Deep Learning Based Image Aesthetic Quality Assessment- A Review

Daryanavard Chounchenani, Maedeh; Shahbahrami, Asadollah; Hassanpour, Reza; Gaydadjiev, Georgi

DOI
10.1145/3716820

Publication date
2025

Document Version
Final published version

Published in
ACM Computing Surveys

Deep Learning Based Image Aesthetic Quality Assessment- A Review

MAEDEH DARYANAVARD CHOUNCHENANI, Computer Science, Groningen University Faculty of Science and Engineering, Groningen, Netherlands
ASADOLLAH SHAHBAHRAMI, Computer Engineering, University of Guilan, Rasht, Iran (the Islamic Republic of)
REZA HASSANPOUR, Computer Science, Groningen University Faculty of Science and Engineering, Groningen, Netherlands
GEORGI GAYDADJIEV, Delft University of Technology Faculty of Electrical Engineering Mathematics and Computer Science, Delft, Netherlands

Image Aesthetic Quality Assessment (IAQA) spans applications such as the fashion industry, AI-generated content, product design, and e-commerce. Recent deep learning advancements have been employed to evaluate image aesthetic quality. A few surveys have been conducted on IAQA models; however, details of recent deep learning models and challenges have not been fully mentioned. This article aims to fill these gaps by providing a review of deep learning IAQA over the past decade, based on input, process, and output phases. Methodologies for deep learning–based IAQA can be categorized into general and task-specific approaches, depending on the type and diversity of input images. The processing phase involves considerations related to network architecture, learning structures, and feature extraction methods. The output phase generates results such as scoring, distribution, attributes, and description. Despite achieving a maximum accuracy of 91.5%, further improvements in deep learning models are still required. Our study highlights several challenges, including adapting models for task-specific methodology, accounting for environmental factors influencing aesthetics, the lack of substantial datasets with appropriate labels, imbalanced data, preserving image aspect ratio and integrity in network architecture design, and the need for explainable AI to understand the causative factors behind aesthetic judgments.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Computing methodologies → Computer vision;**

Additional Key Words and Phrases: Image aesthetic, image aesthetic quality assessment, computer vision, deep learning

ACM Reference Format:
Maedeh Daryanavard Chounchenani, Asadollah Shahbahrami, Reza Hassanpour, and Georgi Gaydadjiev. 2025. Deep Learning Based Image Aesthetic Quality Assessment- A Review. *ACM Comput. Surv.* 57, 7, Article 183 (February 2025), 36 pages. https://doi.org/10.1145/3716820

Authors' Contact Information: Maedeh Daryanavard Chounchenani, Computer Science, Groningen University Faculty of Science and Engineering, Groningen, Groningen, Netherlands; e-mail: m.daryanavard@rug.nl; Asadollah Shahbahrami, Computer Engineering, University of Guilan, Rasht, Gilan, Iran (the Islamic Republic of); e-mail: shahbahrami@guilan.ac.ir; Reza Hassanpour, Computer Science, Groningen University Faculty of Science and Engineering, Groningen, Goningen, Netherlands; e-mail: r.zare.hassanpour@rug.nl; Georgi Gaydadjiev, Delft University of Technology Faculty of Electrical Engineering Mathematics and Computer Science, Delft, Zuid-Holland, Netherlands; e-mail: G.N.Gaydadjiev@tudelft.nl.

ACM 0360-0300/2025/02-ART183
https://doi.org/10.1145/3716820

1 Introduction

Digital media has been utilizing in various domains, such as intelligent transportation systems [138, 143], remote sensing [144], technical diagnostics [50], target tracking [113], advertising. The quality of digital media plays a pivotal role in determining the output of these applications, influencing their attractiveness, utility, and accuracy. Significant efforts have been made in recent years to quantify the quality of digital media, particularly images, using a variety of techniques categorized as **Full-Reference (FR)**, **Reduced-Reference (RR)**, and **No-Reference (NR)** approaches [164]. In the FR technique, the quality of a target image is evaluated by comparing it with the original reference image. The RR approach involves utilizing a set of low-level features such as edges, texture, and color histograms, or high-level features such as objects and concepts, extracted from the original reference image for comparison [4]. In contrast, the NR approach does not rely on either the original reference image or any specific features [19]. Recent researchers have been focusing on aesthetic principles to measure image quality and have increasingly recognized the impact of aesthetics on the overall perception of images [25, 68, 136]. This trend reflects a growing acknowledgment of the multi-faceted nature of image assessment, extending beyond traditional technical parameters. What is aesthetic? According to the Oxford Dictionary, "beauty is a combination of qualities that pleases the aesthetic senses, especially the sight, and aesthetic is the appreciation of beauty." However, aesthetics extends beyond this definition, encompassing critical reflection on art, culture, nature, and style. Although the Oxford Dictionary defines beauty as the appreciation of aesthetics through the senses, on the other side, individual tastes are influenced by culture, history, and personal preferences. Despite these two definitions of aesthetics, it is crucial to note that there are established contractual rules of photography and art for measuring beauty, known as aesthetic features. These rules encompass elements such as symmetry, the rule of thirds, and depth of field [20, 104]. In recent years, these aforementioned aesthetic features have been employed to assess image quality. While numerous studies have explored **Image Aesthetic Quality Assessment (IAQA)** in various aspects [2, 23, 139, 163, 172], the primary focus has often been on traditional methods, while recent research and techniques have often been overlooked. In contrast, the objective of this study is to consolidate essential information using deep learning approaches and systematically categorize the multitude of methods published in the past decade. The pipeline of IAQA using deep learning is clarified through three distinct phases: input, process, and output. Subsequently, the impact of the environment on aesthetics is presented, along with diverse applications that demonstrate practical utility across various domains. Additionally, the challenges associated with IAQA using deep learning approaches in the existing dataset and mentioned phases are explored. The challenge of labeling datasets effectively, limited data, and existing noise in aesthetic scores in training machine learning models potentially hinder models from generalizing effectively, leading to biased judgments. In the input phase, as each domain and diverse genre of images has its particular features, there is a need to shift the focus from a general methodology to a task-specific one. In the process phase, the challenge is related to the consideration of the aspect ratio of images, which influences their aesthetic quality. Accordingly, the focus is on designing a network architecture that preserves the aspect ratio of the input data. In the final phase, there is a requirement to comprehend the causative factors behind aesthetic judgments by **eXplainable AI (XAI)**.

The article is structured as follows: Section 2 discusses methods for assessing the aesthetic quality of images, Section 3 considers the datasets used in IAQA. Section 4 outlines the phases of deep learning–based IAQA, Section 5 reviews of IAQA models using deep learning. Section 6 presents the impact of environmental factors on aesthetics, and applications are explored in Section 7. Section 8 discusses existing challenges. Conclusions are drawn in Section 9.

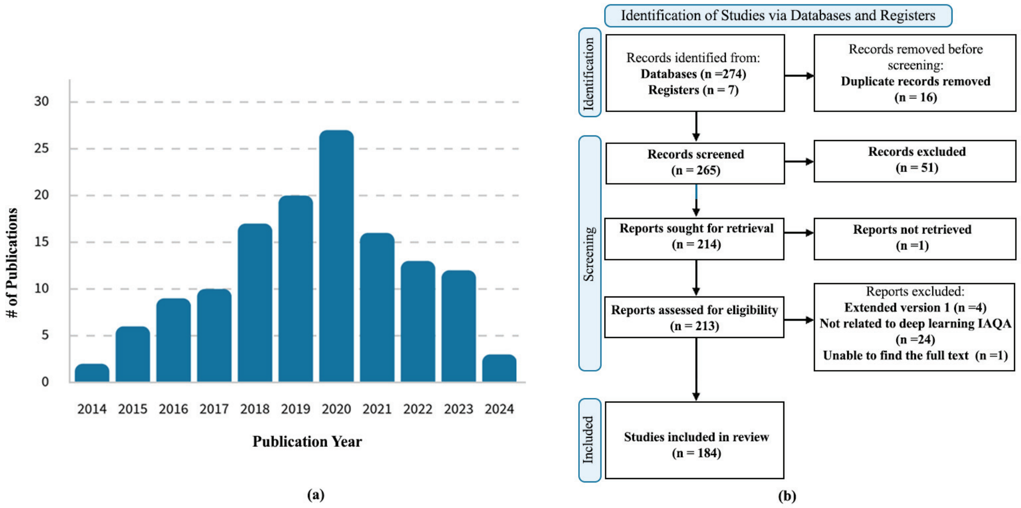

Fig. 1. (a) Distributions of the papers surveyed in the current research work that only focus on the IAQA based on deep learning from 2014–2024 (total: 135 papers up to August 2024). (b) The PRISMA flow diagram showing the researches' inclusion and exclusion criteria focus on a decade of the IAQA method based on deep learning.

2 Image Aesthetics Quality Assessment

The evaluation of image aesthetic quality involves considering low-level features, including color [29, 33], composition, and texture [20] and making connections to high-level features. Color choices can evoke different emotions and carry cultural significance. Composition techniques such as framing and balance guide the viewer's focus. Texture, whether smooth or rough, also influences an image's appeal. Machine learning algorithms use these features to predict aesthetic quality. In general, IAQA techniques involve two main approaches: shallow learning approaches and deep learning approaches. Shallow learning uses hand-crafted features such as color and texture to train a model, relying on traditional computer vision techniques and domain expertise [70, 103, 112, 132, 168]. One of the main problems with using shallow learning is that these models may not be able to capture the complex and high-level visual characteristics that are important for accurately assessing image quality [177].

Image aesthetic quality assessment has been revolutionized by deep learning approaches, which have enabled the development of a range of methods and techniques to extract aesthetic features for quality assessment [13, 52, 53]. Deep learning, particularly **Convolutional Neural Networks (CNNs),** automatically learns complex features from raw image pixels, requiring large amounts of labeled data and computational resources. Deep learning has shown superior performance but comes with higher resource requirements. The choice of approach depends on the application's needs and available resources. Deep learning is often preferred in many scenarios due to its ability to automatically learn hierarchical representations from raw data, superior performance on complex tasks, and its capacity to handle large datasets effectively. This study aims to systematically explore a decade of research in IAQA using deep learning models. The distribution of papers illustrated in Figure 1(a) highlights the trends in research output over the years up to August 2024. The methodology of this study is comprehensively detailed in the PRISMA flow diagram presented in Figure 1(b). Initially, a total of 281 papers were identified through several searching sources such as IEEE, Springer, ACM, Frontiers, Entropy, and Elsevier. After removing

Table 1. Overview of Some Popular Aesthetic Datasets

Ref.	Dataset	Year	# of Images	Images Source	Label	Task
[60]	FAE-Caption	2022	251K	Flickr platform	Comment	Description
[157]	AVA-PCap	2020	8K	AVA dataset	Comment, Discrete 1-10	Scoring, Description
[37]	AVA-Captions	2020	230K	DPChallenge platform	Comment, Discrete 1-10	Scoring, Description
[31]	EAD	2020	25K	Butter Camera platform	Binary	Scoring
[62]	DPCCaptions	2019	154K	DPChallenge platform	Comment, Discrete 1-10	Scoring, Description
[151]	AVA-reviews	2019	52K	DPChallenge platform	Comment, Discrete 1-10	Scoring, Description
[122]	AROD	2018	380K	Flickr platform	Continius [0,1]	Scoring
[118]	FLICKR-AES	2017	40K	Flickr platform	Discrete 1-5	Scoring
[10]	PCCD	2017	4K	Gurushots platform	Comment, Discrete 1-10	Scoring, Attribute, Description
[131]	FACD	2017	28K	Photo.net platform	ACR 1-4	Scoring, Attribute
[92]	Waterloo-IAA	2017	1K	Photo.net platform	Integer[0-100]	Scoring, Attribute
[73]	AADB	2016	10K	Flickr platform	Discrete 1-5	Scoring, Attribute
[121]	Hidden Beauty	2015	15K	Flickr platform	ACR 1-5	Scoring
[96]	IAD	2015	1.5M	DpChallenge,photo.net platforms	Binary	Scoring
[137]	CUHKPQ	2013	17K	Professional photography platform	Binary	Scoring
[109]	AVA	2012	250K	DPChallenge platform	Discrete 1-10	Scoring, Distribution
[55]	Kodak Aesthetics	2010	1.5K	Flickr, Kodak Picture platform	Discrete 1-10	Scoring
[21]	Photo.net	2008	20K	Photo.net platform	Binary, Discrete 1-100	Scoring
[70]	CUHK	2006	12K	DPChallenge platform	Binary	Scoring

ACR: Absolute Category Rating, (sorted by year).

16 duplicate papers, 265 papers remained for screening. During the screening phase, 51 records were excluded for not being related to computational image aesthetic quality assessment, leaving 214 records to be retrieved. However, one study could not be retrieved, resulting in 213 papers being assessed for eligibility. Out of these, 29 papers were excluded for various reasons, including being extended versions, not being related to the IAQA using deep learning, or being early access papers. Consequently, 184 studies were ultimately included in the qualitative synthesis, providing a comprehensive overview of the relevant literature in the field of deep learning.

3 Datasets for Image Aesthetic Quality Assessment

Deep learning approaches are data-driven models, so different datasets should be used to train these models. Depending on the source website or device, images within datasets might have various aesthetic and technical properties. Similarly, the annotations from different data sources for images can vary significantly. Table 1 depicts an overview of some popular aesthetic datasets, consisting of seven columns. The table is sorted by year and includes their corresponding rating systems. The fourth column depicts the approximate number of images in the datasets, and K is equal to 1,000. In the sixth column, **Absolute Category Rating (ACR)** is a rating model that asks human annotators to provide a score based on a fixed set of predefined categories, while the discrete model requires annotators to rate images on a scale of discrete values, for example, ranging from 1 to 5 stars. The last column represents the *task*, which can consist of scoring, distribution, attributes that are numerical outputs, and description, which is textual output.

Fig. 2. Illustrative examples from the AVA dataset, highlighting the diverse range of categories including animals, architecture, cityscapes, children, food and drinks, landscapes, portraits, and still-life, respectively, from (a) to (h).

As observed in the provided table of datasets, most of the datasets are collected for the scoring task and a substantial volume of images belongs to the **Image Aesthetic Dataset (IAD)**, sourced from DPChallenges [26] and Photo.net [117] platforms. One of the most frequently encountered datasets in IAQA is the **Aesthetic Visual Analysis (AVA)** dataset. The AVA dataset encompasses the distribution of raters' scores and various challenges labels and tags, each of which can be regarded as distinct categories. These categories span a broad spectrum, including, but not limited to, animals, architecture, cityscapes, children, food and drinks, landscapes, portraits, and still-life categories, as shown in Figure 2. In contrast, the **Aesthetics and Attributes Database (AADB)** emphasizes not only overall aesthetic scores but also specific visual attributes that contribute to aesthetics, such as balancing elements, color harmony, content, depth of field, light, motion blur, object, repetition, rule of thirds, symmetry, and vivid color. This dataset is particularly valuable for research that aims to understand how individual attributes affect perceived aesthetic quality. However, its smaller size compared to AVA can limit the generalization of models trained on it. Figure 3(a) presents a sample from the AVA dataset, accompanied by the distribution of rater scores. The distribution reveals that most individuals rated the image with a score higher than five. Notably, approximately 77 raters assigned a score of seven to this particular image, indicating a general preference for it. Figure 3(b) illustrates a sample from the AADB dataset. The dark image has a significantly negative score in the Light attribute, reflecting the poor lighting quality.

4 Deep Learning–based Image Aesthetics Quality Assessment

The IAQA based on deep learning involves three important parts: input, process, and output. In the input phase, researchers select an appropriate methodology for the task at hand, and based on that choice, they can determine the most suitable dataset for training and testing the deep learning models. During the process phase, decisions regarding network architecture, such as using CNNs and feature extraction approaches, are made. Finally, in the output phase, researchers can define the specific output tasks and explore applications related to those tasks. These stages are depicted in Figure 4, which are discussed in detail in the following sections.

4.1 Input Phase

In this section, the input phase, which consists of choosing the methodology of input and related dataset as well as performing different pre-processing operations, is discussed.

4.1.1 General or Task-specific. In the initial stage of selecting input data for the 2D-CNNs, a decision has to be made regarding the choice of methodologies. The methodology of IAQA

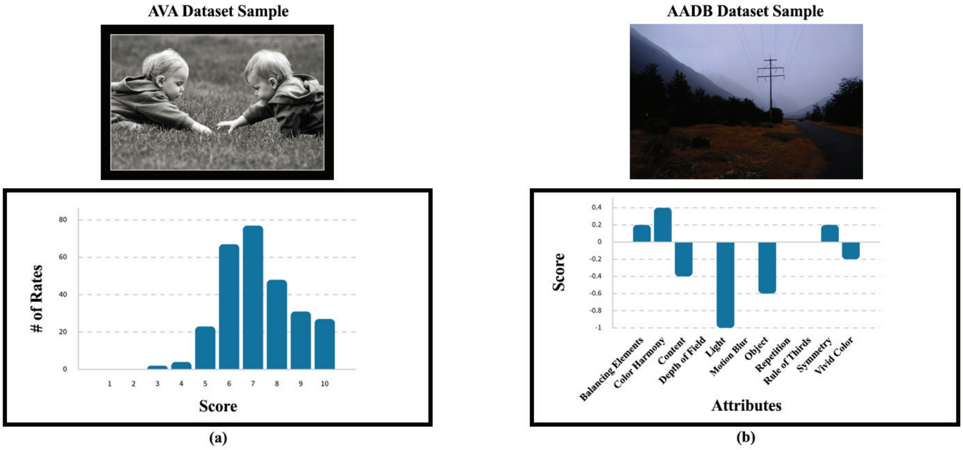

Fig. 3. (a) A sample from the AVA dataset shows the distribution of raters' scores, indicating that the majority rated the image above five, and around 77 raters assigned a score of seven. (b) A sample from the AADB dataset with attributes score, where the dark image received a notably low score in the "Light" attribute, highlighting the poor lighting quality.

