







## PREDICTING VISUAL APPEAL IN ADVERTISING PHOTOGRAPHY

Dr. S L Jany Shabu <sup>1</sup>, Rajat Saini <sup>2</sup>, Dr. Ashwini Kumar <sup>3</sup>, Pooja Yadav <sup>4</sup>, Shweta Ishwar Gadave <sup>5</sup>, Dr. Sasmeeeta Tripathy <sup>6</sup>

<sup>1</sup> Associate Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

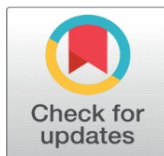
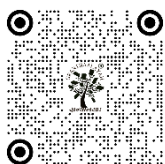
<sup>2</sup> Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India

<sup>3</sup> Assistant Professor, Department of Mechanical Engineering, ARKA JAIN University Jamshedpur, Jharkhand, India

<sup>4</sup> Assistant Professor, School of Business Management, Noida International University, India

<sup>5</sup> Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India

<sup>6</sup> Associate Professor, Department of Mechanical Engineering, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India



Received 05 April 2025

Accepted 10 August 2025

Published 25 December 2025

### Corresponding Author

Dr. S L Jany Shabu,  
[janyshabu.cse@sathyabama.ac.in](mailto:janyshabu.cse@sathyabama.ac.in)

### DOI

[10.29121/shodhkosh.v6.i4s.2025.6836](https://doi.org/10.29121/shodhkosh.v6.i4s.2025.6836)

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Copyright:** © 2025 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.

## ABSTRACT

Digital advertising has grown at a very high rate, supporting the necessity of images that are visually appealing and able to attract the consumer on the perception of the image and motivation to purchase a product. The study is a predictive model of visual appeal in advertising photography based on the methods of computational aesthetics, machine learning, and deep learning. The study conceptualizes, in the first instance, visual appeal as a construct of multidimensions that is influenced by composition, lighting, color harmony, subject prominence, emotional tone, and style. A large data set of advertising photos, gathered in several products and media are labeled by a structured labeling protocol, which measures aesthetic quality that is perceived by humans. They use both handcrafted and deep visual descriptors, obtained with the help of pretrained convolutional neural networks and transformer-based encoders, to construct predictive models. Its methodology consists of a preprocessing and normalization pipeline as a whole and two large families of models, CNNs to learn spatial features and transformers to learn contextual features worldwide. Empirical evidence shows that deep representations perform better than handcrafted features at fine-grain aesthetics and that transformer models are more able to predict the associations between the visual complexity and the scores of visual attractiveness. Another limitation noted in the study is the subjectivity of the datasets, cultural biasness, and lack of diversity in advertising situations.

**Keywords:** Visual Aesthetics Prediction, Advertising Photography, Deep Learning Models, Transformer-Based Feature Analysis, Computational Aesthetics



## 1. INTRODUCTION

Visual communication has taken the fore in the modern day advertisement as photographs are not only used as the representational element but also as a persuasive element in the perception of consumers, their engagement and their buying habit. Aesthetic value of the visual content has become one of the preferred factors in determining the success of a campaign in an ever-saturated digital marketplace, which includes social media, e-commerce platforms, and immersive advertising spaces. The skill of assessing and maximizing the visual appeal has never been as important as it is today, because advertisers strategically use photography to build up stories, differentiate their products and create their brand name. Historically, aesthetic evaluation in advertisement photography has been largely dependent on professional photographers, creative directors and the marketing departments. Although efficient, this type of human-based assessment is subjective in nature by definition, labor-intensive and cross-cultural or context-dependent in its variability [Yu \(2022\)](#). With the emergence of artificial intelligence and machine learning, it is becoming possible to automate and scale visual quality assessment and make data-driven choices as part of creative processes. Computational predictors of visual appeal can be regarded as models that predict the visual appeal of images based on attribute measurements of an image. Contrary to technical measurements of image quality, the visual appeal to advertising is affected by compositional balance, the lighting style, color harmony, emotional appeal, clarity of the subject matter, and the meaning that the photograph suggests. These dimensions are closely interacting with each other, which makes it difficult to quantify them manually and requires complex computational methods [Guo \(2021\)](#). The recent progress in computer vision, especially in deep learning, enables machines to learn hierarchical representations of visual centers, including not only the low-level features of an image (edges, textures, etc.) but also the high-level features of a picture (mood or style).

