Journal of Advertising

ISSN: 0091-3367 (Print) 1557-7805 (Online) Journal homepage: www.tandfonline.com/journals/ujoa20

Computer Vision Models for Image Analysis in Advertising Research

Hairong Li & Nan Zhang

To cite this article: Hairong Li & Nan Zhang (2024) Computer Vision Models for Image Analysis in Advertising Research, Journal of Advertising, 53:5, 771-790, DOI: 10.1080/00913367.2024.2407644

To link to this article: https://doi.org/10.1080/00913367.2024.2407644

JOURNAL OF ADVERTISING
2024, VOL. 53, NO. 5, 771–790
https://doi.org/10.1080/00913367.2024.2407644

Computer Vision Models for Image Analysis in Advertising Research

Hairong Li and Nan Zhang

Michigan State University, East Lansing, Michigan, USA

ABSTRACT
This study introduces computer vision models for image analysis in advertising research. It reviews the literature in social science and computer science and identifies three categories and nine types of image analysis. The study uses these categories and types as a framework to select 12 computer vision models and compare them on their capability, accuracy, availability, and usability. Nine models are single-functional models, and three are multi-functional models; all 12 have been used in advertising research. The study also demonstrates how two models are used to classify a sample of image ads and assess the aesthetic scores of these ads to answer the research question about the relationship between content categories and aesthetic scores in image ads. It outlines several key steps for the use of computer vision models in advertising research and proposes future research directions. The study can serve as a guide to advertising researchers.

Brand-related images of products and services are posted continuously by brands, users, and influencers on social media and have become increasingly crucial in brand communication. They can effectively convey brand messages and personality in a fast-paced digital environment with short attention spans among the audience. Some studies investigated the impact of these images on brand recalls, engagement, and affinity (Bisht 2019; Holiday et al. 2021; Jin, Kerr, and Suh 2019; Li and Xie 2020), and others explored the attributes of these images and the subtle influences of these attributes, such as colors, shapes and symbols, and objects and scenes (Argyris et al. 2020; Li and Xie 2020; Shin et al. 2020; Sutcliffe and Namoune 2008). Advertising researchers have recently started adopting computer vision models to make image analysis more effective and efficient. As computer vision models become increasingly sophisticated and more researchers are interested in using them for brand-related image analysis, this research trend will likely continue.

Computer vision is a discipline that intersects machine learning and image processing and provides sophisticated image recognition and classification capabilities. A computer vision model is designed to interpret and analyze visual content autonomously. In advertising research, the utility of such models is multifaceted. It enables the extraction of meaningful patterns and insights from images that are otherwise difficult for human analysts to perceive because of the sheer volume and complexity of visual data. This model allows advertising researchers to gauge consumer attention with techniques such as eye-tracking, assess the influence of social media advertising, and provide insights into effective content marketing strategies (Chen 2023; Maslowska, Ohme, and Segijn 2021). Furthermore, as emotional responses evoked by ad images play a pivotal role in influencing consumer behaviors, research using facial expression models can reveal how emotional reactions to brand visuals impact ad liking, engagement, and purchase intentions among different demographic groups (Otamendi and Sutil Martín 2020).

Computer vision models significantly outperform manual methods, particularly in efficiency and

CONTACT Nan Zhang zhangn24@msu.edu Information and Media Program, Michigan State University, 404 Wilson Road, Rm 309, East Lansing, MI 48824, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00913367.2024.2407644.

Hairong Li (PhD, Michigan State University) is Professor, Department of Advertising and Public Relations, Michigan State University.
Nan Zhang (MS, Ball State University) is Doctoral Candidate, Information and Media Program, Michigan State University.

scalability. Manual image analysis methods, such as content or semiotic analysis, typically involve human coders who annotate images based on predetermined criteria. These approaches can provide deep sociocultural insights (McQuarrie and Mick 2003) but are limited by their labor-intensive nature, which restricts the number of images that can be analyzed. By contrast, computer vision models can analyze large datasets more rapidly and consistently without fatigue, enabling researchers to explore vast image repositories for patterns and trends that manual analysis might miss (Argyris et al. 2020). For example, emotion recognition models can automate the identification of facial expressions, making it easier to assess how consumers react emotionally to advertisements (Otamendi and Sutil Martín 2020). However, it is important to acknowledge the limitations of these automated systems. Although models are excellent at recognizing patterns, they may struggle with cultural nuances, such as interpreting complex visual narratives that require a deep understanding of sociocultural context, in which manual analysis excels (McQuarrie and Mick 2003).

Although computer vision models are advantageous, appropriate use is challenging for many advertising researchers due to the varying levels of complexity of these models. To promote better use of computer vision models in advertising research, this study reviews the literature in advertising and computer science to identify the types and categories of image analysis; to compare computer vision models in terms of their capability, accuracy, availability, and usability; and to demonstrate examples of using such models in advertising research. It also outlines the steps for designing and implementing image analysis using suitable computer vision models. Thus, this study can guide all advertising researchers interested in brand-related image analysis.

Types of Image Analysis

Brand-related images can be complex, containing various objects such as logos, symbols, products, scenes, humans, faces, and animals. These objects convey meanings, express aesthetics, and project the emotions of the brands. Thus, image analysis decodes the meanings, aesthetics, and emotions of an image. Although image analysis can be done manually, computer vision models are increasingly used to conduct such analysis. These models enable computers to see and understand the content of digital images by breaking images into pixels using machine learning and neural networks (Araujo, Lock, and van de Velde 2020; Nanne et al. 2020; Paolanti et al. 2017; Tous et al. 2018). We conducted a literature review to understand the types of image analysis in advertising research and how computer vision models are used in such analysis.

We selected social science and computer science databases for the search, including Business Source Complete, Academic Search Complete, arXiv, ACM Digital Library, and IEEE Xplore. We used a combination of several keywords, including "advertise*," "image," "deep learning," "computer vision," and "machine learning" to cast a wide net to capture all relevant studies. Our initial search generated a list of 206 articles: 40 from Business Source Complete, 26 from Academic Search Complete, 33 from arXiv, 107 from ACM Digital Library, and 85 from IEEE Xplore. Following the initial search, we refined the list of articles by eliminating those that did not directly address the application of machine learning in brand-related images. For example, studies centered on ad recommendation systems and click-through rates were excluded to maintain a clear focus on image analysis. The final selection comprised 30 articles of direct relevance and served as a foundation for subsequent in-depth analysis and exploration, ensuring that our analysis was grounded in the most pertinent and high-quality literature.

We analyzed these studies in two steps to identify and confirm the types of image analysis. First, our review of these studies identified nine types of image analysis, and we classified these types into three broad categories. The visual conception category includes terms such as "visual conception," "image classification," "object detection," "content understanding," and "advertisement." These terms highlight the primary focus on identifying, classifying, and understanding the visual content in advertising images. The emotion recognition category comprises terms such as "emotion recognition," "sentiment analysis," "facial expression," and "themed sentiment." These terms indicate a focus on recognizing and interpreting the emotional cues in advertising images. The aesthetic appreciation category includes terms like "aesthetic appreciation," "image aesthetic," "style," and "image quality."

Second, we used natural language processing for co-word analysis to confirm the types of image analysis that emerged from our preliminary review of the studies. The abstracts of these articles were standardized using a template format, focusing on the main study, approach, and task, which facilitated the construction of a *co-word network graph*. This graph emphasizes these articles' connections and

Table 1. Types of image analysis and example studies.

Type of image analysis	Study Title
Visual Conception	
Image classification	The Effects of Visual Congruence on Increasing Consumers Brand Engagement
	The Power of Brand Selfies
	Advertisement Image Classification Using Deep Learning With BERT
	On Detection of Advertising Images
	Predicting Social Media Engagement with Computer Vision
	A Multimodal Analysis of Influencer Content on Twitter
	Advertisement Image Classification Using Convolutional Neural Network
	A Novel Machine Learning Technique for Fake Smart Watches Advertisement Detection
Object detection	Exploring Online Ad Images Using a Deep Convolutional Neural Network
	Approach Branding Luxury Hotels: Evidence from The Analysis of Consumers' "Big" Visual Data on TripAdvisor
	Deep Learning for Logo Detection: A Survey
	Deep Learning Approaches for Fashion Knowledge Extraction from Social Media:
	A Review the Use of Computer Vision to Analyze Brand-Related User Generated
	Image Content Machine Learning Techniques for Brand-Influencer Matchmaking on The Instagram Social Network
Content understanding	Automated Coding of Political Campaign Advertisement Videos
	Automatically Detecting Image-Text Mismatch on Instagram
	Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach
Emotion Recognition	
Facial expression detection	Emotional Responses to Tourism Advertisements
	Deep Learning Based Sentiment Analysis on Images
	Visual Sentiment Exploration of Customer Emotions Using Image Analytics
Sentiment analysis	Visual Listening In: Extracting Brand Image Portrayed on Social Media
	Visual And Textual Sentiment Analysis of Brand-Related Social Media Pictures Using Deep Convolutional Networks
	A Survey of Multimodal Sentiment Analysis
Themed sentiment analysis	Going Negative Online?—A Study of Negative Advertising on Social Media
	What You Feel, Is What You Like Influence of Message Appeals on Customer Engagement on Instagram
Aesthetic Appreciation	
Image quality	A Multi-Perspective Approach for Analyzing Long-Running Live Events on social media
Image aesthetic	Visual Representation for Capturing Creator Theme in Brand-Creator Marketplace
	Camera Eats First: Exploring Food Aesthetics Portrayed on Social Media
	Building Brand Authenticity on social media: The Impact of Instagram Ad Model Genuineness and Trustworthiness
Style and design	Style-Aware Image Recommendation for Social Media Marketing

relationships between key terms and reveals three prominent categories: *visual conception*, *emotion recognition*, and aesthetic *appreciation* (see Table 1). These terms emphasize the evaluation of artistic and technical elements that contribute to the visual appeal of advertising images. The co-word analysis and categorizing results validated the types of image analysis (see Supplemental Online Appendix 1 for more information about our co-word analysis). As we use these nine types of image analysis as a base for selecting and comparing computer vision models, these types are detailed in three categories in the following sections.

Visual Conception

Visual conception offers a systematic method for comprehending and examining images, including three interrelated levels of depth and complexity. *Image classification* is central to comprehending what the image represents at a primary level. *Object detection* adds the dimension of identifying and locating multiple specific objects in an image. *Content understanding* advances comprehension by interpreting the image's context, interactions, and narratives.

