International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

Leveraging The Unity In Variety Principle And Deep Learning Model To Enhance Aesthetic Appreciation Predictions For Ceramic Design

Liling Wang[1], Shahriman Zainal Abidin[1], * and Verly Veto Vermol[1]
[1] Faculty of Art & Design, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia.

Abstract
Ceramic design evaluation faces the challenge of capturing consumer aesthetic preferences that balance traditional coherence with innovative variety. This study aims to address this problem by integrating deep learning models with the Unified Model of Aesthetics (UMA) to systematically predict aesthetic features of ceramic designs. Multiple models were compared, and an improved YOLOv11s architecture—incorporating MobileNetv4 as backbone, MPDIoU loss, and a Triple Attention mechanism—achieved the best performance. The proposed model reached 79.4% precision and a mean Average Precision at IoU threshold 0.5 (mAP@50) of 79.7%. Stability was confirmed through K-fold cross-validation, with accuracy fluctuating between 78% and 80%. Furthermore, the model's predictions of unity, variety, and the principle of "Unity in Variety" showed strong alignment with participant evaluations on a 7-point Likert scale. These results demonstrate robust predictive capacity and reliable generalization, though challenges remain in classifying novel, unconventional forms. Overall, this research provides a systematic and objective framework for evaluating visual preferences by combining UMA principles with deep learning, offering both theoretical and practical significance for consumer-centered ceramic design.
Keywords: *Aesthetic preference; Convolutional Neural Network (CNN); UMA model; Unity in Variety*

1. INTRODUCTION

The global ceramic design industry has witnessed a steady resurgence, driven by increasing consumer demand for both functional and aesthetically refined products [1]. Recent research highlights those ceramics, as both utilitarian and cultural artifacts, are gaining renewed attention in international markets, especially in lifestyle and interior design sectors [2]. However, the diversity of consumer expectations, ranging from traditional craftsmanship to innovative contemporary aesthetics, poses significant challenges for accurately predicting design features. For ceramic designers, accurately understanding and classifying consumers' aesthetic cognition can better meet their needs.

The classification of ceramics faces the problem of being time-consuming and heavily dependent on professional knowledge, which has led researchers to explore computer vision (CV) and artificial intelligence (AI) as potential tools to help identify crafts [4]. Among them, image processing is the main classification method, which is the process of converting images into digital form and performing certain operations to obtain some useful information from them [5]. For example, Liu [6] extracted the characteristics of Yaozhou kiln ceramic images from three aspects: shape, type of ornamentation, and inscriptions, to achieve the purpose of advanced technology replacing traditional experts to identify porcelain decorations. This is also echoed by the study of Yi et al. [7], who created metadata from visual elements (shape, material, pattern, and color) in ceramic images and applied it to visual search technology, ultimately achieving the interpretation of visual elements in ceramic images. In addition, some studies focus on distinguishing the visual differences of materials to achieve the purpose of classifying ceramic products [8]. To further screen the methods suitable for this study, we reviewed relevant studies in recent years. Regarding traditional feature extraction methods, Sun et al. [10] used Gradient Vector Flow (GVF) and Local Binary Patterns (LBP) to identify the age of Tang, Song and Yuan porcelain pots. As an important main network of deep learning, convolutional neural networks have attracted widespread attention from researchers due to their strong feature extraction ability [11]. Compared with traditional artificial feature extraction methods, CNN reduces dependence on artificial experience, making the classification process more efficient.

Beyond technical efficiency, however, aesthetic features evaluation requires a theoretical framework that explains how consumers perceive and appreciate ceramics. The UMA provides such a framework, emphasizing two fundamental principles: unity, which relates to coherence, recognizability, and cultural continuity in ceramic forms, and variety, which reflects novelty, curiosity, and creative differentiation through shape, glaze, and decorative motifs [12]. The Unity in Variety Principle further highlights that the balance between coherence and novelty maximizes aesthetic pleasure and functionality [13]. Integrating UMA with AI-based feature extraction not only enhances classification accuracy but also improves interpretability, enabling models to capture structural harmony and stylistic innovation aligned with complex aesthetic cognition.

Consistent with this integration, Wang et al. [14] empirically investigated the influence of typicality (familiarity) and novelty (cognitive interest) on aesthetic preferences in ceramic design using the UMA framework. Their findings revealed that typicality strongly drives aesthetic preference, while novelty provides moderate engagement, aligning with the MAYA (Most Advanced Yet Acceptable) principle. Earlier studies further highlight the cognitive and emotional dimensions of form giving. Abidin et al. [15] demonstrated that designers often rely on embodied and metaphorical thinking, such as hand movements during sketching, reflecting how embodied cognition underpins intuitive form development. Similarly, Jamaludin et al. [17] emphasized the role of product form in eliciting user emotions, showing that visual analogy and form variation can effectively shape affective responses.

More recently, Toyong et al. [18] identified three types of intuition in design, such as affective, heuristic, and holistic, demonstrating that intuitive expertise complements analytical processes in shaping aesthetic outcomes. Together, these studies underscore that beyond quantifiable features, both embodied intuition and the emotional character of form significantly influence design evaluation. Integrating such perspectives with AI-based feature extraction and UMA provides a more comprehensive framework for understanding and predicting consumer aesthetic preferences in ceramic design.

Based on this, this study proposes an improved YOLO11 model. This framework has the following innovations and advantages:

(1) It introduces a novel integration of the UMA framework with deep learning to achieve systematic recognition of aesthetic visual features in ceramic design.

(2) It develops and validates an improved YOLOv11s model with MobileNetv4, MPDIoU, and Triple Attention mechanisms, demonstrating high robustness and generalization ability in aesthetic feature detection.

(3) By aligning model outputs with participant evaluations, this study provides empirical evidence that UMA principles can be computationally operationalized, thus offering a methodological bridge between design theory and AI-based evaluation.