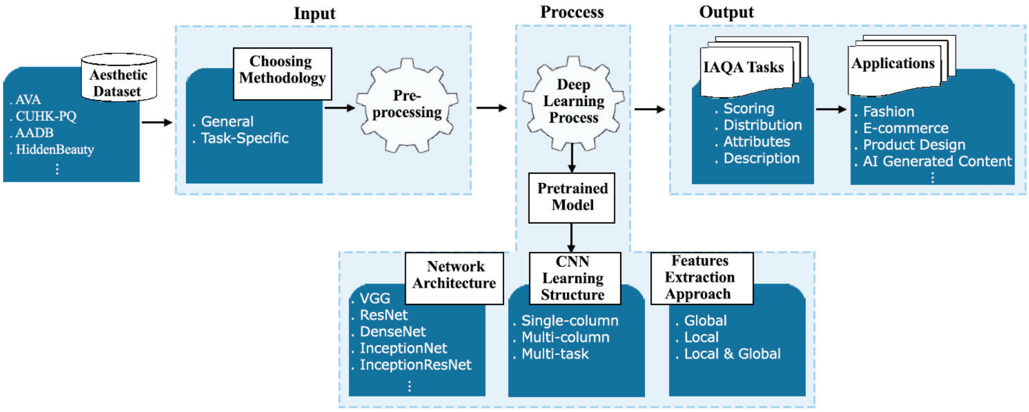

Fig. 4. Pipeline of image aesthetic quality assessment based on deep learning, depicting the input, processing, and output phases.

can be categorized as either general or task-specific, based on the type and diversity of the input images. The former category seeks to predict images' aesthetic quality without making assumptions about the image's content, covering a wide range of items and scenarios, such as natural, man-made, portraits, animals, and so on. In contrast, task-specific categories have prior knowledge of the semantic content of the image, which can significantly improve aesthetic prediction by considering a closed-set classification for image aesthetics [49, 131, 139, 152], which is used in applications such as fashion, interior design, product design, and other task-specific applications. Both aesthetic categories aim to estimate a suitable quality score or details associated with human perception, although this is a difficult task in the field of multimedia signal processing.

4.1.2 Pre-processing. In the field of IAQA, several pre-processing techniques are commonly employed to prepare the images for analysis. These techniques play a crucial role in standardizing the input data and enhancing the performance of the aesthetic assessment models. One important step is image resizing, where images of different sizes in the dataset are resized to a consistent size. Additionally, image normalization is performed by subtracting the mean value of the dataset and dividing by the standard deviation, ensuring consistent input across different images. Data augmentation is another vital technique used to increase the diversity of the training data. Random transformations such as rotations, flips, crops, brightness/contrast adjustments, and Gaussian noise are applied to the images, augmenting the dataset and preventing overfitting. Moreover, image cropping can be beneficial in focusing on the main subject or removing unwanted regions. Manual or automatic cropping techniques such as saliency or object detection can be employed. Enhancing the visual quality of the images through techniques such as histogram equalization, contrast stretching, or adaptive equalization can significantly improve the performance of aesthetic assessment models. Furthermore, converting images to different color spaces, such as grayscale or other color spaces such as Lab or HSV, can capture different aspects of the image and provide better features for aesthetic assessment. Noise reduction techniques, such as Gaussian blurring or median filtering, are used to reduce image noise while preserving important image details [9]. By applying these pre-processing techniques, the input images are standardized, enhanced, and made more suitable for accurate aesthetic assessment. These techniques contribute to improving the overall performance and reliability of IAQA models, leading to a better understanding of image aesthetics and facilitating applications in various domains such as photography, advertising, and digital media.

4.2 Process Phase

Some concepts such as network architecture, learning algorithms, and feature extraction techniques that are in the process phase will be discussed in the following sections.

4.2.1 Network Architecture. The CNN architecture is typically composed of several layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers perform feature extraction by applying a set of filters to the input image, while the pooling layers reduce the dimensionality of the feature maps. The fully connected layers perform the aesthetic tasks by combining the extracted features. There are several well-known CNN architectures used for IAQA, such as AlexNet [75], VGG [130], ResNet [41], and InceptionNet [134] or extension of these networks, which have achieved state-of-the-art performance. While choosing an appropriate neural network architecture is an important consideration when developing a model for image aesthetic assessment, evaluating the aesthetic quality of an image usually involves a subjective measure of its visual appeal and may also take into account other factors beyond the network structure. Designing an IAQA method has two primary approaches when developing models, including using pre-trained networks or designing methods from scratch. Using pre-trained networks involves leveraging models that have been trained on large datasets, such as ImageNet [22] for generic tasks and utilizing known architecture to fine-tune and extract meaningful features from the images in a smaller dataset to overcome the challenge of lack of data. However, designing IAQA methods from scratch involves constructing a neural network architecture that learns from aesthetic labels or negative aesthetic aspects without using pre-trained data [124].

4.2.2 Learning Structure. There exist three distinct categories of CNN learning structure employed for feature extraction from images. These categories encompass the Single-column, Multi-column, and Multi-task approaches [23]. The selection of the appropriate feature extraction methodology is contingent upon the specific inputs and outputs under consideration. Brief descriptions of these approaches are outlined as follows:

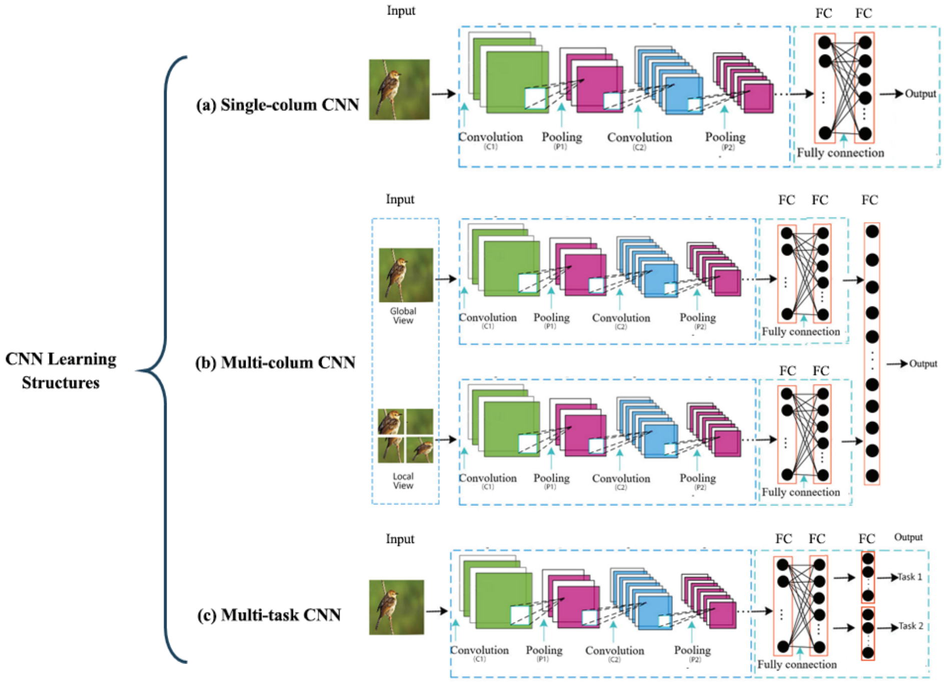

Fig. 5. Different CNNs learning structures: (a) Single-column, (b) Multi-column, (c) Multi-task.

— Single-column: The single-column architecture is the most basic type of CNN, in which a single column of layers is responsible for extracting aesthetic features from the input image. The layers usually include convolutional, pooling, and fully connected layers, as can be seen in Figure 5(a). Single-column CNNs are commonly used for image classification tasks that only require the extraction of global or local aesthetic features.

— Multi-column: Unlike single-column, multi-column CNNs are composed of several columns of layers that extract both global and local features from the input image. Each column is dedicated to extracting a unique set of features, and their outputs are merged to generate the final prediction, as can be seen in Figure 5(b). Multi-column CNNs are commonly utilized for tasks that need more than one extraction approach, such as the extraction of both global and local aesthetic features.

— Multi-task: A multi-task CNN is an architecture that can execute multiple interrelated tasks concurrently. In a multi-task, the layers are shared across all tasks, as illustrated in Figure 5(c), enabling the network to learn representations that are advantageous for all tasks. Multi-task CNNs are beneficial in scenarios where the tasks share some common aesthetic features, for instance, classification and regression scores or distribution and description.

4.2.3 Feature Extraction Approaches. Feature extraction involves capturing relevant visual characteristics from images to evaluate their aesthetic quality. This process is categorized into local and global feature extraction. Deep learning models are then trained to automatically extract these features, enabling them to capture complex patterns and hierarchical representations in the data Local feature extraction involves capturing information from specific regions or patches within an image. This can be done through techniques such as cropping, which involves selecting a specific area of the image to analyze [178]. Local feature extraction can be useful for capturing details such as texture or color, however, can be limited in its ability to capture the overall semantic meaning of the image. Global feature extraction, however, involves analyzing the image as a whole to capture its overall spatial and semantic layout. This can be useful for capturing higher-level concepts such as composition, balance, and harmony, which are important factors in image aesthetics. Global feature extraction can be achieved through techniques such as resizing, which involves changing the overall size of the image while maintaining its aspect ratio. In many models to achieve both global and local feature extraction, a multi-column deep learning approach is necessary [36, 178].

The **Graph Convolution Neural Networks (GCNNs)** are specialized neural networks designed to operate on graph-structured data, where the relationships between data points are as important as the data points themselves. Unlike traditional CNNs, GCNNs generalize the concept of convolution to graphs, making them ideal for tasks where data can be represented as nodes and edges. In GCNNs, feature extraction is achieved by aggregating features from a node's neighborhood and combining them in a way that respects the graph structure. This enables the network to learn representations that capture both the local and global graph topology, making GCNNs powerful tools for IAQA. Attention mechanisms are techniques that allow neural networks to focus on specific parts of the input data when making predictions, enabling the model to dynamically weigh the importance of different features. In feature extraction, attention mechanisms help models selectively highlight important regions of an image or specific features in a sequence, effectively filtering out irrelevant information. This targeted focus enhances the model's ability to capture both local details (such as texture or small objects) and global features (such as the overall structure or context), leading to improved performance in IAQA. Transformer models are another technique for feature extraction in IAQA. The key innovation of transformers is their use of self-attention mechanisms, which allow the model to weigh the importance of different input data elements relative to each other, regardless of their position. This enables transformers to capture long-range dependencies and contextual relationships more effectively than traditional architectures like CNNs. Vision Transformers have been developed to apply the transformer architecture to grid-like image data, dividing images into patches and treating them as sequences. This approach enables the model to extract both local features (within individual patches) and global features (by analyzing the relationships between patches across the entire image). By doing so, transformers can extract high-level features across the entire image, improving performance in IAQA.

4.3 Output Phase

The output phase can be divided into four main tasks: aesthetic scoring, distribution, attribute evaluation, and description. Aesthetic scoring categorizes images by their perceived aesthetic quality, either through binary classification (high or low quality) or regression (continuous scores, such as from 1 to 10). Since image aesthetic scores are unable to reflect the preferences of users entirely, the distribution score is an approach to characterize the disagreement among users' aesthetic preferences regarding the image. Aesthetic attributes evaluate each aspect of images in terms of color, composition, and texture such as the rule of thirds, golden ratio, symmetry, leading lines, balancing element, color harmony, vivid color, depth of field, contrast, lighting, use of shape, viewpoint, framing, and several other attributes involved in aesthetics. In addition to scoring, distribution, and attribute evaluation, descriptions can be submitted to analyze the factors that contribute to an image's attractiveness or the emotions it elicits from users, which can be extracted from either the images themselves or accompanying comments. The architecture for describing image aesthetics employs techniques such as **Long Short-Term Memory (LSTM)** [133] to generate output descriptions. These generated outputs are then compared to the ground-truth human translations existing in the dataset to evaluate the quality of machine translation. The similarity between the generated outputs and reference translations is measured using various factors,

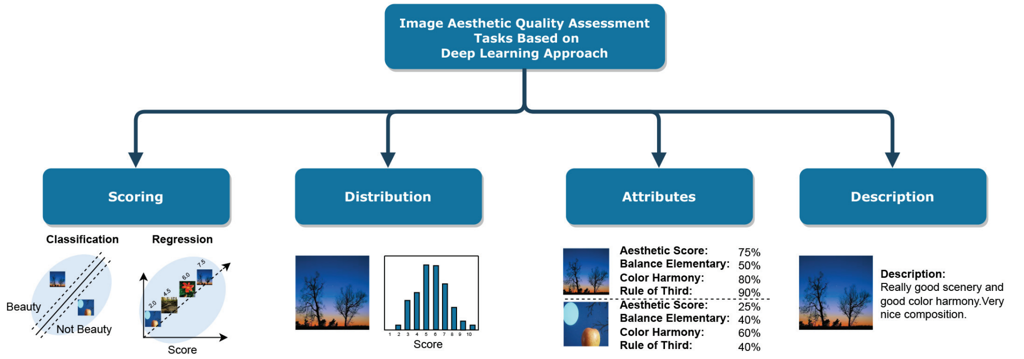

Fig. 6. Various outputs of image aesthetic quality assessment tasks through deep learning approach.

including word order, sentence-level similarity, synonymy, and paraphrasing [20, 103, 112, 172]. These tasks are depicted in Figure 6 with the corresponding examples of each task.

5 A Decade of Image Aesthetic Quality Assessment Model Using Deep Learning

A decade-long study on image aesthetic quality assessment methods explores four key outputs of aesthetic quality: scoring, distribution, attribute, and description output.

5.1 Aesthetic Scoring Methods

There are some methods that use pairwise strategy, which, instead of direct classification, the model learns to compare two images and determine which one is more pleasing by focusing on relative aesthetic [72, 79]. In addition to models that use pairwise strategy in the evolving field of aesthetic quality assessment, various methodologies have emerged to enhance the precision and reliability of scoring systems. This section focuses on several prominent approaches, including attention mechanisms, regional dependencies, minimizing label reliance, personality-driven, user-oriented, and combining algorithms and features.

5.1.1 Attention Mechanism Models. Attention-based multi-patch techniques are widely used to fuse holistic scene information and fine-grained details from different patches to extract features, as demonstrated [126, 178]. A double-subnet gated peripheral-foveal convolutional neural network was proposed in Reference [176]. This network captures holistic information and detects the attended regions using peripheral vision while acquiring details from critical areas with foveal vision. A multi-modal self-and-collaborative attention network was presented in Reference [175], which incorporates a self-attention module to encode the spatial interaction of visual elements and capture the global composition of images. The global composition derives the response at a position by considering all positions in the images. A multi-modal recurrent attention convolutional neural network was introduced in Reference [177], which includes a vision stream and a language stream. The vision stream uses the recurrent attention network to filter out irrelevant data and retrieve visual details from regions, while the language stream employs text convolutional neural network to assess the high-level semantics of user comments. Finally, to merge the visual and textual data successfully, a multimodal factorized bilinear pooling strategy is used.

5.1.2 Regional Dependencies Models. Graph neural networks such as the GCNs are popular in IAQA due to their ability to capture the underlying graph structure of images, with image pixels or regions as nodes and their interactions as edges. Pre-processing techniques such as cropping an image [95] or resizing [56] can lead to the loss of crucial aesthetic information, as these operations may alter the image's original aspect ratio or remove important visual elements. To address the issue of aspect ratio caused by resizing or cropping a method was introduced by using graph attention network [38]. The proposed methods with a two-stage framework are based on graph neural networks that maintain the original aspect ratio and resolution of the input image and capture semantic relationships between different regions of the image using visual attention. Another approach was employed in Reference [90] that utilized a graph neural network to model the mutual dependency of local regions in an image. The approach densely partitions the image into local regions and computes aesthetics-preserving features over them to characterize the aesthetic properties of image content. The region composition graph is built using the feature representation of local regions, and reasoning is performed on the graph via graph convolution. The activation of each node is determined by its highly correlated neighbors, thus naturally uncovering the mutual dependency of local regions in the network training procedure. To capture complex relations among image regions using the GCNs architecture, the researchers designed a new model called **Hierarchical Layout-Aware Graph Convolutional Network (HLA-GCN)** [123]. The HLA-GCN method is a double-subnet neural network consisting of two layout-aware graph convolutional modules that construct aesthetics-related graphs in the coordinate space and perform reasoning over spatial nodes. The output is a hierarchical representation with layout-aware features from both spatial and aggregated nodes for unified aesthetics assessment. Evaluating image aesthetics by considering the thematic context of images by GCN is introduced in Reference [83]. The proposed method involves several key components, a visual attribute analysis network, and a theme understanding network, both of which are pre-trained to extract visual attribute features and theme features, respectively. These networks form the foundation for building two levels of graph-based reasoning. The first level, the attribute-theme graph, explores the relationship between an image's theme and its visual attributes, while the second level, the attribute-aesthetics graph, further examines the connection between these theme-aware visual attributes and general aesthetic features. Together, these components enable the model to predict the aesthetic score, attribute distribution from visual attribute analysis network, and theme probability from the theme understanding network, resulting in a more accurate and context-sensitive aesthetic evaluation.