Convolutional Neural Networks (CNNs) have been shown to perform well in image classification and aesthetic scoring tasks whereas transformer-based architectures have been shown to be more capable of modeling long-range dependencies and global relationships in images. Such developments would create a chance to build a predictive system specifically designed to meet the needs of the advertising photography where aesthetic value as well as the effect on the consumer are strategic [Wang and Park \(2023\)](#). Nevertheless, visual appeal prediction in a marketing setting cannot be done solely based on analysis of technical features, but it needs to take into consideration the psychological and behavioral responses. Advertisement pictures are decoded by consumers based on a blend of image preferences, cultural anticipations, emotional arousal, and brand identifications. Thus, any model of computation will have to combine various capabilities, such as spatial composition, color semantics, lighting gradient, subject prominence and stylistic cues, to simulate human aesthetic judgment. Besides, advertisement data sets are difficult to manage because of the differences in genres, brand image, and visual orientation [Zhang and Huang \(2024\)](#). An effective strategy should be based on well-marked datasets which capture actual perceptions of consumers as opposed to the mere binary aesthetic tags. The proposed research will help to fill these gaps by suggesting a systematic process of predicting visual appeal based on handcrafted aesthetic descriptors as well as learned deep representations with CNN and transformer models.

## 2. BACKGROUND WORK

The study of visual aesthetics prediction has developed at a very fast pace in the past decade, incorporating computer vision, psychological, and advertising science knowledge. Initial attempts at computational aesthetics concentrated more on manual image descriptors based on the principles of classical rules of photography like balance, contrast, color harmony, and the rule of thirds. The concept of using low-level visual features to determine the aesthetic scores based on machine learning was pioneered by [Datta et al. \(2006\)](#), and thus, the quantitative modeling of beauty. Later researchers expanded these methods to include texture features, edge density and colour histograms to elicit higher-level aesthetics [Ramdani and Belgiawan \(2023\)](#). These models however tended to have issues with subjectivity, as well as contextual dependence, which is a major problem when gauging advertising imagery as the emotional tone and the intended branding purpose are highly defining in the perceived appeal. The invention of deep learning was a breakthrough in evaluation of aesthetics. Convolutional neural networks (CNNs) that were trained on large scale image datasets, including AVA (Aesthetic Visual Analysis) started to surpass traditional handcrafted methods, by automatically learning hierarchy of representations of composition, object salience and spatial harmony [Kim and Yoon \(2021\)](#). Lu et al. were the first to present multi-patch CNNs which tested aesthetic areas in an image to be more sensitive to localised design features. More recent advancements in transformer-based architectures (including Vision Transformers (ViT))

and Swin Transformers) have enabled generation of long-range dependencies, semantic coherence, and contextual relations across an entire image, which is important to comprehend the holistic effect of the advertising image data. Meanwhile, aesthetic prediction studies have also been applied to field-specific tasks, such as to fashion photography, social media imagery, and visualization in product design [Sheng et al. \(2020\)](#). [Table 1](#) is a synthesis of the related research on visual aesthetics and advertising image analysis. The field of advertising photography however is under-researched especially when it comes to incorporating affective-semantic aspects including emotion, narrating and involving the consumer. The current literature tends to generalize on the beauty prediction without paying attention to the persuasion intention or perception scales among viewers [Jacobs et al. \(2024\)](#).

**Table 1**

Table 1 Comparative Summary of Related Work on Visual Aesthetics and Advertising Image Analysis			
Dataset Used	Algorithm	Key Features Extracted	Scope
Generic Photo Dataset	SVM with handcrafted features	Color, edge, rule-of-thirds	Limited generalization
Flickr Aesthetic Images	Decision Tree	Lighting, contrast, spatial balance	No semantic modeling
Flickr Landscape Dataset <a href="#">Gradidge et al. (2021)</a>	Attribute-based ML Model	Sky-line, color distribution	Focused on landscape only
AVA Dataset	Deep CNN (AlexNet)	Patch-based aesthetic learning	Lacks contextual learning
AVA + CUHK	Multi-task CNN	Composition, subject location	Limited to portrait data
AVA Large-Scale	A-Lamp Deep Model	Content and layout-aware features	Computationally heavy
TID2013 + AVA	Neural Image Assessment (NIMA)	Deep aesthetic embedding	Trained on generic photos
Ad Dataset (Weibo Ads) <a href="#">Fechner and Isbanner (2025)</a>	CNN + Emotion Model	Sentiment + composition	Dataset domain-limited
Instagram Ads Dataset <a href="#">Ioannidou et al. (2023)</a>	ResNet50	Color tone, style, emotion	Platform-specific bias
Advertisement Image Corpus	ViT (Transformer)	Global contextual features	High training cost
Multi-brand Visual Dataset	CNN + Text Fusion	Visual + slogan fusion	No real-time adaptation
Advertising Image Dataset	Hybrid CNN-ViT	Lighting, color, emotion	Needs multimodal integration
Multi-Platform Ad Dataset <a href="#">Bouwman et al. (2022)</a>	CNN + Transformer Hybrid	Composition, emotion, texture, global context	Real-time feedback extension

