www.cad-journal.net
Computer-Aided Design AND APPLICATIONS

AI-driven Art Learning System: Automated Color Assessment Through Image Recognition

Hanjun Su[1,*], Zhuang Li[2] and Jianhua Hu[3]

[1] Centre for Instructional Technology and Multimedia, 623728442su@gmail.com
[2] China University of Mining and Technology, 939193035@qq.com
[3] China University of Mining and Technology, jianhuahu2001@cumt.edu.cn

Corresponding author: Hanjun Su, 623728442su@gmail.com

Abstract. With the promotion of quality education, the importance of art education is becoming more and more prominent. This study proposes an AI(artificial intelligence)-driven art learning system that automates color assessment through image recognition, aiming to improve the art learning experience and the automation of color assessment. The system takes advantage of the HSV (Hue Saturation Value) color space and combines it with the NIMA (Neural Image Assessment) network framework to perform in-depth color analysis and assessment of artworks. In this paper, an automatic color evaluation system using HSV-NIMA as a model is designed and implemented. The system was trained and evaluated on AVA (aesthetic visual analysis), TID2013(Tampere Image Database 2013) and Flower datasets, and the results show that the system achieves an accuracy of 80.84% on the AVA dataset, which has high color stability, and 81.69% on the Flower dataset, which is rich in color diversity, and the confidence distribution is closest to a score of 7, which indicates that the system is able to effectively understand and evaluate image color. Additionally, the system shows good robustness on the TID2013 dataset. Finally, this paper also explores the impact of AI art learning systems on socio-cultural, art education, and art creation. This study also analyzed and evaluated in conjunction with sketch painting art images. Although some of the paintings were not predicted accurately enough and were biased, the overall evaluation was not far from the actual rating.

Keywords: AI art learning; image recognition; color assessment; HSV color space; NIMA.
DOI: https://doi.org/10.14733/cadaps.2026.147-161

1 INTRODUCTION

As the concept of quality education continues to grow, the topic of arts education has become increasingly hot. Through the study of the arts, learners can expand their worldview and improve their critical thinking skills[1].

In the field of art, color is not only a basic element of visual art, but also an important medium for emotional expression and cultural communication[2]. With the rapid development of digital technology, the way art is created and learned is undergoing a revolution. Traditional art learning methods have limitations in terms of efficiency, accessibility and personalized instruction[3].

In recent years, the emergence of technologies such as virtual reality, image recognition, and artificial intelligence (AI) has created new opportunities for art education. Educators across various disciplines are increasingly integrating these technologies into their teaching and learning processes[4]. Breakthroughs in AI, particularly in image recognition, natural language processing, and machine learning, have enabled machines to understand and process complex visual information, thereby providing new tools and platforms for art education[5]. AI technology can facilitate the automatic analysis of artworks, interactive teaching of color theory, and personalized learning path recommendations, significantly enriching the methods and content of art education[6]. However, art paintings can vary dramatically based on the artist, the depicted scene, and the artistic style employed. Therefore, when processing fine art images, more factors need to be considered compared to conventional image processing[7]. For example, Ayush et al.[8] have leveraged the potential of AI to analyze and learn from a photographer's historical stylistic choices, enabling personalized image enhancement directly in-camera. This innovation provides photographers with a tool to immediately visualize and apply their preferred aesthetics, thereby challenging traditional post-processing workflows.

Despite technological advancements, current art learning systems still face limitations in efficiency, accessibility, and personalized instruction. AI, particularly in image recognition and machine learning, offers new opportunities for interpreting complex visual data and provides innovative tools for art education. However, effectively integrating AI, especially for tasks such as automated color assessment in image recognition, requires further investigation. This study explores the role of AI in art learning, focusing on automated color evaluation through image recognition. The contributions of this paper to the field of art learning are as follows:

(1) Analyzing the current needs and challenges of color assessment in art learning, and exploring the potential application of AI technology in color assessment.

(2) Designing and implementing an automated color assessment system based on image recognition, which can assist art learners in color analysis and learning. The system adopts HSV color space and combines with NIMA network framework to improve the accuracy and efficiency of color assessment.

(3) The system was trained and evaluated on AVA[9]、TID2013[10] and Flower datasets. The experimental results show good performance and generalization ability.

(4) To explore the impact of AI art learning system on social culture, art education and art creation, as well as the future development trend.

2 RELATED WORK

Arts education has a rich history, with its pedagogical approaches evolving significantly over time[11]. Traditional and modern methods of art learning diverge in technology utilization, teaching ideologies, learning approaches, and creative techniques[12].

Traditional art learning typically relies on the master-apprentice system, field sketching, and classical textbooks. In this model, students acquire artistic skills through imitation and repetitive practice. Teachers play a crucial role by demonstrating techniques and providing direct instruction. Students study color, composition, and technique through historical artworks[13]. Additionally, assessing colors, which is essential for artistic development, requires expertise and experience, presenting a significant challenge for novices[14].

With the development of technology, modern art learning methods have begun to incorporate digital tools and artificial intelligence techniques [15]. Modern methods utilize digital drawing boards, graphic software, and 3D modeling as core tools for creation and expression[16]. They promote

interdisciplinary learning, merging art with technology, design, and the humanities to cultivate versatile artists. AI's role is particularly transformative, offering personalized feedback, aiding in color and composition analysis, and supporting the creative process.

