Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
January-June 2026 7(1), 48–59

IMAGE AESTHETICS EVALUATION THROUGH AI ALGORITHMS

Pournima Pande [1], Dr. Ganesh Ramkrishna Rahate [2], Dr. Ashok Rajaram Suryawanshi [3], Dr. Anil Laxmanrao Wakekar [4], Suraj Rajesh Karpe [5], Swati Varma [6]

[1] Assistant Professor, Department of Applied Chemistry, Yeshwantrao Chavan College of Engineering, Nagpur, India
[2] Department of Electronics and Telecommunication Engineering, Pune, India
[3] Pimpri Chinchwad College of Engineering, Department of Electronics and Telecommunication Engineering, India
[4] Principal, Devi Mahalaxmi College of Engineering and Technology, Titwala, Ta. Kalyan, Dist. Thane, Maharashtra, India
[5] Department of Electrical Engineering, CSMSS Chh. Shahu College of Engineering, Chhatrapati Sambhajinagar, Maharashtra, India
[6] Department of Computer Engineering, Vidyavardhini's College of Engineering and Technology, Vasai, Maharashtra, India

Received 14 November 2025
Accepted 16 December 2025
Published 15 January 2026

Corresponding Author
Pournima Pande,
Pournimapande5@gmail.com

DOI
10.29121/shodhkosh.v7.i1.2026.7010

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

ABSTRACT

The evaluation of image aesthetics through automation has become a significant research issue because of the fact that social media, photography and creative industries are expanding at a very fast rate compared to digital imagery. Contrary to typical vision exercises, aesthetic evaluation is multi-dimensional, subjective by nature, and perceptual, semantic, and emotional. In this paper, the analysis of image aesthetics evaluation based on artificial intelligence algorithms is carried out in detail, starting with the traditional methods of the evaluation based on the handwritten feature and concluding with the advanced deep learning ones. We compare convolutional and vision convolutional neural networks, as well as hybrid networks, and point out their advantages in the local visual quality of modeling and the global compositional structure. In order to overcome the inconsistency in the human judgment, the research focuses on subjectivity-sensitive learning, using the distribution-based annotation and the model of the pairwise preferences. An aesthetic scoring system based on the combination of regression, probabilistic distribution learning and ranking objectives is addressed in terms of implementation on the system level. Through experimental analysis and discussion, it has been shown that hybrid models are stronger, easier to interpret and closer to human perception. The results prove the relevance of holistic learning of features, uncertainty modeling, and explainable decision-making to reliable and human-congruent aesthetic evaluation systems.

Keywords: Artificial Intelligence, Deep Learning, Convolutional Neural Networks, Vision Transformers, Subjectivity Modeling, Aesthetic Attributes, Preference Learning

1. INTRODUCTION

The explosion of digital images on social media, mobile photography, and e-commerce, as well as in the creative sector, has further heightened the need of automated ways on how image aesthetics can be assessed in a consistent and scalable way. Image aesthetics assessment can be defined as the computational method of assessing visual appeal, which includes aspects of composition, color harmony, lighting, balance, and semantic relevance Zeng et al. (2020).

Conventionally, aesthetic judgment has been treated as the subjective one relying on human senses, the culture and on the experience of the individual. Nonetheless, the rapid increase in the number of visual data has made the evaluation of the visual data by using human evaluation irrelevant and it has inspired the creation of algorithmic solutions capable of estimating human taste with reasonable accuracy and uniformity He et al. (2022).