The remainder of this paper is organized as follows. Section 2 introduces the theoretical foundation and overall architecture of the proposed model, as well as the experimental design and dataset construction. Section 3 presents and analyzes the experimental results. Section 4 summarizes the main contributions of this study and outlines directions for future research.

2. METHODS

This study adopted a standardized deep learning model development process, which is suitable for image or data prediction tasks, as shown in Figure 1.

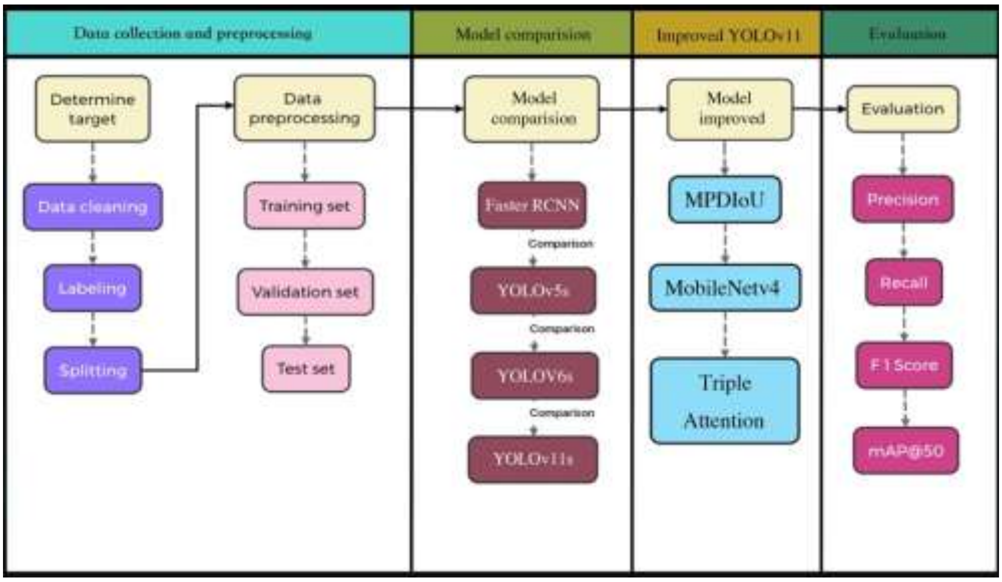

Figure 1. The Research Framework

2.1 Data Collection and Preprocessing

First, we convened a panel of ceramic design experts together with researchers from the UMA group to analyze, classify, and annotate the aesthetic features of ceramic images. These annotated images were then used as the dataset for training and validating the deep learning model. The dataset covers a wide range of ceramic design styles and visual attributes, including variations in vessel shape, rim width, decorative motifs, glazing techniques, surface textures, and color schemes. The images were primarily sourced from museum collections, academic ceramic design repositories, heritage databases, and open-access design platforms. To ensure compliance with intellectual property regulations, all images were collected strictly for academic research purposes and not for commercial use. In total, 3300 images in JPEG format were collected, each standardized to a resolution of 512 × 512 pixels.

To enhance image quality, we applied a bilateral filtering denoising method that accounts for differences in both the spatial domain and the pixel value domain. The core idea is to perform a weighted average of the pixels within the filter window of each pixel, with the weight determined by both spatial distance and color difference. The main advantage of this method is that it effectively reduces noise while avoiding the edge-blurring problem commonly associated with filters such as the Gaussian filter.

To formalize this process, the bilateral filter can be expressed as follows. $I\,(x)$ is the pixel value of the original image at position x. S is the set of all pixels within a window center at x. f_r is a color space kernel function that adjusts weights based on the difference between pixel values, usually a Gaussian function. f_s is a spatial domain kernel function that adjusts weights based on the spatial distance between pixels, usually also a Gaussian function. W_p is a normalization factor that ensures the sum of the weights. The formula 1 for the bilateral filter is as follows:

$$I_{filtered}(x) = \frac{1}{W_P}\sum_{x_i \epsilon S} I(x_i) \cdot f_r(\|I(x_i) - I(x)\|) \cdot f_s(\|x_i - x\|) \qquad (1)$$

Usually, f_r and f_s in a bilateral filter are defined using Gaussian functions as follows:

Color Gaussian function f_r: $f_r(\Delta I) = e^{-\frac{\Delta I^2}{2\sigma_r^2}}$ (2)

Spatial Gaussian function f_s: $f_s(\Delta x) = e^{-\frac{\Delta x^2}{2\sigma_s^2}}$ (3)

Secondly, in terms of data cleaning and quality control, this study employed multiple methods to optimize the quality and usability of the ceramic image dataset. Image processing tools were used to correct lighting and shadow inconsistencies, remove interference from complex or distracting backgrounds, and exclude images that were too low in resolution or excessively blurred, thereby ensuring overall clarity and reliability of the dataset. In addition, for ceramic images with intricate forms or detailed decorative patterns, contrast adjustment and sharpening techniques were applied to enhance the visibility of surface textures, glaze effects, and ornamental features. These steps provided a solid foundation for subsequent feature extraction and classification.

Thirdly, in the data annotation process, emphasis was placed on both accuracy and efficiency. The annotation team consisted of UMA researchers and ceramic design experts, ensuring that each training image was consistently and correctly labeled according to the aesthetic features of unity, variety, and Unity in Variety. We used the LabelImg tool for annotation, which supports hierarchical classification and allows fine-grained labeling of ceramic shapes, patterns, and textures. Furthermore, its magnification and comparison functions enabled annotators to carefully examine details, ensuring precision and reliability in the labeling results.