Image classification is a fundamental analysis that categorizes an entire image into a single class or label. The primary objective is identifying the image's main subject or dominant theme. This task does not focus on the subject's location or how it interacts with other elements. Instead, it emphasizes categorizing an image's overall theme or subject into a single category or class. *Image classification* is the technical term used to describe this process, and it can be likened to categorical data. Previous studies include the classification of images as ads or non-ads (Jain, Taneja, and Taneja 2024; Li et al. 2007; Villegas, Goanta, and Aletras 2023), the classification of the display of online ads as clear or unclear (Vo, Tran, and Le 2017), and the classification of the ads as genuine or fake (Zaheer et al. 2022). Image classification is also used to classify physical activities, Lululemon-style clothing, and pets to examine visual congruence in influencer marketing and brand engagement (Argyris et al. 2020), food images for social media engagement prediction (Philp, Jacobson, and Pancer 2022), and brand-related images and their match with Instagram influencers (Sweet, Rothwell, and Luo 2019).

Object detection is a more advanced analysis that identifies and locates objects within an image. It focuses on "what" and "where" questions about objects in the image, involving identifying multiple objects and their specific locations. Object detection was used for clothes detection, clothes parsing, and product retrieval, highlighting the identification of fashion items (Mameli et al. 2022). It was also deployed to detect and locate hotel room-related objects (Giglio et al. 2020), detect common objects in various image ads to predict ad click-through rates (Fire and Schler 2017), and detect multiple objects in brand-related images for comparison across computer vision models (Nanne et al. 2020).

Content understanding is the most advanced analysis of visual conception. It involves interpreting the context, relationships, and interactions between elements in an image. It goes beyond simply identifying "what" and "where" elements are located and instead seeks to understand "how" and "why" they interact or relate to each other. This level of analysis is focused on grasping the narrative or story that the image is trying to convey. Ha et al. (2021) and Shin et al. (2020) used this analysis to detect mismatches between images and texts, which recognized an image and correlated it with an associated text description. Tarr, Hwang, and Imai (2023) applied content understanding to analyzing political video ads. They broke the video into individual images to generate a comprehensive content summary while underscoring the intricate relationship between video and image analysis in understanding complex content.

Emotion Recognition

Emotion recognition in image analysis focuses on identifying and interpreting emotional cues within images. Unlike image classification, which categorizes images based on their visible content, emotion recognition models analyze emotional expressions and sentiments in images. These models mimic human visual perception to some extent, allowing computers to identify, classify, interpret, and interact with visual data, like how a human would perceive and understand visual scenes. In advertising research, emotion recognition models can adopt ad images to identify the emotion in the ad's content and use customers' images to recognize their reactions. Specifically, researchers expect to label the ad image's emotional attributes and find customer feedback by capturing and understanding their emotions. The research in this area varies from analyzing facial expressions to sensing the sentiments conveyed through the visual cues of an image.

Facial expression detection in computer vision models allows computers to recognize and interpret human facial expressions from images. The process involves face detection, facial landmark detection, feature extraction, and expression classification (Kumar, Kaur, and Kumar 2019). By analyzing facial cues of image ads, the detection analysis allows advertisers to gauge the potential emotional impact of the content. Researchers have used this technique to explore the association of smiles in ads with television viewer engagement (Hernandez et al. 2013); to examine consumers' emotional responses to tourism videos (Hadinejad et al. 2019); to measure the facial expression polarities of images with high accuracy (N et al. 2023); to examine how emotional variation in Super Bowl advertisements influences ad liking (Jones and Hamby 2023); and to identify the amount of emotion used and the impact on engagement by specific discrete emotions (Holiday et al. 2023).

Sentiment analysis simplifies the complex spectrum of human emotions by categorizing them into basic states: *positive*, *negative*, or *neutral*. This approach simplifies analyzing and interpreting the overall sentiment conveyed in image advertising content. For example, Soleymani et al. (2017) provided a comprehensive overview of the state of the art, challenges, and perspectives related to binary sentiment analysis. Paolanti et al. (2017) conducted sentiment analysis in brand-related social media images, and the analysis of brand-related images on Instagram and Flickr advanced sentiment analysis using visual cues and texts. Liu (2019) focused on sentiment analysis in political advertising, employing Google Cloud Vision for image recognition, and offered insights into the impact of negative advertising on social media.

Themed sentiment analysis is an advanced approach to sentiment analysis that goes beyond the basic categorization of sentiments to delve deeper into the specific themes or topics of visual content in combination with text. This technique provides a more detailed understanding of the sentiments of images. Instead of interpreting the image as "positive," "negative," or "neutral," the themed sentiment analysis can attribute the image to more categories or self-defined categories. For example, the customized categories can be "violence," "love," or "happiness." In previous research, Rietveld et al. (2020) utilized the themed sentiment analysis to determine the visual content sentiment with "violence" and "arousal" and then investigated how different sentiments of the ads affect customer engagement on Instagram. Liu, Dzyabura, and Mizik (2020) proposed a visual-textual sentiment analysis, which extracted the themed sentiments by integrating the image and text content in ads together so that they precisely understand the message conveyed in the ads. Srivastava (2021) explored customer emotions in marketing through themed sentiment analysis by focusing on specific emotional themes evoked in marketing materials and analyzing the associated sentiments. These studies indicate the value of themed sentiment analysis in providing in-depth knowledge of sentiments in specific contexts and enhancing the understanding of consumer behavior and preferences in advertising and marketing.

Aesthetic Appreciation

Aesthetic appreciation in image analysis encompasses image quality, aesthetics, and style and design. Unlike visual conception and emotion recognition, which focus on objective analysis, aesthetic appreciation analyzes the image content subjectively. For example, Li and Xie (2020) conducted manual image coding, focusing on facial expressions, image quality, and image-text mismatch, and their findings revealed that professionally shot, high-quality images consistently lead to higher user engagement.

Image quality refers to the technical aspects of an image, such as clarity, resolution, and color accuracy, which all contribute to the overall visual appeal of an image. High-quality images are essential for capturing attention and conveying a sense of professionalism and credibility. For example, Sabet, Brambilla, and Hosseini (2021) explored how image quality, particularly brightness and color, impacts user behavior during live events on Instagram. Hartmann et al. (2021) investigated the influence of image quality in brand selfies on Twitter and Instagram, revealing its significant impact on brand engagement. These studies suggest the importance of maintaining high image quality in digital advertising to enhance brand perception and user engagement.

Image aesthetic refers to an image's artistic and compositional elements, such as balance, harmony, and emotional appeal. These elements are crucial in creating visually appealing advertisements that resonate emotionally with viewers. Studies have highlighted the pivotal role of image aesthetics in several key areas: shaping user interaction and brand identity in digital advertising; as the visual appeals of food images on Yelp (Gambetti and Han 2022); aesthetic alignment between creators and brands for effective brand representation (Duanis et al. 2023); brand authenticity in visual presentation (Yang et al. 2021); and the technological advancement in leveraging image aesthetics (Shin et al. 2020).

Style and design in image analysis deal with an image's overall visual presentation and thematic elements, including specific design styles, color schemes, and visual themes that align with the brand's identity and message. Zhang and Yamasaki (2021) used style and design analysis by exploring the impact of image style on content discovery and brand posts on social media. Focusing on brand concept consistency, the study used object and style vectors to learn brand representation. The findings suggest that the visual style of an image significantly influences how content is perceived and engaged with by social media users and that influencers' recommendations should be tailored to specific brand concepts and styles.

In sum, these three categories and nine types of media analysis we identified from the review of the 30 studies provided a framework for comparing computer vision models used in these studies. The following section focuses on the capability, accuracy, availability, and usability of select computer vision models for advertising research.

Comparison of Computer Vision Models

In the process of identifying the types of image analysis in advertising, we found that 16 of the 30 studies utilized computer vision models that are available for academic use, with some being open-source and others provided through commercial application programming interfaces (APIs) with extensive documentation (e.g., Philp, Jacobson, and Pancer 2022; Rietveld et al. 2020), and the rest used self-developed models by the authors (e.g., Argyris et al. 2020; Vo, Tran, and Le 2017). In our comparison, we used the models accessible for academic research and excluded those self-developed models. Publicly accessible models, like those from open-source repositories, offer transparency and reproducibility, which are paramount in academic settings. In contrast, self-developed, proprietary models may offer advanced capabilities beyond public models or incorporate cutting-edge algorithms and proprietary technologies unavailable in the public domain. For example, proprietary models may leverage unique data sets or advanced machine learning techniques that significantly enhance performance in specific applications. They may also include features beyond the standard set, such as more sophisticated forms of natural language processing and real-time processing capabilities. Whereas these nonpublic models can offer superior performance and advanced features, they also have limitations, such as restricted access and lack of transparency. For academic researchers, the choice between public and proprietary models will depend on the specific needs of their research, the importance of open access and reproducibility, and the potential benefits of accessing advanced, proprietary technologies. For this review, however, no adequate information is available for these nonpublic models.

As a result, 10 existing publicly accessible models were selected from these studies. Seven of the selected models are single-functional models, and three are multifunctional models capable of more than one category of image analysis. To cover all three categories and nine types of image analysis with at least one single-functional model in our comparison, we added two more computer vision models: DeepFace and image quality assessment (IQA). These two models are highly relevant to advertising research; as the model's name indicates, the first model focuses on *recognition of facial emotion*, and the second focuses on *assessment of image quality*. DeepFace not only detects facial expressions but also identifies race, age, and gender, whereas the IQA model encompasses a broad range of contemporary quality assessment architectures. We verified these 12 selected models using reputable online resources known for high-quality, state-of-the-art models. This approach to selection ensures the reliability of our analysis in providing a holistic evaluation of computer vision models in advertising research.