For example, in recent years, image processing techniques have significantly contributed to the advancement of artistic identification in paintings and calligraphic works[17]. Fang et al.[18] proposed a digital teaching system that enhances art education by monitoring students' creative states in real time, providing instant feedback, and facilitating the generation of creative works. Zhang et al.[19] introduced a machine-learning-assisted topology optimization method for architectural design with an artistic touch. Lin et al.[20] proposed a Transformer-based network called ArtFormer for example-based line art coloring, which learns spatial and visual relationships between line art and color images. Falomir et al.[21] designed a QArt-Learn approach for classifying stylized paintings based on qualitative color descriptors (QCD), color similarity (SimQCD), and quantitative global features (e.g., mean values of luminance, hue, saturation, and luminance contrast). Liu et al.[22] proposed a Knowledge Distillation-based OIQA (KD-OIQA) framework, which uses qualitative features from the teacher network to guide the student network's quality feature representation through different projection formats. Li et al.[23] developed a novel color image quality index based on sparse representation and reconstructed residuals (SRRR), incorporating brightness similarity to generate an overall quality score for color images. Bhandari et al.[24] proposed a deep convolutional neural network framework to perform image recognition using TensorFlow and Keras. Advanced networks are used to train and categorize images to accomplish automatic image aesthetic assessment.

Despite the long history of research on image color assessment algorithms, the importance of color in quality assessment has yet to be fully recognized. To address this gap, this paper proposes the HSV-NIMA network framework, which aims to enhance the model's ability to capture color information.

3 IMAGE COLOR RECOGNITION FRAMEWORK

3.1 Color Space Transformation

Given that the acquired image is a multicolored image, prevailing methods for color evaluation typically categorize pixels using either a spectral signature map or color-based classification. Among the various color spaces, the RGB model is heavily reliant on the hardware setup. For tasks that involve identifying or filtering image colors, the HSV color space is often more appropriate[25].

The HSV color space provides a more accurate representation of color characteristics by utilizing the dimensions of hue, saturation, and value. Typically, the HSV color space is depicted in a cone-shaped diagram as shown in Figure 1, offering a more intuitive visual model for users. To obtain a specific color, one can set a constant H (hue) and initiate the saturation and value at their maximum levels. Subsequently, by adjusting the amounts of black and white, the desired color can be achieved.

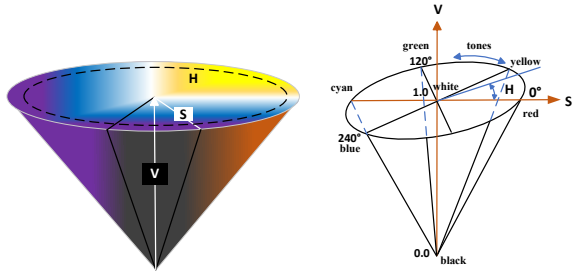

Figure 1: HSV color models.

Therefore, the color extraction method based on HSV color space has better color perception, robustness, and accuracy, which makes full use of the human perception characteristics of color, and can extract more accurate and stable color information, providing a more reliable basis for the subsequent image processing and analysis tasks.

The RGB-HSV color space conversion formula is as follows:

$$H=\begin{cases}0^\circ, & \text{if max} = \text{min}\\ 60^\circ\times\frac{G-B}{\text{max}-\text{min}}+0^\circ, & \text{if max} = R \text{ and } G\geq B\\ 60^\circ\times\frac{G-B}{\text{max}-\text{min}}+360^\circ, & \text{if max} = R \text{ and } G<B\\ 60^\circ\times\frac{G-B}{\text{max}-\text{min}}+120^\circ, & \text{ifmax} = G\\ 60^\circ\times\frac{G-B}{\text{max}-\text{min}}+240^\circ, & \text{if max} = B\end{cases} \tag{3.1}$$

$$S=\begin{cases}0, & \text{if max} = 0\\ \frac{\text{max}-\text{min}}{\text{max}}=1-\frac{\text{min}}{\text{max}}, & \text{otherwise}\end{cases} \tag{3.2}$$

$$V=\text{max} \tag{3.3}$$

Where R, G, B represent the intensity values of red, green, and blue, respectively, and H, S, V represent the hue, saturation, and luminance, respectively. R, G, and B represent the average value of the image in the red, green, and blue channels, respectively. Max represents the maximum value of the three RGB components, i.e., max (R, G, B). Max represents the maximum value of the three RGB components, i.e., max (R, G, B), and Min represents the minimum value of the three RGB components, i.e., min (R, G, B).

3.2 Subsection Headings Use the Capitalization Rules of the Main Title

The NIMA (Neural Image Assessment)[26] leverages an advanced deep learning architecture for object detection to predict the spectrum of human perceptions regarding images, including both technical clarity and aesthetic appeal. The NIMA scoring system closely aligns with human subjective evaluations, making it suitable for assessing image quality. As shown in Figure 2, NIMA's core structure utilizes the MobileNet network, which is optimized for mobile and embedded systems. This framework prioritizes the use of compact CNNs (Convolutional Neural Networks) in such devices, significantly reducing model complexity and computational load while only marginally sacrificing accuracy compared to standard CNNs.