Finally, the preprocessed and annotated dataset was used to train and evaluate multiple object detection and classification models, including Faster R-CNN, YOLOv5s, YOLOv6s, YOLOv11s, and the improved YOLOv11s. All models were trained and tested on the same dataset using the same preprocessing pipeline to ensure fairness in comparison. Faster R-CNN, as a two-stage detector, performed well in fine-grained feature extraction but was relatively slow. YOLOv5s and YOLOv6s, as single-stage detectors, offered a balance between accuracy and speed, while YOLOv11s introduced novel modules (C3K2 and C2PSA) to enhance feature representation. The improved YOLOv11s model, which incorporates MobileNetv4 as the backbone, MPDIoU as the loss function, and the Triple Attention mechanism, achieved superior results in aesthetic feature recognition, demonstrating improvements in accuracy, robustness, and efficiency. Experimental evaluations (including accuracy, recall, precision, F1 score, and inference time) revealed the trade-offs across models, confirming the effectiveness of the proposed approach in ceramic aesthetic feature classification.

2.2 Model Comparison

To effectively identify aesthetic features of ceramics based on the Unified Model of Aesthetics (UMA), this study compares several deep neural network architectures to determine the most suitable backbone for feature extraction. The rationale for conducting such a comparison lies in the complementary strengths of different models: convolutional layers are effective at capturing local visual details, pooling layers reduce computational complexity, and fully connected layers support high-level classification. However, the capacity of each backbone to balance accuracy, efficiency, and robustness varies considerably. By systematically evaluating multiple models, we aim to identify the architecture that best aligns with the recognition of UMA's aesthetic dimensions—unity, variety, and their balance in Unity in Variety.

The comparison covers four representative architectures, such as ResNet50 [19], EfficientRep [20], MobileNetv4 [21], and CSPDarknet [22]. ResNet50 leverages residual connections to capture multi-level structural and semantic information, making it suitable for detecting coherence and proportion in ceramic forms. EfficientRep adopts a reparameterization mechanism to improve both training and inference efficiency, which is advantageous for processing large ceramic datasets. MobileNetv4, with its depthwise separable convolutions, emphasizes efficiency without compromising accuracy, allowing fast recognition of glaze harmony, shape balance, and texture detail. CSPDarknet employs a cross-stage partial strategy to enhance gradient flow and feature representation, making it effective for identifying intricate decorative motifs and irregular designs. To further clarify the advantages, limitations, and suitability of these backbones for UMA-based aesthetic feature recognition in ceramics, a comparative summary is presented in Table 1.

Table 1. Comparison of Backbone Networks for UMA-based Ceramic Aesthetic Feature Recognition.

Backbone	Parameters & Complexity	Strengths	Limitations

ResNet50	Heavy	Multi-level features; residuals	Slow; high computation
EfficientRep	Light	Fast; balanced detail extraction	Less deep feature power
MobileNetv4	Very light	Efficient; fast inference	Weaker in fine details
CSPDarknet	Medium-heavy	Strong gradient flow; robust reps	Resource demanding

The significance of this comparison is twofold. First, it provides a systematic evaluation of how different architectures perform in recognizing aesthetic features, offering a transparent basis for selecting the most effective model. Second, the comparative results inform the design of an improved YOLOv11s model, which incorporates the most advantageous mechanisms from these architectures to enhance both accuracy and generalization in ceramic aesthetic feature recognition. This comparative approach ensures that the final model is not only technically optimized but also theoretically aligned with the UMA framework, thereby bridging computational modeling and aesthetic theory.

2.3 Loss Function and Optimization Algorithm

Considering the specific requirements of UMA-based aesthetic feature recognition, the choice of optimization algorithm and loss function was carefully designed to ensure compatibility with the tasks of classifying and localizing ceramic aesthetic features. The overall loss function was divided into three components:

(1) Classification Loss. Cross-Entropy Loss was adopted to measure the difference between the predicted category distribution and the ground-truth labels. This enables the model to accurately classify ceramic aesthetic features into UMA dimensions such as unity, variety, and unity in variety.

(2) Bounding Box Regression Loss. For locating decorative elements or structural details in ceramic images, different detectors employ distinct loss functions. Faster R-CNN applies Smooth L1 Loss to evaluate the deviation between predicted and true bounding box coordinates. YOLO-based models use CIoU Loss, which jointly considers overlap, distance, and aspect ratio. In our improved model, we adopted MPDIoU Loss (Multi-Perspective Distance IoU), which incorporates IoU, center distance, aspect ratio, and angle alignment. This provides a more precise and stable bounding box prediction, particularly important for complex ceramic forms and irregular decorative patterns.

(3) Object Confidence Loss. In YOLO models, Binary Cross-Entropy Loss is used to optimize objectness scores, ensuring the model distinguishes ceramic features from background noise. In Faster R-CNN, object confidence is handled in two stages: the Region Proposal Network (RPN) generates candidate boxes and assigns objectness scores using binary cross entropy, followed by the ROI head that refines classification and bounding box regression. While Faster R-CNN achieves higher precision in feature localization, its inference is slower compared with the single-stage YOLO models.

Optimization Algorithm. All models were trained using the stochastic gradient descent (SGD) optimizer, combined with a warm-up strategy and cosine annealing learning rate scheduling. SGD iteratively updates model weights to minimize the gap between predictions and labels, ensuring stable convergence. This combination enhances training efficiency and improves the generalization performance of the aesthetic feature recognition models.

2.4 Model Training and Evaluation

Before training began, we optimized key hyperparameters, including learning rate, weight decay, and dropout rate, based on the validation set performance. The choice of these hyperparameters has a significant impact on the convergence speed and final performance of the model. Next, to ensure that the model fully learns the features in the training data, the model traverses the entire training set in multiple iterations. We selected enough training cycles to gradually converge the model's losses on the training set and validation set to avoid overfitting or undertraining.