5.1.3 Minimizing Label Reliance Models. Graph neural networks are also used in a method with a broader set of unlabeled images, enabling personalized assessments without requiring extensive labeled data [87]. The method starts by collecting a small set of images that the user has labeled according to their aesthetic preferences. These labeled images serve as a foundation for understanding the users' specific tastes. The system then uses a transductive learning approach to propagate these individual preferences to a larger set of unlabeled images. This means that the system infers how a user would likely rate other images based on the known preferences from the labeled set. The propagation process is typically implemented using a graph-based method, where images are nodes, and edges represent similarities between images. The users' preferences are propagated through this graph, influencing the aesthetic assessment of similar images and finally assigning personalized aesthetic scores.

A method for evaluating image aesthetics that integrates semi-supervised learning with adversarial techniques aims to reduce dependence on extensive manual attribute annotation by leveraging a semi-supervised learning framework and a partially attribute-annotated dataset [128]. An adversarial training framework is employed to explore the joint distribution of image features, aesthetic attributes, and aesthetic scores. This approach uses aesthetic attributes as privileged information to enhance the performance of a score-attribute generator. Additionally, supervised losses are applied to the networks to predict within commonly occurring ranges, improving the accuracy and robustness of aesthetic assessment. Another technique is the zero-shot method, which allows the model to assess image aesthetics without having been explicitly trained on labeled examples for every possible category or attribute. Instead, the model uses pre-existing knowledge and semantic embedding to generalize its understanding of aesthetic quality to unseen images [142]. This method enables the model to assess aesthetic quality without extensive labeled training data by leveraging both external and internal knowledge. Specifically, the approach involves using an attribute-specific prompt template, where each aesthetic attribute has a unique context to extract relevant features from a pre-trained model. The prompts are then fine-tuned based on the similarity between image features and text features. In addition, a quadruplet set is constructed to capture image relationships, and sentiment polarity is used to select anchor images. The aesthetic score is estimated by integrating information from various attributes. Taking advantage of zero-shot learning to assess image aesthetics involves pre-training vision-language models on image-comment pairs, as presented in Reference [69]. This approach learns rich aesthetic semantics in a self-supervised manner, eliminating the need for expensive labeled datasets. The pre-trained model demonstrates various exceptional tasks, including zero-shot learning for image aesthetics assessment, style classification, and image retrieval. To effectively tailor the model for IAA without diminishing its zero-shot capabilities, the authors introduce a lightweight rank-based adapter module. This module leverages text embedding as anchors and explicitly models ranking, allowing for superior performance with minimal additional parameters.

The IAQA without manual annotations is also possible, as demonstrated in Reference [124]. The authors employed self-supervised learning techniques to minimize the need for manual annotation and extract useful details for image aesthetic assessment. The method aims to develop a feature representation that effectively distinguishes between different expert-designed image manipulations, which are closely related to negative aesthetic effects. By using self-supervised learning, the model autonomously extracts and understands aesthetic features from images through pretext tasks that do not require manual labels. This approach trains the model to recognize and represent aesthetic qualities based on inherent data patterns and structures, enabling effective aesthetic assessment without extensive labeled data.

5.1.4 Personality Driven Models. Incorporating personality features into aesthetic evaluation is another approach to assessing aesthetics. An end-to-end personality-driven multi-task deep learning approach was proposed in Reference [84] to evaluate the aesthetics of an image. An inter-task fusion of personality features driven by Big-five personality and generic aesthetics was employed to personalize the image score. This approach leverages a multi-task learning model to capture the influence of individual personality traits such as agreeableness, conscientiousness, extroversion, neuroticism, and openness on aesthetic preferences. By integrating these personality-driven features, the model aims to improve the accuracy and relevance of aesthetic assessments by acknowledging the subjective nature of visual appeal. Subsequently, the authors expanded on their previous work [85]. They propose an enhanced multi-task learning framework that not only evaluates generic aesthetic qualities but also tailors assessments to individual user's preferences. This model utilizes personality information to refine aesthetic predictions, offering both generalized and personalized assessments. The integration of personality traits in this way allows for a more sophisticated and user-specific evaluation of image aesthetics, advancing the field of personalized aesthetic assessment.

5.1.5 User-oriented Models. To capture the unique preferences of individual users, several frameworks have been proposed that integrate user feedback into the assessment process, allowing for a more personalized experience. Among these, a user-guided personalized image aesthetic assessment framework for scoring and distribution was proposed in Reference [98].

This framework is based on deep reinforcement learning and takes into account the personal preferences of users in its predictions. The proposed method learns from the feedback provided by the users, in fact, allowing it to leverage user interactions by retouching and gradually adapting to their individual tastes and preferences. Another framework called **User-specific Aesthetic Ranking (USAR)** that personalizes the aesthetic ranking of images based on individual user preferences was presented in Reference [161]. Unlike the models that apply generic aesthetic criteria, USAR involves users directly in the assessment process by gathering their feedback on image rankings. Through an interactive loop, the system learns which aesthetic features such as color, contrast, and composition are most important to each user. This approach results in a personalized image ranking, tailored to the unique tastes of the users.

5.1.6 Combining Algorithms and Features Models. Various neural network architectures by combining different algorithms or feature fusions for IAQA using deep learning are developed. For instance, a framework that leverages features derived from image content and composition, such as color, texture, and spatial layout, is proposed to predict aesthetic scores. They utilize a support vector regression model trained on a dataset of images with human-annotated aesthetic ratings. A multi-task CNN network is proposed for assessing the aesthetic quality of images using a hierarchical framework [66]. The authors propose a multi-level approach where images are analyzed and scored based on three primary attributes: scene, object, and texture. Each of these attributes is predicted separately by specialized CNNs trained to recognize patterns associated with aesthetic quality in these specific aspects. The hierarchical framework combines the predictions from these attributes to extract local features from texture, the global features from scene, and saliency detecting from object, then produce a final aesthetic score. Another study focused on a different aspect of aesthetics prediction by leveraging multi-level spatially pooled features from a network is proposed in Reference [43]. The method extracts features at multiple scales and from different convolutional layers to capture a wide range of image attributes. The emphasis is on combining these features from various levels of the network to effectively predict aesthetic quality, considering both global and local image details. A network architecture was designed to extract various image attributes and predict aesthetic rating [86]; similarly, a study proposes a method for assessing the aesthetic quality of images using a regression model [67]. An attempt was made to improve feature elimination and fuse learned features using CNNs in the study conducted in Reference [76]. The utilization of hand-crafted features based on domain expert feature knowledge in photography was considered to enhance image aesthetic inference. Another model integrates various attributes—such as composition, color, and lighting—that contribute to the overall score of images [81]. In Reference [93], researchers introduced an approach to evaluating image aesthetics through deep semantic aggregation. This method utilizes a deep CNN to gather and integrate semantic information from multiple layers, capturing both high-level contextual and low-level visual details for a more nuanced aesthetic assessment. A key component of their approach is the use of the ordered weighted averaging pooling layer, which provides flexibility in how different features are aggregated. The ordered weighted averaging operator can automatically learn the aggregation rule by adjusting the weights assigned to various semantic features based on their relevance to aesthetic quality.

Unlike studies that employ graph neural networks to tackle pre-processing issues and preserve aspect ratio, a different strategy by incorporating adaptive fractional dilated convolution into the network architecture is presented in Reference [12]. The key focus is on maintaining the integrity of an image's aspect ratio and spatial structure during feature extraction, which is crucial for accurate aesthetic evaluation. The adaptive dilation rate adjusts based on the image's aspect ratio, ensuring the spatial structure and composition are preserved. A unified probabilistic formulation for multiple tasks of IAQA, including classification, regression, and distribution, was conducted in Reference [169]. Employing this unified framework allows for the development of effective loss functions. Furthermore, they tackled the issue of learning from noisy raw scores. A deep convolutional neural network is specifically designed to extract hierarchical features that encapsulate both global and local aesthetic attributes of images [95]. A double-column deep convolutional neural network is employed to integrate global and local views, effectively combining feature extraction with classifier training. Additionally, style attributes are incorporated to enhance the accuracy of aesthetic quality categorization. Another method that connected local and global features was introduced by the researchers in Reference [36], resulting in predicting aesthetic ratings. A comparison of various IAQA methods based on deep learning, focusing on feature extraction techniques, architectures, and accuracy, is depicted in Table 2. These accuracies have been obtained from the AVA dataset. As seen in this table, the majority of the classification accuracies are low.

Table 3 focuses on the regression task of various image aesthetic quality assessment methods on the AVA dataset. The evaluation metrics such as **Spearman Rank Correlation Coefficient (SRCC)**, **Pearson Linear Correlation Coefficient (PLCC)**, **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, **Root Mean Squared Error (RMSE)**, and **Standard Deviation (STD)** are calculated to assess the relationship between ground-truth scores and prediction scores. Each metric carries its advantages and limitations, and by examining several metrics, a more refined comprehension of the model's performance across various dimensions is achieved. Although employing a combination of these metrics offers a thorough assessment of the model's effectiveness, the use of all metrics together may be time-consuming, and the most frequently used metrics for regression scores are SRCC and PLCC. For instance, SRCC and PLCC shed light on correlation and monotonic patterns in predictions, while MAE, MSE, and RMSE offer insights into the size and direction of errors.

5.2 Aesthetic Distribution Methods

An initial attempt proposed a CNN-based model that represented the aesthetic distribution of ratings as a histogram, instead of using conventional methods of image classification into low or high scores or estimating the mean score through regression [136]. A framework to predict the distribution of aesthetics was presented in References [45, 46], where the weighting of object-level regions is learned in two stages. Then, shared weights are used for regional and global features extracted to image aesthetic assessment by the attention-based mechanism and graph attention-based aggregation. A deep neural network model that combines low-level visual features and high-level semantic information to predict the aesthetic quality of an image was proposed in Reference [17]. The model uses a hybrid network architecture that consists of a CNN and a recurrent neural network, where the CNN extracts low-level features, and the recurrent neural network incorporates semantic information. The authors predict a rating distribution to determine whether users' aesthetic preferences about the same image differ. An attention-based and context-aware approach to predict aesthetic distribution was proposed in Reference [159]. To generate the long-range perception of images and concern of multi-level aesthetic details, spatial context and hierarchical context are used, respectively. The problem of predicting and defining scores as a degree of consensus among human raters was analyzed in Reference [64]. They also considered several measures of subjectivity and two prediction frameworks motivated by statistic and information theory for aesthetic distribution. A multi-task framework named aesthetics-based saliency network, facilitated by two jointly distinct branches, was presented in Reference [91], which is used for predicting saliency maps and aesthetic distribution, respectively. Table 4 shows the distribution task of various IAQA methods on the AVA dataset. The evaluation metrics, for instance, **Percentage Correctly Evaluated (PCE)**, **Kullback-Leibler (KL)**, **Jensen-Shannon**

Table 2. A Comparison of Different Deep Learning Methods for Image Aesthetic Quality Assessment on the AVA Dataset Based on Classification Accuracy

Ref.	Features extraction	Architecture	Accuracy
[73]	Multimodal fusion image and user preference	AlexNet	77.33
[57]	Multimodal fusion image and user preference	ResNet-101	82.65
[98]	Multimodal fusion image and user preference	Policy Network	85.10
[85]	Multimodal fusion image and user preference	InceptionNet-V3	83.70
[84]	Multimodal fusion image and user preference	DenseNet-121	81.50
[161]	Multimodal fusion image and user preference	AlexNet	78.05
[180]	Multimodal fusion Text with image	Densenet-161	84.32
[111]	Multimodal fusion Text with image	VGG-16	82.85
[177]	Multimodal fusion Text with image	InceptionNet-V3	85.71
[183]	Multimodal fusion Text with image	VGG-16	78.19
[97]	Local and Global	AlexNet	75.41
[8]	Local and Global	EfficentNet-B4	80.75
[38]	Local and Global	Inception-ResNet-V4	81.15
[123]	Local and Global	ResNet-101	84.60
[135]	Local and Global	Xception	78.10
[46]	Local and Global	InceptionResNet-V2	81.93
[45]	Local and Global	Inception-ResNet-V2	81.67
[175]	Local and Global	InceptionNet-V3	86.66
[68]	Local and Global	ResNet-50	81.50
[54]	Local and Global	InceptionNet-V3	82.40
[86]	Local and Global	SERessNet-50	82.82
[12]	Local and Global	ResNet-50	83.24
[90]	Local and Global	DenseNet-121	83.59
[159]	Local and Global	InceptionNet-V3	80.90
[181]	Local and Global	MobileNet-V2	82.35
[176]	Local and Global	InceptionNet-V1	81.81
[178]	Local and Global	VGG-16	82.95
[43]	Local and Global	InceptionResNet-V2	81.72
[148]	Local and Global	ResNet-50	80.55
[36]	Local and Global	InceptionNet-V4	82.66
[23]	Local and Global	VGG-16	78.72
[32]	Local and Global	ResNet-50	78.99
[100]	Local and Global	VGG-16	82.50
[145]	Local and Global	InceptionNet-V3	80.09
[102]	Local and Global	VGG-16	77.40
[128]	Local and Global	ResNet-152	83.72
[95]	Local and Global	Specialized	74.46
[66]	Local and Global	Specialized	74.50
[174]	Global	AestheticNet	80.70
[124]	Global	ResNet-18	82.80
[169]	Global	ResNet-101	80.81
[79]	Global	ResNet-50	91.50
[136]	Global	Inception-ResNet-V2	81.73

(Continued)

Table 2. Continued

[76]	Global	ResNet-50	81.95
[72]	Global	InceptionNet-V1	82.20
[105]	Global	InceptionNet-V3	79.38
[61]	Global	InceptionNet-V1	80.08
[171]	Global	VGG-16	78.87
[122]	Global	ResNet-50	75.83
[108]	Global	ResNet-101	80.30
[65]	Global	ResNet-50	79.08
[149]	Global	VGG-16	76.90
[42]	Global	InceptionNet-V1	82.27
[153]	Separate Local and Global	VGG-16	82.00
[83]	Global	ResNet-50	85.10
[93]	Global	ResNet-152	78.60
[44]	Global	ResNeXt101	82.1
[142]	Global	ViT-B/32	73.30
[24]	Local	AlexNet	78.92
[126]	Local	ResNet-18	83.03
[146]	Local	InceptionNet-V3	79.38
[152]	Local	AlexNet	76.94

(JS), **Chamfer Euclidean Distance (CED)**, **Chamfer Jensen-Shannon (CJS)**, **Chi-squared Statistic** (X^2), **Earth Mover's Distance (EMD)**, **Chebyshev Distance (Cheb)**, **Clark Distance (Clark)**, **Cosine Similarity (Cosine)**, and **Intersec similarity (Intersec)** are calculated to compare the similarity or difference between probability distributions of predicted distribution and ground-truth distribution. Both EMD and KL are frequently employed due to their theoretical foundations, each offering unique insights into distributional dissimilarity. The EMD is advantageous for capturing shape differences and assesses the minimum cost of transforming one distribution into another, focusing on the distance between individual data points, while KL divergence provides a measure of information loss and divergence between probability distributions.

5.3 Aesthetic Attributes Methods

Apart from methods that predict aesthetic attributes as well as aesthetic scoring [81, 83, 86], some methods are designed to predict the aesthetic attribute especially. A deep learning semantic context-based method capable of predicting composition and style attributes alongside an assessment of aesthetic distribution was presented in Reference [8]. The method has a pre-trained network to extract semantic features, a multi-layer perceptron network to indicate aesthetic attributes, and a self-adaptive hypernetwork to predict the parameters of the aesthetic assessment network. In another study with the help of EfficientNet, which estimated eight aesthetic attributes, a multi-task deep convolutional network was designed. Additionally, a visualization technique for comprehending the representation of attributes that determine the key regions was developed using backpropagation of gradients [34].