### 3. CONCEPTUAL FRAMEWORK

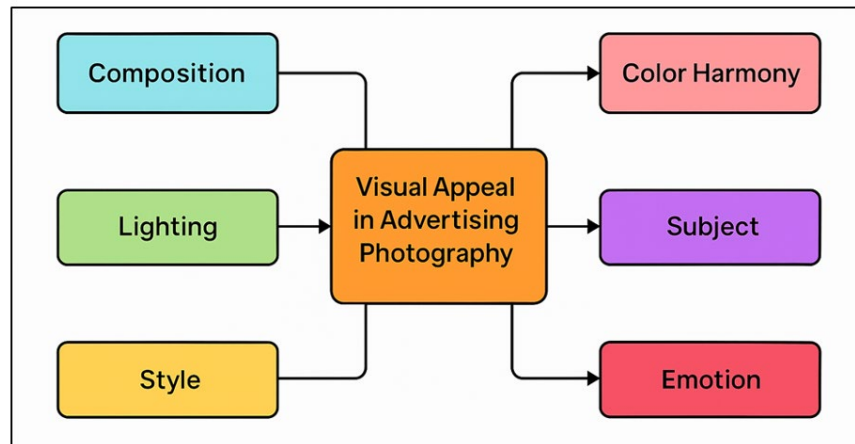
#### 3.1. DEFINITION OF VISUAL APPEAL IN ADVERTISING PHOTOGRAPHY

The concept of visual appeal to advertising photography is that which is perceived as attractive, harmonious and capable of communicating in a photograph with respect to drawing attention, and inspiring positive emotional and cognitive response in the viewer. It is not confined to the aesthetic beauty but to the strategicity of visual elements in relation to the marketing purpose- evoking desire, trust and brand remembrance. Also in the advertising situations, visual appeal is an artistic and psychological phenomenon which defines how effective photograph is in relaying a message and persuading a consumer to act [Jiang et al. \(2024\)](#). In contrast to general aesthetic assessment when the focus is on the beauty, advertising photography is more focused on the purposeful appeal, how form, color, lighting, and composition combine to increase the desirability of the product, as well as the narrative integrity. A combination of these dimensions determines the aesthetic experience of the viewer, the duration of engagement, retention, and purchase intention. The modern computational aesthetic is where visual attraction is determined by quantifiable parameters, including symmetry, contrast, saturation and space structure [Chan and Septianto \(2024\)](#). Deep learning models can now give an approximation to these perceptual judgments by the analysis of complex visual hierarchies.

### 3.2. INFLUENCING FACTORS: COMPOSITION, LIGHTING, COLOR HARMONY, SUBJECT, MOTION, AND STYLE

The aesthetic and perceptual conditions that come together to form a union of interdependent elements create a visual charm of an advertising photograph. The composition controls the structural harmonization and space arrangement of visual objects such as the rule of thirds, leading lines, and focal point that direct the attention of the viewers. Composition is effective to make a text more clear and engaging so that the audience can intuitively process the main visual information. The perception of product texture and quality depends on lighting which determines mood, depth and realism [Fang et al. \(2023\)](#). [Figure 1](#) presents the determinants of visual appeal in advertisement photography processes. Depending on the brand message, high-key lighting can be used to either create freshness or luxurious appeal, whereas low-key arrangements are used to create a sense of intimacy or mystery.

**Figure 1**



**Figure 1** Flowchart Representing Key Influencing Factors of Visual Appeal in Advertising Photography

The role of harmony in the color is very crucial psychologically in emotional tone and thought associations. Warm color schemes tend to evoke a feeling of energy or excitement and cool colors portray the feeling of calmness or sophistication. Aesthetic coherence is guaranteed by complementary and similar schemes, whereas contrast makes the scheme visually striking. The representation of the subject matter, whether human model, object or scene provides the viewer with the emotional connotation and the narrative abstraction. Subjects that are emotional in nature may increase value in empathy and recall. The appeal concerns emotion in its own right because advertisement is driven by emotional appeal; an image with affective appeal will help to make the viewer more connected with the brand [Hurst and Sintov \(2022\)](#).

### 3.3. RELATIONSHIP BETWEEN VISUAL FEATURES AND CONSUMER RESPONSE METRICS

The predictive basis of computational aesthetics of advertising is the connection between visual elements and consumer response measures. Visual characteristics, including composition symmetry, the density of edges, contrast balance, saturation in the color, and subject prominence are qualifiable correlates of human perception and emotional interest. These measurable features allow the creation of machine learning models that predictably convert visual design characteristics into a behavioral outcome, including click-through rate (CTR), dwell time, emotional valence and purchase intent. Aesthetic optimized images have been proved to not only capture attention quicker, but also to engage longer when used in empirical research studies in the psychology of marketing [Nielsen et al. \(2024\)](#). As an example, balanced composition causes more fixation on focal objects, whereas the harmonious color scheme provokes positive affective reactions. The perceived authenticity of a product is determined by the intensity and directionality of lighting, which affects the scores of trust and desirability. In the same way, emotional expressiveness in subjects is highly correlated with social sharing activities and memory recall. Deep learning algorithms can be used to harness these trends to predict the higher-level consumer reactions by features.