Figure 1

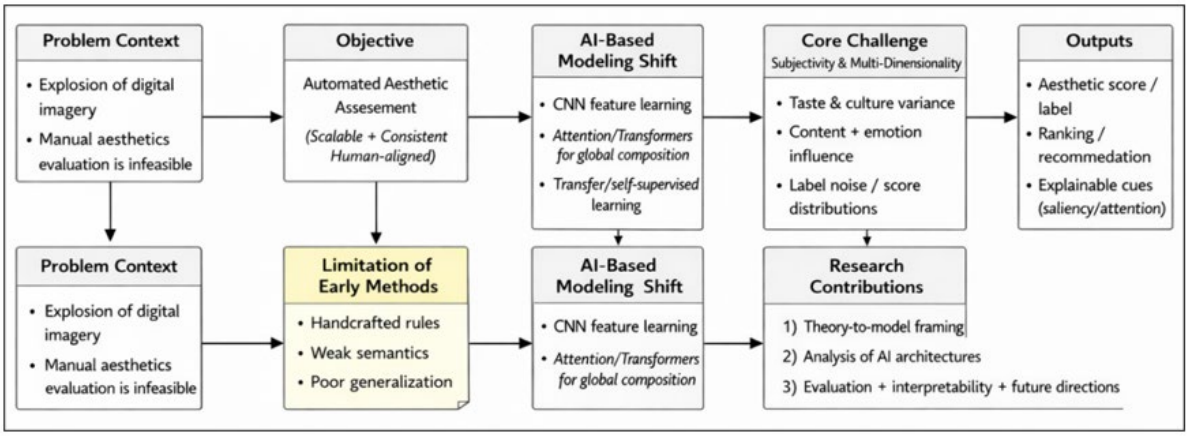

Figure 1 Conceptual Framework of AI-Based Image Aesthetics Evaluation

Initial computational methods of aesthetic judgement had to use hand drawn features based on photographic principles and low-level visual statistics, including color histograms, edge distributions, and compositional rules of thumb Luo (2023). Although these approaches were useful in offering preliminary understanding of aesthetic modeling, their results were poor due to poor ability to generalize and failure to recognize high level semantic and contextual information. This has been completely changed by the introduction of deep learning which has made it possible to end-to-end learn aesthetic representations using data as depicted in Figure 1. Convolutional neural networks have been shown to be highly effective in learning hierarchical visual features, whereas attention-based models have also enhanced the modeling of global composition and long-range associations that apply to the aesthetic perception Chen et al. (2020). Irrespective of these developments, the process of evaluation of image aesthetics is a difficult issue because aesthetic evaluation is a subjective and multi-dimensional process. The subjectivity of individual taste and cultural influences as well as situational purpose give a lot of uncertainty to labeling procedures and assessment regimens Dosovitskiy et al. (2020). Furthermore, the aesthetic quality is not always simply a visual property but it is frequently shaped by semantic information and the emotional appeal, and it is necessary to have models that incorporate both perceptual and cognitive as well as affective indications. These issues imply the need to have strong algorithmic architectures, well-selected data, and metrics that are highly consistent with human judgment.

The purpose of the present paper is to present a systematic inquiry into evaluating aesthetics of images with the help of artificial intelligence algorithms Mehta and Rastegari (2021). The main contributions of this work can be three-fold: first, the conceptual framework that connects aesthetic theory to modern AI-based modeling models is created; second, the current state-of-the-art deep learning architectures and learning strategies that are used to evaluate aesthetics are analyzed; and, third, the evaluation methodologies, interpretability, and future directions of research. Through these viewpoints, the paper aims at contributing to the knowledge of the aesthetic intelligence in computer vision and contribute to the creation of more dependable, transparent, and human-oriented aesthetic evaluation systems.

2. FOUNDATIONS OF IMAGE AESTHETICS

The aesthetics of images is based on the theories of interdiscipline including visual psychology, art theory, photography and cognitive science. Fundamentally, aesthetic perception is a way of human thought and emotional response to visual objects depending on the low-level sensory perception and the high-level cognitive functions Li et al. (2020). It is important to understand these underpinnings to build artificial intelligence models that seek to estimate human aesthetic judgment in a principled and understandable way. The low-level visual properties of color, luminance,

contrast, texture and spatial frequency are put into the spotlight by the early theories of aesthetics, as far as perception is concerned. The Gestalt principles burdened with balance, symmetry, proximity, and figure ground separation are explanations of how viewers structure visual outlay into coherent structures Celona et al. (2021). These principles of perception have been operationalized in visual design and photography using compositional rules, which include the rule of thirds, leading lines, framing, cues of depth, color harmony, etc. Although these guidelines are not dictatorial rules of beauty, they make available a system of terms with which visual order and aesthetic unity can be articulated. These aesthetic human principles should be mapped into quantifiable forms so that they can be processed by algorithms to facilitate computational modeling Horanyi et al. (2022). Table 1 demonstrates a conceptual correspondence between fundamental aesthetic concepts, their perceptual interpretations, and commonly known representation of computational features. This mapping is a conceptualization between aesthetic theory and algorithmic application, emphasising the process of approximating subjective judgments of the visual judgement by quantifiable visual descriptors.

Table 1

Table 1 Mapping of Aesthetic Principles to Computational Features for Image Aesthetics Evaluation

Aesthetic Principle	Human-Centered Interpretation	Computational Feature Representation	Relevance to AI Models
Composition and Balance	Harmonious spatial arrangement of visual elements	Rule-of-thirds grids, saliency maps, object centroid distribution, symmetry measures	Enables learning of global layout and spatial aesthetics
Color Harmony	Pleasing color combinations and tonal consistency	Color histograms, hue–saturation distributions, contrast ratios	Supports mood and stylistic coherence modeling
Lighting and Exposure	Appropriate illumination enhancing clarity and depth	Luminance statistics, exposure histograms, shadow–highlight balance	Distinguishes well-lit images from technical artifacts
Sharpness and Clarity	Perceived focus quality and visual precision	Edge density, Laplacian variance, blur metrics	Separates technical quality from aesthetic appeal
Texture & Detail	Surface richness and visual complexity	Gabor filters, local binary patterns, wavelet features	Captures material quality and fine-grained structure
Depth and Perspective	Sense of spatial realism and immersion	Vanishing point detection, scale variation, depth cues	Supports aesthetic realism assessment
Semantic Content	Meaningful subjects and contextual relevance	Object detection outputs, scene labels	Integrates content awareness into aesthetics

Semantic Content Significant topics and relevance with contexts Scene labels, objects detected in a scene Meaningful subjects.

In addition to perceptual organization, semantic interpretation and contextual interpretation are highly determinants on aesthetic judgment in addition Le et al. (2020). The aesthetic appreciation depends decisively on high-level content, i.e., the presence of humans, natural scene, symbolic elements, or expressions of human emotions. Cognitive theories propose that the viewers draw aesthetic values out of visual order as well as meaning, emotional response, and perceived intentionality. Therefore, two images of similar low-level visual statistics can receive significantly different aesthetic responses when they differ in terms of semantic and contextual information Zhang and Ban (2022). One of the main peculiarities of the image aesthetics is subjectivity. Different individuals possess different aesthetic preferences because of their cultural disparities, tastes and preferences, exposure to art and circumstances. Such subjectivity brings variability and uncertainty in human annotations and in result, tends to have broad distributions of scores, instead of unanimous labels. Contemporary computational (studies are thus turning more towards the aesthetic quality as a continuous or probabilistic entity, as opposed to a binary classification problem).