The Conv3_s2 (convolutional layer with kernel size 3x3 and step size 2) in Figure 2 is a convolutional layer using a 3x3 kernel size and a step size of 2, which is commonly used for extracting image features and reducing the spatial dimensionality of the data. The BN (Batch Normalization) is a technique commonly used to train deep neural networks by reducing the internal covariate offsets to normalize the inputs to the layer, thus improving training speed and stability. The hswish (H-Swish activation function) is an activation function, a variant of the Swish activation function, designed for efficient computation on mobile devices. It combines sigmoid functions and linear transformations to improve the performance and efficiency of the model. The bneck in deep learning, a bottleneck block is a network structure that reduces the amount of computation by decreasing the dimensionality of the intermediate layers while maintaining the expressiveness of the network. The avg_pool (average pooling) is a pooling operation that reduces the spatial dimensionality of the feature map and aggregates the information by calculating the average within a region. The FC (Fully Connected Layer) is a fully connected layer in which each input unit is connected to an output unit. This is typically used at the end of the network to map extracted features to the final output, such as category probabilities in a classification task. The 1*1000 is the output dimension of the fully-connected layer, where 1 means the output is a vector and 1000 means the vector has 1000 elements.

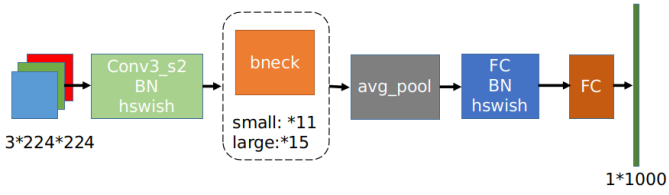

Figure 2: Mobile network.

In this paper, data enhancement is performed using a horizontal flip of the image. The specific formula is as follows:

$$\mu = \sum_1^n \ s_i \times p_{si} \tag{3.4}$$

$$\sigma = \sqrt{\left(\sum_{i=1}^N \ (s_i - \mu)^2 \times p_{si}\right)} \tag{3.5}$$

where μ denotes the expected value of the rating, where $s_i$is the rating scale and p_{si} is the probability of that rating scale. σ denotes the standard deviation of the rating, which measures the dispersion of the rating.

In the case of ordered classes (e.g., aesthetics and quality assessment), the cross-entropy loss commonly used in classification problems lacks internal relationships between score intervals, and training on datasets with intrinsic ordering relationships between classes can benefit from EMD-based losses. See Eq. (3.6) and Eq. (3.7) for specific expressions.

$$\text{EMD}\ (p, \hat{p}) = \sqrt[r]{\left(\frac{1}{N}\sum_{k=1}^N \ \left|\text{CDF}_p\ (k) - \text{CDF}_{\hat{p}}(k)\right|^r\right)} \tag{3.6}$$

$$\text{CDF}_p\ (k) = \sum_{i=1}^k P_{si} \tag{3.7}$$

where $\text{EMD}(p, \hat{p})$ denotes the Earth movement distance between the predicted probability distribution $\hat{p}$ and the true probability distribution CDFP(k)p. It is the cumulative value of the predicted rating probability, instead of independently predicting the probability of obtaining each rating, as a substitute for the distribution.

In this paper, the HSV color space is integrated into the NIMA network framework for image quality assessment to extract color features using the properties of HSV. A branch for color harmony assessment is also added to the NIMA model for calculating color histogram similarity or applying color harmony theory. The enhanced framework, HSV-NIMA, including convolutional and pooling layers, is shown in Figure 3. The CVs in Figure 3 are CV mappings in the HSV color space.

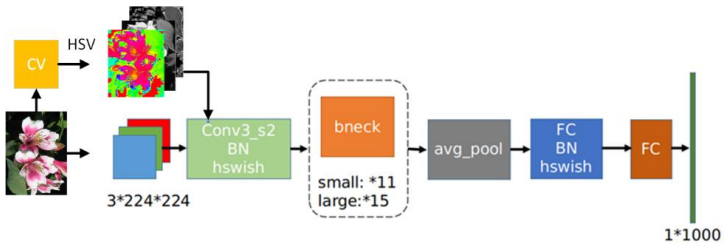

Figure 3: HSV-NIMA network framework.

4 TRAINING RESULTS OF THE AUTOMATED COLOR ASSESSMENT SYSTEM

4.1 Datasets

In neural network model training, dataset quality significantly influences the training outcome. For this paper's focus on automated color assessment, we have selected diverse datasets for training.

Our initial dataset is the AVA database, a comprehensive collection for aesthetic visual analysis developed by Murray et al. at the Center for Computer Vision, Universitat Autònoma de Barcelona (UAB), Spain. This dataset consists of 255,530 images, each rated by an average of 210 human artists from diverse online platforms on a scale from 1 to 10. The number of raters per image fluctuates between 78 and 539, as shown in Figure 4-(a).

The second dataset we employed is TID2013, which includes 25 reference images and 3000 modified versions. It covers 24 types of distortions, such as Gaussian noise addition, color saturation changes, chromatic aberration, quantization, and blurring. The dataset's DMOS scores, ranging from 0 to 9, are based on 524,340 observations by 971 participants, as depicted in Figure 4-(b).

Lastly, we gathered 8047 images of 12 distinct types and colors of flowers, including subjective scores from our lab staff, with an average score labeled on a scale of [0,10], where higher scores indicate superior image quality and lower scores indicate the opposite. This is displayed in Figure 4-(c).