An adversarial learning framework assisted by attributes for IAQA was proposed in Reference [114]. Motivated by the generative adversarial networks framework, the introduction of adversarial learning aims to capture the correlation between aesthetic attributes and aesthetic score, ensuring a close correlation between the distributions of ground truth and prediction. Moreover, a discriminator is used to determine the predictions from the real labels and supervised learning

Table 3. Comparative Analysis of Image Aesthetic Quality Assessment Methods for Score Regression on the AVA Dataset

Ref.	SRCC	PLCC	MAE	SRCC-std	PLCC-std	MAE-std	MSE	RMSE
[101]	0.6930	0.7070	-	-	-	-	-	-
[180]	0.8528	0.8683	0.2928	-	-	-	-	0.3759
[38]	0.7620	0.7640	-	-	-	-	-	-
[28]	0.6890	-	-	-	-	-	-	-
[8]	0.7318	0.7329	0.4011	-	-	-	-	0.5128
[111]	0.7930	0.7930	-	-	-	-	-	-
[54]	0.7736	0.7753	0.7562	0.7562	0.7512	-	0.2305	-
[175]	0.8475	0.8600	0.3106	-	-	-	-	0.3979
[174]	0.6810	0.7020	-	0.2210	0.2160	-	-	-
[123]	0.6650	0.6870	-	-	-	-	0.2550	-
[46]	0.7560	0.7570	-	0.3620	0.3740	-	-	-
[45]	0.7510	0.7530	-	0.3530	0.3630	-	-	-
[135]	0.7070	0.7070	0.6220	0.1500	0.1550	0.1570	-	-
[68]	0.7260	0.7380	-	-	-	-	0.2420	-
[98]	0.6920	-	-	-	-	-	-	-
[145]	0.7004	0.7096	-	-	-	-	0.2637	-
[159]	0.7240	0.7250	-	-	-	-	-	-
[85]	0.6770	-	-	-	-	-	-	-
[129]	0.6578	-	-	-	-	-	-	-
[162]	0.7072	0.7100	-	-	-	-	-	-
[12]	0.6489	0.6711	-	-	-	-	0.2706	-
[181]	0.7480	0.7600	-	-	-	-	-	-
[86]	0.6619	-	-	-	-	-	-	-
[15]	0.4940	0.5200	-	0.1600	0.1710	-	-	-
[177]	0.8318	0.8434	0.3252	-	-	-	-	0.4190
[146]	0.6868	0.6923	0.2330	-	-	0.2330	0.2764	-
[84]	0.6800	0.7020	-	-	-	-	-	0.0160
[176]	0.6900	0.7042	0.4072	-	-	-	-	0.5246
[169]	0.7190	0.7200	-	0.2410	0.2470	-	0.2750	-
[43]	0.7560	0.7570	-	-	-	-	-	-
[114]	0.6313	-	-	-	-	-	-	-
[79]	0.9180	-	-	-	-	-	-	-
[99]	0.5160	-	-	-	-	-	-	-
[136]	0.6900	0.6940	0.2520	0.2120	0.2180	-	0.2930	-
[105]	0.6730	0.6860	-	-	-	-	-	-
[72]	0.8711	-	-	-	-	-	-	-
[161]	0.6002	-	-	-	-	-	-	-
[108]	0.7090	-	-	-	-	-	0.2790	-
[128]	0.7740	0.7880	-	-	-	-	-	-
[83]	0.7250	0.7360	-	-	-	-	-	-
[44]	0.7010	0.7220	-	-	-	-	-	-
[142]	0.4460	0.4540	-	-	-	-	-	1.0780
[69]	0.6570	0.6630	-	-	-	-	-	-

Evaluation metrics are Spearman Rank Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC), SRCC-Standard Deviation (SRCC-std), PLCC-Standard Deviation (PLCC-std), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE).

Table 4. A Comparative Analysis of Image Aesthetic Quality Assessment Methods for Distribution Tasks on AVA Dataset

Ref.	PCE	KL	JS	CED	CJS	X^2	CD	EMD	Cheb	Clark	Cosine	Intersec
[180]	-	-	-	-	-	-	-	0.0520	-	-	-	-
[8]	-	-	-	-	-	-	-	0.0439	-	-	-	-
[57]	2.6400	0.0650	0.0190	0.1220	0.0130	0.0280	-	0.0230	-	-	-	-
[54]	-	0.0850	0.0210	-	-	0.0390	0.0390	0.0390	-	-	-	-
[175]	-	-	-	-	-	-	-	0.0354	-	-	-	-
[174]	-	-	-	-	-	-	-	0.0490	-	-	-	-
[177]	-	-	-	-	-	-	-	0.0369	-	-	-	-
[159]	-	-	-	-	-	-	-	0.0520	-	-	-	-
[45]	-	-	-	-	-	-	-	0.0520	-	-	-	-
[85]	-	-	-	-	-	-	-	0.0470	-	-	-	-
[12]	-	-	-	-	-	-	-	0.0447	-	-	-	-
[15]	-	-	-	-	-	-	-	0.0504	-	-	-	-
[145]	-	-	-	-	-	-	-	0.0640	-	-	-	-
[146]	-	-	-	-	-	-	-	0.0650	-	-	-	-
[176]	-	-	-	-	-	-	-	0.0450	-	-	-	-
[169]	-	0.1010	-	-	-	-	0.0650	-	-	-	-	-
[136]	2.6930	0.0810	0.0280	0.1370	0.0290	0.0440	-	0.0500	-	-	-	-
[17]	-	0.0940	-	-	-	-	0.0420	-	0.0860	1.2280	0.9580	0.8490
[61]	2.7600	0.3810	0.0370	0.2600	0.0400	0.0680	-	-	-	-	-	-
[30]	-	0.1200	-	-	-	-	0.0560	-	0.0960	1.2830	0.9440	0.8270
[108]	-	0.1030	-	-	-	-	-	0.0610	-	-	-	-

Evaluation metrics are Percentage Correctly Evaluated (PCE), Kullback-Leibler (KL), Jensen-Shannon (JS), Chamfer Euclidean Distance (CED), Chamfer Jensen-Shannon (CJS), Chi-squared Statistic (X^2), Earth Mover's Distance (EMD), Chebyshev Distance (Cheb), Clark Distance (Clark), Cosine Similarity (Cosine), and Intersec similarity.

for rating the network. With the help of objective and subjective levels, a system was designed to predict the aesthetic quality of images and determine effective features [71]. In addition, semantic information of users' comments is analyzed to comprehend the level of subjectivity. This research showed hand-crafted features related to composition and color harmony highly correlate with standard deviation and mean aesthetic scores.

Another investigation introduced a deep learning framework for evaluating image aesthetics using a model composed of five specialized sub-networks [59]. These sub-networks were designed for classification, regression prediction by combining pseudo-labelling with meta reweighting learning. In a subsequent effort, the authors focus on assessing the aesthetic attributes of images numerically across mixed multi-attribute datasets [58]. Similar to their initial approach, they employ five specialized sub-networks designed for classification, regression, and attribute prediction, concentrating on aspects such as composition, color, and lighting.

5.4 Aesthetic Description Methods

The first attempt at the aesthetic description as multi-aspect captioning was presented in Reference [10], in which different photography aspects are characterized, including subject-contrast aspects, composition, and color arrangement. They propose two approaches named **aspect-oriented (AO)** and **aspect fusion (AF)**. The AO approach is considered the baseline CNN-LSTM model applied to divided training data to create image descriptions. In the AF approach, the descriptions learned from the individual aspects are fused. A soft attention mechanism and vanilla CNN-LSTM caption are used to create image descriptions that could perform better than learning a direct CNN-LSTM model. Another attempt following the first attempt [10] trained the network of a standard CNN-LSTM framework, which was named "clean data and weakly-supervised". This approach exploited the latent aesthetic information in images and applied the Latent Dirichlet Allocation (LDA) topic model presenting a weakly-supervised approach for image aesthetic description [37]. A personalized aesthetic image description method, which consists of two sub-models, Aesthetic feature **Extraction Network (AEN)** and **User Encoder Network (UEN)**, was presented in Reference [157]. The AEN utilizes multi-patch processing to extract detailed local features from images, enhancing its ability to capture nuanced information. Simultaneously, AEN integrates multi-level spatial perception network to broaden perception, prevent overfitting, and efficiently manage memory resources. The combination of these techniques allows AEN to extract both local intricacies and globally relevant multi-level features, ensuring a comprehensive understanding of images. Another sub-model, the UEN, was designed to learn limited semantic information and extract the user preferences from the personalized dataset. A description generator is used in another study for creating aesthetic-based descriptions for images in Reference [167]. They presented a multi-encoder framework that consists of two encoders named Encfact and Encaest. The Encfact pre-trained a CNN on ImageNet, and the Encaest is an encoder trained to perform the assessment, while the decoder part contains the LSTM layer and soft visual attention. Word mover's distance is used to measure the semantic differences between generated descriptions and ground-truth descriptions. In a different approach, Reference [151] employed a CNN and recurrent neural network for image aesthetic classification. This utilized a single-column CNN and a vision-to-language generator to create descriptions for images via LSTM. Similarly, in Reference [62], authors offered an aesthetic multi-attribute network, which is a two-stage training process consisting of channel and spatial attention network, multi-attribute feature network, and language generation network. The proposed method can employ five aesthetic attributes including color and lighting, composition, depth and focus, impression and subject, and the use-of-camera for a description task beside the numerical score to assess the aesthetic quality of images. A model with two examinations that used the LDA for aesthetic topics and active learning to screen sentences is proposed in Reference [60]; the first model of this method is implemented on ResNet-101 architecture with the LSTM model for generating sequences. This model used sequence generation models employing an Encoder to convert input into a fixed form and a Decoder to generate sequences word-by-word. The Encoder uses CNNs to encode input images with three color channels. The ResNet-101 encoder progressively reduces the image size, creating smaller representations with more learned features, ultimately producing a final encoding of size 14 × 14 with 2,048 channels. Moreover, they implemented their methods with the Bottom Up Attention method; in this case, they obtained image features through ResNet-101, referred to as grid-based features, and handled description data. Subsequently, they utilized the Bottom Attention Up Model for description generation, which they found in this situation performance can be increased. Table 5 depicts a performance comparison of methods evaluated by metrics, such as **Bilingual Evaluation Understudy (BLEU)** [115], **Metric for Evaluation of Translation with Explicit ORdering (METEOR)** [78], **Recall-Oriented Understudy for Gisting Evaluation (ROUGE)** [89], and **Consensus-based Image Description Evaluation (CIDEr)** [140], which are algorithms for evaluating the quality of text translated by machine.

6 Impact of Environment on Aesthetics

Aesthetics are influenced by environmental factors such as culture, history, and religion, which shape personal preferences. Consequently, cultural disparities contribute to aesthetic judgment, as historical and ecological influences shape individual tastes and impact fashion and art styles over time [120]. Generally, the impact of the environment on aesthetics can be categorized into two

Table 5. Performance Comparison of Different Aesthetic Image Description Methods Based on Deep Learning Using Different Datasets

Ref.	Year	Dataset	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE	CIDEr
[60]	2022	FAE-Captions	0.503	0.252	0.154	0.113	-	-	-
[167]	2021	AVA-Captions	0.464	0.238	0.122	0.063	0.109	0.262	0.051
[157]	2020	AVA-Pcap	0.142	0.069	0.035	0.015	0.058	0.135	0.094
[37]	2019	AVA-Captions	0.535	0.282	0.150	0.074	0.107	0.254	0.059
[151]	2019	AVA-Reviews	0.495	0.264	0.145	0.074	0.115	0.261	0.060

Evaluation metrics are Bilingual Evaluation Understudy (BLEU), Metric for Evaluation of Translation with Explicit ORdering (METEOR), Recall-Oriented Understudy for Gisting Evaluation (ROUGE), and Consensus-based Image Description Evaluation (CIDEr).

groups: *non-acquired identity and aesthetic perception* and *acquired identity and aesthetic perception*, which will be discussed in the following sections.

6.1 Non-acquired Identity and Aesthetic Perception

Non-acquired factors such as age and gender can play a role in how people evaluate aesthetic qualities in images. Taking into account these demographic differences, researchers are able to improve the accuracy of their predictive model for estimating aesthetic preferences in images. A study analyzed on AVA-dataset to investigate how demographic factors such as age, gender, and ethnicity influence peoples' aesthetic preferences for portrait images [63]. Researchers found that younger participants tended to prefer brighter images with higher contrast and saturation, while the older participants preferred images with more muted colors and less contrast. In addition, women tended to prefer images with warmer tones and smoother textures, while men preferred sharper images with more contrast. This study indicates that aesthetic choices can be influenced by age and gender, which highlights the impact of these demographic factors on individuals' preferences.

6.2 Acquired Identity and Aesthetic Perception

The acquired factors can be categorized into regional attachments of aesthetics, the impact of color and symbolism in cultural identity on aesthetics, and experience and history on aesthetics. The following subsections discuss these categories.

6.2.1 Regional Attachments of Aesthetics. In terms of ethnicity and the impact of culture on aesthetics, a study found that people from East Asian countries tended to prefer images with higher contrast and saturation compared to those from Western countries [63].

Another study on visual perception and aesthetic valuation of natural landscapes in Russia and Japan presented findings that, despite their shared border and similar natural environments, these countries differ significantly in their cultural and historical contexts [116]. This study reveals that cultural traditions and familiar natural environments significantly influence aesthetic judgments. Japanese participants tend to appreciate landscapes that align with traditional Japanese aesthetics, such as simplicity, balance, and seasonal changes. Russian participants, however, favor landscapes that reflect the vastness and diversity of Russian scenery, emphasizing grandeur and natural wilderness. Differences emerged in the evaluation of seacoasts, rivers, forests, and swampy plains, with Russian respondents favoring exotic landscapes more than Japanese respondents. These findings illustrate that while some aesthetic principles may be universal, local cultural and environmental contexts play a crucial role in individual preferences. Neural responses to art appreciation are shaped by cultural background, as explored in Reference [165]. Using functional magnetic resonance imaging, the researchers observed brain activity in Eastern (Chinese) and Western (European) participants as they viewed traditional landscape paintings. Eastern participants showed stronger activation in areas related to holistic and contextual processing when viewing Chinese paintings, while Western participants exhibited stronger activation in areas associated with object-focused and analytical processing when viewing European paintings. Participants tended to prefer paintings from their own culture, indicating that cultural context strongly influences aesthetic perception, affecting neural processes.

To investigate how cultural identity influences aesthetic preferences, a study was conducted with participants from Eastern (e.g., Chinese) and Western (e.g., American) backgrounds [5]. Participants were asked to evaluate traditional art from both cultural traditions. The findings indicated that participants generally favored art from their own culture. Both explicit (self-reported) and implicit (response time tasks) measures showed this cultural bias, though implicit measures indicated subtler biases, suggesting an unconscious influence of cultural identity on aesthetic judgments.

6.2.2 Color and Cultural Identity on Aesthetics. Related to culture and color, authors in Reference [3] explored how the meanings of colors can vary significantly across cultures, impacting their effectiveness as marketing tools. For example, red is associated with excitement and love in Western cultures, but in Eastern cultures like China, it symbolizes luck and prosperity. Similarly, blue conveys trust and calm in the West, while in the Middle East, it represents safety and protection. Other colors also discuss how green, white, and black carry different connotations in various cultural contexts. In Western cultures, green often signifies nature and health, whereas in China, it can have mixed meanings related to fertility and infidelity. White is seen as pure and peaceful in the West but is associated with mourning in Eastern cultures. Black generally represents sophistication in the West but can be linked with mourning and loss in many cultures. Colors such as yellow, purple, and pink further illustrate cultural diversity in color meanings. Yellow conveys optimism and caution in the West but is a symbol of power and prosperity in China. Purple represents royalty and luxury in Western contexts, while in Thailand, it signifies mourning. Pink is linked with femininity and love in the West, whereas in Japan, it symbolizes youth and good health. Understanding these cultural nuances helps marketers create more culturally sensitive and effective strategies.

6.2.3 Symbolism and Cultural Identity on Aesthetics. The symbolic meanings of flowers, birds, plants and trees, numbers, and other elements that convey meaning are influenced by cultures, as noted in Reference [16]. According to this study, the rose is a universal symbol of love and romance, with its colors conveying different sentiments—red for deep love, yellow for friendship, and white for purity and new beginnings. The lotus, revered in many Eastern traditions, symbolizes purity, spiritual awakening, and rebirth, with its emergence from muddy waters representing the journey from ignorance to enlightenment. The lily is associated with purity and innocence, often linked to the Virgin Mary in Christian symbolism, while also representing death and resurrection in various cultural contexts. Meanwhile, the orchid stands for exotic beauty, strength, and refinement, symbolizing love and the delicate balance of emotions. These diverse meanings illustrate how flowers serve as powerful symbols, reflecting cultural values, emotions, and spiritual concepts and highlighting their importance in artistic and ritualistic expressions across different societies. The symbol of birds, in Reference [16] explores the rich symbolic meanings associated with various birds, revealing their deep cultural and spiritual significance. The eagle, for instance, stands as a powerful symbol of freedom and strength, revered for its majestic flight and keen vision. In many traditions, it represents a high perspective and independence, while in Native American cultures, the eagle is seen as a spiritual messenger embodying divine insight and connection to the higher realms. The dove, however, is universally recognized as a symbol of peace, purity, and love. In Christian iconography, it represents the Holy Spirit and divine peace, often depicted in the context of the Annunciation and Baptism, and is also associated with hope and reconciliation through its role in the Noah's Ark story. The owl is another bird with profound symbolic meaning, representing wisdom and knowledge. In Greek mythology, the owl was the sacred bird of Athena, the goddess of wisdom. Additionally, owls are symbols of mystery and transition, associated with the processes of change and transformation. The phoenix, with its mythological association with rebirth and immortality, symbolizes the cyclical nature of life. Meanwhile, the sparrow symbolizes joy and the beauty of simplicity, representing the small pleasures in life and the vitality of everyday existence. It is also associated with community and loyalty, highlighting the importance of social bonds. The raven symbolizes mystery, magic, and change, acting as a guide to the spirit world. It also represents intelligence and adaptability, known for its problem-solving skills and resourcefulness. These diverse bird symbols reflect their varied roles in art, religion, and cultural traditions, showcasing their significance in conveying powerful messages and embodying key aspects of the human experience.

6.2.4 Experience and History on Aesthetics. Individual tastes are influenced by history, as examined in Reference [51], which explores how aesthetic appreciation is shaped by a complex interplay of cognitive neuroscience, cultural context, historical background, and individual differences. By integrating insights from psychology, neuroscience, cultural studies, and art history, the study highlights that the brain's processing of aesthetic experiences involves key regions such as the reward system and the visual cortex. These neural mechanisms, however, are significantly influenced by the cultural environment and personal experiences, leading to varied standards and criteria for beauty across different cultures and historical periods. The subjectivity of aesthetic preferences is emphasized, with individual traits such as personality, education, and personal history identified as modulating factors.