In addition, the batch size has a direct impact on the training speed and stability. We comprehensively considered the computing power and video memory limitations of the training hardware and selected a suitable batch size. Finally, we used the SGD optimizer to update the model weights to gradually reduce the prediction error of the training data.

After completing the training and verification of deep learning models, the overall effect of multiple numerical reaction models will produce multiple numerical reaction models, including recall, precision, F1 score, and mAP@50. These values will be used to evaluate the model.

(1) Recall. In ceramics image recognition, the recall rate is particularly critical because if the ceramics image with unique characteristics is omitted, it may cause the loss of important information. A high recall rate means that the model can identify the characteristics of the ceramics as much as possible.

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (4)$$

(2) F1 Score. In ceramics image recognition, F1 scores help researchers evaluate the model to improve the recall rate while maintaining a high accuracy. In other words, the accuracy rate is not sacrificed while increasing the recall rate. This is very important for optimizing the overall performance of the model, especially in the data set with unbalanced categories.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

(3) Precision. In ceramics picture recognition, high precision means the identification results of deep learning models are the same as expected. The precision rate refers to the ratio of the model prediction as a positive sample, which is the proportion of the positive sample.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (6)$$

(4) mAP@50. In ceramics picture recognition, a high mAP@50 score indicates that the model effectively balances the accuracy and recall of different categories. It can capture both the details and overall characteristics of different ceramics designs while maintaining reliable classification performance.

$$mAP@50 = \frac{1}{C}\sum_{i=1}^{C} AP_i \quad (7)$$

C: Total number of classes.

AP_i: Average Precision for class *i*.

When completing the comprehensive evaluation of the model, we observed which deep learning architecture was more effective in recognizing the UMA aesthetic dimensions of unity, variety, and unity in variety. To confirm the reliability of the results, cross-validation techniques were applied, specifically the K-Fold Cross-Validation method. In this procedure, the ceramic dataset was randomly divided into five subsets. Each subset was used once as the validation set, while the remaining four were used for training. This process was repeated five times until every subset had served as a validation set. Since unity, variety, and unity in variety are complex multidimensional features, the model may be overly sensitive to certain distributions if trained on a single partition. Cross-validation therefore reduced the risks of overfitting and underfitting, improving the robustness and predictive stability of aesthetic feature recognition.

After model training and validation, we conducted human–machine comparison experiments to further examine the effectiveness of the proposed approach. Ten representative ceramic images were selected as test samples, and an online questionnaire was administered to evaluate their aesthetic features, as shown in Figure 2. The survey followed established UMA-related procedures and adhered to ethical standards. Participants rated each image on a 7-point Likert scale (“1” = strongly disagree; “7” = strongly agree) according to evaluation items adapted from Blijlevens et al. [23], such as: “This design is unified,” “This design shows variety,” and “This design is visually pleasing.” Participants’ ratings were averaged to form a comprehensive aesthetic evaluation index for each image. To ensure the academic robustness of the human validation design, this study followed prior UMA-related experiments, where sample sizes

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

typically ranged from 8 to 16 stimuli and 100–150 participants [24]. The present study used 120 participants and ten ceramic stimuli, which falls within this reasonable range. The stimuli were selected to represent varying degrees of typicality and novelty, thus directly operationalizing the UMA principles of unity and variety.

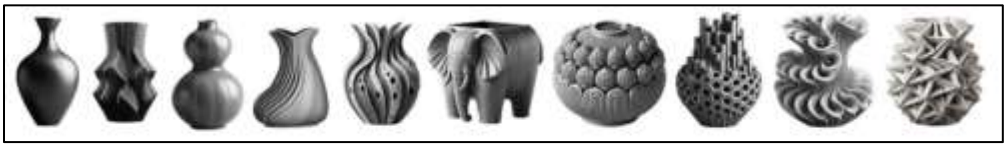

Figure 2. Ten representative ceramic images for test.

Finally, we compared the model's predicted aesthetic feature scores with the participants' evaluations using estimated marginal means. The aim was to verify the extent to which the deep learning model's recognition of UMA features corresponded with human judgments. This validation procedure enhanced the scientific rigor of the study and provided empirical support for the practical application of deep learning in ceramic aesthetic evaluation.

3. RESULTS

3.1 Performance Comparison in Different Deep Learning Model

Table 2 summarizes the performance metrics of different methods and backbone networks in aesthetic features prediction. The improved YOLOv11s model achieved the highest mAP@50 score (79.7%), indicating that its detection accuracy is better than other models. This model also has the highest Precision (79.4%) and F1 score (76.9%), showing a balanced ability in detection and aesthetic prediction. However, the recall rate of this model is relatively low at 74.6%. This means that it is not sensitive to certain types of objects, such as small objects or objects with complex backgrounds.

In contrast, YOLOv6s using the EfficientRep backbone network performed the worst among all models, with the lowest mAP@50 score (35.3%), poor precision (42.2%) and F1 score (43.3%). This shows that YOLOv6s cannot fully extract these complex features when identifying complex and high-level visual features, resulting in a decline in model performance.

In general, models using more advanced backbone networks such as MobileNetv4 and CSPDarknet53 outperform other backbone networks. This is due to their powerful feature extraction capabilities and their ability to capture high-order visual features related to aesthetic preferences. The improved YOLOv11s model has significant improvements in all indicators compared to the regular version of YOLOv11s, which shows that architecture optimization and backbone network improvement play an important role in improving detection performance. If accuracy and overall precision are the priorities, the improved YOLOv11s model is the best choice. For applications that require a trade-off between speed and accuracy, YOLOv5s is a strong candidate model with high precision and F1 score.

Table 2. Comparative test results of Faster R-CNN, YOLOv5s, v6s, v11s and improved model.