Our comparison of these models evaluates their capability, accuracy, availability, and usability. *Capability* refers to a model's ability to perform specific tasks, such as categorizing and analyzing visual content. This feature is exemplified by the automated visual content analysis approach, which utilizes pretrained models for large-scale image classification, aiding in theory building and research (Araujo, Lock, and van de Velde 2020). Similarly, a multimodal visual communication system demonstrates a model's capacity to personalize video content based on user emotions, enhancing engagement and satisfaction (Fang and Gong 2023). *Accuracy* is another critical criterion, reflecting a model's precision in identifying and classifying visual elements. For instance, models designed to detect e-cigarette products on social media or recognize facial emotions are vital for targeted advertising and content moderation. High accuracy ensures that these models can reliably interpret and respond to visual data, which is crucial for the success of advertising campaigns (Vassey et al. 2024; Asresa et al. 2023). Furthermore, accurate models are essential for maintaining trust and validity in research findings (Borji et al. 2015). *Availability* and *usability* are particularly important in multidisciplinary studies because only reachable computational resources and data can help researchers of different disciplines. The availability of models and datasets affects the scope of their evaluation and application (Xu et al. 2022). They ensure that sophisticated tools are accessible to researchers and practitioners across various fields, facilitating the implementation of these technologies in diverse applications (Fang and Gong 2023; Araujo, Lock, and van de Velde 2020). These evaluations also establish that the selected models meet technical requirements and confirm their practicality and effectiveness for the intended applications.

For model capability, as presented in Table 2, these models are grouped according to the categories and types of image analysis. Nine single-functional models are grouped by the three categories of image analysis: *visual conception*, *emotion recognition*, and *aesthetic appreciation*. Three multi-functional models follow them. Some single-functional models can have more than one type of image analysis in a category, so we listed them first in that group. We conducted a multiple correspondence analysis (MCA) to visually depict the relationships among these 12 computer vision models in Figure 1. The MCA plot provides a detailed visualization of the capabilities of various computer vision models across three principal axes (visual conception, emotion recognition, and aesthetic appreciation). This three-dimensional plot effectively highlights the specialization and versatility of these models based on their categorical features.

Figure 1 reveals distinct groupings among the models. Neural image assessment (NIMA), IQA, and style-aware image recommendation (S-AIR) are closely grouped on the visual conception axis, indicating their focus on image quality and aesthetics, which is crucial for fields like marketing and design. FaceReader, DeepFace, and BrandImageNet form another cluster, emphasizing their capabilities in recognizing facial expressions and emotions. Finally, YOLO (You Only Look Once), Inception, and ImageIdentify are positioned together, suggesting similar functionalities and suitability for certain image analysis tasks.

Moreover, the plot illustrates the versatility of multi-functional models such as Vertex AI, Clarifai, and CLIP (contrastive language–image pretraining), which are located near each other, indicating their ability to handle diverse tasks across different areas of computer vision. This visualization aids in

Table 2. Capability of computer vision models by category of image analysis.

	Visual Conception			Emotion Recognition			Aesthetic Appreciation		
Computer vision models	IC	OD	CU	FED	SA	TSA	IQ	IA	SD
YOLO	N	Y	N	N	N	N	N	N	N
Inception	Y	Y	N	Y	N	N	N	N	N
ImageIdentify	N	Y	N	N	N	N	N	N	N
FaceReader	N	N	N	Y	N	N	N	N	N
DeepFace[a]	N	N	N	Y	N	N	N	N	N
BrandImageNet	N	N	N	N	N	Y	N	N	N
NIMA	N	N	N	N	N	N	N	Y	N
IQA[a]	N	N	N	N	N	N	Y	Y	N
S-AIR	N	N	N	N	N	N	N	N	Y
Vertex AI	Y	Y	Y	Y	Y	Y	N	N	N
Clarifai	Y	Y	Y	Y	Y	Y	N	N	N
CLIP	Y	Y	Y	Y	Y	Y	Y	Y	N

Note. [a]These models were chosen to enhance the models, and all others were sourced from the 30 articles used in developing the type of image analysis.
CLIP = contrastive language–image pretraining; CU = content understanding; FED = facial expression detection; IA = image aesthetic; IC = image classification; IQ = image quality; IQA = image quality assessment; NIMA = neural image assessment; OD = object detection; SA = sentiment analysis; TSA = themed sentiment analysis; S-AIR = style-aware image recommendation; SD = style and transfer design; YOLO = you only look once.

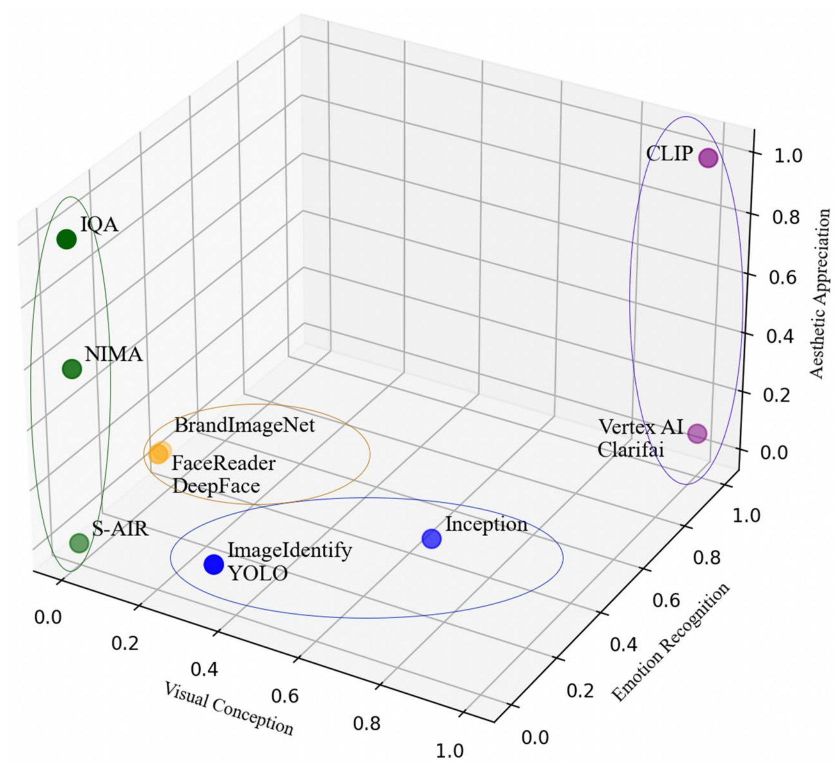

Figure 1. Multiple correspondence analysis (MCA) pivot of computer vision models.

understanding the range of model capabilities from specialized functions to more generalized applications, providing valuable insights for selecting appropriate models for specific tasks based on their strengths and areas of effectiveness.

The model's accuracy, availability, and usability are summarized in Table 3. We assess the accuracy of each model based on self-report data. Because not all authors used the same metrics, we classified these metrics into three levels: *low*, *moderate*, and *high accuracy*. Specifically, following the convention of computer vision model analysis, the models with an accuracy or performance metric of 85% or higher were classified to be high in accuracy; those with 60% to 84% were considered moderate; and those below 60% were classified as low. This classification allows for a simplified comparison and easy understanding of the models' effectiveness in their respective applications. Finally, we examined these models' availability and usability to discuss where users can find them, whether there is any cost, and the required skills for using these models.

Table 3. Image analysis models: Accuracy, availability, and usability.

Model	Accuracy	Availability	Usability
Visual Conception			
YOLOv7	Moderate	Open-source model available on GitHub	Strong programming skills in Python and experience with deep learning frameworks
Inception	High	Available through various deep learning frameworks like TensorFlow Pre-trained on large datasets and can be fine-tuned	Good understanding of deep learning concepts and proficiency in Python or a similar language
ImageIdentify	—	Offers a user-friendly interface	Basic understanding of image classification
Emotion Recognition			
FaceReader	Varies	Commercial software with a user-friendly interface	Basic understanding of facial expression analysis
DeepFace	Varies	Python library available from viso.ai	Strong programming skills and a deep understanding of facial recognition algorithms
BrandImageNet	High	Research-oriented model accessible through published papers and code implementations	Good understanding of deep learning concepts and proficiency in Python or a similar language
Aesthetic Appreciation			
NIMA	Moderate	Open-source model available on GitHub	Strong Python programming skills and familiarity with PyTorch
IQA	Varies	Comprehensive framework available as a Python library (IQA-PyTorch) on GitHub	Strong Python programming skills and familiarity with PyTorch
S-AIR	High	Research-oriented model accessible through published papers and code implementations	Strong programming skills and a deep understanding of recommendation systems and machine learning algorithms
Multi-functional			
Vertex AI	Varies	Accessible through the Google Cloud platform, offering both pre-trained models and custom model development User-friendly interface with options for varying levels of expertise	Python (beneficial for customization and integration)
Clarifai	Varies	User-friendly platform with a range of pre-trained models Offers a visual interface and APIs	Basic coding skills (helpful for complex integrations)
CLIP	Varies	Accessible through its open-source codebase	Strong understanding of Python and machine learning frameworks like PyTorch

Note. Some accuracy varies by specific attributes or models; details are presented in the paper.
API = application programming interface; CLIP = contrastive language–image pretraining; IQA = image quality assessment; NIMA = neural image assessment; S-AIR = style-aware image recommendation; YOLO = you only look once.

Visual Conception Models

Three models in this visual conception category are YOLO, Inception, and ImageIdentify.

YOLO (vision 7) is a state-of-the-art, real-time object detection model that allows advertising researchers to analyze visual content. Its strength lies in the speed and accuracy, enabling rapid identification and classification of objects within images and videos. For advertising researchers, this translates to a streamlined process for pinpointing crucial elements such as products, brand logos, or specific objects featured in advertisements. The model opens doors to automating content analysis, allowing researchers to efficiently derive insights from vast amounts of visual advertising data. Boasting an average precision of 56.8%, it demonstrates high proficiency in correctly detecting and categorizing objects across diverse scenarios. Moreover, the model's remarkable speed, operating at 30 frames per second or higher, makes it indispensable for real-time applications for which rapid insights are essential (Wang, Bochkovskiy, and Liao 2023).

YOLO is an open-source model readily accessible on GitHub, which democratizes researchers' use. Although its availability is a major advantage, effectively utilizing YOLO requires a solid foundation in programming skills (particularly Python) and familiarity with deep learning frameworks like PyTorch or TensorFlow. Researchers possessing these skills can harness the capabilities of YOLO to streamline workflows, analyze visual advertising content in depth, and uncover valuable insights that inform strategic decision-making. By leveraging its speed, accuracy, and open-source nature, researchers can unlock a new level of efficiency and precision in their investigations, leading to more informed strategies and enhanced advertising campaigns.