3. AI AND DEEP LEARNING FOUNDATIONS FOR AESTHETIC EVALUATION

The weaknesses of handcrafted feature-based methods have prompted a shift of paradigm to artificial intelligence-based methods where aesthetic representations are trained in direct interaction with the data. Deep learning models have shown great ability in learning the complex, hierarchical and abstract patterns underlying human aesthetical perceptions Ataer-Cansizoglu et al. (2019). The majority of AI-based aesthetics evaluation systems are based on convolutional neural networks (CNNs). The manner of CNNs learning hierarchical features representations is in a

hierarchical order, those early layers detect low-level visual features like edges and textures, mid-level compositional features like mid-level layers, and high-level semantic features like deeper layers Yu and Chung (2023). This development is quite analogous to steps of human visual perception and has allowed CNN based architectures to be especially useful in aesthetic classification as well as regression tasks. Large scale visual recognition datasets are also used to transfer learning to ameliorate the paucity and noise of aesthetic annotations. When compared to CNNs, the latter is very efficient in extracting local features, but aesthetic decisions tend to require global composition and long-range spatial relations among the visual entities. The approach of attention based architectures and vision transformers overcomes this limitation by explicitly attempting to model the interactions between more distant regions of an image by self-attention based mechanisms Lindenthal and Johnson (2021). Transformer-based models have thus been receiving growing interest in the field of aesthetic evaluation study, especially in the fine-grained and holistic judgment tasks.

Table 2

Table 2 Comparison of Deep Learning Architectures for Image Aesthetics Evaluation

Aspect	CNN-Based Models	Transformer-Based Models	Hybrid CNN–Transformer Models
Feature Learning	Hierarchical local feature extraction	Global feature modeling via self-attention	Joint local–global representation learning
Spatial Context Modeling	Implicit, localized receptive fields	Explicit long-range dependency modeling	Balanced local and global context
Composition Awareness	Moderate	Strong	Very strong
Semantic Understanding	High with deep networks	High with sufficient data	High with enhanced contextual alignment
Data Requirements	Moderate	High	Moderate to high
Computational Complexity	Low to moderate	High	Moderate
Robustness to Noise	High	Moderate	High
Suitability for Aesthetic Regression	Good	Very good	Excellent
Interpretability	Feature maps, Grad-CAM	Attention maps	Combined saliency and attention
Typical Use Cases	Real-time scoring, mobile systems	Fine-art and composition analysis	High-fidelity, explainable aesthetics

In order to unify these architectural views, Table 2, provides a comparative study of CNN-based, Transformer-based, and hybrid CNN Transformer models when it comes to the evaluation of image aesthetics. Their strengths, computational properties, and the appropriateness to the various aesthetic modeling scenarios are pointed out in the table and thus explains the design trade-offs incurred in choosing the right architecture. The hybrid architectures have become a logical extension of the purely convolutional or purely attention-based designs. Hybrid CNN-Transformer models combine the use of convolutional layers to extract local features in an efficient way with those of the transformer block to reason the global features, which provide a more detailed representation of the aesthetic qualities. These architectures are specifically useful in those cases when technical quality evaluation is needed, as well as balanced examination of compositions, but with manageable computational prices.

4. AI ALGORITHMS FOR IMAGE AESTHETICS EVALUATION

It is based on the architectural background presented in the previous section that this section reviews the major artificial intelligence algorithms utilized in image aesthetics evaluation. These algorithms vary in terms of their learning purpose, representational approach and capability of simulating subjectivity but are all aimed at estimating human aesthetic judgments in automated and scalable forms. Initial deep learning methods mostly used convolutional neural networks, which were either trained to be classifiers or regressors. When using classification-based formulations, images are categorized into discrete aesthetic values, e.g. high quality or low quality, which allows making decisions efficiently with respect to the application of content filtering and ranking. By comparison, regression based models are models that

predict continuous aesthetic scores which are more representative of the grading characteristic of human perception. The CNN-based regression models have demonstrated better correlation with human ratings especially when using functions of loss that are consistent with ranking consistency and score distribution properties.