7 Applications of Image Aesthetic Quality Assessment

Image aesthetic quality assessment algorithms can be applied in various fields, including but not limited to automatic image generation, fashion, content creation, product design, e-commerce, social networks, image enhancement, image recommendation, and art criticism. These applications are primarily used in two distinct scenarios: First, when images do not exist and there is a need to generate them while considering aesthetic aspects; and second, when existing images require aesthetic improvement or critical evaluation. To enhance clarity, the following section explains some studies related to the mentioned IAQA applications.

7.1 Content Generation with Aesthetic Considerations

The IAQA algorithms offer significant benefits in situations where images are non-existent and there is a desire to generate images while considering aesthetic aspects such as fashion industry [14, 48], automatic image generation [6, 74, 107, 170], gaming, content generation, and product design [1, 80, 125, 127, 147, 182, 184]. The IAQA algorithms in artificial intelligence image generation, such as text-to-image generative models, produce synthetic images based on human preferences that create high-quality, aesthetically pleasing images. This approach enhances models by integrating detailed human feedback and personal preferences, ensuring that the generated images resonate more effectively with users [88, 156, 158]. The role of IAQA in automatic image generation, gaming, content generation, and product design is to analyze aesthetic features such as color harmony, balance, and symmetry to ensure the creation of high-quality, aesthetically pleasing images that not only meet technical standards but also possess unique aesthetic appeal. Likewise, these algorithms are beneficial for the creation and recommendation of fashion images to consumers. These algorithms work behind the scenes to sort or create through vast collections of fashion imagery, ensuring that only the most visually appealing content reaches our screens. Beyond this, these methods play a role in understanding and predicting fashion trends. By analyzing patterns and aesthetics in fashion images, these algorithms assist designers and marketers in making informed decisions about styles, colors, and visual presentations. These algorithms not only enhance the visual appeal of fashion-related content but also empower consumers to make style choices that resonate with their personal tastes. Imagine scrolling through your favorite shopping app and only seeing clothing items and accessories that align perfectly with the aesthetic preferences; or installing an application on your mobile that creates the best design, depending on your style.

7.2 Aesthetic Enhancement

The second scenario belongs to when existing images require improvement or need to be criticized from an aesthetic perspective. This application extends to areas such as image enhancement [27, 106, 110, 119, 166], automatic image cropping [40, 82, 94, 150, 179], image recommendation [11, 47, 77, 160], and art criticism [39, 141, 173]. Individuals capture a multitude of images during various life events, ceremonies, and everyday situations. However, a significant portion of these images are often unused due to perceived quality issues. This abundance of useless images not only leads to cluttered digital storage but also strains the memory capacities of our devices. The IAQA algorithms, applied in applications such as image enhancement, automatic image cropping, and image recommendation, offer a solution to this challenge and significantly reduce the number of low-quality images produced. This not only optimizes digital storage but also ensures that the images captured hold aesthetic value, making each photograph more meaningful and memorable in the vast digital archives of our personal and collective experiences. Correspondingly, in art criticism applications, the usage of IAQA algorithms holds notable benefits, particularly in the context of photography competitions. By incorporating these algorithms into the evaluation process, a significant advancement is made in mitigating individual biases and expediting the assessment of entries. Traditional art criticism processes are susceptible to subjective interpretations, influenced by individual preferences. However, the IAQA algorithms provide an objective and standardized framework for analyzing the aesthetic quality of images. This not only ensures a fair and impartial evaluation but also streamlines the judging process, making it more efficient. However, there are instances where certain applications can derive benefits from both categories, exemplified by social media platforms [18] and e-commerce [154, 155].

8 Existing Challenges in IAQA Using Deep Learning Model

The IAQA based on deep learning faces several challenges that impact its effectiveness in these approaches. These challenges include issues related to datasets, which can affect assessment of image aesthetic quality. Additionally, there are difficulties encountered at each phase of the deep learning process—input, processing, and output. These challenges will be discussed in detail in the following sections.

8.1 Challenges in Dataset

One of the key challenges in image aesthetic quality assessment lies in effectively labeling datasets. Using binary classification, where images are labeled either “0” for low aesthetic quality or “1” for high, can be overly simplistic, as it fails to capture the complexity of aesthetic judgments. Continuous scoring offers a more nuanced approach, but it raises issues around determining the threshold that separates high aesthetic quality from low. Establishing a meaningful boundary becomes subjective and can vary, depending on the dataset or task. For instance, datasets such as EAD, IAD, and the **Chinese University of Hong Kong Photo Quality (CUHKPQ)** use a binary

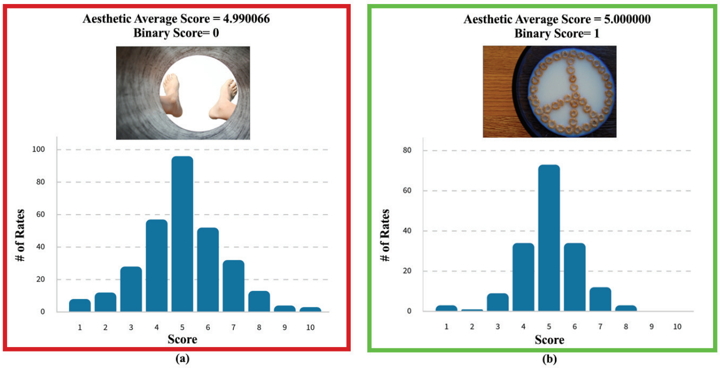

Fig. 7. Illustration of the threshold problem in a binary scoring system for image aesthetics assessment. (a) Images with average scores just below threshold 5 are classified as “0” or “not beautiful,” and (b) Images with average scores just above threshold 5 are classified as “1” or “beautiful,” despite having nearly indistinguishable aesthetic qualities.

scoring system, classifying images simply as beautiful or not. In some datasets, like the AVA dataset, images are scored by thresholding the average score just above or below a set threshold (e.g., >5 is considered beautiful and <5 is not). However, this distinction is often too narrow, as depicted in Figure 7, and fails to capture the subtleties of aesthetic appeal. Another challenge is the distribution of aesthetic scores, which complicates the use of statistical measures and makes it difficult to understand the reasons behind the differences in raters’ scores. This variability often reflects subjective preferences, making it harder to establish consistent evaluation criteria. While datasets like AVA provide a distribution of scores, relying on the average score of these distributions can be misleading. Different distributions can have the same average but vary significantly in their consensus among raters. As shown in Figure 8, two images might share an identical average aesthetic score, but the underlying rater distributions could vary. As we mentioned the problem of binary aesthetic score and average aesthetic score, some approaches assess the images from the distribution of scores using entropy. Two distributions of scores may have the same entropy, as depicted in Figure 9, and one may exhibit a higher degree of subjectivity. This variation highlights the need for more sophisticated metrics that account for the full diversity of subjective responses. While some metrics have been developed to address this issue, it remains a significant challenge in evaluating aesthetic distributions [57, 61, 64]. Assessing image aesthetic quality using binary scores, average scores, or entropy is often inadequate. These methods fail to capture the nuanced factors that make an image pleasing from different raters’ perspectives. Additionally, relying on photographic rules or attributes (e.g., color balance, lighting) may not fully capture aesthetic appeal, as aesthetic judgment is subjective and context-dependent. Understanding what raters perceive and translating their thoughts into labels or descriptions is difficult. To address the challenge of scoring labels, some research used semi-supervised learning with adversarial techniques [128], or techniques like zero-shot learning are employed to assess image aesthetics without requiring explicit training on labeled examples [142]. However, assessing aesthetics without labeled data can lead to challenges in ensuring accuracy and consistency.

Furthermore, the challenge of limited data in training machine learning models is notable, as a small dataset may not sufficiently encompass the diverse array of aesthetics found in real-world situations. Inadequate data may lead to models that struggle to generalize effectively to unseen examples, potentially introducing biases due to the limited perspectives represented in the

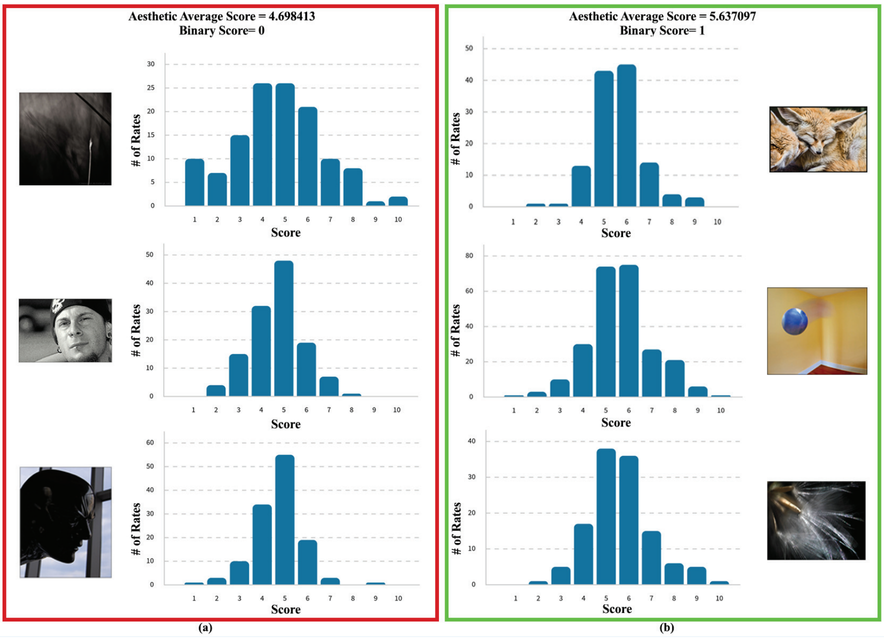

Fig. 8. Illustration of the aesthetic average score problem for image aesthetics assessment. (a) Images with the same average aesthetic scores, while distribution varies, belong to less aesthetic class, and (b) Images with the same average aesthetic scores in high aesthetic class while distribution varies.

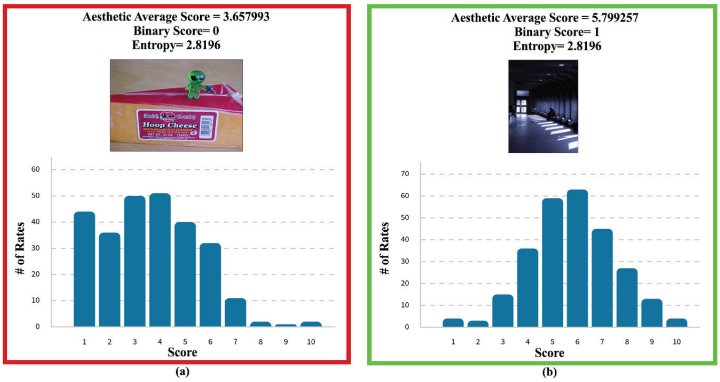

Fig. 9. Visualization of two distributions of scores with similar entropy (2.8196): (a) reflects a tendency toward lower scores, while (b) shows a shift toward higher ratings, indicating a more favorable evaluation. Although both distributions share the same entropy, (b) has a higher degree of subjectivity, aligning with its bias toward more positive scores.

training set. The presence of noise in aesthetic scores and the inherent uncertainty in aesthetic evaluation accentuate the urgent need for expansive, reliable datasets. As well, it is also crucial to recognize the potential imbalance in popular datasets like AVA, where the number of images in each category may not be equal, contributing to imbalanced data. To tackle these challenges, it is recommended to develop a larger, clean, and balanced dataset enriched with diverse user annotations. This strategy involves gathering a more extensive and diverse collection of images, paired with subjective annotations from various users, thereby capturing a broader spectrum of aesthetic preferences. Moreover, it is essential to acknowledge that the origin of the images can significantly influence the data. Images from amateur photographers might differ significantly in quality from those taken by professionals, leading to inconsistencies in aesthetic judgments across different datasets.

The type of images within a dataset—whether real, edited, or a combination of both—can have a profound impact on how aesthetics are perceived and graded. For example, the AVA dataset includes a mix of real and edited photos, providing a diverse range of visual content. However, the presence of both real and edited photos can also introduce complexity in grading, as the criteria for assessing aesthetics might differ between these two types. Datasets such as AADB, **Aesthetic Ratings from Online Data (AROD)**, and CUHKPQ consist of real photos. These datasets focus on capturing and evaluating the natural aesthetic qualities found in unaltered images. On the other end of the spectrum, the **Edited photo Aesthetic Dataset (EAD)** exclusively features edited photos. This dataset is particularly valuable for studying how various editing techniques, such as color correction, retouching, or compositional adjustments, affect aesthetic perception. Considering whether an image is edited or real is crucial, as this distinction influences the assessment of image aesthetic quality.

8.2 Challenges in Input Phase

In the input phase of aesthetic assessment, it is notable that the majority of studies have predominantly employed a *general* methodology, with the utilization of task-specific approaches remaining relatively scarce. As illustrated in Table 2, most of the methods that employed the AVA dataset were unable to achieve high accuracy. One of the challenges lies in the AVA dataset, which encompasses various categories. This distinction becomes particularly significant when we consider the amalgamation of images from diverse categories. Combining images from various categories can pose a substantial challenge, given that the perception of aesthetic appeal is inherently different across these categories. A study attempted to solve this challenge by a technique called knowledge distillation, where knowledge from object classification models, which deal with more concrete and objective labels, is transferred to an aesthetic assessment model. This helps mitigate the abstract nature of aesthetic labels by grounding aesthetic assessment in features that are meaningful for image classification [44]. However, this approach alone was not sufficient to fully capture the complexity of aesthetics, as aesthetic judgment extends beyond categorization inside a dataset. Moreover, due to the diversity of aesthetic features and different applications, the methods need to be task-specific. For example, in the fashion industry, in addition to beauty features such as color and texture, people's movement features are also involved, while in advertising images or images related to faces, these types of features do not have much effect.

8.3 Challenges in Process Phase

The challenges related to the process phase demand careful consideration for advancements in deep learning methods. Simply focusing on architectural changes and fine-tuning methods is insufficient; striking a balance between network architecture, learning structure, and feature extraction approaches is vital, with a keen awareness of the tradeoff between time and accuracy; the impact of parameters on real-time applications must be systematically evaluated. Not all methods exhibit the capability to design independently without relying on transfer learning, and this poses a challenge, because transfer learning, while useful in many scenarios, can introduce dependencies on pre-existing models or datasets. Methods that can be designed independently are valuable, as they are not contingent upon external knowledge or frameworks, showcasing a more self-sufficient and adaptable approach. Another challenge involves considering the aspect ratio in images, which has a significant impact on the aesthetic quality of images. The fixed-size constraint in CNNs can lead to distortions or information loss, particularly in images with diverse aspect ratios. Adapting techniques to account for aspect ratios becomes crucial to ensure that the aesthetic assessment model can effectively handle the varying proportions present in real-world images. Overcoming this challenge involves developing methods that can dynamically adjust to different aspect ratios, preserving the visual integrity of images during the assessment process. Failure to address this challenge may result in biased or inaccurate aesthetic evaluations, especially when dealing with a wide range of image dimensions and proportions.

8.4 Challenges in Output Phase

As previously mentioned, we can categorize aesthetic assessment output into four distinct domains. However, it is important to recognize that these assessments may provide aesthetic scores or descriptions, but they do not inherently furnish us with the causative factors underlying these aesthetic judgments. To explore deeper into the underlying reasons for aesthetic scores, an exploration within the layers of deep learning models becomes necessary. This exploration aims to pinpoint the specific layers and mechanisms within these models that contribute to aesthetic judgments, thereby enhancing our understanding of the **eXplainable Artificial Intelligence (XAI)** aspect. The XAI is an emerging field that attempts to break the black-box nature of machine learning models. Particularly, the ability to explain promotes end-user trust and helps developers to ensure that the system is working well or not [7]. This might involve using appropriate comparison metrics or uncovering unknown features or additional concepts that have not yet been thoroughly investigated. Challenges in image aesthetic quality assessment is the inherent complexity and diversity of the processes underlying image aesthetics. Beauty, often defined as qualities that please the senses or elevate the mind, is a central aspect of aesthetics, but it is highly subjective. Many computational methods for assessing aesthetics adopt an objective approach, treating beauty as an intrinsic property of the image itself, focusing on measurable factors such as symmetry, color balance, or contrast. However, beauty is deeply influenced by individual perspectives shaped by a wide range of factors, including historical and cultural backgrounds, religious beliefs, and even variables such as age and gender. This subjectivity makes it difficult for computational systems to fully capture the nuanced and personal nature of aesthetic experiences. [3, 5, 16, 35, 63, 116, 165].

9 Conclusions

Aesthetics follows certain established photography and art rules, known as aesthetic features such as symmetry, the rule of thirds, and several others. In recent years, these aesthetic features have been leveraged to evaluate image quality using deep learning methods. **Image Aesthetic Quality Assessment (IAQA)** models have diverse applications, including content generation and image enhancement considering aesthetic metrics. In this work, an overview of IAQA through deep learning approaches over the past decade was provided. The functionality of IAQA was organized into input, processing, and output phases. During the input phase, users can select from either all types of images or specific domain images. The processing phase involves convolutional neural network architectures and feature extraction methods, which extract aesthetic features and produce outputs tailored to application requirements, such as scoring, distribution, attributes, and description. Furthermore, we discussed the challenges associated with IAQA when utilizing deep learning. Limitations related to data, such as proper labeling and imbalances, can introduce biases into the models. The widespread use of general methodology in the input phase potentially diminishes the significance of aesthetics across diverse genres of images. In the processing phase, significant challenges arise, such as the crucial tradeoff between time and accuracy, the need for independently designed methods, and the consideration of aspect ratios in IAQA techniques. Moreover, individual tastes are influenced by environmental factors, which can either be acquired or non-acquired identities. Although assessments yield aesthetic outputs, the underlying factors influencing these outputs often remain ambiguous due to the complexities of deep learning and the inherent subjectivity of aesthetics. This highlights the need for a deeper exploration of the layers within deep learning models, emphasizing the growing importance of XAI in this context.