Inception is an ensemble model that harnesses the power of multiple architectures, including Inception (version 4) and residual networks, to achieve exceptional accuracy in image classification tasks. As a study by Szegedy et al. (2017) demonstrated, Inception achieved a remarkable 3.08% top-5 error rate on the ImageNet classification challenge test set, highlighting its prowess in recognizing and categorizing a vast array of objects and scenes. For advertising researchers, this capability translates to a potent tool for understanding the visual content of advertisements, identifying key objects, themes, and even subtle nuances that might influence consumer perception. The ensemble nature of Inception contributes to its robustness and high accuracy. Combining different model architectures leverages their strengths while mitigating their weaknesses, resulting in a more comprehensive and reliable classification system. Inception's ability to discern fine-grained details and classify images into thousands of categories makes it an asset for analyzing advertising visuals, particularly in complex scenes with multiple elements.

Inception models are available through popular deep-learning frameworks like TensorFlow. These pretrained models, trained on extensive datasets, can be fine-tuned for specific advertising research tasks, such as identifying product categories, analyzing brand imagery, or understanding visual trends. However, effectively utilizing Inception requires a good understanding of deep-learning concepts and proficiency in Python or a similar programming language. Researchers with these skills can leverage Inception's capabilities to delve deeper into the visual aspects of advertising, uncovering insights that drive more effective campaigns.

ImageIdentify is a sophisticated neural network-based model developed by Wolfram Research, and it is designed for the robust identification and classification of a wide array of objects and scenes in images. Specific details regarding its accuracy and performance metrics are not readily available. However, it was trained on an extensive array of more than 4,000 object classes. This model was meticulously designed to balance classification accuracy, computational size, and evaluation speed, ensuring efficient and effective image recognition capabilities (Villalobos Alva 2021). Although specific accuracy metrics are not readily available, its comprehensive training in many object classes suggests a robust capability to recognize and label various elements within advertising visuals. This capability can be particularly valuable for researchers seeking to understand the composition of ads, identify prominent objects or themes, and even detect subtle details that might influence consumer responses.

The uniqueness of ImageIdentify lies in its ability to process images while maintaining high accuracy and efficiency. This capability makes it well-suited for analyzing large datasets of advertising visuals, promptly providing researchers with actionable insights. The model's ability to handle diverse object classes and scenes further enhances its utility in advertising research, enabling comprehensive and nuanced analyses. Accessing and utilizing ImageIdentify is facilitated by its integration within the Wolfram Language ecosystem. While specific technical skills required might vary depending on the implementation, a general understanding of image analysis principles and familiarity with the Wolfram Language can enable researchers to leverage the power of this model for their advertising research endeavors.

Emotion Recognition Models

FaceReader, DeepFace, and BrandImageNet are three models in the emotion recognition category.

FaceReader is a powerful model in the arsenal of advertising researchers, specializing in the real-time analysis of facial expressions to gauge audience responses to various stimuli, including advertisements. As Hadinejad et al. (2019) demonstrated, FaceReader can measure how people react emotionally to different ad creatives, providing invaluable insights into the effectiveness of emotional appeals, humor, or shock value. With an 89% recognition rate for static images and 80% for dynamic expressions, FaceReader offers a reliable way to assess the emotional impact of advertising images.

For advertising researchers, the user-friendly interface of FaceReader is desirable. They can easily analyze facial expressions frame by frame, tracking emotional changes throughout an ad's duration. These data can reveal which moments resonate most with viewers, pinpoint areas where engagement drops, and identify specific emotions elicited by different ad elements. FaceReader is a commercial software designed for researchers and practitioners across various fields. Although it requires no coding, a basic understanding of facial expression analysis is beneficial for interpreting results and drawing meaningful conclusions.

DeepFace is a cutting-edge facial analysis library that has become a valued tool for advertising researchers seeking to understand audience responses and demographics. It can analyze photographs to identify human faces with high accuracy. Researchers have observed that this facial recognition model surpasses the human accuracy benchmark (Serengil and

Ozpinar 2020, 2021). Beyond facial recognition, DeepFace analyzes various facial attributes, including age, gender, and emotion. For specific attribute predictions, the model demonstrates a mean absolute error of 4.65 in age estimation, suggesting it can predict age with an accuracy range of about plus or minus 4.65 years. On the same dataset used for age estimation, the model achieves a 97.44% accuracy in gender prediction, 68% in race/ethnicity classification, and 57.42% in emotion detection (Serengil and Ozpinar 2021). DeepFace is a research-oriented model primarily accessible through published papers and code implementations. It requires strong programming skills and a deep understanding of facial recognition algorithms. It is more suitable for researchers and developers with expertise in this domain.

By identifying various facial attributes, this model enables researchers to gauge audience attention to faces within ads, measure emotional reactions to specific characters or scenarios, and even conduct demographic analysis based on facial types, and thus, tailor advertising campaigns for maximum impact and resonance. DeepFace is available as a Python library, offering a wide range of features and capabilities for facial analysis. While it requires programming skills and familiarity with machine learning concepts, its extensive documentation and active community support make it accessible to researchers and developers with varying levels of expertise. With its exceptional accuracy and comprehensive facial analysis capabilities, DeepFace has become an essential tool for advertising researchers seeking to unlock the hidden potential of facial data and gain a deeper understanding of consumer behavior.

BrandImageNet, a specialized model trained to identify and analyze brand imagery attributes, offers a unique capability for advertising researchers. Unlike FaceReader and DeepFace, which focus on facial analysis, BrandImageNet takes ad images as input and outputs a range of brand-related attributes, such as "glamorous," "healthy," "rugged," or "fun." With an average accuracy rate of 90% (Liu, Dzyabura, and Mizik 2020), this model provides a reliable way to assess brand image consistency across different campaigns. Thus, BrandImageNet provides a window into the minds of consumers, revealing how a brand is being perceived in the real world. This information is invaluable for assessing the effectiveness of brand positioning strategies, identifying potential misalignments between intended and perceived brand images, and tailoring marketing messages to evoke specific emotions or associations.

BrandImageNet is available primarily through academic research papers and code implementations, making it a valuable resource for researchers with a strong foundation in machine learning and image analysis. While it offers immense potential for brand-related insights, utilizing BrandImageNet effectively requires familiarity with machine learning concepts and potential coding skills to adapt and apply the model to specific research questions. Researchers with this expertise can use BrandImageNet to better understand brand perception, refine brand messaging, and ultimately enhance brand equity.

Aesthetic Appreciation Models

We have three models in the aesthetic appreciation category: NIMA, IQA, and S-AIR.

NIMA represents a paradigm shift in aesthetic evaluation, enabling researchers to quantify the visual appeal of advertising images through a lens that aligns with human perception. Predicting the aesthetic quality of images allows researchers to understand the visual elements that resonate most effectively with consumers. It has demonstrated an accuracy of 81.5% in predicting image appeal, a testament to its ability to mimic human judgment in evaluating aesthetic qualities. This accuracy offers researchers a reliable metric for assessing the effectiveness of visual elements within advertisements (Talebi and Milanfar 2018).

In the context of advertising research, NIMA provides a valuable tool for optimizing visual content. Researchers can utilize NIMA to identify high-performing visuals, analyze the impact of different artistic styles, and tailor creative strategies to maximize viewer engagement. Advertisers can craft visually compelling campaigns that leave a lasting impression by understanding the aesthetic qualities that appeal to target audiences. NIMA is an open-source model accessible to researchers and developers proficient in Python and deep learning frameworks. Although technical expertise is necessary for implementation and fine-tuning, the insights from NIMA's analysis can provide a significant competitive advantage in advertising research.

IQA is a comprehensive framework that offers the reimplementation of a diverse set of models and metrics designed to evaluate images' technical and perceptual quality. This toolbox enables advertising researchers to assess various aspects of image quality, including sharpness, noise, color accuracy, and overall aesthetic appeal. The performance evaluation protocol of this toolbox is methodically structured. When official models are available, they are utilized for evaluation purposes. When official models are unavailable, the toolbox adopts a uniform set of training and evaluation settings to ensure simplicity and consistency across different models. This approach allows for a fair and standardized comparison of model performance. All benchmark results can be assessed at the GitHub site (https://github.com/chaofengc/IQA-PyTorch/tree/main/tests).

IQA can play a crucial role in the visual integrity of advertising campaigns. By rigorously evaluating the quality of images used in advertisements, researchers can identify and address potential issues that may detract from the overall impact, such as compression artifacts, blurriness, or color inconsistencies. IQA is available as a Python library (IQA-PyTorch), requiring strong programming skills and familiarity with the PyTorch framework. Its comprehensive nature and extensive evaluation tools make it an indispensable resource for researchers and professionals dedicated to upholding the highest visual quality standards in advertising.

S-AIR is a content-based recommendation system that leverages machine learning to understand and predict the visual preferences of specific brands or influencers. By analyzing vast amounts of visual data, this model learns a particular entity's unique style and aesthetic characteristics, enabling it to recommend images that align with their established brand identity or personal style. The evaluation is made using the metric area under the ROC curve (AUC). The higher AUC indicates a higher probability of achieving a high true positive rate while maintaining a low false positive rate. Here, the true positive is counted by the model correctly recommending associated ad images to the brands, and the false positive is calculated by the model falsely recommending a related ad image to the brands. Their method achieved an AUC of 0.946 (maximum = 1) on the brand dataset and 0.918 on the influencer dataset, further affirming its effectiveness in aligning image recommendations with specific influencer styles (Zhang and Yamasaki 2021).

This model is a powerful tool for optimizing visual content strategies for advertising research. By understanding the specific visual preferences of target audiences, researchers can curate and recommend images that resonate with them, leading to increased engagement, brand recall, and positive consumer interactions. The model's ability to learn and adapt to evolving styles ensures that recommendations remain relevant and effective. This model is available through research papers and code implementations, requiring strong programming skills and expertise in recommendation systems and machine learning algorithms.

Multi-Functional Models

Three multi-functional models, Vertex AI, Clarifai, and CLIP, were used in our reviewed studies. These models represent the forefront of computer vision technology and can carry out image analysis in more than one category. The accuracy and performance of these platforms are highly contingent on the choice between using their sophisticated pre-trained models, which are adept in tasks such as object detection and sentiment analysis, and the option to train custom models, which allows for tailored solutions specific to unique datasets and requirements.