Figure 2

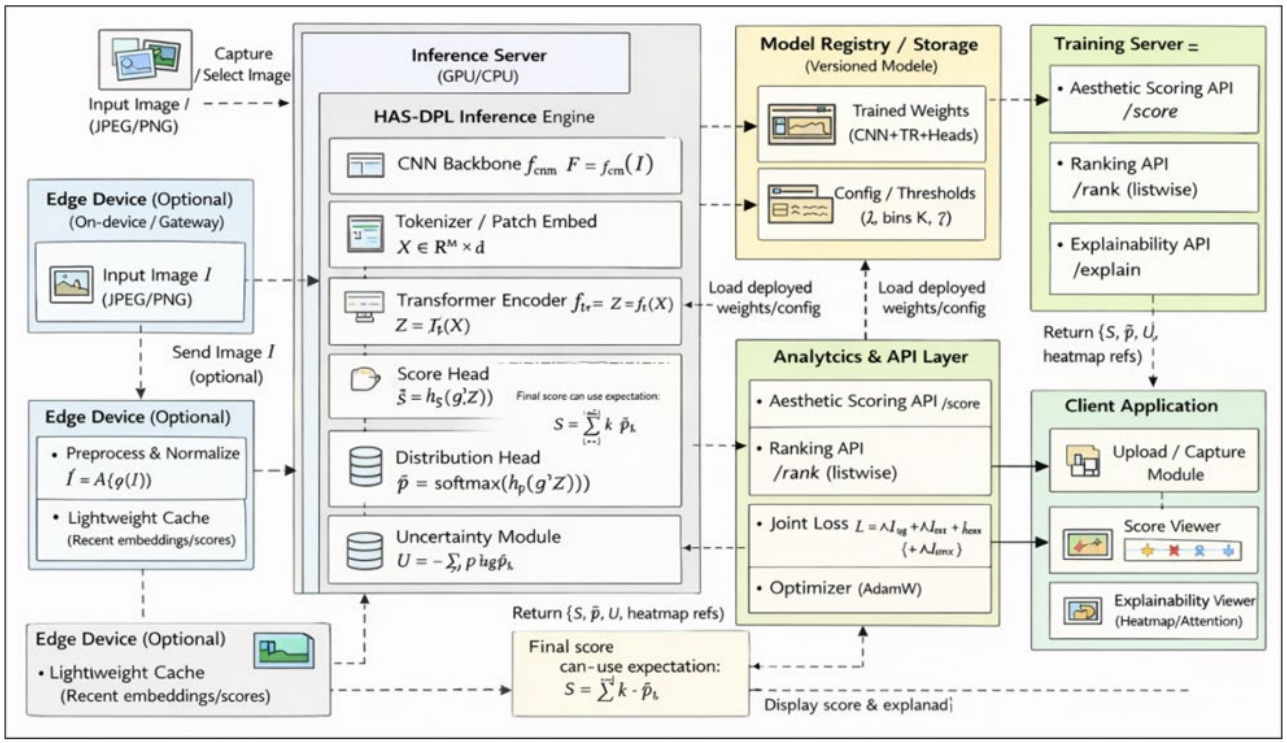

Figure 2 System Design Diagram for HAS-DPL

Vision transformers use self-attention techniques to learn the world relationship between the regions of an image enabling aesthetic judgments to take into account the holistic information contained in compositions. Transformer model aesthetic models are particularly useful in the situations when the aesthetic perceptions are dominated by global harmony, symmetry, and semantic completeness. Their scalability to large datasets and large computational cost however have inspired the search to develop more efficient variations and hybrid designs. Hybrid CNN-Transformer algorithms are the major development in the aesthetic evaluation. Convolutional layers are used in these models to extract local features that are robust and then are processed more by transformer blocks to learn the long-range dependencies and global context as illustrated in Figure 2. The two-stage representation allows the correct modeling of technical quality of images and upper level compositional aesthetics. These hybrid architectures have always shown better performance on cross-dataset testing and better noise resistance in subjective labelling.

Subjective aesthetics and enhanced ranking accuracy with extensive image collections Pairwise ranking of losses and list wise optimization Pairwise ranking is especially effective in the modeling of subjective aesthetics and rank enhancement in large-scale image collections. Aesthetic evaluation and optimization activity has also been investigated through reinforcement learning. The aesthetic quality in this case is considered as a reward signal which directs an agent in producing visually pleasing results. Although reinforcement learning is not as widely applied to the direct aesthetic scoring, it has its useful role in the area of sequential decision-making, such as automatic image enhancement, cropping, and aesthetic optimization.

5. HYBRID AESTHETIC SCORING WITH DISTRIBUTION + PREFERENCE LEARNING (HAS-DPL)

Train a hybrid CNN-Transformer to not only predict a discrete score (i) of continuous aesthetic score along with (ii) of an opinion-centered score distribution, but also pairwise preferences (ranking), in order to deal with subjectivity and annotation noise. At inference, one score and explainable attention/saliency.

Step 1: Input and standardization

This step seeks to guarantee the input of the samples utilized in the study and their standardization.

$$I\sim = A(\phi(I)).$$

$$\phi(I) = \sigma Resize(I) - \mu,$$

It enhances strength without tainting the esthetic indications.

Step 2: Hybrid feature extraction (local + global)

Pass (I) such a CNN backbone (e.g. ResNet / EfficientNet) to get multi-scale local feature maps, which capture the texture, sharpness, edges and fine composition details.

$$F = fcnn(I\sim), F \in Rh \times w \times c.$$

$$xi = Wp \cdot vec(Fpi), X \in RN \times d.$$

Convert these feature maps into a sequence of tokens (via patch embedding or feature pooling), then feed tokens into a Transformer encoder to model global composition and long-range dependencies (balance, symmetry, subject–background relationships).

$$Z = ftr(X), Z \in RN \times d.$$

Step 3: Multi-head prediction (score + distribution + attributes)

From the Transformer output, compute three heads:

1) Score head creates a scalar aesthetic score regression.

$$ttn(Q, K, V) = softmax(dkQK\top)V.$$

2) In order to simulate human disagreement, distribution head generates a probability distribution across rating bins (10).

$$p\hat{} = softmax(hp(g(Z))) \in [0,1]K, k = 1\sum Kp\hat{}k = 1.$$

3) Attribute head predicts explainable aesthetic qualities (composition, color harmony, lighting) to assist in the explanation and analysis.

