Cover Article

VLA-Driven Robotic Painting: Towards Human-Level Creativity Through Parallel Art

Yue Lu, Yong Dou, Chao Guo, Qinghua Ni, Tianxiang Bai, Tom Gedeon, and Thales S. W. Theseus

***Abstract*—Robotic autonomous artistic creation represents an emerging frontier at the intersection of artificial intelligence (AI) and human creativity, pushing embodied AI systems toward more human-like perceptual, cognitive, and decision-making capabilities. However, the artistic creation process inherently involves complex action sequences, intricate motion trajectories, and subtle aesthetic representations, presenting substantial challenges for robotic modeling and execution. This paper proposes a vision-language-action (VLA) driven robotic painting framework that introduces advanced VLA models into the Parallel Art architecture to provide a more effective and systematic solution to these challenges. Building upon parallel systems theory and the ACP (artificial systems, computational experiments, and parallel execution) methodology, the framework integrates VLA models into the process of artificial systems, computational experiments, and parallel execution. It constructs virtual artistic creation environments for efficient model training, develops VLA-based artistic creation models through iterative computational experiments, and enables autonomous painting perception and decision-making through creative virtual-real interaction, thereby forming an integrated workflow for robotic artistic creation. The robotic painting results demonstrate the effectiveness of the proposed method and highlight its advantages over the existing approaches.**

***Index Terms*—Robotic painting, vision-language-action (VLA) models, Parallel Art, parallel intelligence**

I. Introduction

In recent years, with the rapid advancement of multimodal large models in perception and understanding, artificial intelligence (AI) has achieved remarkable progress across a wide range of domains [1–3]. These developments have significantly expanded the boundaries of AI's capabilities in achieving human-level intelligence, leading to substantial breakthroughs in various intelligent systems and applications [4–7]. Among these diverse activities, artistic creation represents a particularly significant domain, as it encapsulates both human consciousness and executive ability [8–11]. Consequently, research on AI-driven artistic creation offers an essential pathway for investigating the extent to which AI can simulate human creativity and consciousness. Despite the considerable advancements achieved, current AI art generation methods primarily focus on directly producing final outcomes [12, 13]. In contrast, human artistic creation is inherently a progressive and multi-stage process carried out in the physical world. In this context, robotic painting has emerged as a compelling research direction, as it creates paintings through physical brush manipulation and real pigment interaction [14]. Unlike outcome-oriented generation methods, robotic painting intrinsically integrates perception, planning, and fine-grained embodied control to reproduce human-like artistic workflows.

Researchers have conducted a series of explorations in robotic painting, achieving preliminary results. Early efforts explored whether robots could physically execute painting procedures, using robotic manipulators to hold brushes and produce simplified portraits or line drawings [15–17]. While these pioneering systems verified the feasibility of robotic painting, they were limited by simplified color models, heuristic stroke-planning rules, and the absence of high-level semantic reasoning. As research progresses toward richer color palettes, diverse brushwork, closed-loop visual feedback, and multimodal interaction [18, 19], the inherent challenges of robotic painting become more apparent. The task involves complex action spaces, nuanced brush-surface interactions, and multi-step procedures such as color decomposition, pigment mixing, and layer-by-layer application [20–22].

Researchers have also introduced the framework of Parallel Art [23–25], a human-AI collaborative creation paradigm grounded in the theory of parallel systems [26, 27] and the ACP (artificial systems, computational experiments, and parallel execution) methods. Parallel Art first constructs artificial art creation systems that emulate real-world artistic processes, thereby providing efficient and controllable experimental platforms for developing and validating creative models. Leveraging a limited amount of real-world creative data, the computational experiments can then expand and simulate diverse creative strategies, generating large-scale

Manuscript received: 3 November 2025; revised: 10 December 2025; accepted: 12 December 2025. *(Corresponding author: Chao Guo.)*

Citation: Y. Lu, Y. Dou, C. Guo, Q. Ni, T. Bai, T. Gedeon, and T. S. W. Theseus, VLA-driven robotic painting: Towards human-level creativity through parallel art, *Int. J. Intell. Control Syst.*, 2025, 30(4), 274–286.

Yue Lu is with School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China (e-mail: luyue@sut.edu.cn).

Yong Dou and Qinghua Ni are with Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China (e-mail: douyong@fortunevc.com; qinghua.ni@ieee.org).

Chao Guo is with Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: chao.guo@ia.ac.cn).