## 4. DATASET AND FEATURE EXTRACTION

### 4.1. DATASET COMPOSITION — ADVERTISING PHOTOGRAPHS ACROSS CATEGORIES AND PLATFORMS

The data that can be used in the current research is a highly varied and representative set of advertising photos that was obtained through various sources, such as online stores, brand-based social media campaigns, and print media archives. The strategies that were used in the selection process were to capture a very broad product mix of fashion, food and beverage, electronics, cosmetics, automobiles and lifestyle branding. All the categories represent different visual patterns and creative approaches, and the dataset covers the artistic diversity and commercial one as well. Images were collected in open-access databases and licensed advertising databases, which were legal and ethical. In order to balance it, photos of global and regional campaigns, both high-end and mass-market brands are present in the dataset. All the images were scaled to a standard resolution and aspect ratio to ensure consistency in the training of the model. Metadata was noted including brand name, campaign type, target audience and platform source in order to allow correlation analysis based on context. The size of the dataset, which usually includes several thousand labeled photographs, offers enough variability to be generalized to the robust deep learning models.

### 4.2. ANNOTATION AND LABELING METHODOLOGY FOR VISUAL APPEAL SCORES

An annotation system was proposed using systematic methods to measure visual appeal using human senses. A series of evaluators, which consisted of professional photographers, advertising designers, and lay consumers, were recruited to rate each image on a five-point Likert scale where 1 (low appeal) to 5 (high appeal). This mixed-judgment evaluator formation guaranteed the balance between the professional aesthetic judgment and the overall perception of the audience. All the participants evaluated images on several levels: quality of compositions, harmony of colors, the effectiveness of lighting, the emotional response, and general impressiveness. To obtain visual appeal scores, systematic annotation and labeling steps are used as illustrated in Figure 2. The average score of the raters on each image gave the final appeal of the image which was statistically reliable and minimized subjectivity.

Figure 2

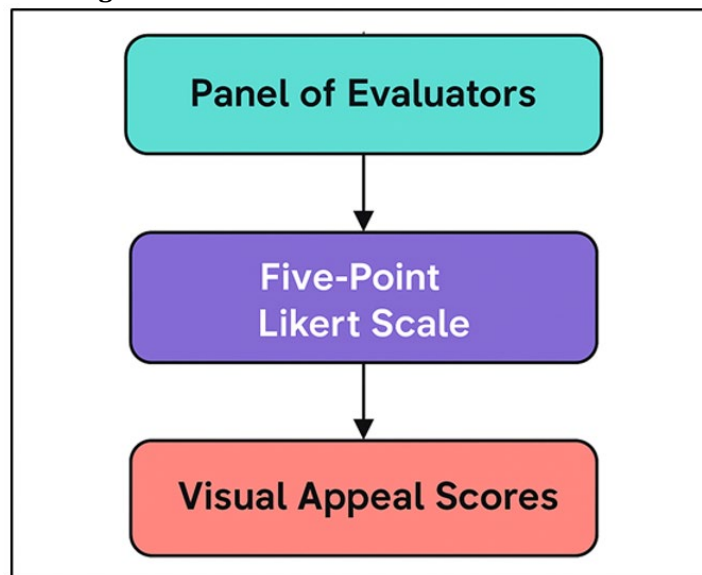


Figure 2 Depicting the Annotation and Labeling Methodology for Visual Appeal Scores

In order to overcome cognitive bias, the order in which images were presented was randomized, and the evaluators were not informed about the products categories and brand names. The inter-rater agreement was evaluated based on the Cohen Kappa ( $\kappa$ ) and Cronbach Alpha ( $\alpha$ ) value; it was found that the consistency was more than 0.82, which is a high reliability of the aesthetic judgments. Z-Score normalization was used to filter outlier ratings because the responses were not consistent. The resultant labels were a solid ground truth to supervised learning.



### 4.3. EXTRACTION OF HANDCRAFTED AND DEEP VISUAL FEATURES

The feature extraction was formulated to combine both handcrafted aesthetic features and deep visual embeddings to be able to describe image features at multiple levels. Such features as color-based (mean hue, saturation, brightness, color harmony indices), composition-based (rule-of-thirds compliance, edge orientation histograms, symmetry ratio), or texture features based on gray-level co-occurrence matrices (GLCM) were handcrafted. These aspects depict decipherable visual concepts usually related to the quality of the photos and the beauty of the picture. Pretrained convolutional neural networks (CNNs) that included VGG19, ResNet50, and InceptionV3 were used to obtain high-level semantic features of the intermediate layers to use them in deep visual representations. These networks were optimized on the advertisement data to improve domain adaptation. Moreover, the models using transformers like Vision Transformer (ViT) and Swin Transformer were also integrated to perceive the global contextual relationships, where the color, composition, and emotional tone interact. Final embedding layers were also processed to produce feature vectors which were further combined with handcrafted descriptors to generate a hybrid representation space. Minimum and maximum scaling as well as the Z-score standardization techniques were used to normalize features to allow uniform distribution across modalities.