Method	Backbone	Recall (%)	Precision (%)	F1 score (%)	mAP$_{50}$ (%)
Faster RCNN	ResNet50	79.4	58.1	65.0	66.9
YOLOv5s	CSPDarknet53	76.7	73.9	75.2	78.0
YOLOV6s	EfficientRep	47.9	42.2	43.3	35.3
YOLOv11s	ResNet50	79.2	61.9	69.4	77.6
YOLOv11s (Improved)	MobileNetv4	74.6	79.4	76.9	79.7

Figure 3 shows the precision-recall curves of the improved version of YOLOv11s, analyzing the prediction performance of the deep learning model on the core features that affect aesthetic preference, such as Unity in Variety principle, unity, and variety. In addition, the results also include the overall performance of the model in handling Unity in Variety, unity, and variety. Among them, the mAP@0.5 of the unity category is 0.875, indicating that the model has a strong ability to distinguish this category. The model's mAP@0.5 in the unity category reached 0.820, indicating that it is stable in identifying aesthetic features related to unity. The curve is relatively smooth, showing the consistency of the model's performance at different thresholds. In the "unity in variety" category, the mAP@0.5 is the lowest, only 0.696. The curve is unstable, and the precision drops rapidly as the recall rate increases. This indicates that the model performs poorly in identifying aesthetic features that contain both unity and variety, which may be due to feature overlap or ambiguity between categories. The overall mAP@0.5 of the model is 0.797, and the overall curve performance is relatively balanced, which reflects its strong comprehensive ability in detecting and classifying aesthetic categories.

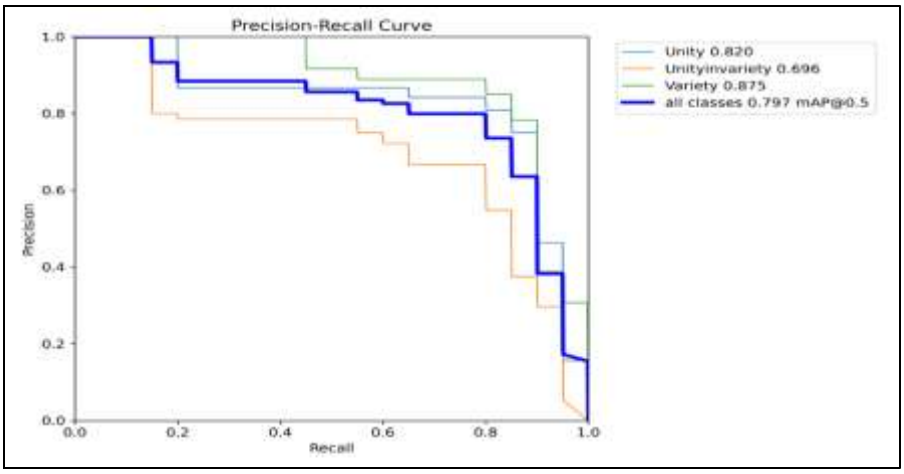

Figure 3. Precision-Recall Curve Analysis of YOLOv11s (Improved)

Figure 4 shows the F1-Confidence Curve of improved YOLOv11s, which analyzes the confidence thresholds of the deep learning model with Unity in Variety principle, unity, and variety. It also includes the relationship of the overall performance of all categories. The features of variety show the highest and most stable F1 scores in most confidence ranges, indicating a robust and balanced classification performance in terms of precision and recall. The Unity in Variety category has the lowest F1 score, indicating that this category has challenges in balancing precision and recall, possibly due to large variability or misclassification.

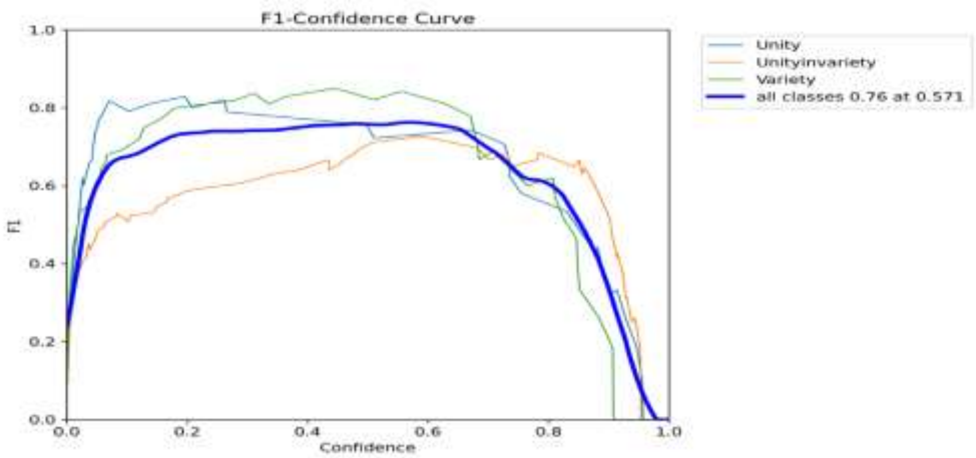

Figure 4. F1-Confidence Curve Analysis of YOLOv11s (Improved)

3.2 The Result of K-Fold Cross-Validation

After the model performance evaluation is completed, we use the K-Fold Cross-Validation method to confirm the reliability of the results. As shown in Table 3. The dataset is divided into five discounts, each time one of the discounts is used as the verification set, and the remaining four discounts are used as a training set. The recall values range from 73% to 76%, with the highest recall (76%) for fold 2. This indicates that the model is consistently able to correctly identify aesthetic features in all folds. The precision ranges from 78% to 80%, with the highest precision (80%) for folds 1 and 4. However, the lower precision compared to the recall indicates that there are moderate levels of false positives in the model's predictions. The F1 score, which is the harmonic mean of precision and recall, ranges from 76% to 77%. The highest F1 score (77%) was obtained for folds 2 and 3, indicating that this subset has the best balance between precision and recall. mAP@50 values range from 78% to 80%, with Fold 3 achieving the highest mAP@50 (80%), reflecting excellent performance in detecting and classifying PC features. This metric considers a stricter IoU threshold and illustrates the model's ability to generalize to more complex scenes.

