we acknowledge that design practice relies heavily on inductive approaches and visual intuition, as emphasized in semiotic frameworks [75]. Thus, our tool is best viewed as a complementary aid that provides data-driven insights rather than a prescriptive solution, allowing designers to refine their intuitions with empirical evidence. Moreover, our experimental dataset primarily consists of controlled and synthetic images, which ensures systematic evaluation but may not fully capture the complexities of real-world packaging—such as variations in shape, angle, and minor damage. Future work could focus on incorporating more diverse, real-world datasets and employing eye-tracking experiments to further validate and cross-validate the brand attention score. Such efforts will enhance the robustness and external validity of the model.

Author contributions Conceptualization: A.H.; Methodology: A.H, K.H.; Formal analysis and investigation: A.H, K.H.; Software: A.H.; Data curation: Re.T.; Validation: A.H, K.H, R.T.; Visualization: A.H, K.H.; Writing - original draft preparation: A.H, K.H, P.O.; Writing - review and editing: A.H, K.H, R.T, Z.E, M.A.A.; Supervision: M.A.A.

Funding No funding was received to assist with the preparation of this manuscript.

Data availability The research data supporting the results of this manuscript are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate This research did not involve human participants, and hence, informed consent is not applicable.

Competing interests The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Bloch P. Seeking the ideal form: product design and consumer response. *J Market.* 1995;59:16–29. <https://doi.org/10.2307/1252116>.
2. Ampuero O, Vila N. Consumer perception of product packaging. *J Consum Market.* 2006;23:100–12. <https://doi.org/10.1108/07363760610655032>.
3. Méndez J, Oubiña J, Rubio N. The relative importance of brand-packaging, price and taste in affecting brand preferences. *Br Food J.* 2011;113:1229–51. <https://doi.org/10.1108/00070701111177665>.
4. Stewart B. Packaging as an effective marketing tool. Pira Packaging Guide Series. Pira International, Surrey, UK 1995. <https://books.google.com/books?id=1Rro1I2aGxIC>.
5. Shukla P, Singh J, Wang W. The influence of creative packaging design on customer motivation to process and purchase decisions. *J Bus Res.* 2022;147:338–47. <https://doi.org/10.1016/j.jbusres.2022.04.026>.
6. Riaz T, Ghafoor M. Strategic logo placement on packaging—using conceptual metaphors of power in packaging—evidence from pakistan. *Procedia Comput Sci.* 2019;158:582–9. <https://doi.org/10.1016/j.procs.2019.09.092>.
7. Dong R, Gleim M. High or low: the impact of brand logo location on consumers product perceptions. *Food Qual Prefer.* 2018. <https://doi.org/10.1016/j.foodqual.2018.05.003>.
8. Rebollar R, Lidón I, Martín Vallejo F, Puebla M. The identification of viewing patterns of chocolate snack packages using eye-tracking techniques. *Food Qual Prefer.* 2015;39:251–8. <https://doi.org/10.1016/j.foodqual.2014.08.002>.
9. Piqueras-Fiszman B, Velasco C, Salgado-Montejo A, Spence C. Using combined eye tracking and word association in order to assess novel packaging solutions: a case study involving jam jars. *Food Qual Prefer.* 2013;28(1):328–38. <https://doi.org/10.1016/j.foodqual.2012.10.006>.
10. Raheem AR, Vishnu P, Ahmed AM. Impact of product packaging on consumer's buying behavior. *Eur J Sci Res.* 2014;122(2):125–34.
11. Clement J. Visual influence on in-store buying decisions: an eye-track experiment on the visual influence of packaging design. *J Market Manag.* 2007;23(9–10):917–28.
12. Shimizu Y, Uleman JS. Attention allocation is a possible mediator of cultural variations in spontaneous trait and situation inferences: eye-tracking evidence. *J Exp Soc Psychol.* 2021;94(104115):10–1016.