References

[1] Hadi Alzayer, Hubert Lin, and Kavita Bala. 2021. AutoPhoto: Aesthetic photo capture using reinforcement learning. In *International Conference on Intelligent Robots and System*s. 944–951. DOI: https://doi.org/10.1109/IROS51168.2021.9636788

[2] Abbas Anwar, Saira Kanwal, Muhammad Tahir, Muhammad Saqib, Muhammad Uzair, Mohammad Khalid Imam Rahmani, and Habib Ullah. 2022. Image aesthetic assessment: A comparative study of hand-crafted and deep learning models. *IEEE Access* 10 (2022), 101770–101789. DOI: https://doi.org/10.1109/ACCESS.2022.3209196

[3] Mubeen M. Aslam. 2006. Are you selling the right colour? A cross-cultural review of colour as a marketing cue. *J. Market. Commun.* 12, 1 (2006), 15–30. DOI: https://doi.org/10.1080/13527260500247827

[4] Abdolrahman Attar, Asadollah Shahbahrami, and Reza Moradi Rad. 2016. Image quality assessment using edge based features. *Multimedia Tools Applic.* 75 (2016), 7407–7422. DOI: https://doi.org/10.1007/s11042-015-2663-9

[5] Yan Bao, Taoxi Yang, Xiao-Xiong Lin, Yuejing Fang, Yi Wang, Ernst Pöppel, and Quan Lei. 2016. Aesthetic preferences for eastern and western traditional visual art: Identity matters. *Front. Psychol.* 7 (2016), 1–8. DOI: https://doi.org/10.3389/fpsyg.2016.01596

[6] Samah Saeed Baraheem and Tam V. Nguyen. 2020. Aesthetic-aware text to image synthesis. In *Annual Conference on Information Sciences and Systems*. 1–6. DOI: https://doi.org/10.1109/CISS48834.2020.1570617383

[7] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins et al. 2020. Explainable artificial intelligence: Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* 58 (2020), 82–115. DOI: https://doi.org/10.1016/j.inffus.2019.12.012

[8] Luigi Celona, Marco Leonardi, Paolo Napoletano, and Alessandro Rozza. 2021. Composition and style attributes guided image aesthetic assessment. *IEEE Trans. Image Process.* 31 (2021), 5009–5024. DOI: https://doi.org/10.1109/TIP.2022.3191853

[9] Jyotismita Chaki and Nilanjan Dey. 2018. *A Beginner's Guide to Image Pre-processing Techniques.* Chemical Rubber Company. DOI: https://doi.org/10.1201/9780429441134

[10] Kuang-Yu Chang, Kung-Hung Lu, and Chu-Song Chen. 2017. Aesthetic critiques generation for photos. In *International Conference on Computer Vision.* 3534–3543. DOI: https://doi.org/10.1109/ICCV.2017.380

[11] Ling Chen, Dandan Lyu, Shanshan Yu, and Gencai Chen. 2023. Multi-level visual similarity based personalized tourist attraction recommendation using geo-tagged photos. *Assoc. Comput. Machin. Trans. Knowl. Discov. Data* 17, 7 (2023), 1–18. DOI: https://doi.org/10.1145/3582015

[12] Qiuyu Chen, Ning Zhou, Peng Lei, Yi Xu, Yu Zheng, and Jianping Fan. 2020. Adaptive fractional dilated convolution network for image aesthetics assessment. In *Conference on Computer Vision and Pattern Recognition.* 14102–14111. DOI: https://doi.org/10.48550/arXiv.2004.03015

[13] Yanxiang Chen, Yuxing Hu, Luming Zhang, Ping Li, and Chao Zhang. 2018. Engineering deep representations for modeling aesthetic perception. *IEEE Trans. Cybern.* 48, 11 (2018), 3092–3104. DOI: https://doi.org/10.1109/TCYB.2017.2758350

[14] Wen-Huang Cheng, Sijie Song, Chieh-Yun Chen, Shintami Chusnul Hidayati, and Jiaying Liu. 2021. Fashion meets computer vision: A survey. *Assoc. Comput. Machin. Comput. Surv.* 54, 4 (2021), 1–41. DOI: https://doi.org/10.1145/3447239

[15] June Hao Ching, John See, and Lai-Kuan Wong. 2020. Learning image aesthetics by learning inpainting. In *International Conference on Image Processing*. 2246–2250. DOI: https://doi.org/10.1109/ICIP40778.2020.9191130

[16] Farrin Chwalkowski. 2016. *Symbols in Arts, Religion and Culture: The Soul of Nature.* Cambridge Scholars Publishing.

[17] Chaoran Cui, Huihui Liu, Tao Lian, Liqiang Nie, Lei Zhu, and Yilong Yin. 2019. Distribution-oriented aesthetics assessment with semantic-aware hybrid network. *IEEE Trans. Multimedia* 21, 5 (2019), 1209–1220. DOI: https://doi.org/10.1109/TMM.2018.2875357

[18] Chaoran Cui, Wenya Yang, Cheng Shi, Meng Wang, Xiushan Nie, and Yilong Yin. 2020. Personalized image quality assessment with social-sensed aesthetic preference. *Inf. Sci.* 512 (2020), 780–794. DOI: https://doi.org/10.1016/j.ins.2019.10.011

[19] Maedeh Daryanavard and Asadollah Shahbahrami. 2021. Non-distortion-specific no-reference image quality assessment using statistical features. *Sig. Data Process.* 18, 2 (2021), 115–134. DOI: https://doi.org/10.52547/jsdp.18.2.115

[20] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Wang. 2006. Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision*, Vol. 3. 288–301. DOI: https://doi.org/10.1007/11744078_23

[21] Ritendra Datta and James Wang. 2010. ACQUINE: Aesthetic quality inference engine—Real-time automatic rating of photo aesthetics. In *International Conference on Multimedia Information Retrieval.* 421–424. DOI: https://doi.org/10.1145/1743384.1743457

[22] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *Conference on Computer Vision and Pattern Recognition.* 248–255. DOI: https://doi.org/10.1109/CVPR.2009.5206848

[23] Yubin Deng, Chen Change Loy, and Xiaoou Tang. 2017. Image aesthetic assessment: An experimental survey. *IEEE Sig. Process. Mag.* 34, 4 (2017), 80–106. DOI: https://doi.org/10.1109/MSP.2017.2696576

[24] Zhe Dong, Xu Shen, Houqiang Li, and Xinmei Tian. 2015. Photo quality assessment with DCNN that understands image well. In *Multimedia Modeling*. Springer International Publishing, 524–535. DOI: https://doi.org/10.1007/978-3-319-14442-9_57

[25] Zhe Dong and Xinmei Tian. 2015. Multi-level photo quality assessment with multi-view features. *Neurocomputing* 168 (2015), 308–319. DOI: https://doi.org/10.1016/j.neucom.2015.05.095

[26] Dpchallenge. (n.d.). *A Digital Photography Contest.* Retrieved August, 2024 from https://www.dpchallenge.com/

[27] Xiaoyu Du, Xun Yang, Zhiguang Qin, and Jinhui Tang. 2019. Progressive image enhancement under aesthetic guidance. In *International Conference on Multimedia Retrieval.* 349–353. DOI: https://doi.org/10.1145/3323873.3325055

[28] Jiachen Duan, Pengfei Chen, Leida Li, Jinjian Wu, and Guangming Shi. 2022. Semantic attribute guided image aesthetics assessment. In *International Conference on Visual Communications and Image Processing.* 1–5. DOI: https://doi.org/10.1109/VCIP56404.2022.10008896

[29] Andrew Elliot. 2015. Color and psychological functioning: A review of theoretical and empirical work. *Front. Psychol.* 6, 368 (2015), 1–8. DOI: https://doi.org/10.3389/fpsyg.2015.00368

[30] Huidi Fang, Chaoran Cui, Xiang Deng, Xiushan Nie, Muwei Jian, and Yilong Yin. 2018. Image aesthetic distribution prediction with fully convolutional network. In *Multimedia Modeling*. Springer International Publishing, 267–278. DOI: https://doi.org/10.1007/978-3-319-73603-7_22

[31] Na Fang, Yuan Zhang, Qinglan Wei, and Yuanhang Guo. 2020. Composition-aware learning for aesthetic assessment of edited photos. In *International Conference on Computer and Communications.* 2151–2155. DOI: https://doi.org/10.1109/ICCC51575.2020.9345309

[32] Xin Fu, Jia Yan, and Cien Fan. 2018. Image aesthetics assessment using composite features from off-the-shelf deep models. In *International Conference on Image Processing.* 3528–3532. DOI: https://doi.org/10.1109/ICIP.2018.8451133

[33] John Gage. 1999. *Color and Meaning: Art, Science, and Symbolism.* University of California Press.

[34] Viswanatha Gajjala, Snehasis Mukherjee, and Mainak Thakur. 2020. Measuring photography aesthetics with deep CNNs. *Instit. Eng. Technol. Image Process.* 14, 8 (2020), 1561–1570. DOI: https://doi.org/10.1049/iet-ipr.2019.1300

[35] Philip Galanter. 2012. *Computational Aesthetic Evaluation: Past and Future.* Springer Berlin, 255–293. DOI: https://doi.org/10.1007/978-3-642-31727-9_10

[36] Shiming Ge, Xin Jin, Le Wu, Xiaodong Li, Xiaokun Zhang, Jingying Chi, Siwei Peng, Geng Zhao, and Shuying Li. 2018. ILGNet: Inception modules with connected local and global features for efficient image aesthetic quality classification using domain adaptation. *Instit. Eng. Technol. Comput. Vis.* 13 (2018), 206–212. DOI: https://doi.org/10.1049/iet-cvi.2018.5249

[37] Koustav Ghosal, Aakanksha Rana, and Aljosa Smolic. 2019. Aesthetic image captioning from weakly-labelled photographs. In *International Conference on Computer Vision Workshop.* 4550–4560. DOI: https://doi.org/10.1109/ICCVW.2019.00556

[38] Koustav Ghosal and Aljosa Smolic. 2022. Image aesthetics assessment using graph attention network. In *International Conference on Pattern Recognition.* 3160–3167. DOI: https://doi.org/10.1109/ICPR56361.2022.9956162

[39] Artyom M. Grigoryan and Sos S. Agaian. 2020. Evidence of golden and aesthetic proportions in colors of paintings of the prominent artists. *Multimedia* 27, 1 (2020), 8–16. DOI: https://doi.org/10.1109/MMUL.2019.2908624

[40] Guanjun Guo, Hanzi Wang, Chunhua Shen, Yan Yan, and Hong-Yuan Mark Liao. 2018. Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression. *IEEE Trans. Multimedia* 20, 8 (2018), 2073–2085. DOI: https://doi.org/10.1109/TMM.2018.2794262

[41] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition.* 770–778. DOI: https://doi.org/10.1109/CVPR.2016.90

[42] Yong-Lian Hii, John See, Magzhan Kairanbay, and Lai-Kuan Wong. 2017. Multigap: Multi-pooled inception network with text augmentation for aesthetic prediction of photographs. In *International Conference on Image Processing.* 1722–1726. DOI: https://doi.org/10.1109/ICIP.2017.8296576

[43] Vlad Hosu, Bastian Goldlucke, and Dietmar Saupe. 2019. Effective aesthetics prediction with multi-level spatially pooled features. In *Conference on Computer Vision and Pattern Recognition*. 9367–9375. DOI: https://doi.org/10.1109/CVPR.2019.00960

[44] Jingwen Hou, Henghui Ding, Weisi Lin, Weide Liu, and Yuming Fang. 2022. Distilling knowledge from object classification to aesthetics assessment. *IEEE Trans. Circ. Syst. Vid. Technol.* 32, 11 (2022), 7386–7402. DOI: https://doi.org/10.1109/TCSVT.2022.3186307

[45] Jingwen Hou, Sheng Yang, and Weisi Lin. 2020. Object-level attention for aesthetic rating distribution prediction. In *ACM International Conference on Multimedia*. 816–824. DOI: https://doi.org/10.1145/3394171.3413695

[46] Jingwen Hou, Sheng Yang, Weisi Lin, Baoquan Zhao, and Yuming Fang. 2021. Learning image aesthetic assessment from object-level visual components. *Comput. Sci. Multimedia* 1 (2021), 1–13. DOI: https://doi.org/10.48550/arXiv.2104.01548

[47] Lei Hou and Xue Pan. 2023. Aesthetics of hotel photos and its impact on consumer engagement: A computer vision approach. *Tour. Manag.* 94 (2023), 1–13. DOI: https://doi.org/10.1016/j.tourman.2022.104653

[48] Shih-Wen Hsiao, Fu-Yuan Chiu, and Hsin-Yi Hsu. 2008. A computer-assisted colour selection system based on aesthetic measure for colour harmony and fuzzy logic theory. *Color Res. Applic.* 33 (2008), 411–423. DOI: https://doi.org/10.1002/col.20417

[49] Bogdan Ionescu, Wilma A. Bainbridge, and Naila Murray. 1955. *Human Perception of Visual Information: Psychological and Computational Perspectives*. Springer Cham. DOI: https://doi.org/10.1007/978-3-030-81465-6

[50] Somayeh Iranpak, Asadollah Shahbahrami, and Hassan Shakeri. 2021. Remote patient monitoring and classifying using the internet of things platform combined with cloud computing. *J. Big Data* 8, 120 (2021), 1–22. DOI: https://doi.org/10.1186/s40537-021-00507-w

[51] Thomas Jacobsen. 2010. Beauty and the brain: Culture, history and individual differences in aesthetic appreciation. *J. Anat.* 216, 2 (2010), 184–191. DOI: https://doi.org/10.1111/j.1469-7580.2009.01164.x

[52] Hyeongnam Jang and Jong-Seok Lee. 2021. Analysis of deep features for image aesthetic assessment. *IEEE Access* 9 (2021), 29850–29861. DOI: https://doi.org/10.1109/ACCESS.2021.3060171

[53] Hyeongnam Jang, Yeejin Lee, and Jong-Seok Lee. 2023. Probabilistic modeling of image aesthetic assessment toward measuring subjectivity. *IEEE Access* 11 (2023), 145772–145780. DOI: https://doi.org/10.1109/ACCESS.2023.3345921

[54] Gengyun Jia, Peipei Li, and Ran He. 2023. Theme aware aesthetic distribution prediction with full resolution photographs. *Trans. Neural Netw. Learn. Syst.* 34, 11 (2023), 8654–8668. DOI: https://doi.org/10.1109/TNNLS.2022.3151787

[55] Wei Jiang, Alexander C. Loui, and Cathleen Daniels Cerosaletti. 2010. Automatic aesthetic value assessment in photographic images. In *International Conference on Multimedia and Expo*. 920–925. DOI: https://doi.org/10.1109/ICME.2010.5582588

[56] Bin Jin, Maria V. Ortiz Segovia, and Sabine Süsstrunk. 2016. Image aesthetic predictors based on weighted CNNs. In *International Conference on Image Processing*. 2291–2295. DOI: https://doi.org/10.1109/ICIP.2016.7532767

[57] Xin Jin, Xinning Li, Heng Huang, Xiaodong Li, Chaoen Xiao, and Xiqiao Li. 2022. A deep drift-diffusion model for image aesthetic score distribution prediction. In *International Conference on Multimedia and Expo Workshops*. 1–6. DOI: https://doi.org/10.1109/ICMEW56448.2022.9859450

[58] Xin Jin, Xinning Li, Hao Lou, Chenyu Fan, Qiang Deng, Chaoen Xiao, Shuai Cui, and Amit Kumar Singh. 2023. Aesthetic attribute assessment of images numerically on mixed multi-attribute datasets. *ACM Trans. Multimedia Comput., Commun. Applic.* 18, 3s (2023), 1–16. DOI: https://doi.org/10.1145/3547144

[59] Xin Jin, Hao Lou, Heng Huang, Xinning Li, Xiaodong Li, Shuai Cui, Xiaokun Zhang, and Xiqiao Li. 2022. Pseudo-labeling and meta reweighting learning for image aesthetic quality assessment. *IEEE Trans. Intell. Transport. Syst.* 23, 12 (2022), 25226–25235. DOI: https://doi.org/10.1109/TITS.2022.3207152

[60] Xin Jin, Jianwen Lv, Xinghui Zhou, Chaoen Xiao, Xiaodong Li, and Shu Zhao. 2022. Aesthetic image captioning on the FAE-Captions dataset. *Comput. Electric. Eng.* 101 (2022), 1–7. DOI: https://doi.org/10.1016/j.compeleceng.2022.107866

[61] Xin Jin, Le Wu, Chenggen Song, Xiaodong Li, Geng Zhao, Siyu Chen, Jingying Chi, Siwei Peng, and Shiming Ge. 2017. Predicting aesthetic score distribution through cumulative Jensen-Shannon divergence. *Proc. Assoc. Advanc. Artif. Intell. Conf. Artif. Intell.* 32 (2017), 77–84. DOI: https://doi.org/10.1609/aaai.v32i1.11286