Table 3: The Result of K-Fold Cross-Validation

Discount	Recall (%)	Precision (%)	F1 score (%)	mAP_{50} (%)
1st discount	74	80	76	78
2nd discount	76	78	77	79
3rd discount	75	79	77	80
4th discount	73	80	76	78
5th discount	75	78	76	79

3.3 Aesthetic Preference Prediction from Human

As shown in Figure 5, this study collected participants' scores for variety and unity on ten ceramic stimuli using a 7-point Likert scale. The mean scores were calculated for each stimulus to capture participants' aesthetic judgments. According to the UMA framework, unity reflects coherence and recognizability, whereas variety represents novelty and visual stimulation.

In the variety dimension, Stimulus 9 achieved the highest score (M = 4.86), followed by Stimulus 8 (M = 4.44). These designs incorporated more unconventional or decorative features, suggesting that participants recognized them as highly innovative and distinct. In contrast, Stimulus 10 (M = 3.37) and Stimulus 2 (M = 3.58) received the lowest variety ratings, indicating that these designs were perceived as more typical and less innovative.

In the unity dimension, Stimulus 7 received the highest score (M = 4.98), followed by Stimulus 2 (M = 4.81) and Stimulus 8 (M = 4.33). These designs demonstrated clear coherence, proportion, and recognizability, which are characteristic of traditional or well-structured ceramic forms. On the other hand, Stimulus 4 (M = 3.87) and Stimulus 10 (M = 3.60) were rated the lowest in unity, reflecting that their irregular or unconventional shapes were perceived as less coherent.

More balanced results were observed for Stimuli 3, 4, 5, 6, and 8, where Unity (≈ 3.9–4.4) and Variety (≈ 3.9–4.3) were both above the mid-point and relatively close in value. These stimuli best exemplify the Unity in Variety principle, showing that participants favored designs combining recognizability with innovative variation.

Overall, the results show that participants differentiated clearly between unity-dominant and variety-dominant designs. Stimuli such as 2 and 7 emphasize unity, achieving high coherence scores, while stimuli such as 8 and 9 emphasize variety, achieving high novelty scores. These findings highlight the trade-off between coherence and innovation: while variety can stimulate curiosity, unity contributes strongly to recognizability and comfort. In line with the Unity in Variety principle of the UMA model, the balance of these two dimensions remains key to maximizing aesthetic appreciation.

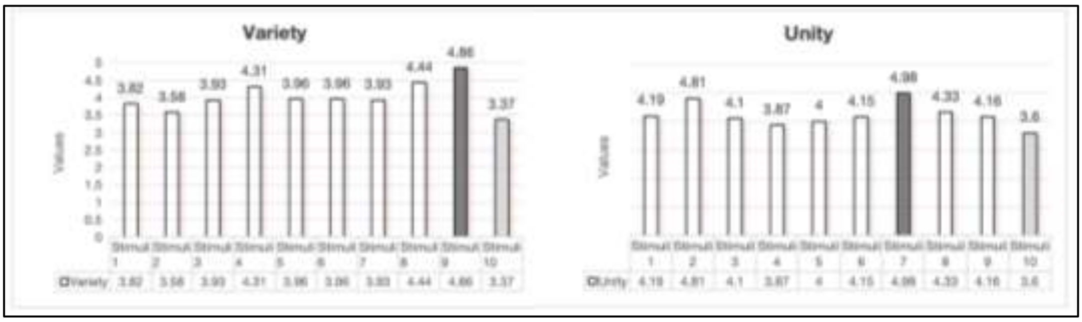

Figure 5. The Estimated Marginal Means for the Unity and Variety.

As shown in Figure 6, participants' liking scores for the ten ceramic stimuli displayed clear variation, reflecting the relative influence of unity and variety on aesthetic preference. Stimulus 10 achieved the highest liking score (M ≈ 5.2), indicating that when high levels of variety are combined with sufficient structural coherence, the design can strongly engage participants and be perceived as most aesthetically pleasing. Stimulus 1 (M ≈ 4.8) and Stimulus 6 (M ≈ 4.7) also performed well, showing that both unity-dominant designs (Stimulus 1) and balanced unity–variety designs (Stimulus 6) were positively received.

In contrast, Stimulus 2 (M ≈ 3.3) and Stimulus 4 (M ≈ 3.5) obtained the lowest liking scores. Stimulus 2, with excessive structural rigidity, shows strong unity but limited variety, reducing overall appeal. Stimulus 4, while introducing variety, lacked sufficient coherence, resulting in weaker liking evaluations.

Overall, these results demonstrate that participants' liking does not simply increase with unity or variety alone. Instead, the highest aesthetic preference emerges when unity and variety are integrated, in line with the Unity in Variety principle of the UMA model.

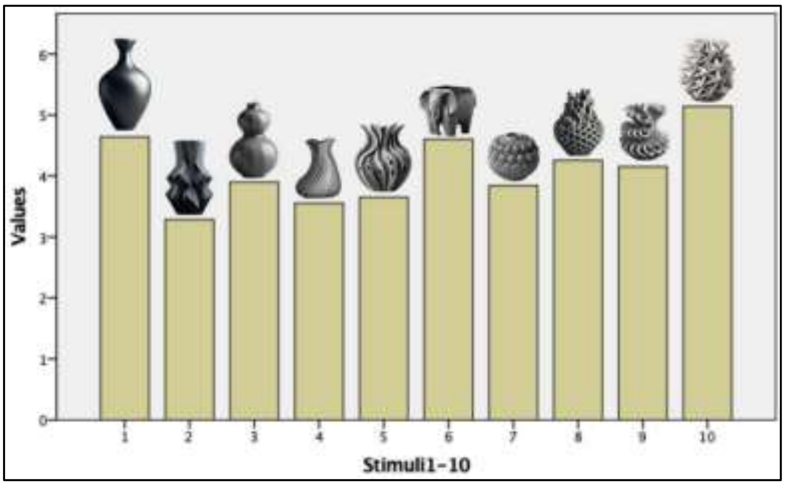

Figure 6. Aesthetic preference prediction results of Human.