13. Riswanto AL, Kim S, Williady A, Ha Y, Kim H-S. How visual design in dairy packaging affects consumer attention and decision-making. *Dairy.* 2025;6(1):4.
14. Girard T, Anitsal MM, Anitsal I. The role of logos in building brand awareness and performance: Implications for entrepreneurs. *Entrepreneurial Executive.* 2013;18:7.
15. Krishna A, Cian L, Aydinoglu N. Sensory aspects of package design. *J Retail.* 2017;93:43–54. <https://doi.org/10.1016/j.jretai.2016.12.002>.
16. Otterbring T, Shams P, Wästlund E, Gustafsson A. Left isn't always right: placement of pictorial and textual package elements. *Br Food J.* 2013. <https://doi.org/10.1108/BFJ-08-2011-0208>.
17. Hou S, Li J, Min W, Hou Q, Zhao Y, Zheng Y, Jiang S. Deep learning for logo detection: a survey. *ACM Trans Multimed Comput Commun Appl.* 2023;20(3):1–23.
18. Borji A, Itti L. State-of-the-art in visual attention modeling. *IEEE Trans Pattern Anal Mach Intell.* 2012. <https://doi.org/10.1109/TPAMI.2012.89>.
19. Jocher, G., Chaurasia, A., Qiu, J.: YOLO by ultralytics. <https://github.com/ultralytics/ultralytics>
20. Hubert M, Baecke S, Kenning P. What they see is what they get? an fmri-study on neural correlates of attractive packaging. *J Consum Behav.* 2008;7:342–59. <https://doi.org/10.1002/cb.256>.
21. Alvino L, Constantinides E, Lubbe RH. Consumer neuroscience: attentional preferences for wine labeling reflected in the posterior contralateral negativity. *Front Psychol.* 2021;12: 688713.
22. Maynard O, McClernon F, Oliver J, Munafò M. Using neuroscience to inform tobacco control policy. *NicotineTobacco Res.* 2018. <https://doi.org/10.1093/ntr/nty057>.
23. Gofman A, Moskowitz H, Fyrbjork J, Moskowitz D, Mets T. Extending rule developing experimentation to perception of food packages with eye tracking. *Open Food Sci J.* 2009;3:66–78. <https://doi.org/10.2174/1874256400903010066>.
24. Pertzov Y, Avidan G, Zohary E. Accumulation of visual information across multiple fixations. *J Vis.* 2009;9(10):2–2. <https://doi.org/10.1167/9.10.2>.
25. Nagel R, Reutskaja E, Camerer C, Rangel A. Search dynamics in consumer choice under time pressure: an eye-tracking study. *Am Econ Rev.* 2011;101:900–26. <https://doi.org/10.1257/aer.101.2.900>.
26. Ares G, Deliza R. Studying the influence of package shape and colour on consumer expectations of milk desserts using word association and conjoint analysis. *Food Qual Prefer.* 2010;21:930–7. <https://doi.org/10.1016/j.foodqual.2010.03.006>.
27. Rettie R, Brewer C. The verbal and visual components of package design. *J Prod Brand Manag.* 2000. <https://doi.org/10.1108/10610420010316339>.
28. Boia R, Florea C, Florea L. Elliptical asift agglomeration in class prototype for logo detection. In: *Proceedings of the British Machine Vision Conference*, 2015;115–111512

29. Sahbi H, Ballan L, Serra G, Bimbo A. Context-dependent logo matching and recognition. *IEEE Trans Image Process*. 2013;22(3):1018–31.
30. Revaud J, Douze M, Schmid C. Correlation-based burstiness for logo retrieval. In: *Proceedings of the 20th ACM International Conference on Multimedia*, 2012;965–968.
31. Girshick RB, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 2014;580–587.
32. Girshick RB. Fast r-cnn. In: *IEEE International Conference on Computer Vision*, 2015;1440–1448.
33. Ren S, He K, Girshick RB, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2015;39(6):1137–49.
34. Hoi SCH, Wu X, Liu H, Wu Y, Wang H, Xue H, Wu Q. Logo-net: Large-scale deep logo detection and brand recognition with deep region-based convolutional networks. *arXiv preprint [arXiv:1511.02462](https://arxiv.org/abs/1511.02462)* 2015.