[62] Xin Jin, Le Wu, Geng Zhao, Xiaodong Li, Xiaokun Zhang, Shiming Ge, Dongqing Zou, Bin Zhou, and Xinghui Zhou. 2019. Aesthetic attributes assessment of images. In *ACM International Conference on Multimedia*. 311–319. DOI: https://doi.org/10.1145/3343031.3350970

[63] Magzhan Kairanbay, John See, and Lai-Kuan Wong. 2018. Towards demographic-based photographic aesthetics prediction for portraitures. In *Springer Multimedia Modeling*. Springer, 531–543. DOI: https://doi.org/10.1007/978-3-319-73603-7_43

[64] Chen Kang, Giuseppe Valenzise, and Frédéric Dufaux. 2019. Predicting subjectivity in image aesthetics assessment. In *International Workshop on Multimedia Signal Processing*. 1–6. DOI: https://doi.org/10.1109/MMSP.2019.8901716

[65] Yueying Kao, Ran He, and Kaiqi Huang. 2017. Deep aesthetic quality assessment with semantic information. *Trans. Image Process.* 26, 3 (2017), 1482–1495. DOI: https://doi.org/10.1109/TIP.2017.2651399

[66] Yueying Kao, Kaiqi Huang, and Steve Maybank. 2016. Hierarchical aesthetic quality assessment using deep convolutional neural networks. *Image Commun.* 47 (2016), 500–510. DOI: https://doi.org/10.1016/j.image.2016.05.004

[67] Yueying Kao, Chong Wang, and Kaiqi Huang. 2015. Visual aesthetic quality assessment with a regression model. In *International Conference on Image Processing*. 1583–1587. DOI: https://doi.org/10.1109/ICIP.2015.7351067

[68] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang. 2021. MUSIQ: Multi-scale image quality transformer. In *International Conference on Computer Vision*. 5128–5137. DOI: https://doi.org/10.1109/ICCV48922.2021.00510

[69] Junjie Ke, Keren Ye, Jiahui Yu, Yonghui Wu, Peyman Milanfar, and Feng Yang. 2023. VILA: Learning image aesthetics from user comments with vision-language pretraining. In *Conference on Computer Vision and Pattern Recognition*. 10041–10051. DOI: https://doi.org/10.1109/CVPR52729.2023.00968

[70] Yan Ke, Xiaoou Tang, and Feng Jing. 2006. The design of high-level features for photo quality assessment. In *Conference on Computer Vision and Pattern Recognition*, Vol. 1. 419–426. DOI: https://doi.org/10.1109/CVPR.2006.303

[71] Won-Hee Kim, Jun-Ho Choi, and Jong-Seok Lee. 2020. Objectivity and subjectivity in aesthetic quality assessment of digital photographs. *IEEE Trans. Affect. Comput.* 11, 3 (2020), 493–506. DOI: https://doi.org/10.1109/TAFFC.2018.2809752

[72] Keunsoo Ko, Jun-Tae Lee, and Chang-Su Kim. 2018. PAC-Net: Pairwise aesthetic comparison network for image aesthetic assessment. In *International Conference on Image Processing*. 2491–2495. DOI: https://doi.org/10.1109/ICIP.2018.8451621

[73] Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. 2016. Photo aesthetics ranking network with attributes and content adaptation. In *European Conference on Computer Vision*, Vol. 9905. Springer International Publishing, 662–679. DOI: https://doi.org/10.1007/978-3-319-46448-0_40

[74] Megumi Kotera, Ren Togo, Takahiro Ogawa, and Miki Haseyama. 2019. Aesthetic style transfer through text-to-image synthesis and image-to-image translation. In *Global Conference on Consumer Electronics*. 483–484. DOI: https://doi.org/10.1109/GCCE46687.2019.9015508

[75] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Commun. Assoc. Comput. Machin.* 60, 6 (2017), 84–90. DOI: https://doi.org/10.1145/3065386

[76] Michal Kucer, Alexander C. Loui, and David W. Messinger. 2018. Leveraging expert feature knowledge for predicting image aesthetics. *IEEE Trans. Image Process.* 27, 10 (2018), 5100–5112. DOI: https://doi.org/10.1109/TIP.2018.2845100

[77] Dmitry Kuzovkin, Tania Pouli, Rémi Cozot, Olivier Le Meur, Jonathan Kervec, and Kadi Bouatouch. 2018. Image selection in photo albums. In *ACM International Conference on Multimedia Retrieval*. Association for Computing Machinery, 397–404. DOI: https://doi.org/10.1145/3206025.3206077

[78] Alon Lavie and Abhaya Agarwal. 2007. Meteor: An automatic metric for MT evaluation with high levels of correlation with human judgments. In *Workshop Statistical Machine Translation*. Association for Computational Linguistics, 228–231. Retrieved from https://aclanthology.org/W07-0734

[79] Jun-Tae Lee and Chang-Su Kim. 2019. Image aesthetic assessment based on pairwise comparison a unified approach to score regression, binary classification, and personalization. In *International Conference on Computer Vision*. 1191–1200. DOI: https://doi.org/10.1109/ICCV.2019.00128

[80] Zhenyu Lei, Yejing Xie, Suiyi Ling, Andreas Pastor, Junle Wang, and Patrick Le Callet. 2021. Multi-modal aesthetic assessment for mobile gaming image. arXiv:2101.11700 [cs.CV]

[81] Marco Leonardi, Paolo Napoletano, Alessandro Rozza, and Raimondo Schettini. 2021. Modeling image aesthetics through aesthetics-related attributes. *London Imag. Meet.* 2021 (09 2021), 11–15. DOI: https://doi.org/10.2352/issn.2694-118X.2021.LIM-11

[82] Debang Li, Huikai Wu, Junge Zhang, and Kaiqi Huang. 2019. Fast A3RL: Aesthetics-aware adversarial reinforcement learning for image cropping. *Trans. Image Process.* 28, 10 (2019), 5105–5120. DOI: https://doi.org/10.1109/TIP.2019.2914360

[83] Leida Li, Yipo Huang, Jinjian Wu, Yuzhe Yang, Yaqian Li, Yandong Guo, and Guangming Shi. 2023. Theme-aware visual attribute reasoning for image aesthetics assessment. *IEEE Trans. Circ. Syst. Vid. Technol.* 33, 9 (2023), 4798–4811. DOI: https://doi.org/10.1109/TCSVT.2023.3249185

[84] Leida Li, Hancheng Zhu, Sicheng Zhao, Guiguang Ding, Hongyan Jiang, and Allen Tan. 2019. Personality driven multi-task learning for image aesthetic assessment. In *International Conference on Multimedia and Expo*. 430–435. DOI: https://doi.org/10.1109/ICME.2019.00081

[85] Leida Li, Hancheng Zhu, Sicheng Zhao, Guiguang Ding, and Weisi Lin. 2020. Personality-assisted multi-task learning for generic and personalized image aesthetics assessment. *IEEE Trans. Image Process.* 29 (2020), 3898–3910. DOI: https://doi.org/10.1109/TIP.2020.2968285

[86] Xuewei Li, Xueming Li, Gang Zhang, and Xianlin Zhang. 2020. A novel feature fusion method for computing image aesthetic quality. *IEEE Access* 8 (2020), 63043–63054. DOI: https://doi.org/10.1109/ACCESS.2020.2983725

[87] Yaohui Li, Yuzhe Yang, Huaxiong Li, Haoxing Chen, Liwu Xu, Leida Li, Yaqian Li, and Yandong Guo. 2022. Transductive aesthetic preference propagation for personalized image aesthetics assessment. In *ACM International Conference on Multimedia*. 896–904. DOI: https://doi.org/10.1145/3503161.3548244

[88] Youwei Liang, Junfeng He, Gang Li, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang et al. 2024. Rich human feedback for text-to-image generation. *ArXiv preprint arXiv:2312.10240* (2024), 19401–19411.

[89] Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*. Association for Computational Linguistics, 74–81. Retrieved from https://aclanthology.org/W04-1013

[90] Dong Liu, Rohit Puri, Nagendra Kamath, and Subhabrata Bhattacharya. 2020. Composition-aware image aesthetics assessment. In *Winter Conference on Applications of Computer Vision*. 3558–3567. DOI: https://doi.org/10.1109/WACV45572.2020.9093412

[91] Jing Liu, Jincheng Lv, Min Yuan, Jing Zhang, and Yuting Su. 2020. ABSNet: Aesthetics-based saliency network using multi-task convolutional network. *IEEE Sig. Process. Lett.* 27 (2020), 2014–2018. DOI: https://doi.org/10.1109/LSP.2020.3035065

[92] Wentao Liu and Zhou Wang. 2017. A database for perceptual evaluation of image aesthetics. In *International Conference on Image Processing*. 1317–1321. DOI: https://doi.org/10.1109/ICIP.2017.8296495

[93] Kung-Hung Lu, Kuang-Yu Chang, and Chu-Song Chen. 2016. Image aesthetic assessment via deep semantic aggregation. In *Global Conference on Signal Information Processing*. 232–236. DOI: https://doi.org/10.1109/GlobalSIP.2016.7905838

[94] Peng Lu, Hao Zhang, Xujun Peng, and Xiaofu Jin. 2021. Learning the relation between interested objects and aesthetic region for image cropping. *IEEE Trans. Multimedia* 23 (2021), 3618–3630. DOI: https://doi.org/10.1109/TMM.2020.3029882

[95] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z. Wang. 2014. RAPID: Rating pictorial aesthetics using deep learning. In *ACM International Conference on Multimedia*. 457–466. DOI: https://doi.org/10.1145/2647868.2654927

[96] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James. Z. Wang. 2015. Rating image aesthetics using deep learning. *IEEE Trans. Multimedia* 17, 11 (2015), 2021–2034. DOI: https://doi.org/10.1109/TMM.2015.2477040

[97] Xin Lu, Zhe Lin, Xiaohui Shen, Radomír Mech, and James Z. Wang. 2015. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In *International Conference on Computer Vision*. 990–998. DOI: https://doi.org/10.1109/ICCV.2015.119

[98] Pei Lv, Jianqi Fan, Xixi Nie, Weiming Dong, Xiaoheng Jiang, Bing Zhou, Mingliang Xu, and Changsheng Xu. 2023. User-guided personalized image aesthetic assessment based on deep reinforcement learning. *IEEE Trans. Multimedia* 25 (2023), 736–749. DOI: https://doi.org/10.1109/TMM.2021.3130752

[99] Ning Ma, Alexey Volkov, Aleksandr Livshits, Pawel Pietrusinski, Houdong Hu, and Mark Bolin. 2019. An universal image attractiveness ranking framework. In *Winter Conference on Applications of Computer Vision*. 657–665. DOI: https://doi.org/10.1109/WACV.2019.00075

[100] Shuang Ma, Jing Liu, and Chang Wen Chen. 2017. A-Lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. In *Conference on Computer Vision and Pattern Recognition*. 722–731. DOI: https://doi.org/10.1109/CVPR.2017.84

[101] Xiaoyu Ma, Suiyu Zhang, Yaqi Wang, Rong Li, Xiaodiao Chen, and Dingguo Yu. 2023. ASCAM-Former: Blind image quality assessment based on adaptive spatial and channel attention merging transformer and image to patch weights sharing. *Expert Syst. Applic.* 215, 1 (2023), 1–14. DOI: https://doi.org/10.1016/j.eswa.2022.119268

[102] Long Mai, Hailin Jin, and Feng Liu. 2016. Composition-preserving deep photo aesthetics assessment. In *Conference on Computer Vision and Pattern Recognition*. 497–506. DOI: https://doi.org/10.1109/CVPR.2016.60

[103] Long Mai, Hoang Le, Yuzhen Niu, and Feng Liu. 2011. Rule of thirds detection from photograph. In *IEEE International Symposium on Multimedia*. 91–96. DOI: https://doi.org/10.1109/ISM.2011.23

[104] Eftichia Mavridaki and Vasileios Mezaris. 2015. A comprehensive aesthetic quality assessment method for natural images using basic rules of photography. In *IEEE International Conference on Image Processing*. 887–891. DOI: https://doi.org/10.1109/ICIP.2015.7350927

[105] Xuantong Meng, Fei Gao, Shengjie Shi, Suguo Zhu, and Jingjie Zhu. 2018. MLANs: Image aesthetic assessment via multi-layer aggregation networks. In *International Conference on Image Processing Theory Tools and Application*. 1–6. DOI: https://doi.org/10.1109/IPTA.2018.8608132

[106] Sean Moran, Pierre Marza, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. 2020. DeepLPF: Deep local parametric filters for image enhancement. In *Conference on Computer Vision and Pattern Recognition*. 12823–12832. DOI: https://doi.org/10.1109/CVPR42600.2020.01284

[107] Naila Murray. 2019. PFAGAN: An aesthetics-conditional GAN for generating photographic fine art. In *International Conference on Computer Vision Workshop*. 3333–3341. DOI:https://doi.org/10.1109/ICCVW.2019.00415

[108] Naila Murray and Albert Gordo. 2017. A deep architecture for unified aesthetic prediction. *CoRR* abs/1708.04890 1 (2017), 1–10. arXiv:1708.04890 http://arxiv.org/abs/1708.04890

[109] Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *Conference on Computer Vision and Pattern Recognition*. 2408–2415. DOI:https://doi.org/10.1109/CVPR.2012.6247954

[110] Zhangkai Ni, Wenhan Yang, Shiqi Wang, Lin Ma, and Sam Kwong. 2020. Towards unsupervised deep image enhancement with generative adversarial network. *IEEE Trans. Image Process.* 29 (2020), 9140–9151. DOI:https://doi.org/10.1109/TIP.2020.3023615

[111] Daniel Vera Nieto, Luigi Celona, and Clara Fernandez-Labrador. 2022. Understanding aesthetics with language: A photo critique dataset for aesthetic assessment. In *Conference on Neural Information Processing Systems* 35 (2022), 34148–34161. arXiv:2206.08614 [cs.CV]

[112] Masashi Nishiyama, Takahiro Okabe, Imari Sato, and Yoichi Sato. 2011. Aesthetic quality classification of photographs based on color harmony. In *Conference on Computer Vision and Pattern Recognition*. 33–40. DOI:https://doi.org/10.1109/CVPR.2011.5995539

[113] Hamid Nodehi and Asadollah Shahbahrami. 2022. Multi-metric re-identification for online multi-person tracking. *IEEE Trans. Circ. Syst. Vid. Technol.* 32, 1 (2022), 147–159. DOI:https://doi.org/10.1109/TCSVT.2021.3059250

[114] Bowen Pan, Shangfei Wang, and Qisheng Jiang. 2019. Image aesthetic assessment assisted by attributes through adversarial learning. In *AAAI Conference on Artificial Intelligence*. 679–686. DOI:https://doi.org/10.1609/aaai.v33i01.3301679

[115] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *Annual Meeting on Association for Computational Linguistics*. 311–318. DOI:https://doi.org/10.3115/1073083.1073135

[116] Elena G. Petrova, Yury V. Mironov, Yoji Aoki, Hajime Matsushima, Satoshi Ebine, Katsunori Furuya, Anastasia Petrova, Norimasa Takayama, and Hirofumi Ueda. 2015. Comparing the visual perception and aesthetic evaluation of natural landscapes in Russia and Japan: Cultural and environmental factors. *Prog. Earth Planet. Sci.* 2 (2015), 1–12. DOI:https://doi.org/10.1186/s40645-015-0033-x

[117] Photo.net. (n.d.). Retrieved August, 2024 from http://photo.net

[118] Jian Ren, Xiaohui Shen, Zhe Lin, Radomír Mech, and David J. Foran. 2017. Personalized image aesthetics. In *International Conference on Computer Vision*. 638–647. DOI:https://doi.org/10.1109/ICCV.2017.76

[119] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. 2019. Low-light image enhancement via a deep hybrid network. *IEEE Trans. Image Process.* 28, 9 (2019), 4364–4375. DOI:https://doi.org/10.1109/TIP.2019.2910412

[120] George Santayana. 1955. *The Sense of Beauty: Being the Outline of Aesthetic Theory*. Dover Publications.