3.4 Human–Model Consistency in Aesthetic Feature Recognition

As shown in Figure 7, the comparison between human evaluations and the improved YOLO model predictions demonstrates a high degree of consistency in classifying ceramic designs into Unity, Variety, and Balance (Unity in Variety) categories. Specifically, both human raters and the model identified Stimuli 1, 2, and 7 as Unity-dominant, Stimuli 9 as Variety-dominant, and Stimuli 3, 4, and 6 as balanced examples. Minor discrepancies were observed for Stimuli 5, 8, and 10, where the model tended to classify complex or highly decorative designs as Variety, while human participants perceived them as Balance. Overall, the model achieved an 70% agreement rate with human judgments, confirming its effectiveness in capturing aesthetic features under the UMA framework, while also indicating that further improvements in global feature extraction and structural coherence recognition are needed.

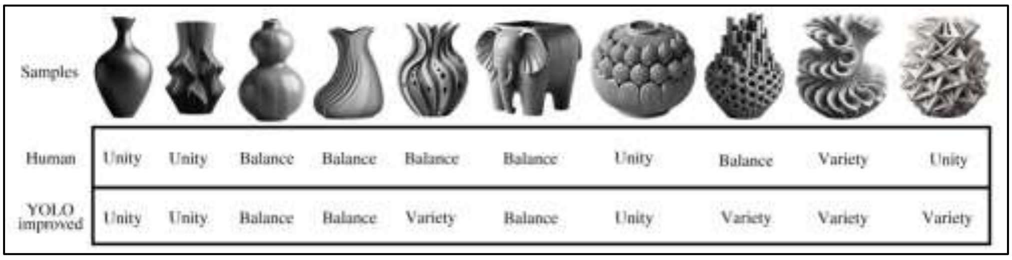

Figure 7. Comparison of aesthetic feature classification between human evaluation and the improved YOLO model across ten ceramic stimuli.

4. DISCUSSION

The findings of this study highlight the potential of deep learning models in predicting aesthetic preferences within ceramic design. The improved YOLOv11s model, which integrates MobileNetv4, MPDIoU, and Triple Attention, demonstrated strong performance with a precision of 79.4% and mAP@50 of 79.7%. These results confirm that advanced backbone networks and loss optimization strategies are effective in capturing multi-dimensional features such as unity, variety, and the Unity in Variety principle. Nevertheless, the relatively low recall (74.6%) suggests that the model still misses certain positive cases, underscoring the need for further improvement in sensitivity to complex forms.

The K-Fold Cross-Validation results further demonstrate the stability of the proposed model. Accuracy fluctuations across folds remained within a narrow range (78–80%), confirming reliable generalization. However, slightly lower recall under stricter IoU thresholds indicates that bounding box predictions may still lack precision when dealing with highly irregular shapes or unconventional decorative elements.

From an aesthetic perspective, the alignment between model predictions and participants' evaluations supports the applicability of UMA principles in computational modeling. Both humans and the model showed consistent recognition of unity-dominant, variety-dominant, and balanced designs, although discrepancies emerged in highly decorative stimuli. This indicates that while the model effectively operationalizes UMA principles, its global feature perception remains limited, particularly when novelty and irregularity dominate the design.

Overall, this study contributes to bridging aesthetic theory with AI-driven methods. However, limitations remain, especially in handling highly novel forms. Future research could focus on developing adaptive loss functions or integrating multimodal data (e.g., tactile and semantic features) to improve recognition of unconventional designs.

5. CONCLUSION

This study demonstrates the feasibility of integrating deep learning with the Unified Model of Aesthetics (UMA) to predict aesthetic features in ceramic design. The improved YOLOv11s model achieved robust performance and stable generalization, with predictions aligning closely with human judgments of unity, variety, and Unity in Variety. By computationally operationalizing UMA, the study establishes a methodological bridge between design theory and AI, providing both theoretical validation and practical tools for consumer-centered design evaluation.

The contributions are threefold: (1) proposing an improved deep learning model for aesthetic feature detection, (2) validating the UMA framework in computational prediction, and (3) demonstrating the practical potential for accelerating design evaluation and reducing reliance on subjective judgments. Nonetheless, challenges remain in classifying highly novel and unconventional forms. Future work should address these limitations by refining feature extraction strategies and exploring optimization approaches to enhance global perception.

In conclusion, combining UMA principles with deep learning provides a systematic and objective pathway for aesthetic evaluation in ceramic design, offering significant implications for both design practice and consumer research.

REFERENCE

[1] Hong, L., Wang, L., and Song, W., 2020, "Service Design and Upgrade of Domestic-Ceramic Consumption Idea: Service Design for Customized Domestic-Ceramic," *Proc. International Conference on Human-Computer Interaction*, Springer, Cham, pp. 618–632.

[2] Ewins, N., 2017, *Ceramics and Globalization: Staffordshire Ceramics, Made in China*, Bloomsbury Publishing, London, UK.

[3] Ma, Y., and Roosli, R., 2025, "Research on the Adaptive Development of Traditional Architectural Decorative Crafts: A Case Study of Chaozhou Inlaid Porcelain," *Sustainability*, 17(4), 1541.