Vertex AI, a comprehensive machine learning platform offered by Google Cloud (previously known as Google Vision), empowers advertising researchers with a diverse suite of pre-trained and customizable models for various visual analysis tasks. Its capabilities span brand logo detection, object identification within ad creatives, sentiment analysis, and image text extraction. This versatility stems from its extensive collection of pre-trained models and the ability to develop and deploy custom models tailored to specific research needs. Vertex AI offers a robust solution for advertising research for comprehensive visual content analysis. Its ability to identify brand logos enables researchers to track brand visibility and measure the effectiveness of product placements. Object detection facilitates the analysis of visual elements within ads, providing insights into composition, messaging, and potential biases. Sentiment analysis helps gauge audience reactions to advertising content, and text extraction can be used to analyze textual elements embedded within images. Vertex AI caters to users with varying levels of expertise, offering a user-friendly interface for beginners and more advanced tools for experienced practitioners. Whereas proficiency in Python benefits customization and integration with other tools, the platform's intuitive design allows researchers to leverage its capabilities without extensive coding knowledge.

Clarifai is a user-friendly platform renowned for its wide array of pre-trained models for visual recognition tasks, including image classification, object detection, and facial recognition. Its intuitive visual interface and API integrations make it accessible to technical and non-technical users, enabling seamless integration into existing research workflows. For advertising research, the real-time analysis capabilities of Clarifai are particularly advantageous. By quickly identifying objects, demographics, and potentially harmful content within images and videos, researchers can gain immediate insights into audience composition, brand safety, and the effectiveness of visual elements. This capability enables timely adjustments to advertising strategies and ensures that brand messaging aligns with ethical standards. The Clarifai pre-trained models are readily available, and the platform also offers the option to train custom models for specific research needs. Whereas basic coding skills can be helpful for more complex integrations, the user-friendly design for Clarifai allows researchers to harness its power without extensive programming knowledge. With its focus on real-time analysis and ease of use, Clarifai has become a valuable tool for advertising researchers seeking efficient and actionable insights from visual content.

CLIP, developed by OpenAI, represents a breakthrough in bridging the visual and textual content gap. Its unique capability to understand and interpret images and text allows it to perform various tasks with remarkable efficiency, including image classification, object detection, and sentiment analysis. The zero-shot learning capabilities of CLIP enable it to generalize its understanding to new, unseen data, making it a versatile tool for advertising research. In advertising, the strength of CLIP lies in analyzing the alignment between ad copy and visuals, content moderation, and sentiment analysis. It can assess whether the visual elements of an advertisement effectively convey the intended message, identify potential biases or inconsistencies, and gauge audience reactions to both textual and visual components (Ramesh et al. 2022). The capability of CLIP to understand the interplay between language and images offers valuable insights into how these elements create meaning and influence consumer behavior. CLIP is primarily accessible through its open-source codebase, requiring a strong understanding of Python and machine learning frameworks like PyTorch. The unique approach of CLIP to multi-modal understanding opens new avenues for advertising research, allowing for a deeper exploration of the complex relationship between text and image in advertising communication.

Single- Versus Multi-Functional Models

Multi-functional models offer numerous benefits, including performing various tasks in a single framework. This versatility streamlines the workflow by reducing the need for multiple models, simplifies integration processes, and provides a consistent approach to various types of analysis. These models are particularly useful in settings that require a comprehensive analysis because they can handle everything from image classification and object detection to emotion recognition and aesthetic assessment. Furthermore, multi-functional models often come with pre-trained capabilities and customizable options, making them accessible and adaptable for various research and practical applications. However, there are situations in which choosing single-functional models is more advantageous. Single-functional models are often preferred when a task requires high precision and specialized features that multi-functional models may lack. They tend to have optimized performance for particular applications. Due to their focused scope, single-functional models can be more resource-efficient and easier to train, fine-tune, and debug. They are also better when computational resources are limited or the deployment environment demands minimal resource consumption. Thus, whereas multi-functional models offer broad capabilities, single-functional models provide superior performance and efficiency for specialized tasks.

Example Use of Select Computer Vision Models

To demystify the use of computer vision models in advertising research, we present an example study that used single-functional and multi-functional models in image analysis. Our research question is: What content categories strongly correlate with high aesthetic advertisement scores? We conducted two interrelated studies using Vertex AI for image classification and NIMA for aesthetic evaluation.

We utilized the Pitts Ad dataset (Kovashka 2024), which is available online (https://people.cs.pitt.edu/~kovashka/ads/). This dataset is a large collection of visual advertisements representing various products labeled by human coders. This dataset is ideal for training and evaluating machine learning models for image classification and aesthetic evaluation tasks. We visited the "readme" page to access this dataset and followed the instructions to download the images. The readme file contains links to the image files and additional details about the dataset. Once the images were downloaded, we randomly selected 500 advertisements for this study.

We went to Vertex AI dashboard on Google Cloud to create our machine-learning model for classifying the sample image advertisements. In the dashboard, we clicked "Prepare Data" to create the dataset and selected "Image Classification." It used the single label to predict the correct label the user wanted assigned to an image. We uploaded the downloaded Pitts Ad images from our computer, and once the import was done, we got an email from Google support. Then, we split our 500 images into 90 images as the training set and 410 images as the testing set. We labeled the training set with various categories of advertisements, such as food, beauty, clothing, auto, public service announcements (PSAs), and others. We used those labels just for demonstration purposes; researchers can label those ads they deem the most suitable for analysis.

After labeling the training set, we returned to the dashboard and found the "Model Development" tab. We selected to create a model and train a new model. Our model training process lasted for 1 hour and 53 minutes. Returning to the dashboard, we selected "Deploy & Use" and "Online Prediction—Deploy models for online predictions." The trained model was then applied to the testing dataset to predict the content categories of the image ads, and we received the predicted labels and some accuracy results. Completing the training model cost us $18, which would escalate with larger datasets and extended training durations. Also, additional costs would arise when deploying the model on the cloud. For instance, storing and deploying the training model on Google Cloud for approximately 20 hours would cost about $200.

The precision and recall metrics provide valuable insights when evaluating the model's performance across various advertisement categories like food, beauty, clothing, auto, and PSAs. The model demonstrates high reliability with an average precision of 0.869, and it achieves a precision of 100% at a higher confidence threshold of 0.75, which indicates that the model accurately identifies when an advertisement fits a specific category. Therefore, when the model categorizes an ad as belonging to one of these groups, there is a high likelihood it is correct, particularly when operating above the confidence threshold. The two example testing images and their predicted category with a confidence score are presented in Figure 2.

The recall rate of 50% indicates a limitation in the model's ability to identify all relevant ads across the various categories. This metric shows that the model only successfully identifies 50% of the advertisements that should be classified under the categories it has been trained to recognize. This lower recall rate could stem from several factors. First, each advertisement category has a significant diversity. For example, beauty ads could range from minimalist to highly elaborate. If the training data did not capture this wide variation, the model might fail to recognize less typical ads. Second, ads across different categories might share visual or thematic similarities, such as color schemes or imagery, which could confuse the model if it has not been adequately trained to distinguish these nuances. Lastly, different categories may rely on distinct features, such as textual content in PSAs versus visual imagery in food ads. If the model is not well-tuned to these specifics, its performance might lag in certain categories.

Now that we had all the content category labels for our 500 images, we moved on to assess the aesthetic appeal of the categorized ads using the NIMA model from the IQA PyTorch library. NIMA was chosen for its robust training on extensive datasets specifically rated for aesthetic quality, making it a reliable tool for evaluating the aesthetic appeal of images. To utilize the IQA-PyTorch library, we first need to install PyTorch and the IQA PyTorch library. Then, we must create a Python script to load the NIMA model and evaluate the images. Due to their objective and structured nature, generative AI tools, including ChatGPT, can efficiently handle these scripts. After running the script, we received a column called "Image Scores", which contained the aesthetic scores for each image ad. We further analyzed these scores to examine correlations between advertisement categories and aesthetic appeal scores to answer our research question (see Supplemental Online Appendix 2 for a description of our example study, including the procedures and research findings).

Summary and Discussion

We reviewed the literature on social science and computer science and identified three categories and nine types of image analysis of brand-related images on social media. We then used these categories and types to select 12 computer vision models and compare them in terms of their capability, accuracy, availability, and usability. As a quick demonstration, we gathered a sample of image ads from a publicly available dataset. We used two computer vision models to analyze the sample and answer our research question. These analysis, comparison, and example use of select models should facilitate advertising scholars and practitioners in selecting and using the models in research.

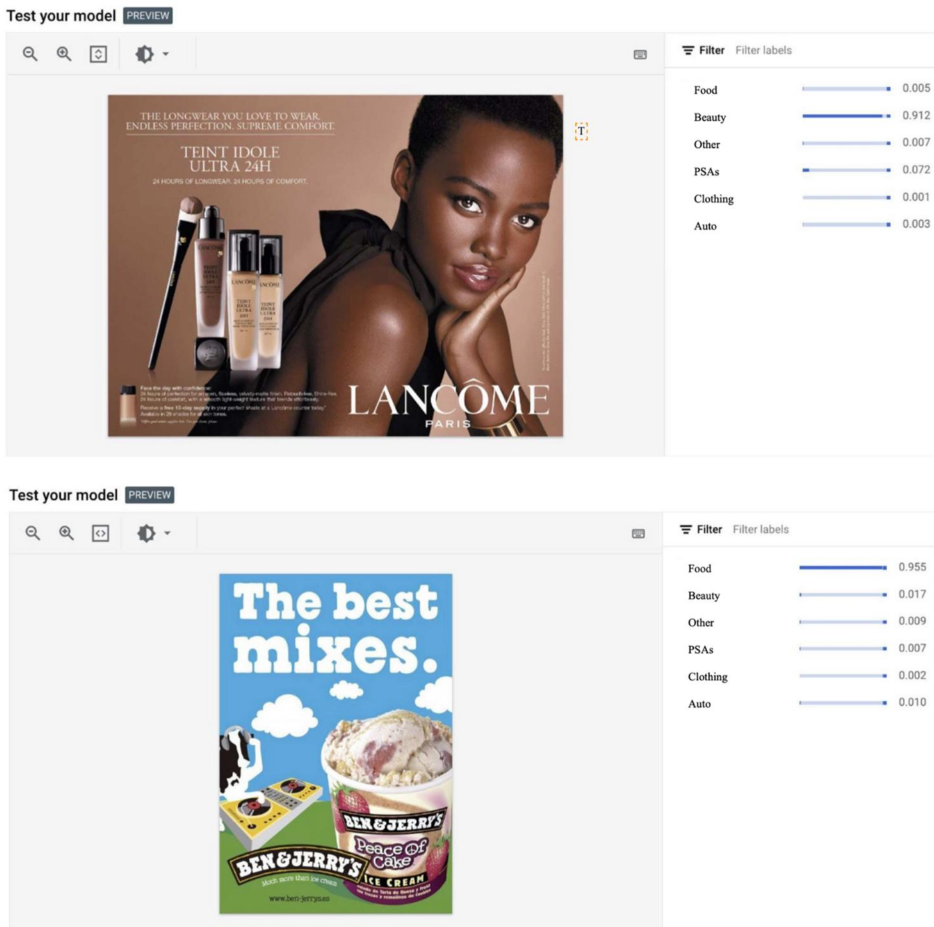

Figure 2. Example of Vertex AI performance.