Step 4: Preference learning (pairwise ranking)

Sample image pairs (Ia,Ib) are made to match with the human comparative judgment, the ground-truth preference based on mean scores or voting majority based on annotations.

$$P(a \succ b) = \sigma\left(s^a - s^b\right).$$

$$Lrank = -[ylogP(a \succ b) + (1 - y)log(1 - P(a \succ b))].$$

This improves ranking consistency in real-world deployment (feeds, retrieval, selection).

Step 5: Joint training objective (handles subjectivity)

Train the model using a weighted combination of losses:

- Regression loss Lregrobust loss (Huber) between (s) and ground truth mean score (s).

$$s(i) = k = 1\sum Kkpk(i).$$

- Loss of distribution Ldistcross-entropy or KL divergence between the empirical human rating histogram (p) and (hatp).
- Ranking loss Lrankhinge or logistic pairwise loss applying preference order.

$$S = s\hat{}exp = k = 1\sum Kkp\hat{}k(or\ S = s\hat{})$$

- Attribute lossLattr: BCE/CE for aesthetic attributes (if labels exist or can be weakly derived).

$$L = \lambda 1Lreg + \lambda 2Ldist + \lambda 3Lrank + \lambda 4Lattr$$

The paradigms of modern AI-based solutions thus focus on the hierarchical and attribute-sensitive feature learning to close the divide between the perceptual theory and computational modeling. At the bottom, learning aesthetic features starts with the retrieval of fundamental visual primitives e.g. edges, textures, color distributions and luminance patterns.

6. DATASETS AND ANNOTATION STRATEGIES

The Quality and the structure of datasets utilized to train and evaluate the model is the key to reliable image aesthetics evaluation. The aesthetic quality, in contrast to classical computer vision tasks, in which labels are usually objective and deterministic, is by definition subjective and multi-dimensional. Therefore, dataset design should be sensitive to the visual diversity and scale, as well as disagreement, uncertainty, and bias in human judgements. Aesthetic datasets: There are crowdsourced annotation campaigns and curated image repositories as well as mostly online photography platforms which collect most image aesthetics datasets. Photographs are traditionally characterized by a discrete quality scale or a numerical score of several human judges.

Massive datasets can also be used to train models or architectures that have a large capacity (hybrid CNNTransformer models, etc.), whereas smaller datasets can be used to analyze aesthetics on an attribute level (composition, lighting, color harmony, etc.) to allow interpretation and diagnostic analysis. Another critical design option in dataset construction is the annotation strategy because it directly influences the method of computing human aesthetic perception. The three most used strategies that are mean-score annotation, distribution-based annotation and pairwise preference annotation vary in many aspects in terms of the capability to reflect subjectivity and variability. The comparative summary of these annotation approaches is given in Table 3; their representations of labels, their strong and weak sides in the framework of aesthetic modeling are noted.

Table 3

Table 3 Comparison of Annotation Strategies for Image Aesthetics Datasets

Annotation Strategy	Advantages	Limitations	Suitable Use Cases
Mean-Score Annotation	Simple and storage-efficient; directly compatible with regression models; widely used in early benchmarks	Discards inter-rater disagreement; sensitive to outliers; masks polarized opinions	Baseline aesthetic scoring; lightweight systems; initial benchmarking
Distribution-Based Annotation	Preserves subjectivity and uncertainty; supports probabilistic learning; enables confidence estimation	Requires more ratings per image; higher annotation and modeling complexity	Opinion-aware learning; uncertainty modeling; explainable AI systems
Pairwise Preference Annotation	Aligns with natural human comparison; reduces cognitive load; robust to rating-scale bias	Does not provide absolute scores; requires many comparisons for global ranking	Ranking, recommendation, retrieval, and personalization systems

Mean-score annotation is also appealing because it is simple and annotating it is not expensive; nevertheless, it usually cannot capture the multifacetedness of human judgement. To overcome this shortcoming, distribution-based annotation maintains the entire rating histogram of each image enabling models to also learn perceived uncertainty and disagreement. Pairwise preference annotation has a different formulation that can be closely related to human comparative judgment and perform learning in the form of ranking, but it needs special caution in the experimental design in order to maintain transitivity and consistency of preference. One solution to this problem would be dataset curation that would focus on variety in terms of content type, cultural background, capture tools, and artistic styles. The cross dataset evaluation and domain adaptation protocols are thus fundamental parts of sound aesthetic assessment investigations. The quality control systems are also vital. Techniques in common use are rater qualification tests, consistency, repeated item checks, outliers and minimum vote per image. In settings that are preference-based, redundant pair comparisons as well as transitivity enhance label reliability. All of these methods can be used to increase the fidelity of databases and to improve the correspondence between the computational prediction and the aesthetic sense of the human observer.

7. FEATURE LEARNING AND AESTHETIC ATTRIBUTE MODELING

The feature learning feature is at the center of image aesthetics consideration because the quality of aesthetics is the result of interaction between low-level visual messages and high-level semantic messages and affective reactions. Contrary to the traditional vision tasks, where the use of object-centric features is the main feature, aesthetic evaluation demands representations, which capture composition, balance, emotional coloring and contextual pertinence. With deep learning systems such cues are implicitly learned through the initial convolutional layers that can be viewed as adaptive filters that pick up local gradients and spatial frequency information. These low level characteristics help in the evaluation of technical image quality, (sharpness, exposure, and noise) which is a necessary, but not sufficient condition of aesthetic image. Compositional structure and spatial organization are at the level of mid-level feature representations.