[121] Rossano Schifanella, Miriam Redi, and Luca Aiello. 2015. An image is worth more than a thousand favorites: Surfacing the hidden beauty of Flickr pictures. In *International Association for the Advancement of Artificial Intelligence Conference on Web Social Media*, Vol. 9. 397–406. DOI:https://doi.org/10.1609/icwsm.v9i1.14612

[122] Katharina Schwarz, Patrick Wieschollek, and Hendrik Lensch. 2018. Will people like your image? Learning the aesthetic space. In *Winter Conference on Applications of Computer Vision*. 2048–2057. DOI:https://doi.org/10.1109/WACV.2018.00226

[123] Dongyu She, Yu-Kun Lai, Gaoxiong Yi, and Kun Xu. 2021. Hierarchical layout-aware graph convolutional network for unified aesthetics assessment. In *Conference on Computer Vision and Pattern Recognition*. 8471–8480. DOI:https://doi.org/10.1109/CVPR46437.2021.00837

[124] Kekai Sheng, Weiming Dong, Menglei Chai, Guohui Wang, Peng Zhou, Feiyue Huang, Bao-Gang Hu, Rongrong Ji, and Chongyang Ma. 2020. Revisiting image aesthetic assessment via self-supervised feature learning. *Proc. Assoc. Advanc. Arti. Intell. Conf. Artif. Intell.* 34 (2020), 5709–5716. DOI:https://doi.org/10.1609/aaai.v34i04.6026

[125] Kekai Sheng, Weiming Dong, Huang Haibin, Menglei Chai, Yong Zhang, Chongyang Ma, and Bao-Gang Hu. 2020. Learning to assess visual aesthetics of food images. *Computat. Vis. Media* 7 (2020), 139–152. DOI:https://doi.org/10.1007/s41095-020-0193-5

[126] Kekai Sheng, Weiming Dong, Chongyang Ma, Xing Mei, Feiyue Huang, and Bao-Gang Hu. 2018. Attention-based multi-patch aggregation for image aesthetic assessment. In *ACM International Conference on Multimedia*. 879–886. DOI:https://doi.org/10.1145/3240508.3240554

[127] Lei Shi and Baiyuan Ding. 2022. Design of packaging design evaluation architecture based on deep learning. *Scient. Program.* 2022 (2022), 1–8. DOI:https://doi.org/10.1155/2022/4469495

[128] Yangyang Shu, Qian Li, Lingqiao Liu, and Guandong Xu. 2024. Semi-supervised adversarial learning for attribute-aware photo aesthetic assessment. *IEEE Trans. Multimedia* 26 (2024), 4086–4096. DOI:https://doi.org/10.1109/TMM.2021.3117709

[129] Yangyang Shu, Qian Li, Shaowu Liu, and Guandong Xu. 2020. Learning with privileged information for photo aesthetic assessment. *Neurocomputing* 404, 1 (2020), 304–316. DOI: https://doi.org/10.1016/j.neucom.2020.04.142

[130] Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*. 1–14. Retrieved from https://arxiv.org/abs/1409.1556

[131] Wei-Tse Sun, Ting-Hsuan Chao, Yin-Hsi Kuo, and Winston Hsu. 2016. Photo filter recommendation by category-aware aesthetic learning. *IEEE Trans. Multimedia* 19, 8 (2016), 1870–1880. DOI: https://doi.org/10.1109/TMM.2017.2688929

[132] Xiaoshuai Sun, Hongxun Yao, Rongrong Ji, and Shaohui Liu. 2009. Photo assessment based on computational visual attention model. In *ACM International Conference on Multimedia*. 541–544. DOI: https://doi.org/10.1145/1631272.1631351

[133] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *International Conference on Neural Information Processing Systems*. 3104–3112. DOI: https://doi.org/10.48550/arXiv.1409.3215

[134] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2015. Going deeper with convolutions. In *Conference on Computer Vision and Pattern Recognition*. 1–9. DOI: https://doi.org/10.1109/CVPR.2015.7298594

[135] Hironori Takimoto, Fumiya Omori, and Akihiro Kanagawa. 2020. Image aesthetics assessment based on multi-stream CNN architecture and saliency features. *Appl. Artif. Intell.* 35 (2020), 1–16. DOI: https://doi.org/10.1080/08839514.2020.1839197

[136] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment. *IEEE Trans. Image Process.* 27, 8 (2018), 3998–4011. DOI: https://doi.org/10.1109/TIP.2018.2831899

[137] Xiaoou Tang, Wei Luo, and Xiaogang Wang. 2013. Content-based photo quality assessment. *IEEE Trans. Multimedia* 15, 8 (2013), 1930–1943. DOI: https://doi.org/10.1109/TMM.2013.2269899

[138] Ali Tourani, Asadollah Shahbahrami, Sajjad Soroori, Saeed Khazaee, and Ching Yee Suen. 2020. A robust deep learning approach for automatic Iranian vehicle license plate detection and recognition for surveillance systems. *IEEE Access* 8 (2020), 201317–201330. DOI: https://doi.org/10.1109/ACCESS.2020.3035992

[139] Giuseppe Valenzise, Chen Kang, and Frédéric Dufaux. 2022. Advances and challenges in computational image aesthetics. In *Human Perception of Visual Information: Psychological and Computational Perspectives*. Springer, 133–181. DOI: https://doi.org/10.1007/978-3-030-81465-6_6

[140] Ramakrishna Vedantam, C. Lawrence Zitnick, and Devi Parikh. 2015. CIDEr: Consensus-based image description evaluation. In *Conference on Computer Vision and Pattern Recognition*. 4566–4575. DOI: https://doi.org/10.1109/CVPR.2015.7299087

[141] Yitian Wan, Weijie Li, Xingjiao Wu, Junjie Xu, and Jing Yang. 2023. Automatic image aesthetic assessment for human-designed digital images. In *International Workshop Multimedia Content Generation and Evaluation: New Methods and Practice*. Association for Computing Machinery, 1–8. DOI: https://doi.org/10.1145/3607541.3616810

[142] Guolong Wang, Yike Tan, Hangyu Lin, and Chuchun Zhang. 2024. Keep knowledge in perception: Zero-shot image aesthetic assessment. In *International Conference on Acoustics, Speech and Signal Processing*. 8311–8315. DOI: https://doi.org/10.1109/ICASSP48485.2024.10447301

[143] Guangming Wang, Chi Zhang, Hesheng Wang, Jingchuan Wang, Yong Wang, and Xinlei Wang. 2020. Unsupervised learning of depth, optical flow and pose with occlusion from 3D geometry. *IEEE Trans. Intell. Transport. Syst.* 23 (2020), 308–320. DOI: https://doi.org/10.1109/TITS.2020.3010418

[144] Junjue Wang, Yanfei Zhong, Zhuo Zheng, Ailong Ma, and Liangpei Zhang. 2021. RSNet: The search for remote sensing deep neural networks in recognition tasks. *IEEE Trans. Geosci. Rem. Sens.* 59, 3 (2021), 2520–2534. DOI: https://doi.org/10.1109/TGRS.2020.3001401

[145] Lijie Wang, Xueting Wang, and Toshihiko Yamasaki. 2022. Image aesthetics prediction using multiple patches preserving the original aspect ratio of contents. *Multimedia Tools Applic.* 82, 22 (2022), 2783–2804. DOI: https://doi.org/10.1007/s11042-022-13333-w

[146] Lijie Wang, Xueting Wang, Toshihiko Yamasaki, and Kiyoharu Aizawa. 2019. Aspect-ratio-preserving multi-patch image aesthetics score prediction. In *Conference on Computer Vision and Pattern Recognition Workshops*. 1833–1842. DOI: https://doi.org/10.1109/CVPRW.2019.00234

[147] Tao Wang, Wei Sun, Xiongkuo Min, Wei Lu, Zicheng Zhang, and Guangtao Zhai. 2021. A multi-dimensional aesthetic quality assessment model for mobile game images. In *International Conference on Visual Communications and Image Processing*. 1–5. DOI: https://doi.org/10.1109/VCIP53242.2021.9675430

[148] Weining Wang, Rui Deng, Lemin Li, and Xiangmin Xu. 2019. Image aesthetic assessment based on perception consistency. In *Pattern Recognition and Computer Vision*. Springer International Publishing, 303–315. DOI: https://doi.org/10.1007/978-3-030-31723-2_26

[149] Wenguan Wang and Jianbing Shen. 2017. Deep cropping via attention box prediction and aesthetics assessment. In *International Conference on Computer Vision*. 2205–2213. DOI: https://doi.org/10.1109/ICCV.2017.240

[150] Wenguan Wang, Jianbing Shen, and Haibin Ling. 2019. A deep network solution for attention and aesthetics aware photo cropping. *Trans. Pattern Anal. Mach. Intell.* 41, 7 (2019), 1531–1544. DOI: https://doi.org/10.1109/TPAMI.2018.2840724

[151] Wenshan Wang, Su Yang, Weishan Zhang, and Jiulong Zhang. 2018. Neural aesthetic image reviewer. *Instit. Eng. Technol. Comput. Vis.* 13, 8 (2018), 749–758. DOI: https://doi.org/10.1049/iet-cvi.2019.0361

[152] Weining Wang, Mingquan Zhao, Li Wang, Jiexiong Huang, Chengjia Cai, and Xiangmin Xu. 2016. A multi-scene deep learning model for image aesthetic evaluation. *Sig. Process.: Image Commun.* 47 (2016), 511–518. DOI: https://doi.org/10.1016/j.image.2016.05.009

[153] Yeqing Wang, Yi Li, and Fatih Porikli. 2016. Finetuning convolutional neural networks for visual aesthetics. In *International Conference on Pattern Recognition*. 3554–3559. DOI: https://doi.org/10.1109/ICPR.2016.7900185

[154] Zhoufutu Wen, Xinyu Zhao, Zhipeng Jin, Yi Yang, Wei Jia, Xiaodong Chen, Shuanglong Li, and Lin Liu. 2023. Enhancing dynamic image advertising with vision-language pre-training. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, 3310–3314. DOI: https://doi.org/10.1145/3539618.3591844

[155] Jianfeng Wu, Baixi Xing, Huahao Si, Jian Dou, Jin Wang, Yuning Zhu, and Xiaojian Liu. 2020. Product design award prediction modeling: Design visual aesthetic quality assessment via DCNNs. *IEEE Access* 8 (2020), 211028–211047. DOI: https://doi.org/10.1109/ACCESS.2020.3039715

[156] Xiaoshi Wu, Keqiang Sun, Feng Zhu, Rui Zhao, and Hongsheng Li. 2023. Human preference score: Better aligning text-to-image models with human preference. *International Conference on Computer Vision*. 2096–2105. DOI: https://doi.org/10.1109/ICCV51070.2023.00200

[157] Kun Xiong, Liu Jiang, Xuan Dang, Guolong Wang, Wenwen Ye, and Zheng Qin. 2020. Towards personalized aesthetic image caption. In *International Joint Conference on Neural Networks*. 1–8. DOI: https://doi.org/10.1109/IJCNN48605.2020.9206953

[158] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. 2023. ImageReward: Learning and evaluating human preferences for text-to-image generation. In *Conference on Neural Information Processing Systems*. DOI: https://doi.org/10.48550/arXiv.2304.05977

[159] Munan Xu, Jia-Xing Zhong, Yurui Ren, Shan Liu, and Ge Li. 2020. Context-aware attention network for predicting image aesthetic subjectivity. In *ACM International Conference on Multimedia*. 798–806. DOI: https://doi.org/10.1145/3394171.3413834

[160] Qianqian Xu, Xinwei Sun, Zhiyong Yang, Xiaochun Cao, Qingming Huang, and Yuan Yao. 2019. *iSplit LBI: Individualized Partial Ranking with Ties via Split LBI.* Curran Associates Inc., 3901–3911. DOI: https://doi.org/10.48550/arXiv.1910.05905

[161] Yongbo Xu, Meng Wang, Pei Lv, Ze Peng, Junyi Sun, Shimei Su, Bing Zhou, and Mingliang Xu. 2018. USAR: An interactive user-specific aesthetic ranking framework for images. In *ACM International Conference on Multimedia*. 1328–1336. DOI: https://doi.org/10.1145/3240508.3240635

[162] Ying Xu, Yi Wang, Huaixuan Zhang, and Yong Jiang. 2020. Spatial attentive image aesthetic assessment. In *International Conference on Multimedia and Expo*. 1–6. DOI: https://doi.org/10.1109/ICME46284.2020.9102804

[163] Hongtao Yang, Ping Shi, Saike He, Da Pan, Zefeng Ying, and Ling Lei. 2019. A comprehensive survey on image aesthetic quality assessment. In *International Conference on Computer and Information Science*. 294–299. DOI: https://doi.org/10.1109/ICIS46139.2019.8940355

[164] Jie Yang, Mengjin Lyu, Zhiquan Qi, and Yong Shi. 2023. Deep learning based image quality assessment: A survey. *Proced. Comput. Sci.* 221 (2023), 1000–1005. DOI: https://doi.org/10.1016/j.procs.2023.08.080

[165] Taoxi Yang, Sarita Silveira, Arusu Formuli, Marco Paolini, Ernst Pöppel, Tilmann Sander, Yan Bao, Norimasa Takayama, and Hirofumi Ueda. 2019. Aesthetic experiences across cultures: Neural correlates when viewing traditional eastern or western landscape paintings. *Front. Psychol.* 10 (2019), 1–10. DOI: https://doi.org/10.3389/fpsyg.2019.00798

[166] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Conference on Computer Vision and Pattern Recognition*. 3060–3069. DOI: https://doi.org/10.1109/CVPR42600.2020.00313

[167] Yong-Yaw Yeo, John See, Lai-Kuan Wong, and Hui-Ngo Goh. 2021. Generating aesthetic based critique for photographs. In *International Conference on Image Processing*. 2523–2527. DOI: https://doi.org/10.1109/ICIP42928.2021.9506385

[168] Junyong You, Andrew Perkis, Miska Hannuksela, and Moncef Gabbouj. 2009. Perceptual quality assessment based on visual attention analysis. In *ACM Multimedia Conference*. 561–564. DOI: https://doi.org/10.1145/1631272.1631356

[169] Hui Zeng, Zisheng Cao, Lei Zhang, and Alan C. Bovik. 2020. A unified probabilistic formulation of image aesthetic assessment. *IEEE Trans. Image Process.* 29 (2020), 1548–1561. DOI: https://doi.org/10.1109/TIP.2019.2941778

[170] Cunjun Zhang, Kehua Lei, Jia Jia, Yihui Ma, and Zhiyuan Hu. 2018. AI painting: An aesthetic painting generation system. In *ACM International Conference on Multimedia.* 1231–1233. DOI: https://doi.org/10.1145/3240508.3241386

[171] Chao Zhang, Ce Zhu, Xun Xu, Yipeng Liu, Jimin Xiao, and Tammam Tillo. 2018. Visual aesthetic understanding: Sample-specific aesthetic classification and deep activation map visualization. *Sig. Process. Image Commun.* 67 (2018), 12–21. DOI: https://doi.org/10.1016/j.image.2018.05.006

[172] Jiajing Zhang, Yongwei Miao, and Jinhui Yu. 2021. A comprehensive survey on computational aesthetic evaluation of visual art images: Metrics and challenges. *IEEE Access* 9 (2021), 77164–77187. DOI: https://doi.org/10.1109/ACCESS.2021.3083075

[173] Jiajing Zhang, Yongwei Miao, Junsong Zhang, and Jinhui Yu. 2020. Inkthetics: A comprehensive computational model for aesthetic evaluation of chinese ink paintings. *IEEE Access* 8 (2020), 225857–225871. DOI: https://doi.org/10.1109/ACCESS.2020.3044573

[174] Lingyun Zhang and Pingjian Zhang. 2021. Research on aesthetic models based on neural architecture search. *Int. J. Intell. Fuzz. Syst.* 41, 2 (2021), 2953–2967. DOI: https://doi.org/10.3233/JIFS-210026

[175] Xiaodan Zhang, Xinbo Gao, Lihuo He, and Wen Lu. 2020. MSCAN: Multimodal self-and-collaborative attention network for image aesthetic prediction tasks. *Neurocomputing* 430 (2020), 14–23. DOI: https://doi.org/10.1016/j.neucom.2020.10.046

[176] Xiaodan Zhang, Xinbo Gao, Wen Lu, and Lihuo He. 2019. A gated peripheral-foveal convolutional neural network for unified image aesthetic prediction. *IEEE Trans. Multimedia* 21, 11 (2019), 2815–2826. DOI: https://doi.org/10.1109/TMM.2019.2911428

[177] Xiaodan Zhang, Xinbo Gao, Wen Lu, Lihuo He, and Jie Li. 2021. Beyond vision: A multimodal recurrent attention convolutional neural network for unified image aesthetic prediction tasks. *IEEE Trans. Multimedia* 23 (2021), 611–623. DOI: https://doi.org/10.1109/TMM.2020.2985526

[178] Xiaodan Zhang, Xinbo Gao, Wen Lu, Ying Yu, and Lihuo He. 2019. Fusion global and local deep representations with neural attention for aesthetic quality assessment. *Sig. Process.: Image Commun.* 78 (2019), 42–50. DOI: https://doi.org/10.1016/j.image.2019.05.021

[179] Xiaoyan Zhang, Zhuopeng Li, and Jianmin Jiang. 2021. Emotion attention-aware collaborative deep reinforcement learning for image cropping. *IEEE Trans. Multimedia* 23 (2021), 2545–2560. DOI: https://doi.org/10.1109/TMM.2020.3013350

[180] Xiaodan Zhang, Qiao Song, and Gang Liu. 2022. Multimodal image aesthetic prediction with missing modality. *Mathematics* 10 (2022), 1–19. DOI: https://doi.org/10.3390/math10132312

[181] Lin Zhao, Meimei Shang, Fei Gao, Rongsheng Li, Fei Huang, and Jun Yu. 2020. Representation learning of image composition for aesthetic prediction. *Comput. Vis. Image Underst.* 199 (2020), 103024. DOI: https://doi.org/10.1016/j.cviu.2020.103024

[182] Aimin Zhou, Hongbin Liu, Shutao Zhang, and Jinyan Ouyang. 2021. Evaluation and design method for product form aesthetics based on deep learning. *IEEE Access* 9 (2021), 108992–109003. DOI: https://doi.org/10.1109/ACCESS.2021.3101619

[183] Ye Zhou, Xin Lu, Junping Zhang, and James Wang. 2016. Joint image and text representation for aesthetics analysis. In *ACM International Conference on Multimedia.* 262–266. DOI: https://doi.org/10.1145/2964284.2967223

[184] Xiaohan Zou, Cheng Lin, Yinjia Zhang, and Qinpei Zhao. 2020. To be an artist: Automatic generation on food image aesthetic captioning. In *International Conference on Tools with Artificial Intelligence.* 779–786. DOI: https://doi.org/10.1109/ICTAI50040.2020.00124

Received 2 February 2024; revised 14 October 2024; accepted 26 January 2025