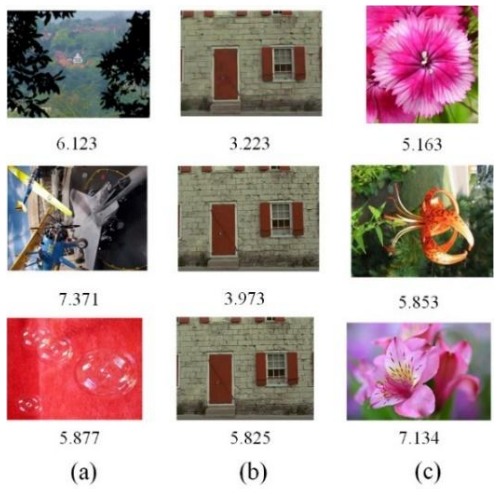

Figure 4: AVA, TID2013, and Flower datasets.

The dataset used in this paper is divided into a training set, a validation set, and a test set in the ratio of 6:1:1.

4.2 Experimental Environment Configuration

The specific environment configuration of the HSV-NIMA network designed in this paper for training the artwork dataset is shown in Table 1. The original NIMA network framework, CNN, was implemented using the TensorFlow architecture. In this paper, the required environment configuration is discussed after the Pytorch deep learning architecture was used to implement it. The baseline CNN weights are initialized by training on ImageNet, and the last fully connected layer is randomly initialized.

Categories	*parameters*
operating system	Windows 11
CPU version	Intel(R)Core(TM)i9-14900k CPU
GPU version	NVIDIA GeForce RTX 4090 GPU
video memory	24G
Deep Learning Framework	pytorch1.8.1
programming language	Python3.9.18
CUDA	11.1

Table 1: Model training environment configuration.

To optimize the training of the HSV-NIMA network, the following training parameters were configured for the homemade dataset and environment used in this study: The training epoch was set to 150, the batch size was set to 32, and the image input size was resized to 640×640. The initial learning rate α was set to 0.01, β was set to 0.05, and the number of working threads was set to 8.

4.3 Evaluation Metrics

To study the performance of the methodology of this research, various metrics were used to evaluate its performance. The main ones are Accuracy, LCC(linear correlation coefficient), and MSE(Mean Square Error).

The accuracy in this paper is calculated using the normalization error. This method is able to quantify the closeness between the predicted scores and the actual scores. The specific formula is as follows:

$$\text{Accuracy}_i = 1 - \left| \frac{\text{Predicted Score}_i - \text{Actual Score}_i}{\text{Max Score} - \text{Min Score}} \right| \tag{4.1}$$

where Accuracy_i is the accuracy of the i-th image, Predicted Score_i is the predicted score of the i-th image, Actual Score_i is the actual score of the i-th image, and Max Score and Min Score are the maximum and minimum values of the score range.

Next, the average of the accuracy of all the images is calculated to get the average accuracy (Accuracy) of the whole dataset:

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \text{Accuracy}_i \tag{4.2}$$

where N is the total number of images in the dataset, $\sum_{i=1}^{N} \text{Accuracy}_i$ is the sum of the accuracy of all images.

LCC is used to measure the strength and direction of the linear relationship between two variables. Its value is between -1 and 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no linear correlation. The LCC is calculated using the formula:

$$\text{LCC} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{4.3}$$

where x_i and y_i are the predicted and actual values, respectively. $\bar{x}$ and $\bar{y}$ are the average of the predicted and actual values, respectively. n is the number of data points.

MSE is the average of the squares of the difference between the predicted and actual values and is used as a measure of the accuracy of the predicted values. MSE is calculated by the formula:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - x_i)^2 \tag{4.4}$$

where x_i is the i-th predicted value. y_i is the i-th actual value. n is the number of data points.

Among these evaluation metrics, a higher Accuracy indicates better performance. The LCC ranges from -1 to 1. The closer the LCC value is to 1, the stronger the linear relationship between the predicted and actual values, indicating that the model's predictions are more consistent with the actual results. Conversely, an LCC value close to -1 indicates a strong negative linear relationship between the predicted and actual values. For MSE, lower values are preferable. A smaller MSE indicates a smaller difference between the predicted and actual values, suggesting higher prediction accuracy. Conversely, a larger MSE indicates a larger prediction error and poorer model performance.

4.4 Experimentation and Discussion

Prior to color extraction, we validated the effectiveness of colorblind color features. The color space conversion is shown in Figure 5. Comparative analysis reveals that HSV color extraction significantly outperforms RGB, especially in preserving the original color's integrity and central color accuracy. By leveraging the hue, saturation, and value channels in the HSV color space, our method achieves more precise color differentiation, avoiding the ambiguity of approximate color extraction. This approach is particularly suitable for intricate color patterns.

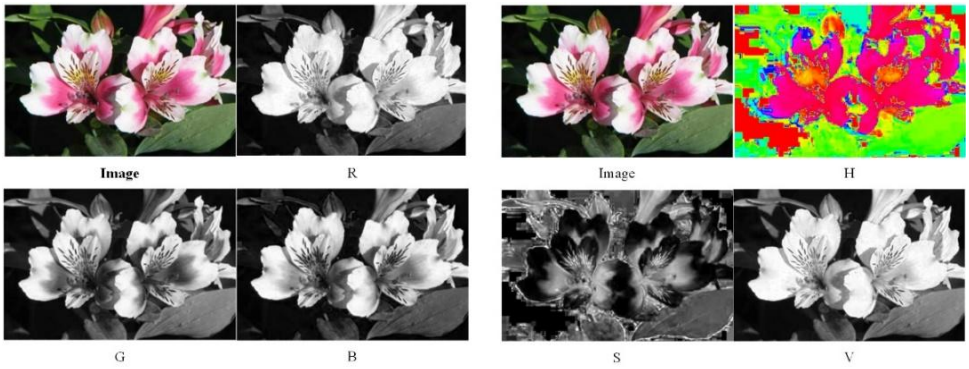

Figure 5: Color space conversion chart.