These are object placement, symmetry, depth, foreground-background distance and visual saliency. The CNN-based architectures represent these properties in a sequence of mid-level layers with increasing receptive fields, whereas attention-based models explicitly represent the spatial connections between remote image regions. At this level, feature maps and attention weights tend to be highly related to classical photographic principles like the rule of thirds, balance and leading lines and are therefore specifically relevant to aesthetic modeling. High level feature learning obtains semantic and contextual features, which have profound effect on aesthetic judgment, indicated in Figure 3. Deep CNNs and transformer encoders learn the representations that relate to the type of the scene, object identity, human presence, and symbolic content. These semantic features allow one to distinguish between pictures with different meanings or emotional appeal that are visually similar.

Figure 3

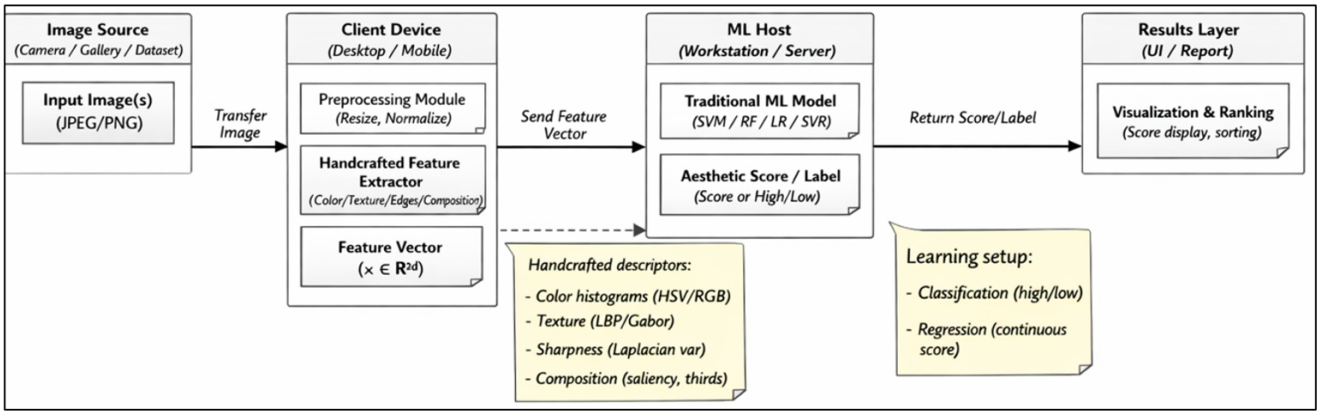

Figure 3 Deployment Diagram for Traditional (Handcrafted + ML) Aesthetic Assessment Pipeline

A photograph of a human face or a melodramatic view of nature can be given a higher aesthetic rating because of the richness of semantics, but the low-level visual qualities can be equivalent to less appealing content. In addition to generic representations, explicit aesthetic attribute modeling has become more and more popular. Aesthetic qualities,

e.g., quality of composition, harmony of colors, lighting, depth, emotional coloring, etc., can offer interpretable dimensions through which the aesthetic quality can be broken down. Attribute-aware models either explicitly predict these attributes with the help of auxiliary prediction heads or learn disentangled latent representation that is represented by interpretable aesthetic factors. This multi-task formulation is more effective as it not only provides a better performance by facilitating shared representation learning but also increases the transparency and diagnostic potential. Aesthetic representation is extended further by the emotional and affective feature modeling which involves emotion and viewer reaction cues. Color temperature, contrast dynamics, facial expression and scene semantics are some of the visual cues that make part of affective embeddings which are correlated to emotional valence and arousal. By adding affective elements, the aesthetic models can cease to be based on structural quality and more reflective of the human experience.

8. DISCUSSION

The findings and methodological advances made in the current research pinpoint the advances, as well as the issues that remain unaddressed concerning the assessment of image aesthetics with the help of AI. The move to the deep, hybrid learning architecture as opposed to the handcrafted features has significantly enhanced the predictive capabilities of computational models in the approximation of human aesthetic judgment. Specifically, the combination of convolutional and attention-based models makes it possible to model both local visual quality and global compositional coherence as the important elements of aesthetic perception simultaneously. These results testify the idea that aesthetic assessment cannot be suitably dealt with in isolation; however, it involves holistic and multi-level representations. One of the most notable findings made based on the suggestion of the framework is the necessity to explicitly model subjectivity. Distribution-sensitive learning and preference-based ranking are more consistent with human aesthetic judgment as a probabilistic process as opposed to single-point regression. This is particularly true in real life application where aesthetic may differ among users, cultures and application conditions. Further, the uncertainty measures based on expected distributions of scores also offer a principled mechanism of determining images with ambiguous or polarized aesthetic values.