Advantages and Limitations of Computer Vision Models

Before embarking on computer vision models, researchers should understand the advantages and limitations of these models compared with manual coding and analyzing images. The application of computer vision models in image analysis brings several clear advantages, particularly in advertising research. One of the most significant advantages is the ability to analyze large datasets efficiently. Unlike manual image analysis, which is time-consuming and labor-intensive, automated methods allow researchers to quickly handle and process vast amounts of visual data. This efficiency opens new possibilities for asking research questions that were previously unfeasible due to resource constraints (Araujo, Lock, and van de Velde 2020). Additionally, automated models reduce the potential for human coder bias, thereby offering more consistent and objective results across different datasets (Argyris et al. 2020). This objectivity is crucial when conducting large-scale because it helps establish reliable and reproducible findings.

Despite these advantages, the use of computer vision models for image analysis has several disadvantages. A key limitation is the inability of automated models to capture the rich qualitative insights that manual methods can accomplish, such as content analysis or semiotic analysis. Manual approaches allow researchers to delve deeper into images' cultural and societal meanings, interpreting visual rhetoric and uncovering nuances that automated systems may overlook (McQuarrie and Mick 2003). When interpreting and analyzing visual data, it is important to recognize that visual representations can be ambiguous and open to multiple interpretations (Patricia 2020) because images can hold multiple meanings depending on the cultural, social, and historical contexts in which they are viewed. In addition, computer vision models can struggle with cultural variation, potentially misinterpreting images from different sociocultural contexts due to biases in the training data (Serengil and Ozpinar 2020). Lastly, fully automated systems may lack the capacity to understand the subjective and context-dependent meanings central to understanding visual content's cultural impact

(Patricia 2020). Therefore, while computer vision models provide valuable quantitative insights, they should complement qualitative approaches to offer a more holistic and nuanced visual data analysis.

Selection of Suitable Models

With so many types of image analysis and computer vision models, researchers may start to think about three basic questions about image analysis: 1) *what is in an image ad?*; 2) *what emotion does an image ad convey?*; and 3) *what is the quality of an image ad?*. These three questions can be further explored considering the three categories of image analysis using relevant computer vision models (see Table 4 for a summary). For example, if a study intends to investigate what products virtual influencers are likely to present in their posts on social media, a sample of such posts can be collected, and some of them can be used to train an image classification model, like ImageIdentify. After classifying the products in the sample, an emotional model, like BrandImageNet, can be used to reveal brand-related attributes of these products. If the study is also interested in the facial impression of the virtual influencers, a model like FaceReader or DeepFace can be used to generate data from these image posts. Finally, if the study needs to assess the aesthetics of these posts, models like NIMA can be easily trained to analyze these posts. Ultimately, what computer vision model to use in a study will depend on the research questions, types of images, and the researcher's programming skills.

Key Steps in Using a Model

Although brief, our example study includes several key steps in using a computer vision model, summarized in the following section.

Developing Research Questions

The first step in applying computer vision models in advertising research is clearly stating the research questions your study aims to answer. For example, "Do certain ads evoke positive emotions?" or "How does the ad's aesthetic align with its brand identity?" These questions will guide the choice of models and the overall research design. At this stage, researchers must ensure the questions are well-defined, contextually relevant, and feasible for analysis with a computer vision model.

Selecting Appropriate Models

Once the research questions are defined, the next step is to select the most suitable computer vision model. Choose a model that fits best with the objective of analysis—for instance, using emotion recognition

Table 4. Relevance of computer vision models to advertising issues.

Model	Relevance to Advertising Issues
Visual Conception	
YOLOv7	Real-time object detection for tracking ad placements in videos, analyzing product placement effectiveness, measuring viewer attention to specific elements within an ad
Inception	Transfer learning for custom image classification tasks (e.g., identifying ad types, categorizing brand imagery), feature extraction for building custom models
ImageIdentify	Automatic identification of objects and scenes in ads, providing insights into visual content and its potential impact
Emotion Recognition	
FaceReader	Emotion detection in response to ads, analyzing the effectiveness of facial expressions in ad creatives, understanding audience engagement through facial cues
DeepFace	Facial recognition for identifying celebrities or influencers in ads, measuring audience attention to faces in ads, demographic analysis based on facial types
BrandImageNet	Analyzing brand image consistency across different ad campaigns, measuring the effectiveness of brand imagery, understanding how consumers perceive brands based on visual cues
Aesthetic Appreciation	
NIMA	Assessing the aesthetic quality of ad creatives, predicting consumer preferences based on visual appeal, optimizing ad design for maximum impact.
IQA	Ensuring high-quality visuals in ad campaigns, evaluating the technical quality of images and videos, comparing the visual quality of different ad variations
S-AIR	Recommending visually similar ad creatives to users, personalizing ad experiences based on visual preferences, generating new ad variations based on existing styles
Multi-functional	
Vertex AI	Brand logo detection, object detection in ad creatives, sentiment analysis of images, optical character recognition for text analysis
Clarifai	Object recognition and classification, visual search for similar ad creatives, demographic analysis of people in ads, NSFW content detection
CLIP	Text-image pairing for analyzing the relationship between ad copy and visuals, content moderation for brand safety, sentiment analysis of images

CLIP = contrastive language–image pretraining; IQA = image quality assessment; NIMA = neural image assessment; NSFW = not safe for work; S-AIR = style-aware image recommendation; YOLO = you only look once.

models like FaceReader or DeepFace to gauge emotional cues and aesthetic appreciation models like NIMA or IQA to evaluate the visual appeal of image advertisements. When in doubt, engaging domain experts in computer science is helpful to refine the research focus and ensure that the model's outputs align with analytical tasks. Refer to Tables 2 and 3 for guidance on the capability, accuracy, availability, and usability of each of the 12 models, ensuring that your selection is well-informed and suitable to your research needs.

Collecting and Preparing Data

Gather or create a dataset of relevant images or videos relevant to the research questions. The dataset should be properly formatted and labeled according to the requirements of the chosen computer vision model. For instance, if using a model that requires labeled emotional states, ensure that each image or video segment is annotated with the corresponding emotional labels. Human validation is vital during data labeling to ensure that annotations accurately reflect the intended categories because errors in labeling can lead to misinterpretation of the results. Proper data preparation is crucial for the accuracy and reliability of the model's output.

Training the Model If Applicable

Training a customizable model on a specific dataset can significantly optimize the performance for the intended tasks. For example, fine-tuning a facial expression model to recognize subtle emotional nuances better can enhance its accuracy. This step involves adjusting the model's parameters and training it on the labeled dataset to ensure it performs well in the research context. During this process, human validation is key to verifying that the model is learning appropriately and not overfitting or underfitting the data.

Analyzing Images

Apply the selected computer vision model to the dataset to extract relevant information. For instance, a model can be used to identify facial expressions and analyze emotional valence scores or to assess aesthetic ratings of ad creatives. This step involves running the images or videos through the model and interpreting the outputs based on the research questions. Human validation is critical to confirm the accuracy and relevance of the model's predictions, ensuring that the extracted insights are correct and meaningful within the research context.

Reporting and Interpreting Findings

The final step is to report the results and interpret their implications. This step involves synthesizing the data outputs into meaningful insights that address the research questions. The findings should be presented clearly and concisely, with key takeaways and their relevance to advertising theories and practices. Researchers should contextualize the findings and warrant that the conclusions drawn from the model align with established theoretical frameworks in advertising research. Effective reporting ensures that the insights derived from the computer vision models are actionable and can inform future advertising research.

Ethical Considerations

Incorporating ethical considerations into computer vision models is imperative, especially when dealing with sensitive data like images of individuals. One of the primary concerns is privacy. Researchers must be certain they have proper consent for using images, particularly those gathered from public platforms such as social media. The potential for misuse, such as in surveillance or manipulative advertising, raises significant ethical questions. Researchers should be cautious and transparent about how the data are collected and processed, emphasizing the importance of safeguarding individual privacy and consent (Joo and Steinert-Threlkeld 2022).

Another major ethical issue is bias in the models. Computer vision algorithms, particularly those trained on biased datasets, can perpetuate or amplify preexisting societal biases. For instance, emotion detection models may interpret facial expressions differently based on race or gender (Serengil and Ozpinar 2020, 2021). Researchers must validate their model outputs with human judgments and ensure their training data are representative and diverse. Raising awareness of these ethical challenges can be achieved through transparent reporting, collaboration with ethical experts, and publishing clear guidelines for the responsible use of these technologies.

Lastly, transparency and accountability are fundamental when using automated tools, especially in research contexts. Researchers should be transparent about their models' limitations and potential errors to guarantee that the outcomes are trustworthy and ethically sound. This transparency involves documenting the model's design, data sources, and the assumptions embedded in its algorithms. By providing transparency, researchers make it easier for others to scrutinize and verify the findings, promoting a culture of openness and collaboration in the scientific community (GOV.UK 2023). Additionally, they should advocate for external accountability mechanisms, ensuring that their models undergo thorough ethical reviews and regular audits by independent experts. Accountability measures include oversight, risk assessments, and ensuring the system is auditable and can trace errors or biases to prevent unintended consequences (Balasubramaniam et al. 2022).