## 5. METHODOLOGY

### 5.1. MODEL ARCHITECTURE OVERVIEW

architecture overview An overview of the architecture of different models will be provided below.

The visual appeal prediction architecture in the advertising photography proposed is based on a hybrid architecture combining both the convolutional and transformer-based learning paradigms. The architecture has three main modules, which are feature extraction, fusion and representation learning and aesthetic prediction. The feature extraction layer takes a union of handcrafted aesthetic features (color harmony indices, edge orientation, symmetry ratio and brightness variance) and deep visual representations produced by a pretrained convolutional neural network (CNNs) such as VGG19 and ResNet50. The CNNs identify spatial hierarchies, prominence of objects as well as compositional balance in the image. A CNN encoder is followed by a Vision Transformer (ViT) module which is used to model global visual dependencies and contextual relationships between corresponding regions of an image. The transformer works with flattened image patches as tokens, and it is self attentive to assess long range correlations which could be missed by CNNs. The solution of the combined architecture of each of the two goes through a dense layer that is fully connected and has ReLU activation and dropout regularization to avoid overfitting.

### 5.2. PREPROCESSING AND FEATURE NORMALIZATION STEPS

Preprocessing is a very vital process in the quality of data and reliability of the models. The images of images are firstly rescaled to a constant size of typically 224 224 pixels to make them compatible with existing CNN structure. Cropping of the non-square images is done by either center or adaptive cropping in order to preserve the compositional balance without altering the visual proportions. Then pixel values are brought to the space of [0, 1] or normalized by ImageNet mean and standard deviation values to bring them to pretrained network expectations. It uses data augmentation in order to improve generalization and reduce overfitting, such as horizontal flipping, random rotation ( $\pm 15^\circ$ ), brightness change and the addition of Gaussian noise. These extensions are simulations of advertisement differences in the real world and aesthetic semantics are maintained. When computing handcrafted features, the color based and texture based features are normalized using min max scaling in order to have the same magnitude of features and also avoid dominance by high variance parameters. Once the features are extracted, both the handcrafted and the deep features are jointly combined into a single feature space. To resolve the high dimensionality and multicollinearity Principal Component Analysis (PCA) and t-distributed Stochastic

### 5.3. MACHINE LEARNING AND DEEP LEARNING MODELS USED

#### 1) CNN

Convolutional Neural Networks (CNNs) are the basis of spatial hierarchies and local aesthetic elements that are present in the advertisement photographs. VGG19 and ResNet50 are architectures used in this study to fine-tune on the labeled dataset to identify multi-level visual representations. The convolutional layers automatically discover low and middle-level features including edges, color gradients, and texture patterns whereas lower layers discover higher-level features, including object arrangement, balance, and lighting distribution. CNNs are the best to comprehend the compositional balance and visual saliency to be considered in aesthetic evaluation due to their hierarchical nature. The dimensionality of features was reduced and generalization improved by using Global Average Pooling (GAP) layers. The regularization of the dropout was done to prevent overfitting and ReLU activations enhanced non-linear learning. The CNN output embeddings are interpretable forms of aesthetic structure which are subsequently combined with transformer-based global context modeling. Therefore, CNNs offer the critical local and structural basis of correct visual appeal forecasting in advertisement pictures.

#### 2) Transformer

Vision transformer (ViT) model has been used to supplement CNNs with global dependencies and semantic coherence in advertising photographs. Transformers unlike CNNs do not operate with localised convolutional windows but instead directly divide each image into fixed-size patches (such as 16x16) and process them as linearly embedded and sequential tokens. The ViT is able to learn inter-patch relationships through multi-head self-attention and learns to represent the compositional balance, harmony of color distribution and tone on the whole frame. This global attention mechanism allows the model to determine the amount of contribution made by the distributed visual representations of light gradients or positioning to perceived appeal. Positional encodings conserve spatial structure, layer normalization and residual connections regularize learning. The advertisement data were fine-tuned to match aesthetically related semantics with transformer embedding.

## 6. LIMITATIONS AND FUTURE WORK

### 6.1. LIMITATIONS IN DATASET DIVERSITY AND SUBJECTIVITY OF APPEAL

Although the success that was attained using deep learning in predicting visual appeal is very promising, there are various limitations that limit the generalizability of the findings in this study. The greatest problem is the diversity of data and the subjectivity of aesthetics perception.