Subsequently, we isolated the prominent colors at the color epicenters of the images to derive the numerical features listed in Table 2. We then reconstructed the images within the color space framework. Our observations reveal that while both color spaces can reconstruct the images on a 1:1 basis using the extracted hues, the HSV color space more accurately preserves the original color distribution, whereas the RGB color space introduces discrepancies during the reconstruction phase.

Color Category	*RGB*	*HSV*
Red	3.01%	3.27%
Green	0.20%	0.28%
Blue	0.30%	0.34%

Table 2: Reconstructing the color center of an image with the percentage of each color.

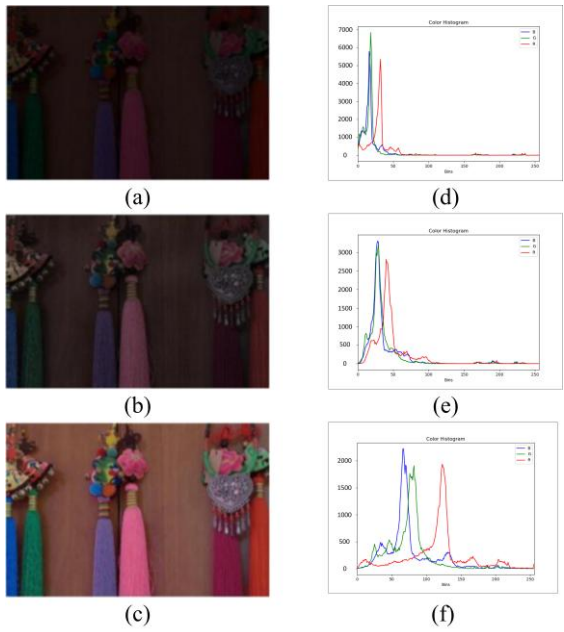

Figure 6: Images with different colors and sharpness and their color histograms.

Figure 6 shows images with varying levels of color and sharpness, along with their corresponding color histograms. The images are presented in panels (a)–(c), while the histograms are shown in panels (d)–(f). It is evident that higher image sharpness and richer color content are associated with better image quality. The histograms reveal a more uniform and diverse distribution of colors as sharpness increases. In contrast, lower-resolution images exhibit more centralized histogram peaks, indicating a limited color palette and a tendency towards a few dominant hues. As image quality improves, the histogram peaks become more dispersed, reflecting a greater variety of color combinations and a more uniform distribution of pixels across different colors. This dispersion signifies enhanced color contrast and saturation, making the image more visually appealing. The broader histogram range, from darker to lighter shades, highlights the increased level of detail within the image.

The above analysis once again proves that the classification of color into hue, saturation, and luminance is advantageous for the extraction of color in the dataset. Therefore, in another channel of the image color evaluation network framework, we chose to use the HSV approach for color feature extraction.

In this paper, we first use NIMA (MobileNet), NIMA (VGG16), and HSV-NIMA for training. The initial network model of NIMA is trained under the AVA dataset, and the results obtained are shown in Figure 7. The numbers in parentheses in Figure 7 represent the actual scores, and the numbers outside the parentheses represent the predicted scores. We can observe that the evaluation results of the initial model are very much related to the distortion, the image noise points, and the comfort level, but there is not much attention to color richness as well as color matching.

Table 3 summarizes the training results of mainstream methods on the flower dataset, revealing distinct characteristics in image quality evaluation among different models. The MobileNet-based NIMA model performs poorly in terms of LCC, MSE, and accuracy, with 0.518, 0.152, and 80.21%, respectively, which indicates its limitation in capturing the color features of the dataset. In contrast, the VGG16-based NIMA model achieves 0.610 in LCC, 0.114 in MSE, and 80.75% in accuracy, which shows high correlation with the actual data and small prediction bias.

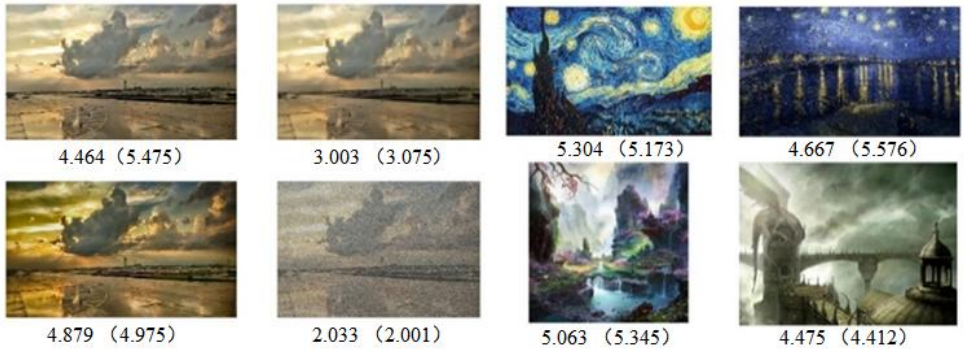

Figure 7: Initial network model training results under AVA dataset.