Future Research Directions

Using computer vision models for image analysis is still in its early stages, and researchers can use such models to open new areas and explore new issues in advertising research. One direction is to explore the relationships between attributes of brand-related images and consumer responses. Image attributes gleaned from computer vision models can serve as independent variables. In contrast, dependent variables may include many factors, such as brand recall, brand perception, brand engagement, purchase intent, and actual purchases. While some of these relationships have been studied, little is known regarding consumer responses to new forms of brand-related images, such as virtual influencers' posts on social media and AI-generated brand-related images. Computer vision models can play an important role in new studies of these issues. Although most computer vision models we reviewed in this study are designed to analyze images, some can decode videos for object detection, like YOLO. As more brands have established a presence in the metaverse, more models are expected to emerge that can effectively and efficiently analyze visuals in virtual, augmented, and mixed-reality settings. Using computer vision models in advertising research in the metaverse represents a brand domain worth exploring.

When selecting computer vision models, researchers should understand the currency of these models. Our review examined the release dates, update frequencies, and current relevance of these 12 models. For visual conception models, the YOLO family began with its original release in 2016 and has seen continuous development, with YOLOv7 being the latest major iteration released in July 2022. These models are highly regarded for real-time object detection, with newer versions like YOLOv7-Tiny and YOLOv7-W6 further improving speed and accuracy, particularly useful for applications like tracking ad placements and analyzing viewer attention. Inception, integrated into the Google TensorFlow framework, has been a cornerstone in image classification tasks since its introduction, benefiting from regular updates and extensive use in transfer learning. While ImageIdentify may not receive frequent updates, they remain specialized and user-friendly, effectively serving specific use cases.

For emotion recognition models, FaceReader, released in 2005, provides reliable emotion detection capabilities that are useful in analyzing facial expressions in advertisements. DeepFace, launched in 2014, is widely employed for facial recognition, including identifying celebrities and conducting demographic analysis. BrandImageNet analyzes brand image consistency across campaigns and is more research-oriented; hence, it is updated less frequently. For aesthetic appreciation models, NIMA, introduced in 2018, plays a critical role in assessing the aesthetic quality of images, predicting consumer preferences, and optimizing ad design. The broader field of IQA has seen various models developed and refined through ongoing research, ensuring the maintenance of high-quality visuals in advertising campaigns. S-AIR, released in 2021, represents a cutting-edge development in recommending images based on stylistic elements.

For multi-functional models, Vertex AI, part of the Google Cloud platform, supports tasks like brand logo detection and sentiment analysis, offering pre-trained models and custom development options. Clarifai, launched in 2013, is a versatile platform known for object recognition and visual search, continually updated to meet diverse needs. CLIP, introduced by OpenAI in early 2021, excels in analyzing the relationship between ad copy and visuals, providing content moderation and sentiment analysis. Although the regular updates and adaptability of these models ensure their continued relevance in the rapidly evolving field of computer vision, researchers should attend to newer models that are likely superior in capability, accuracy, availability, and usability.

Like any algorithm in machine learning, computer vision models may have biases due to the data on which they are trained. Thus, researchers who use such models should be mindful and test them in the research process. Addressing algorithmic biases in computer vision models and establishing ethical guidelines for their use in advertising research is paramount. Bias mitigation and ethical considerations ensure equitable and responsible deployment, safeguarding against potential misuse and ensuring that the benefits of the models are accessible to all.

Finally, breakthrough findings are often the results of multidisciplinary collaboration. Thus, fostering interdisciplinary collaboration between advertising researchers, computer scientists, psychologists, and neuroscientists can lead to more comprehensive and nuanced models of consumer response to visual content. Such collaboration can integrate diverse perspectives and expertise, enhancing the depth and breadth of using computer vision models research in advertising research.

Managerial Implications

While this review focuses on the academic use of computer vision models, it does not mean they are less valuable in advertising practice. Computer vision models are transforming the advertising industry by offering innovative tools to enhance customer engagement and personalize marketing strategies. These models enable brands to recognize visual content that resonates with specific audiences, allowing for more targeted and engaging advertisements (Adweek n.d.). They also detect logos and products in social media posts, allowing brands to monitor their presence and gauge customer sentiment effectively (Medium 2024). Future more, computer vision models can help brands understand the context of visual content, ensuring that ads are not only relevant to the user but also aligned with the content they are viewing, thus aiding in improving ad placements and increasing overall engagement by ensuring that the advertisements complement the content consumed (GumGum n.d.). Creative use of these models in advertising should boost the effectiveness of various campaigns involving brand-related images and achieve better consumer responses.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Hairong Li http://orcid.org/0000-0001-6098-1689
Nan Zhang http://orcid.org/0000-0002-7091-708X

References

Adweek. n.d. "Do You Have a Visual Strategy for Your Digital Advertising?" https://www.adweek.com/sponsored/do-you-have-a-visual-strategy-for-your-digital-advertising/

Araujo, T., I. Lock, and B. van de Velde. 2020. "Automated Visual Content Analysis (AVCA) in Communication Research: A Protocol for Large Scale Image Classification with Pre-Trained Computer Vision Models." *Communication Methods and Measures* 14 (4): 239–265. https://doi.org/10.1080/19312458.2020.1810648.

Argyris, Y. A., Z. Wang, Y. Kim, and Z. Yin. 2020. "The Effects of Visual Congruence on Increasing Consumers' Brand Engagement: An Empirical Investigation of Influencer Marketing on Instagram Using Deep-Learning Algorithms for Automatic Image Classification." *Computers in Human Behavior* 112: 106443. https://doi.org/10.1016/j.chb.2020.106443.

Asresa, T., G. Tigistu, and M. Bayih. 2023. "Convolutional Neural Network Driven Computer Vision Based Facial Emotion Detection and Recognition." *Indian Journal of Computer Graphics and Multimedia* 3 (2): 8–11. https://doi.org/10.54105/ijcgm.D6601.083223.

Balasubramaniam, N., M. Kauppinen, K. Hiekkanen, and S. Kujala. 2022. "Transparency and Explainability of AI Systems: Ethical Guidelines in Practice." In International Working Conference on Requirements Engineering Foundation for Software Quality, 3–18. Cham: Springer International Publishing.

Bisht, S. 2019. "Impact of Cause Marketing on Brand Affinity with Young Mothers." *International Journal of Research in Humanities Arts and Literature* 7 (5): 305–314.

Borji, A., M. M. Cheng, H. Jiang, and J. Li. 2015. "Salient Object Detection: A Benchmark." *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 24 (12): 5706–22. https://doi.org/10.1109/TIP.2015.2487833.

Chen, L. 2023. "How Advertising in Social Media Shapes Users' Consumption Behavior: The Example of the Red Marketing Layout of Chi Forest." *BCP Business & Management* 42: 136–141. https://doi.org/10.54691/bcpbm.v42i.4572.

Duanis, S., K. Gaiger, R. Cohen, S. Zychlinski, and A. Greenstein-Messica. 2023. "Visual Representation for Capturing Creator Theme in Brand-Creator Marketplace." Proceedings of the 17th ACM Conference on Recommender Systems, Singapore, Singapore. https://doi.org/10.1145/3604915.3610237.

Fang, J., and X. Gong. 2023. "Application of Visual Communication in Digital Animation Advertising Design Using Convolutional Neural Networks and Big Data." *PeerJ. Computer Science* 9: e1383. https://doi.org/10.7717/peerj-cs.1383.

Fire, M., and J. Schler. 2017. "Exploring Online Ad Images Using a Deep Convolutional Neural Network Approach." In *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 1053–1060. Exeter, UK: IEEE. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2017.160.

Gambetti, A., and Q. Han. 2022. "Camera Eats First: Exploring Food Aesthetics Portrayed on Social Media Using Deep Learning." *International Journal of Contemporary Hospitality Management* 34 (9): 3300–3331. https://doi.org/10.1108/IJCHM-09-2021-1206.

Giglio, S., E. Pantano, E. Bilotta, and T. C. Melewar. 2020. "Branding Luxury Hotels: Evidence from the Analysis of Consumers' "Big" Visual Data on TripAdvisor." *Journal of Business Research* 119: 495–501. https://doi.org/10.1016/j.jbusres.2019.10.053.

GOV.UK. 2023. "Ethics, Transparency and Accountability Framework for Automated Decision-Making." https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making.

GumGum. n.d. "Seen & Unseen: How Marketers Use Computer Vision." https://gumgum.com/guides/seen-unseen-contextual-relevance-in-advertising.

Ha, Y., K. Park, S. J. Kim, J. Joo, and M. Cha. 2021. "Automatically Detecting Image–Text Mismatch on

Instagram with Deep Learning." *Journal of Advertising* 50 (1): 52–62. https://doi.org/10.1080/00913367.2020.1843091.

Hadinejad, A., B. D. Moyle, N. Scott, and A. Kralj. 2019. "Emotional Responses to Tourism Advertisements: The Application of FaceReader™." *Tourism Recreation Research* 44 (1): 131–135. https://doi.org/10.1080/02508281.2018.1505228.

Hartmann, J., M. Heitmann, C. Schamp, and O. Netzer. 2021. "The Power of Brand Selfies." *Journal of Marketing Research* 58 (6): 1159–1177. https://doi.org/10.1177/00222437211037258.

Hernandez, J., L. Zicheng, G. Hulten, D. DeBarr, K. Krum, and Z. Zhang. 2013. "Measuring the Engagement Level of TV Viewers." In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–7. Shanghai, China: IEEE.

Holiday, S., J. L. Hayes, B. C. Britt, and Y. Lyu. 2021. "The Cause Effect: The Impact of Corporate Social Responsibility Advertising on Cause Consumer Engagement Behavior after Brand Affiliation Ceases." *International Journal of Advertising*40 (2): 199–224. https://doi.org/10.1080/02650487.2020.1769408.

Holiday, S., J. L. Hayes, H. Park, Y. Lyu, and Y. Zhou. 2023. "A Multimodal Emotion Perspective on Social Media Influencer Marketing: The Effectiveness of Influencer Emotions, Network Size, and Branding on Consumer Brand Engagement Using Facial Expression and Linguistic Analysis." *Journal of Interactive Marketing* 58 (4): 414–439. https://doi.org/10.1177/10949968231171104.

Jain, P., K. Taneja, and H. Taneja. 2024. "Advertisement Image Classification Using Deep Learning with BERT: A Novel Approach Exploiting Textual." In *Proceedings of Data Analytics and Management: ICDAM 2023*, Vol. 2, 786, 443. London, UK: London Metropolitan University.

Jin, H. S., G. Kerr, and J. Suh. 2019. "Impairment Effects of Creative Ads on Brand Recall for Other Ads." *European Journal of Marketing* 53 (7): 1466–1483. https://doi.org/10.1108/EJM-10-2017-0674.