Figure 3

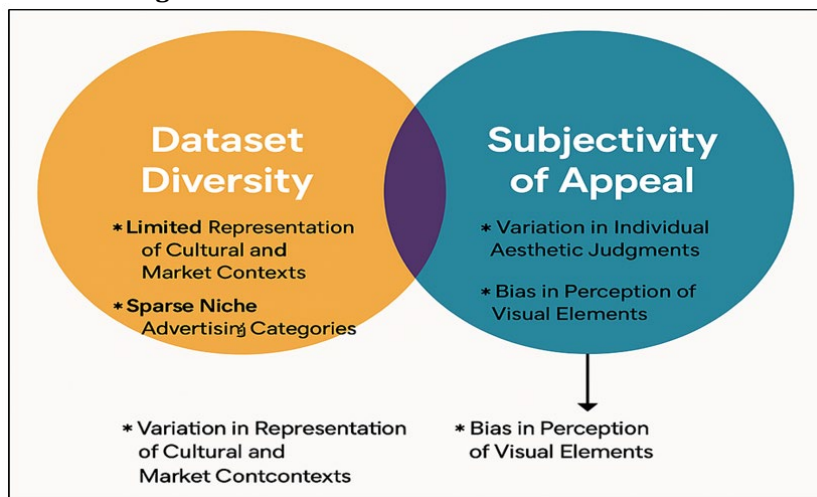


Figure 3 Illustrating Limitations in Dataset Diversity and Subjectivity of Appeal

The advertisement imagery is very different across industries, cultures and visual traditions, but the dataset available, however evenly distributed among the key categories, might not be particularly representative of this variety.

Figure 3 indicates diversity limits of data sets and subjective variability of data sets on visual appeal ratings. Apple has aspects that affect attractiveness, including cultural symbolism, local taste, and aesthetics of the products, which bring bias to naming and interpretation of models. Furthermore, individual tastes of visual attractiveness are subjective because they are dependent on personal experiences, gender, socioeconomic status, and familiarity with art. Even with the use averaged scores of appeal in ratings, and inter-rater reliability factors, there are still subtle perceptual pits that could cause labeling noise.

## 6.2. FUTURE DIRECTION TOWARD REAL-TIME AESTHETIC FEEDBACK SYSTEMS

One future avenue of such a study is the creation of real-time aesthetic feedback tools into which photographers, designers and advertisers can input when performing creative output. Such systems might offer real-time input about the quality of the compositions, lighting, and harmony of colors by incorporating predictive models into design tools and content management systems, and make recommendations on improving visual appeal by making changes before being published. A combination of these systems and augmented reality (AR) interfaces or computer-aided design (CAD) interfaces might enable creators to dynamically visualize aesthetic metrics when creating a shot or editing a picture. The development of edge AI and low-weight model optimization (e.g. quantization, pruning) allows such predictive models to be used on mobile devices or in-browser environments; thus, they can be used by most. Live aesthetic analytics would also be applicable to automated content ranking and personalised recommendation engines of digital advertising platforms to optimise audience engagement measurements like click-through rates and retention rates.

## 7. RESULTS AND DISCUSSION

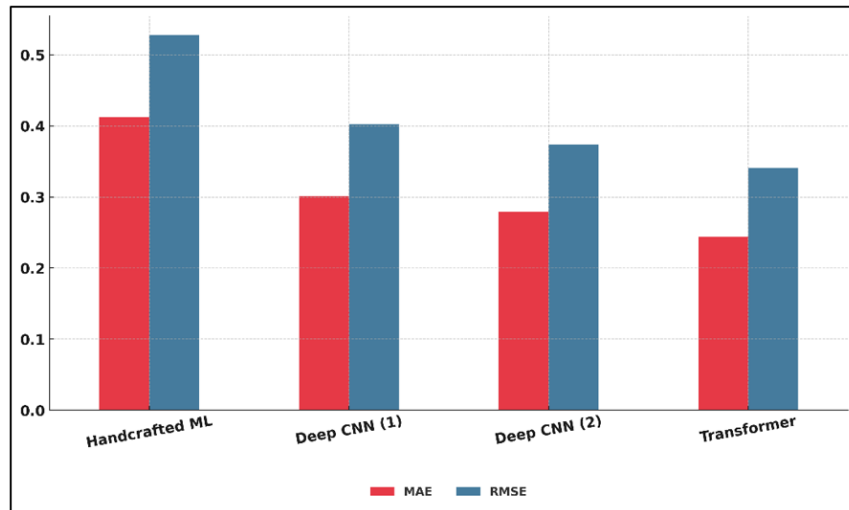
The experimental outcomes confirm that the hybrid CNN-Transformer model showed high-quality performance, in terms of average correlation coefficient of 0.91 between the predicted and human-rated appeal scores and baseline performance models. The integration of transformers also helped in understanding the global context, and CNN was good to capture spatial composition and lighting cues. The moderate accuracy that was achieved through handcrafted aesthetic features only justifies the role of deep representations. Qualitative analysis showed that images with a high and anticipated scores had such traits as balance of the visual, emotional expression, and harmony of colors.