[4] Ling, Z., Delnevo, G., Salomoni, P., and Mirri, S., 2024, "Findings on Machine Learning for Identification of Archaeological Ceramics – A Systematic Literature Review," *IEEE Access*.

[5] Huang, T. S., Schreiber, W. F., and Tretiak, O. J., 2005, "Image Processing," *Proc. IEEE*, 59(11), pp. 1586–1609.

[6] Liu, Q., 2023, "Technological Innovation in the Recognition Process of Yaozhou Kiln Ware Patterns Based on Image Classification," *Soft Computing*, pp. 1–10.

[7] Yi, J. H., Lee, H., and Kim, S., 2022, "An Analysis of the Appearance Characteristics of Korean Ceramics per Era through Statistical Analysis of Metadata Annotated with a Visual Element Classification System of Ceramics," *Heritage Science*, 10(1), 52.

[8] Lv, Z., Yu, H., Sun, L., and Zhang, P., 2022, "Composition Analysis of Ceramic Raw Materials Using Laser-Induced Breakdown Spectroscopy and Autoencoder Neural Network," *Analytical Methods*, 14(13), pp. 1320–1328.

[9] Rahman, M. S., and Metselaar, H. S. C., 2025, "XRF Data-Based Statistical Clustering of Ancient Ceramic Artifacts and Analysing Provenance with Thermo-Mineralogical Analysis," *Spectrochimica Acta Part B: Atomic Spectroscopy*, 224, 107090.

[10]Sun, Y., 2023, "Application-Specific Convolutional Neural Networks for Brain Tumor Detection," *Proc. SPIE*, 3, https://doi.org/10.1117/12.2691696.

[11]Purwono, P., Ma'arif, A., Rahmaniar, W., Fathurrahman, H., Frisky, A., and Haq, Q., 2023, "Understanding of Convolutional Neural Network (CNN): A Review," *Int. J. Robotics and Control Systems*, 2(4), pp. 739–748.

[12]Post, R. A. G., Blijlevens, J., and Hekkert, P., 2013, "The Influence of Unity-in-Variety on Aesthetic Appreciation of Car Interiors," *Proc. 5th Int. Congress of the International Association of Societies of Design Research*, Tokyo, pp. 1–6.

[13]Post, R., Nguyen, T., and Hekkert, P., 2017, "Unity in Variety in Website Aesthetics: A Systematic Inquiry," *Int. J. Human–Computer Studies*, 103, pp. 48–62.

[14]Wang, L., Abidin, S. Z., and Vermol, V. V., 2024, "Balancing Typicality and Novelty in Ceramic Design: A Study on Consumer Aesthetic Preferences," *Environment and Social Psychology*, 9(11), https://doi.org/10.59429/esp.v9i11.3223.

[15]Abidin, S. Z., Bjelland, H. V., and Øritsland, T. A., 2008, "The Embodied Mind in Relation to Thinking about Form Development," *Proc. NordDesign 2008 Conference*, Tallinn, Estonia, Aug. 21–23, 2008.

[16]Abidin, S. Z., Sigurjonsson, J., Liem, A., and Keitsch, M., 2008, "On the Role of Formgiving in Design," *Proc. 10th Int. Conf. on Engineering and Product Design Education (E&PDE 2008)*, Barcelona, Spain, Sept. 4–5, pp. 365–370.

[17]Jamaludin, M. S., Zulkapli, M. F., and Zainal Abidin, S., 2013, "The Characteristics of Form in Relation to Product Emotion," *Proc. 15th Int. Conf. on Engineering and Product Design Education (E&PDE 2013)*, Dublin, Ireland, Sept. 5–6, pp. 716–721.

[18]Toyong, N., Abidin, S. Z., and Mokhtar, S. H., 2021, "A Case for Intuition-Driven Design Expertise," *Proc. ICoRD 2021 – Design for Tomorrow, Vol. 3*, Springer, Singapore, pp. 117–131.

[19]Koonce, B., and Koonce, B. E., 2021, *Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization*, Apress, New York, NY, USA, pp. 109–123.

[20]Nguyen, K. H., Nguyen, H. V. N., Tran, H. N., and Quach, L. D., 2023, "Combining Autoencoder and YOLOv6 Model for Classification and Disease Detection in Chickens," *Proc. 8th Int. Conf. on Intelligent Information Technology*, Feb., pp. 132–138.

[21]Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., et al., 2025, "MobileNetV4: Universal Models for the Mobile Ecosystem," *Proc. European Conference on Computer Vision*, Springer, Cham, pp. 78–96.

[22]Albahli, S., Nida, N., Irtaza, A., Yousaf, M. H., and Mahmood, M. T., 2020, "Melanoma Lesion Detection and Segmentation

Using YOLOv4-DarkNet and Active Contour," *IEEE Access*, 8, pp. 198403–198414.

[23]Blijlevens, J., Hekkert, P., Leder, H., Thurgood, C., Chen, L. L., and Whitfield, T. W. A., 2017, "The Aesthetic Pleasure in Design Scale: The Development of a Scale to Measure Aesthetic Pleasure for Designed Artifacts," *Psychology of Aesthetics, Creativity, and the Arts*, 11(1), pp. 86–98.

[24]Thurgood, C., Hekkert, P., and Blijlevens, J., 2014, "The Joint Effect of Typicality and Novelty on Aesthetic Pleasure for Product Designs: Influences of Safety and Risk," *Proc. 23rd Biennial Congress of the International Association of Empirical Aesthetics*, New York, USA, Aug. 22–24.

[25]Post, R. A. G., Blijlevens, J., Hekkert, P., Saakes, D., and Arango, L., 2023, "Why We Like to Touch: Consumers' Tactile Esthetic Appreciation Explained by a Balanced Combination of Unity and Variety in Product Designs," *Psychology & Marketing*, 40(6), pp. 1249–1262.

[26]