35. Oliveira G, Frazao X, Pimentel A, Ribeiro B. Automatic graphic logo detection via fast region-based convolutional networks. In: *International Joint Conference on Neural Networks*, 2016;985–991.
36. Li Y, Shi Q, Deng J, Su F. Graphic logo detection with deep region-based convolutional networks. In: *IEEE Visual Communications and Image Processing*, 2017;1–4.
37. Lin TY, Dollár P, Girshick RB, He K, Hariharan B, Belongie SJ. Feature pyramid networks for object detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 2017;936–944.
38. Meng Y, Hou S, Wang J, Jia W, Zheng Y, Karim A. An adaptive representation algorithm for multi-scale logo detection. *Displays*. 2021;70:102090.
39. Jin X, Su W, Zhang R, He Y, Xue H. The open brands dataset: Unified brand detection and recognition at scale. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020;4387–4391.
40. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: *European Conference on Computer Vision*, 2020;213–229.
41. Velazquez DA, Gonfaus JM, Rodríguez P, Roca FX, Ozawa S, Gonzalez J. Logo detection with no priors. *IEEE Access*. 2021;9:106–998107011.
42. Hou Q, Min W, Wang J, Hou S, Zheng Y, Jiang S. Foodlogodet-1500: a dataset for large-scale food logo detection via multi-scale feature decoupling network. In: *Proceedings of the 29th ACM International Conference on Multimedia*, 2021;4670–4679.
43. Redmon J, Divvala SK, Girshick RB, Farhadi A. You only look once: unified, real-time object detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016;779–788.
44. Redmon J, Farhadi A. Yolo9000: better, faster, stronger. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 2017;6517–6525.
45. Bochkovskiy A, Wang CY, Liao H. Yolov4: optimal speed and accuracy of object detection. *arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934)* 2020.
46. Wang C-Y, Bochkovskiy A, Liao H-YM. Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023*.
47. Paleček K, Chaloupka J. Logo detection and identification in system for audio-visual broadcast transcription. In: *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, 2021;357–360.
48. Kroner A, Senden M, Driessens K, Goebel R. Contextual encoder-decoder network for visual saliency prediction. *Neural Netw*. 2020;129:261–70.
49. Jia S, Bruce ND. Eml-net: sn expandable multi-layer network for saliency prediction. *Image Vis Comput*. 2020;95:103887.
50. Aydemir B, Hoffstetter L, Zhang T, Salzmann M, Süsstrunk S. Tempsal-uncovering temporal information for deep saliency prediction. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023;6461–6470.
51. Kümmerer M, Wallis TS, Bethge M. Deepgaze ii: Reading fixations from deep features trained on object recognition. *arXiv preprint [arXiv:1610.01563](https://arxiv.org/abs/1610.01563)* 2016.
52. Linardos A, Kümmerer M, Press O, Bethge M. Deepgaze iie: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021;12919–12928.
53. Droste R, Jiao J, Noble JA. Unified image and video saliency modeling. In: *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 2020*;16, 419–435. Springer
54. Cao G, Tang Q, Jo K-h. Aggregated deep saliency prediction by self-attention network. In: *Intelligent Computing Methodologies: 16th International Conference, ICIC 2020, Bari, Italy, October 2–5, 2020, Proceedings, Part III 2020*;16, 87–97. Springer
55. Lou J, et al. Transalnet: towards perceptually relevant visual saliency prediction. *Neurocomputing*. 2022;494:455–67.
56. Lévêque L, Liu H. An eye-tracking database of video advertising. In: *2019 IEEE International Conference on Image Processing (ICIP)*, 2019;425–429. <https://doi.org/10.1109/ICIP.2019.8802989>
57. Liang S, Liu R, Qian J. Fixation prediction for advertising images: dataset and benchmark. *J Vis Commun Image Represent*. 2021;81:103356.