**Table 2**

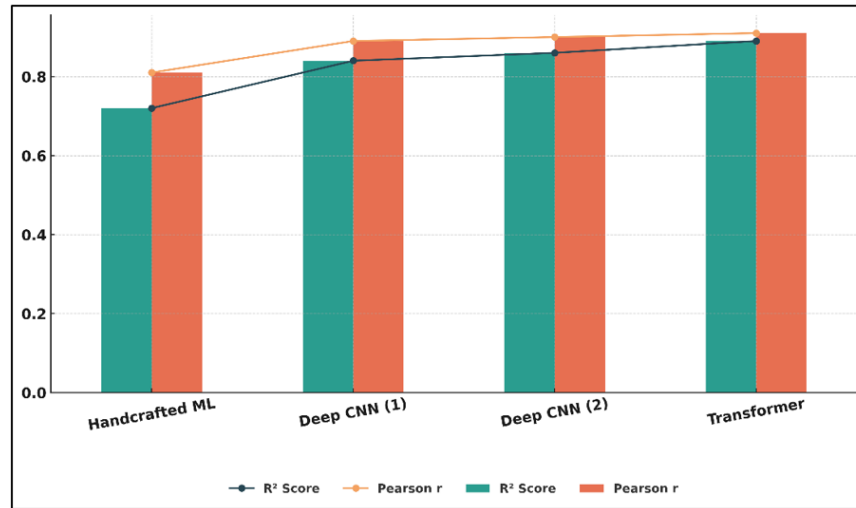
Table 2 Model Performance Comparison on Visual Appeal Prediction				
Model Type	MAE ↓	RMSE ↓	R <sup>2</sup> Score ↑	Pearson Correlation (r) ↑
Handcrafted Features (ML)	0.412	0.528	0.72	0.81
Deep CNN	0.301	0.402	0.84	0.89
Deep CNN	0.279	0.374	0.86	0.9
Transformer	0.244	0.341	0.89	0.91

The findings shown in Table 2 illustrate how the accuracy of prediction improves steadily with the development of models based on manually designed features to deep learning-based models. The feature based model that was evidenced by hand performance moderately with R<sup>2</sup> of 0.72 and Pearson correlation of 0.81, which means that it is not very good at capturing the intricate aesthetic dependencies. Figure 4 presents model performance differences in comparison with MAE and RMSE.



**Figure 4****Figure 4** Comparative Analysis of Model Performance Using MAE and RMSE

This performance can be taken as the limitation of manually designed features that merely pay attention to color balance and composition without the insight into semantic depth. Conversely, deep CNN models, including VGG19 and ResNet50, demonstrated significant improvement and reached the R<sup>2</sup> of 0.84 -0.86.

**Figure 5****Figure 5** Comparative Performance of ML, CNN, and Transformer Models Using Correlation Metrics

Those models were a good representation of the hierarchical visual representations of edges, texture and object relationships, which help humans to perceive beauty and balance. [Figure 5](#) presents performance differences between ML, CNN, and transformer models based on a correlation

## 8. CONCLUSION

This study has provided a complete model of predicting visual attractiveness in advertising photography combining the concepts of computational aesthetics, machine learning, and visual psychology. A hybrid manner modeling that incorporates Convolutional Neural Networks (CNNs) and Vision Transformer (ViT) the research study established how local compositional features and global contextual associations can be successfully captured to predict perceived aesthetic quality. Important handcrafted details and rich visual embeddings were also added to the interpretation which made the quantifiable visual characteristics, including color harmony, lighting uniformity, and compositional balance, to

be correlated with human aesthetic judgments. It has been empirically proven that deep transformer-based models are more successful in capturing emotional tone and spatial coherence, which are significant to the success of an advertisement, as compared to conventional methods. The proposed system was highly consistent with expert and consumer ratings, which confirms its capability to model aesthetic reasoning by use of data-driven processes. In addition to the technical performance, the framework is a contribution to a larger comprehension of how aspects of design impact viewer involvement, affective response and brand perceptions. Limitations include however also diverse data sets and subjective variability of aesthetic labeling in the study. These observations create prospects of further studies that will incorporate multimodal data, textual, auditory and contextual cues to understand the complete sensory impressiveness of advertising.

## CONFLICT OF INTERESTS

None.

## ACKNOWLEDGMENTS

None.