Jones, N., and A. Hamby. 2023. "It's Written on Your Face: How Emotional Variation in Super Bowl Advertisements Influences Ad Liking." *Journal of Advertising* 53 (3): 463–472. https://doi.org/10.1080/00913367.2023.2218896.

Joo, J., and Z. C. Steinert-Threlkeld. 2022. "Image as Data: Automated Content Analysis for Visual Presentations of Political Actors and Events." *Computational Communication Research* 4 (1): 11–67. https://doi.org/10.5117/CCR2022.1.001.JOO.

Kovashka, A. 2024. "Pitts Ad Dataset. Automatic Understanding of Image and Video Advertisements." University of Pittsburgh. https://people.cs.pitt.edu/~kovashka/ads/.

Kumar, A., A. Kaur, and M. Kumar. 2019. "Face Detection Techniques: A Review." *Artificial Intelligence Review* 52 (2): 927–948. https://doi.org/10.1007/s10462-018-9650-2.

Li, D., B. Wang, Z. Li, N. Yu, and M. Li. 2007. "On Detection of Advertising Images." In *2007 IEEE International Conference on Multimedia and Expo*, 1758–1761. Beijing, China: IEEE. https://doi.org/10.1109/ICME.2007.4285011.

Li, Y., and Y. Xie. 2020. "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement." *Journal of Marketing Research* 57 (1): 1–19. https://doi.org/10.1177/0022243719881113.

Liu, H. 2019. "Going Negative Online?—A Study of Negative Advertising on Social Media." arXiv preprint arXiv:1911.05497.

Liu, L., D. Dzyabura, and N. Mizik. 2020. "Visual Listening in: Extracting Brand Image Portrayed on Social Media." *Marketing Science* 39 (4): 669–686. https://doi.org/10.1287/mksc.2020.1226.

Mameli, M., M. Paolanti, R. Pietrini, G. Pazzaglia, E. Frontoni, and P. Zingaretti. 2022. "Deep Learning Approaches for Fashion Knowledge Extraction from Social Media: A Review." *IEEE Access* 10: 1545–1576. https://doi.org/10.1109/ACCESS.2021.3137893.

Maslowska, E., J. Ohme, and C. M. Segijn. 2021. "Attention to Social Media Ads: The Role of Consumer Recommendations and Smartphones." *Journal of Interactive Advertising*21 (3): 283–296. https://doi.org/10.1080/15252019.2021.1997675.

McQuarrie, E. F., and D. G. Mick. 2003. "The Contribution of Semiotic and Rhetorical Perspectives to the Explanation of Visual Persuasion in Advertising." In *Persuasive Imagery*, edited by L. M. Scott & R. Batra, 191–221. Routledge. https://doi.org/10.4324/9781410609689.

Medium. 2024. "Brand Mark and Logo Recognition: Applications, Challenges, and Cutting-Edge Solutions." May 19. https://medium.com/@API4AI/brand-mark-and-logo-recognition-applications-challenges-and-cutting-edge-solutions-b728902b575c.

Nanne, A. J., M. L. Antheunis, C. G. Van Der Lee, E. O. Postma, S. Wubben, and G. Van Noort. 2020. "The Use of Computer Vision to Analyze Brand-Related User Generated Image Content." *Journal of Interactive Marketing* 50 (1): 156–167. https://doi.org/10.1016/j.intmar.2019.09.003.

N., N., S. Bala, V. Satheeswaran, M. Janani, S. Selvanayaki, and S. Ramasami. 2023. "Deep Learning Based Sentiment Analysis on Images." Paper presented at the 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 568–575. https://doi.org/10.1109/ICECAA58104.2023.10212171.

Otamendi, F. J., and D. L. Sutil Martín. 2020. "The Emotional Effectiveness of Advertisement." *Frontiers in Psychology* 11: 2088. https://doi.org/10.3389/fpsyg.2020.02088.

Paolanti, M., C. Kaiser, R. Schallner, E. Frontoni, and P. Zingaretti. 2017. "Visual and Textual Sentiment Analysis of Brand-Related Social Media Pictures Using Deep Convolutional Neural Networks." Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11–15, 2017, Proceedings, Part I 19.

Patricia, A. L. 2020. "Visual Research Methods: Qualifying and Quantifying the Visual." *Beijing International Review of Education* 2 (1): 35–53.

Philp, M., J. Jacobson, and E. Pancer. 2022. "Predicting Social Media Engagement with Computer Vision: An Examination of Food Marketing on Instagram." *Journal of Business Research* 149: 736–747. https://doi.org/10.1016/j.jbusres.2022.05.078.

Ramesh, A., P. Dhariwal, A. Nichol, C. Chu, and M. Chen. 2022. "Hierarchical Text-Conditional Image Generation with Clip Latents." arXiv Preprint arXiv:2204.06125.

Rietveld, R., W. van Dolen, M. Mazloom, and M. Worring. 2020. "What You Feel, Is What You Like Influence of Message Appeals on Customer Engagement on Instagram." *Journal of Interactive Marketing* 49 (1): 20–53. https://doi.org/10.1016/j.intmar.2019.06.003.

Sabet, A. J., M. Brambilla, and M. Hosseini. 2021. "A Multi-Perspective Approach for Analyzing Long-Running Live Events on Social Media. A Case Study on the "Big Four" International Fashion Weeks." *Online Social Networks and Media* 24: 100140. https://doi.org/10.1016/j.osnem.2021.100140.

Serengil, S. I., and A. Ozpinar. 2020. "Lightface: A Hybrid Deep Face Recognition Framework." In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 1–5. Istanbul, Turkey: IEEE. https://doi.org/10.1109/ASYU50717.2020.9259802.

Serengil, S. I., and A. Ozpinar. 2021. "Hyperextended Lightface: A Facial Attribute Analysis Framework." In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, 1–4. Istanbul, Turkey: IEEE. https://doi.org/10.1109/Iceet53442.2021.9659697.

Shin, D., S. He, G. M. Lee, A. B. Whinston, S. Cetintas, and K. C. Lee. 2020. "Enhancing Social Media Analysis with Visual Data Analytics: A Deep Learning Approach." *MIS Quarterly* 44 (4): 1459–1492. https://doi.org/10.25300/MISQ/2020/14870.

Soleymani, M., D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic. 2017. "A Survey of Multimodal Sentiment Analysis." *Image and Vision Computing* 65: 3–14. https://doi.org/10.1016/j.imavis.2017.08.003.

Srivastava, R. 2021. "Visual Sentiment Exploration of Customer Emotions Using Image Analytics." *International Journal of Business Analytics & Intelligence (IJBAI)* 9: 47–52.

Sutcliffe, A., and A. Namoune. 2008. "Getting the Message Across: Visual Attention, Aesthetic Design and What Users Remember." In *Proceedings of the 7th ACM conference on Designing interactive Systems (DIS '08)*, 11–20. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/1394445.1394447

Sweet, T., A. Rothwell, and X. Luo. 2019. "Machine Learning Techniques for Brand-Influencer Matchmaking on the Instagram Social Network." arXiv preprint arXiv: 1901.05949.

Szegedy, C., S. Ioffe, V. Vanhoucke, and A. Alemi. 2017. "Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning." *Proceedings of the AAAI Conference on Artificial Intelligence* 31 (1). https://doi.org/10.1609/aaai.v31i1.11231.

Talebi, H., and P. Milanfar. 2018. "NIMA: Neural Image Assessment." *IEEE Transactions on Image Processing* 27 (8): 3998–4011. https://doi.org/10.1109/TIP.2018.2831899.

Tarr, A., J. Hwang, and K. Imai. 2023. "Automated Coding of Political Campaign Advertisement Videos: An Empirical Validation Study." *Political Analysis* 31 (4): 554–574. https://doi.org/10.1017/pan.2022.26.

Tous, R., M. Gomez, J. Poveda, L. Cruz, O. Wust, M. Makni, and E. Ayguadé. 2018. "Automated Curation of Brand-Related Social Media Images with Deep Learning." *Multimedia Tools and Applications* 77 (20): 27123–27142. https://doi.org/10.1007/s11042-018-5910-z.

Vassey, J., C. J. Kennedy, H. C. Herbert Chang, A. S. Smith, and J. B. Unger. 2024. "Scalable Surveillance of E-Cigarette Products on Instagram and Tiktok Using Computer Vision." *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco* 26 (5): 552–60. https://doi.org/10.1093/ntr/ntad224.

Villalobos Alva, J. 2021. "Neural Network Framework." In *Beginning Mathematica and Wolfram for Data Science*. Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-6594-9_10.

Villegas, D. S., C. Goanta, and N. Aletras. 2023. "A Multimodal Analysis of Influencer Content on Twitter." arXiv preprint arXiv:2309.03064.

Vo, A. T., H. S. Tran, and T. H. Le. 2017. "Advertisement Image Classification Using Convolutional Neural Network." In *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*, 197–202. Hue, Vietnam: IEEE. https://doi.org/10.1109/KSE.2017.8119458.

Wang, C. Y., A. Bochkovskiy, and H. Y. M. Liao. 2023. "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors." In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7464–7475. Vancouver, BC: IEEE. https://doi.org/10.1109/CVPR52729.2023.00721.

Xu, J., Z. Li, F. Huang, C. Li, and P. S. Yu. 2022. ""Visual Sentiment Analysis With Social Relations-Guided Multiattention Networks." *IEEE Transactions on Cybernetics* 52 (6): 4472–84. https://doi.org/10.1109/TCYB.2020.3027766.

Yang, J., C. Teran, A. F. Battocchio, E. Bertellotti, and S. Wrzesinski. 2021. "Building Brand Authenticity on Social Media: The Impact of Instagram Ad Model Genuineness and Trustworthiness on Perceived Brand Authenticity and Consumer Responses." *Journal of Interactive Advertising* 21 (1): 34–48. https://doi.org/10.1080/15252019.2020.1860168.

Zaheer, A., S. Tahir, M. Humayun, M. F. Almufareh, and N. Z. Jhanjhi. 2022. "A Novel Machine Learning Technique for Fake Smart Watches Advertisement Detection." In *2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, 1–5. Karachi, Pakistan: IEEE. https://doi.org/10.1109/Macs56771.2022.10023151.

Zhang, Y., and T. Yamasaki. 2021. "Style-Aware Image Recommendation for Social Media Marketing." In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21)*, 3106–3114. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3474085.3475453.