58. Kou Q, Liu R, Lv C, Jiang H, Cheng D. Advertising image saliency prediction method based on score level fusion. *IEEE Access*. 2023;11:8455–66. <https://doi.org/10.1109/ACCESS.2023.3236807>.
59. Jiang L, Li Y, Li S, Xu M, Lei S, Guo Y, Huang B. Does text attract attention on e-commerce images: A novel saliency prediction dataset and method. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022;2088–2097.
60. Liao M, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion. *IEEE Trans Pattern Anal Mach Intell*. 2022;45(1):919–31.
61. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016;770–778.
62. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: *2021 International Conference on Learning Representations (ICLR) 2021*.
63. Shen Z, et al. Efficient attention: attention with linear complexities. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision 2021*.
64. Che Z, Borji A, Zhai G, Min X, Guo G, Callet PL. Why is gaze influenced by image transformations? dataset and model. *IEEE Trans Image Process*. 2020;29:2287–300.
65. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2019*.

66. Wang J, Min W, Hou S, Ma S, Zheng Y, Jiang S. Logodet3k: A large-scale image dataset for logo detection. *ACM Trans Multimed Comput Commun Appl.* 2022;18(1):1–19.
67. Jiang M, Huang S, Duan J, Zhao Q. Salicon: Saliency in context. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015;1072–1080.
68. Borji A, Itti L. Cat2000: a large scale fixation dataset for boosting saliency research. 2015 arXiv preprint [arXiv:1505.03581](https://arxiv.org/abs/1505.03581).
69. Judd T, Ehinger K, Durand F, Torralba A. Learning to predict where humans look. In: *2009 IEEE 12th International Conference on Computer Vision*, 2009;2106–2113. IEEE.
70. Judd T, Durand F, Torralba A. A benchmark of computational models of saliency to predict human fixations. 2012.
71. Underwood RL. The communicative power of product packaging: creating brand identity via lived and mediated experience. *J Market Theory Pract.* 2003;11(1):62–76.
72. Lautenbacher OP. From still pictures to moving pictures. *Eye-tracking in Audiovisual Translation*, 2012;135–155.
73. Wolfe JM. Guided search 2.0 a revised model of visual search. *Psychonom Bull Rev.* 1994;1: 202–238
74. Gelade G. A feature-integration theory of attention. *Visual perception: Essential Readings*, 2001;347.
75. Kress GR, Leeuwen T. *Reading images: the grammar of visual design*. Routledge, London; New York 1996. <https://books.google.com/books?id=vh07i06q-9AC>
76. Cerf M, Harel J, Einhäuser W, Koch C. Predicting human gaze using low-level saliency combined with face detection. *Adv Neural Inf Process Syst.* 2007;20.
77. Vuilleumier P, Armony J, Clarke K, Husain M, Driver J, Dolan RJ. Neural response to emotional faces with and without awareness: event-related FMRI in a parietal patient with visual extinction and spatial neglect. *Neuropsychologia.* 2002;40(12):2156–66.
78. Ampuero O, Vila N. Consumer perceptions of product packaging. *J Consum Market.* 2006;23(2):100–12.
79. Ware C. *Information visualization: perception for design*. San Francisco, CA: Morgan Kaufmann; 2019.
80. Yu D, Park H, Gerold D, Legge GE. Comparing reading speed for horizontal and vertical English text. *J Vis.* 2010;10(2):21–21.
81. Itti L, Koch C. Computational modeling of visual attention. *Nat Rev Neurosci.* 2001;2(3):194–203.
82. Singh S. Impact of color on marketing. *Manag Decis.* 2006;44(6):783–9.
83. Wheeler A. *Designing brand identity: an essential guide for the whole branding team*. Hoboken, NJ: John Wiley & Sons; 2017.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.