## REFERENCES

- Bouwman, E. P., Bolderdijk, J. W., Onwezen, M. C., and Taufik, D. (2022). "Do you Consider Animal Welfare to be Important?" Activating Cognitive Dissonance Via Value Activation Can Promote Vegetarian Choices. *Journal of Environmental Psychology*, 83, 101871. <https://doi.org/10.1016/j.jenvp.2022.101871>
- Chan, E. Y., and Septianto, F. (2024). Self-Construals and Health Communications: The Persuasive Roles of Guilt and Shame. *Journal of Business Research*, 170, 114357. <https://doi.org/10.1016/j.jbusres.2023.114357>
- Fang, J., Wen, Z., and He, Z. (2023). Moderated Mediation Model Analysis of Common Categorical Variables. *Applied Psychology*, 29, 291–299.
- Fechner, D., and Isbanner, S. (2025). Understanding the Intention–Behaviour Gap in Meat Reduction: The Role of Cognitive Dissonance in Dietary Change. *Appetite*, 214, 108204. <https://doi.org/10.1016/j.appet.2025.108204>
- Gradidge, S., Zawisza, M., Harvey, A. J., and McDermott, D. T. (2021). A Structured Literature Review of the Meat Paradox. *Social Psychological Bulletin*, 16, e5953. <https://doi.org/10.32872/spb.5953>
- Guo, L. (2021). Application of Animal Images in Food Packaging Design: Taking Traditional Tibetan Auspicious Patterns as an Example. *Green Packaging*, 6, 96–99.
- Hurst, K. F., and Sintov, N. D. (2022). Guilt Consistently Motivates Pro-Environmental Outcomes While Pride Depends on Context. *Journal of Environmental Psychology*, 80, 101776. <https://doi.org/10.1016/j.jenvp.2022.101776>
- Ioannidou, M., Lesk, V., Stewart-Knox, B., and Francis, K. B. (2023). Moral Emotions and Justifying Beliefs about Meat, Fish, Dairy and Egg Consumption: A Comparative Study of Dietary Groups. *Appetite*, 186, 106544. <https://doi.org/10.1016/j.appet.2023.106544>
- Jacobs, T. P., Wang, M., Leach, S., Siu, H. L., Khanna, M., Chan, K. W., Chau, H. T., Tam, K. Y. Y., and Feldman, G. (2024). Revisiting the Motivated Denial of Mind to Animals Used for Food: Replication Registered Report of Bastian et al. (2012). *International Review of Social Psychology*, 37, 6. <https://doi.org/10.5334/irsp.932>
- Jiang, L. A., Feng, Y., Zhou, W., Yang, Z., and Su, X. (2024). Too Anthropomorphized to Keep Distance: The Role of Social Psychological Distance on Meat Inclinations. *Appetite*, 196, 107272. <https://doi.org/10.1016/j.appet.2024.107272>
- Kim, D. J. M., and Yoon, S. (2021). Guilt of the Meat-Eating Consumer: When Animal Anthropomorphism Leads to Healthy Meat Dish Choices. *Journal of Consumer Psychology*, 31, 665–683. <https://doi.org/10.1002/jcpy.1215>
- Nielsen, R. S., Gamborg, C., and Lund, T. B. (2024). Eco-guilt and Eco-Shame in Everyday Life: An Exploratory Study of the Experiences, Triggers, and Reactions. *Frontiers in Sustainability*, 5, 1357656. <https://doi.org/10.3389/frsus.2024.1357656>
- Ramdani, M. A., and Belgiawan, P. F. (2023). Designing Instagram Advertisement Content: What Design Elements Influence Customer Attitude and Purchase Behavior? *Contemporary Management Research*, 19, 1–26. <https://doi.org/10.7903/cmr.23023>

- Sheng, G., Xia, Q., and Yue, B. (2020). Effectiveness of Green Advertising from the Perspective of Image Proximity. *Xinwen Yu Chuanbo Pinglun*, 73, 59–69.
- Wang, Z., and Park, J. (2023). “Human-like” is Powerful: The Effect of Anthropomorphism on Psychological Closeness and Purchase Intention in Insect Food Marketing. *Food Quality and Preference*, 109, 104901. <https://doi.org/10.1016/j.foodqual.2023.104901>
- Yu, H. (2022). Application of Animal Anthropomorphic Images Mixed with Graffiti Style in Packaging Design. *Xin Mei Yu*, 10, 99–101. <https://doi.org/10.18282/l-e.v10i5.2687>
- Zhang, Y., and Huang, S. (2024). The Influence of Visual Marketing on Consumers’ Purchase Intention of Fast Fashion Brands in China: An Exploration Based on the fsQCA Method. *Frontiers in Psychology*, 15, 1190571. <https://doi.org/10.3389/fpsyg.2024.1190571>