Figure 4

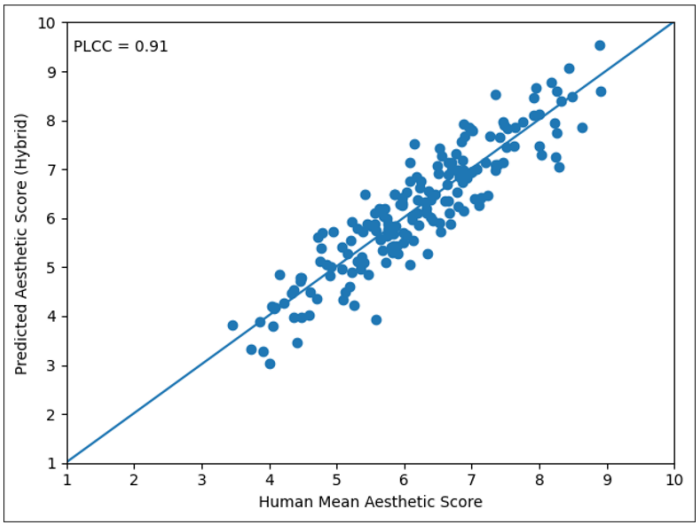

Figure 4 Predicted Aesthetic Scores Versus Mean Human Ratings for the Hybrid Model.

Figure 4 illustrates the performance of the hybrid model as compared to the mean human ratings in the evaluation set. The scatter distribution and trendline modeled shows the extent to which predictions abide by the human scoring scale and the value of PLCC summarizes the linear consistency. The fact that both scores were concentrated around the diagonal trend shows that the scores are well characterized by calibration whereas the broader spread indicates that subjects on which the model and the human raters disagree because of subjective reasons or the semantic context. Interpretability as a part of the aesthetic AI systems is highlighted in the conversation, as well. Explainable visualization

methods like saliency and attention analysis as well as attribute-aware modeling provide useful information about why a model has deemed a specific aesthetic score. This openness is essential not only to the trust of users, but also to the process of diagnosing the cases of failure and unintended bias. Nevertheless, the interpretability has been left as an open question because aesthetic qualities tend to be abstract and contextual and it is hard to determine the full agreement between human explanation and model reasoning.

Figure 5

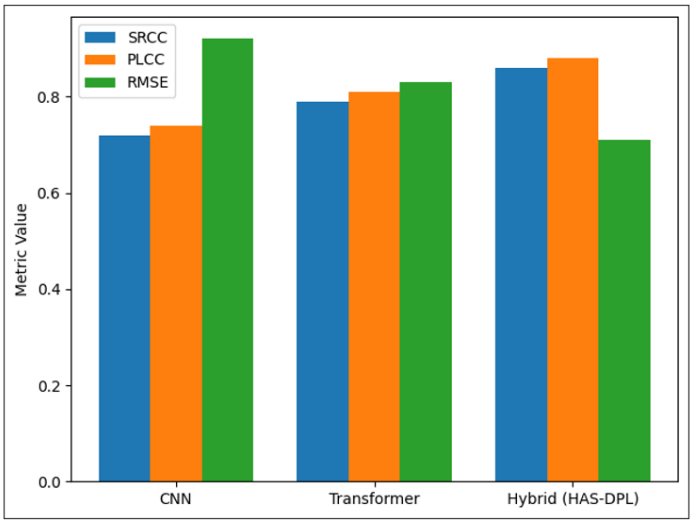

Figure 5 Comparative performance of CNN, Transformer, and hybrid models using SRCC, PLCC, and RMSE.

The trade-off between ranking consistency and score accuracy is compared in Figure 5 which compares representative deep architectures using both correlation-based and error-based measures. The hybrid model is generally having better SRCC/PLCC and lower RMSE, which is in line with its capability to integrate local learning of visual features and global compositional reasoning. This analogy leads to the thesis that holistic modeling of aesthetics enjoys the benefits of architectures that both store fine-grained technical information and long-range relational structure. As much as these advances have been made, there are a number of limitations. Models with high capacity, especially transformer-based architecture and hybrid models require large and varied datasets that are expensive to annotate and may be culturally biased. Additionally, most of the existing benchmarks focus on general photographic aesthetics, which makes the generalization of the models to the specialized field of fine art, medical, or culturally-specific visual aesthetics challenging. These gaps will involve more varied data, dynamic learning approaches, and custom modeling models.

Figure 6

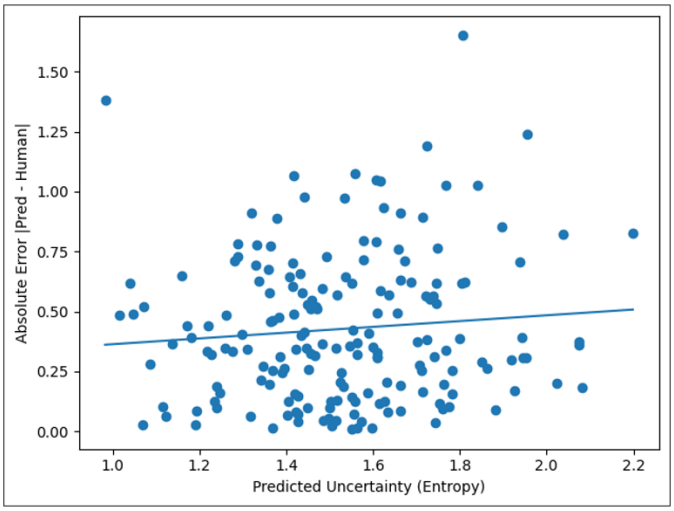

Figure 6 Uncertainty vs Prediction Error (Hybrid)

The relations between uncertainty estimates (entropy of the predicted rating distribution) and prediction error are studied in Figure 6, as it provides a convenient perspective on subjectivity-sensible modeling. There is a positive trend indicating that the images that are tagged to be of high-uncertainty, typically those of polarized or ambiguous aesthetic considerations, are even more difficult to determine correctly. This is a desirable behavior that can be deployed as it allows systems to expose the presence of so-called subjective cases to the review of people or a human-based preference management instead of displaying overconfident results.