The SRRR model is slightly higher than the VGG16-based NIMA in LCC, 0.616, and 0.113 in MSE, but with a slightly lower accuracy of 80.03%, which may imply some errors on some classification boundaries. The KD-OIQA model is similar to SRRR on LCC at 0.607, but has a higher MSE of 0.131, and the accuracy is the highest of all models at 81.12%, which indicates that despite a slightly lower correlation between the predicted color distribution and the actual distribution, it has the best performance in classification tasks. best performance. In contrast, the HSV-NIMA model proposed in this paper performs the best on all metrics, especially on the accuracy, which reaches 81.69%, which emphasizes the importance of the color space in the assessment of image quality and demonstrates that the model's ability to assess datasets with rich colors and high image quality can be significantly improved by integrating the HSV color space into the NIMA model.

Model	*LCC*	*MSE*	*Accuracy*
NIMA (MobileNet)	0.518	0.152	80.21%
NIMA (VGG16)	0.610	0.114	80.75%
SRRR	0.616	**0.113**	80.03%
KD-OIQA	0.607	0.131	81.12%
HSV-NIMA	**0.618**	0.121	**81.69%**

Table 3: Training results of mainstream methods on the flower dataset.

Figure 8 presents the training results for each dataset under the HSV-NIMA network model. Panels (a), (b), and (c) correspond to the AVA, TID2013, and Flower datasets, respectively. The numbers in parentheses indicate the actual scores, while the numbers outside the parentheses indicate the predicted scores. Table 4 summarizes the average scores and accuracy of the training.

Upon evaluating the HSV-NIMA network model's training outcomes across the AVA, TID2013, and Flower datasets, it becomes apparent that color stability has a significant influence on model efficacy. The uniformity of colors in the AVA dataset offers a distinct learning pattern for the model, enabling it to attain an 80.84% accuracy rate and an average score of 5.351 on this dataset.

Conversely, the TID2013 dataset is characterized by images that are not vivid enough (containing multiple types of distorted images, such as noise, blur, color distortion, etc.), and these distortions complicate image quality assessment. This is because the model needs to deal with multiple factors, such as sharpness, color accuracy, and distortion types simultaneously, which challenges the predictive power of the model. Nonetheless, the model still achieves a 79.37% accuracy rate and an average score of 3.512, demonstrating a degree of resilience. Notably, the

model excels on the Flower dataset, which is rich in color variety and image features, with an accuracy of 81.69% and an average score of 7.245.

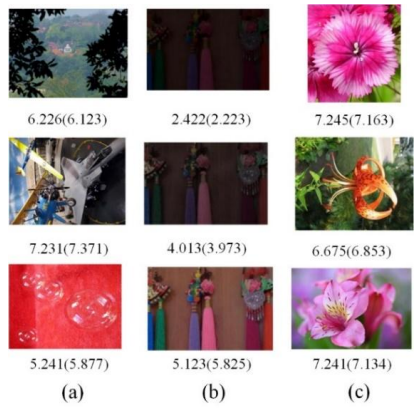

Figure 8: Training scores under each dataset.

dataset	*Accuracy*	*AS*
TID2013	79.37%	3.512
AVA	80.84%	5.351
Flower	81.69%	7.245

Table 4: TID, AVA, and training results under the Flower dataset.

This performance may be due to the model's capacity to assimilate a broader spectrum of colors and image features from the dataset, thereby enhancing its generalization capabilities. These findings indicate that color uniformity and variety are pivotal for the efficacy of image quality assessment models. Models are more adept at capturing the aesthetic attributes of images when trained on datasets with high color stability, which in turn enhances the precision of their assessments. To better observe the model's prediction confidence for each dataset. We summarize the confidence in the model's prediction scores as a histogram. This is shown in Fig. 9.

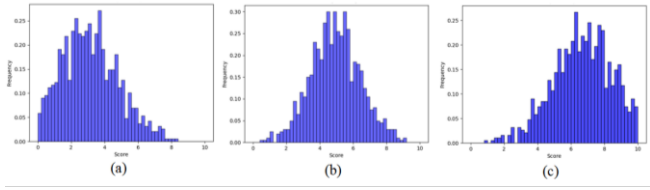

Figure 9: Distribution of image score confidence on AVA, TID13, and Flower datasets(a) Distribution of confidence of training results on AVA dataset, (b) Distribution of confidence of training results on TID13 dataset, (c) Distribution of confidence of training results on Flower dataset.

After analyzing the confidence distribution plots of image scores on the AVA, TID2013, and Flowers datasets, we can observe differences in model performance across datasets. The confidence distribution of the AVA dataset is more concentrated, indicating that the model has a high level of

predictive confidence and consistency when dealing with color-stable images. In contrast, the TID2013 dataset has a relatively dispersed confidence distribution, which may reflect the diversity of image qualities in this dataset, including unclear and color-poor images, leading to fluctuations in model predictive confidence. The Flowers dataset, on the other hand, exhibits a concentrated and generally high-confidence distribution, likely due to the nature of the homemade dataset, which allows the model to better adapt and learn from the image features, thereby improving prediction accuracy and confidence. Overall, the model performs better on datasets with higher color consistency and richer data features, highlighting the importance of high-quality training data in enhancing model prediction confidence.

4.5 Color Art Discussion

Figure 10: Grading of students' drawings.

In the teaching of art, the system designed in this study can be utilized to score work based on assessment guidelines for teacher instruction, as well as scoring outlines for each instructional task. The data set and the score labels included can be determined for scoring the work. For example, Figure 10 shows the scoring of student work for a painting task.

Utilizing the color quality assessment system developed in this study, the quality of student paintings can be rapidly evaluated without the need for face-to-face interaction with instructors. As shown in Figure 10, the system's assessment results exhibit a high degree of agreement with actual scores for some painting images. However, despite the system's ability to provide predictions that closely match actual scores in most cases, significant discrepancies were observed in certain specific instances, such as in Figure 10(b). This discrepancy arises because sketch paintings, which are typically presented in grayscale or monochrome with limited color information, pose challenges for color quality assessment. In such cases, the system relies more heavily on color diversity and saturation, which are less relevant for evaluating sketch paintings. For example, while Figure 10(b) is an excellent work overall, its limited hue variation compared to other works results in a large deviation between the predicted and actual scores.

After validating the automated color assessment model designed in this study, the experimental results demonstrate that the AI system performs well on the AVA dataset, characterized by high color stability, and also achieves excellent performance on the Flowers dataset, which is rich in color diversity. These results validate the potential of AI technology for understanding and assessing image color, particularly in the fields of art education and creation. The automated color assessment system provides instant color analysis and recommendations, helping artists make more precise

decisions regarding composition and color selection. For example, by analyzing color harmony and saturation, artists can better grasp the overall tone and emotional expression of their works, thereby enhancing the artistic effect.

In addition, the real-time feedback mechanism of the AI system provides a new way of interaction in art education and stimulates learners' creativity. However, the system has some limitations in the art of drawing and painting, with art educators and art academics stating that the scoring is not accurate enough in some complex works. Additional modules are needed to better understand and evaluate the color characteristics of sketch painting.

5 CONCLUSIONS

In this paper, through in-depth research and experimental validation, an AI-driven art learning system is successfully developed, which utilizes image recognition technology to achieve automated color assessment. By combining the HSV color space and the NIMA image quality assessment network framework, this system demonstrates significant application potential and practical value in the field of art learning. Experimental results show that the system achieves satisfactory performance on AVA, TID2013, and Flower datasets. On the AVA dataset, the system demonstrated a stable performance with 80.84% accuracy and an average score of 5.351; on TID2013, the accuracy was 79.37%, showing good robustness, and on the Flower dataset, the accuracy reached 82.69%, highlighting its ability to handle complex colors. Especially on the AVA dataset, which has high color stability, and the Flower dataset, which is rich in color diversity, the system shows high accuracy and linear correlation, indicating its potential effectiveness for art color assessment in similar contexts. In addition, the system's ability to handle color consistency and diversity provides artists and learners with a powerful color analysis tool that helps to improve the quality of art creation and learning. The contribution of this study is not only in proposing a new art learning aid, but also in advancing the application of AI technology in the field of art education. The design and realization of this system provide new ideas and a technical basis for the development of future art learning tools.

6 FUTURE WORK

In this study, while the system performs well on the colorful parts of the dataset, it lacks flexibility when evaluating tasks involving complex images with distortions (e.g., noise, blurring, color distortion, etc.), which is one of the system's limitations. Additionally, current research has not fully explored the ethical issues and social implications of the model, nor has it delved into the integration with VR (virtual reality) and AR (augmented reality) technologies. Therefore, future work will aim to use more complex networks to improve the ability to capture complex image features and explore multi-task learning frameworks to enhance comprehensive performance. Secondly, dataset diversity will be enhanced by mixing multiple datasets for training and designing targeted data augmentation strategies to improve the model's generalization ability. Finally, future work will explore integration with VR/AR technologies to provide a more immersive learning experience and facilitate innovative applications across disciplines.

Jianhua Hu, https://orcid.org/0009-0007-7049-7322

REFERENCES

[1] Qi, A.: The effectiveness of using IT in art education to develop learning motivation and psychological well-being, Educ Inf Technol 2024, 29, 19537 – 19552. https://doi.org/10.1007/s10639-024-12601-6

[2] Csillag, P.: The visual communication impacts of Itten's color contrasts investigated and empirically tested as basic principles for use in art and design, Color Res Appl, 2022, 47(2). https://doi.org/10.1002/col.22778