On the whole, the results indicate that the optimal evaluation of image aesthetics is based on the combination of the strong representation learning, subjectivity-conscious annotation, and explainable decision-making. Further development and advancement in these directions will be important to implement aesthetic AI systems that are precise, transparent, as well as in accordance with human visual experience.

9. CONCLUSION

The paper has explored the issue of automated image aesthetics scoring by analyzing the current artificial intelligence algorithms, datasets and modeling strategies in depth. The advancement of the aesthetic foundations to deep learning-guided architectures shown in the piece illustrated the role of aesthetic judgement as it can be well-approximated through representations of data which include perceptual, semantic, and affective signals. One of the contributions of this work is the focus on the subjectivity-conscious modeling. The new hybrid framework with a convolutional and attention-based learning and distributional and preference-based objective is much better-positioned to deal with the fact of human aesthetic judgments becoming more variable than regression task frameworks with single-point prediction. Assimilation of uncertainty estimation and explainable aesthetic features also increase robustness and transparency that makes the system appropriate to implement in the real world. In a comparative analysis that is based on experimental analysis, the benefits of hybrid architectures are identified in capturing the local visual quality and global compositional structure. Such results support the significance of the holistic learning of features and the method of annotation development to get a high level of correlation with human perception. Besides, the discourse emphasizes the need to have varying datasets and strict assessment procedures that will facilitate cross-domain and cultural contexts. Finally, AI systems are useful in image aesthetics assessment not only in their accuracy but also in regard to interpretability and compatibility with human subjectivity. These approaches and ideas introduced in this paper contribute to the future studies of personalized aesthetic modeling, multimodal analysis, and responsible application of aesthetic intelligence to creative and decision-making systems.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Ataer-Cansizoglu, E., Liu, H., Weiss, T., Mitra, A., Dholakia, D., Choi, J. W., and Wulin, D. (2019, December). Room Style Estimation for Style-Aware Recommendation. In Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR) (267–273). https://doi.org/10.1109/AIVR46125.2019.00062

Celona, L., Ciocca, G., and Napoletano, P. (2021). A Grid Anchor-Based Cropping Approach Exploiting Image Aesthetics, Geometric Composition, and Semantics. Expert Systems with Applications, 186, Article 115852. https://doi.org/10.1016/j.eswa.2021.115852

Chen, Q., Zhang, W., Zhou, N., Lei, P., Xu, Y., Zheng, Y., and Fan, J. (2020, June). Adaptive Fractional Dilated Convolution Network for Image Aesthetics Assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (14114–14123). https://doi.org/10.1109/CVPR42600.2020.01412

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv.

He, S., Zhang, Y., Xie, R., Jiang, D., and Ming, A. (2022, July). Rethinking Image Aesthetics Assessment: Models, Datasets and Benchmarks. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI) (942–948). https://doi.org/10.24963/ijcai.2022/132

Horanyi, N., Xia, K., Yi, K. M., Bojja, A. K., Leonardis, A., and Chang, H. J. (2022). Repurposing Existing Deep Networks for Caption- and Aesthetic-Guided Image Cropping. Pattern Recognition, 126, Article 108485. https://doi.org/10.1016/j.patcog.2021.108485

Le, Q. T., Ladret, P., Nguyen, H. T., and Caplier, A. (2020). Study of Naturalness in Tone-Mapped Images. Computer Vision and Image Understanding, 196, Article 102971. https://doi.org/10.1016/j.cviu.2020.102971

Li, D., Zhang, J., Huang, K., and Yang, M.-H. (2020, June). Composing Good Shots by Exploiting Mutual Relations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (1–10). https://doi.org/10.1109/CVPR42600.2020.00427

Lindenthal, T., and Johnson, E. B. (2021). Machine Learning, Architectural Styles and Property Values. Journal of Real Estate Finance and Economics, 1–32. https://doi.org/10.1007/s11146-021-09815-9

Luo, P. (2023). Social Image Aesthetic Classification and Optimization Algorithm in Machine Learning. Neural Computing and Applications, 35, 4283–4293. https://doi.org/10.1007/s00521-022-07128-1

Mehta, S., and Rastegari, M. (2021). MobileViT: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer. arXiv.

Yu, L., and Chung, W. (2023). Analysis of Material and Craft Aesthetics Characteristics of Arts and Crafts Works Based on Computer Vision. Journal of Experimental Nanoscience, 18, Article 2174693. https://doi.org/10.1080/17458080.2023.2174693

Zeng, H., Cao, Z., Zhang, L., and Bovik, A. C. (2020). A Unified Probabilistic Formulation of Image Aesthetic Assessment. IEEE Transactions on Image Processing, 29, 142–149. https://doi.org/10.1109/TIP.2019.2941778

Zhang, Z., and Ban, J. (2022). Aesthetic Evaluation of Interior Design Based on Visual Features. International Journal of Mobile Computing and Multimedia Communications, 13(1), 1–12. https://doi.org/10.4018/IJMCMC.2022010101