[3] Hong, T.; Zhao, X.: Traditional arts and crafts elements and modern art design teaching relieve student attention deficit. CNS Spectrums, 2023, 28(S2), S48. https://doi.org/10.1017/S1092852923003711
[4] Wang, J.; Mokmin, N. A. M.: Virtual reality technology in art education with visual communication design in higher education: a systematic literature review, Educ Inf Technol, 2023, 28, 15125–15143. https://doi.org/10.1007/s10639-023-11845-y
[5] Chen, R.; Aghdam, M. R. G.; Khishe, M.: Correction: Utilization of Artificial Intelligence for the automated recognition of fine arts. PLOS ONE. 2025, 20(1): e0318114. https://doi.org/10.1371/journal.pone.0318114
[6] Li, X.: Research on color quality classification algorithms in computer art education. Proc. SPIE 13063, Fourth International Conference on Computer Vision and Data Mining (ICCVDM 2023). 2024, 130631E. https://doi.org/10.1117/12.3021422
[7] Fumanal-Idocin, J.; Andreu-Perez, J.; Cordón, O.; Hagras, H.; Bustince, H.: ARTxAI: Explainable Artificial Intelligence Curates Deep Representation Learning for Artistic Images using Fuzzy Techniques, arXiv:2308.15284, (2023). https://doi.org/10.48550/arXiv.2308.15284
[8] Majumdar, A.: Individualized Real Time Automation of Color Grading and Editing Using CNNs (Convolutional Neural Networks), and/or GANs (Generative Adversarial Networks) WITHIN Mirrorless Cameras based on user history and stylistic choices, while maintaining photographic and artistic integrity. TechRxiv, April 18, 2024. https://doi.org/10.36227/techrxiv.171340698.86488804/v1
[9] Murray, N.; Marchesotti, L.; Perronnin, F.: AVA: A large-scale database for aesthetic visual analysis, 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence. RI, USA, 2012, 2408-2415. https://doi.org/10.1109/CVPR.2012.6247954
[10] Ponomarenko, N.; Ieremeiev, O.; Lukin, V.; Egiazarian, K.; Jin, L.; Astola, J.; et al.: Color image database TID2013: Peculiarities and preliminary results, European Workshop on Visual Information Processing (EUVIP), Paris, France, 2013, 106-111. https://ieeexplore.ieee.org/document/6623960
[11] Gadsden, V. L.: The Arts and Education: Knowledge Generation, Pedagogy, and the Discourse of Learning, Review of Research in Education, 2008, 32(1), 29-61. https://doi.org/10.3102/0091732X07309691
[12] Chen, J.; Mokmin, N. A. M.: Enhancing primary school students' performance, flow state, and cognitive load in visual arts education through the integration of augmented reality technology in a card game, Educ Inf Technol (2024). https://doi.org/10.1007/s10639-024-12456-x
[13] Xie, N.; Yang, Y.; Shen, H.; Zhao, T.: Stroke-based stylization by learning sequential drawing examples, Journal of Visual Communication and Image Representation, 2018, 51, 1047-3203, 29-39. https://doi.org/10.1016/j.jvcir.2017.12.012
[14] Niu, Y.; Zhang, H.; Guo, W.; Ji, R.: Image Quality Assessment for Color Correction Based on Color Contrast Similarity and Color Value Difference, in IEEE Transactions on Circuits and Systems for Video Technology, 2018 28(4), 849-862. https://doi.org/10.1109/TCSVT.2016.2634590
[15] Valencia, J.; Pineda, G. G.; Pineda, V. G.; et al.: Using machine learning to predict artistic styles: an analysis of trends and the research agenda, Artif Intell Rev, 2024, 57(118). https://doi.org/10.1007/s10462-024-10727-0
[16] Frodon, J.-M.: An Interview with Adam Lowe: To Re-produce Works of Art, Theory, Culture & Society, 2024 0(0). https://doi.org/10.1177/02632764241275837
[17] Gao, P.; Gu, G.; Wu, J.; Wei, B.: Chinese calligraphic style representation for recognition, IJDAR, 2017, 20, 59–68. https://doi.org/10.1007/s10032-016-0277-z
[18] Fang, F.; Jiang, X.: The Analysis of Artificial Intelligence Digital Technology in Art Education Under the Internet of Things. in IEEE Access, 2024, 12, 22928-22937. https://doi.org/10.1109/ACCESS.2024.3363655
[19] Zhang, W.; Wang, Y.; Du, Z.; Liu, C.; Youn, S.; Guo, X.: Machine-learning assisted topology optimization for architectural design with artistic flavor, Computer Methods in Applied

Mechanics and Engineering, 2023, 413(116041), 0045-7825. https://doi.org/10.1016/j.cma.2023.116041

[20] Lin, J.; Zhao, W.; Wang, Y.: Visual Correspondence Learning and Spatially Attentive Synthesis via Transformer for Exemplar-Based Anime Line Art Colorization, in IEEE Transactions on Multimedia, 2024, 26, 6880-6890. https://doi.org/10.1109/TMM.2024.3358027

[21] Falomir, Z.; Museros, L.; Sanz, I.; Gonzalez-Abril, L.: Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (QArt-Learn), Expert Systems with Applications. 2018, 97, 0957-4174, 83-94. https://doi.org/10.1016/j.eswa.2017.11.056

[22] Liu, L.; Ma, P.; Wang, C.; Xu, D.: Omnidirectional Image Quality Assessment With Knowledge Distillation. in IEEE Signal Processing Letters, 2023, 30, 1562-1566. https://doi.org/10.1109/LSP.2023.3327908

[23] Li, L.; Xia, W.; Fang, Y.; Gu, K.; Wu, J.; Lin, W.; Qian, J.: Color image quality assessment based on sparse representation and reconstruction residual, Journal of Visual Communication and Image Representation, 2016, 38, 1047-3203, 550-560. https://doi.org/10.1016/j.jvcir.2016.04.006

[24] Bhandari, P.; Jaiswal, M.; Puranik, V.; Kirtane, S.; Pukale, D. D.: Image Aesthetic Assessment using Deep Learning for Automated Classification of Images into Appealing or Not-Appealing, 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 187-192. https://doi.org/10.1109/ICESC48915.2020.9156003

[25] Lyu, Z.; Chen, Y.; Hou, Y.; MCPNet: Multi-space color correction and features prior fusion for single image dehazing in non-homogeneous haze scenarios, Pattern Recognition, 2024, 150(110290), 0031-3203. https://doi.org/10.1016/j.patcog.2024.110290

[26] Talebi, H.; Milanfar, P.: NIMA: Neural Image Assessment, in IEEE Transactions on Image Processing, 2018, 27(8), pp. 3998-4011. https://doi.org/10.1109/TIP.2018.2831899