Tianxiang Bai is with Aerospace Science and Industry Group Intelligent Technology Research Institute Co., Ltd., Beijing 100048, China (e-mail: flyinskybtx@sina.com).

Tom Gedeon is with School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, Perth 6102, Australia, and also with University of OBuda, Budapest 1034, Hungary (e-mail: tom.gedeon@curtin.edu.au).

Thales S. W. Theseus is with DeSci Center of Parallel Intelligence, Obuda University, Budapest 1034, Hungary.

Digital Object Identifier 10.62678/IJICS202512.10267

virtual art data for model training. Finally, through parallel execution between artificial and actual systems, Parallel Art enables real-time optimization of creative strategies by using artificial systems to forecast the outcomes of potential decision steps. Therefore, Parallel Art provides a data-efficient virtual-real interactive framework and computational experiments platform for robotic painting.

Although the above approaches have made initial progress, they remain limited in meeting the increasing demand for higher-quality painting outcomes, stronger controllability, and richer expressive capabilities, calling for further in-depth investigation [28]. The ultimate goal of robotic artistic creation is to approximate human-level creativity, which inherently requires multimodal integration across visual perception, linguistic instruction, stroke-level planning, and trajectory control. Achieving such abilities necessitates leveraging large models with broad world knowledge and the capacity to fuse heterogeneous modalities for direct and context-aware decision-making. Recent advances in vision-language-action (VLA) models [29, 30] provide a promising foundation in this regard. VLA methods naturally support joint reasoning over visual observations, textual guidance, and embodied action sequences, offering a unified mechanism to connect high-level semantic understanding with fine-grained and physically executable brushstroke behaviors. These capabilities enable new possibilities for enhancing task comprehension, brushstroke planning, and control precision within robotic painting systems.

The main contributions of this paper are as follows:

(1) We construct a VLA-driven robotic painting framework that integrates the advanced VLA models with the Parallel Art paradigm to enable interactive, multimodal, and feasible integration of state observations, language guidance, and creative actions in robotic painting.

(2) We assess the feasibility of the proposed framework through comparisons of representative related painting results, highlighting its advancement in robotic painting.

The remainder of this paper is organized as follows. Section II reviews related works on robotic painting. Section III outlines the evolution of Parallel Art. Section IV introduces the proposed VLA-driven robotic painting framework and its core components. Section V discusses the key features of the framework. Section VI presents an example case and representative experimental results. Finally, Section VII concludes the paper and highlights potential directions for the future research.

II. Related Works

Research on robotic painting has evolved from early rule-based drawing systems to contemporary multimodal and learning-driven frameworks. Early explorations in the art and robotics communities focused primarily on validating the feasibility of physical robotic painting. Representative systems such as AARON [15], portrait-drawing robot Paul [16], and humanoid portrait-sketching platform [17] demonstrate the capability of robots to execute structured drawing procedures, albeit under some algorithmic simplifications that limit color complexity and action space.

As computational methods advanced, researchers begin to incorporate richer color representations, more diverse brushwork strategies, and feedback-driven stroke planning [31]. Systems such as the robot artist [18], ArtCybe [14], Cloudpainter [32], and E-David [19] introduce structured color decomposition, autonomous pigment mixing, style-transfer algorithms, and closed-loop visual feedback for brushstroke refinement [21]. Further extensions explore non-conventional tools through palette-knife-based rendering [20] and watercolor painting via tailored non-photorealistic rendering pipelines in busker robot [33].

Beyond tool-level improvements, substantial attention has been given to expanding robot input modalities and enabling richer human-robot interaction. Gaze-controlled drawing systems [34, 35] demonstrate early efforts toward intuitive artistic control. Subsequently, FRIDA [36] introduces differentiable simulation and continual re-planning to support dynamic and goal-driven artistic adaptation. Building upon this direction, speech-conditioned robotic painting [37] and collaboratively editable text-to-image pipelines such as CoFRIDA [38] further extend multimodal and interactive artistic control.

A parallel research direction focuses on improving physical realism, material modeling, and artistic fidelity. Recent works have introduced data-driven pigment-mixing models [39], physically grounded brush-surface interaction models [40], brushstroke-coherence filtering [41], and quantitative frameworks for analyzing brushstroke-based non-photorealistic rendering [42]. At the planning and learning level, efforts such as content-aware stroke optimization [43] and human-like motion modeling via trajectory autoencoding [28] address the gap between robotic execution and human artistic processes.

Despite these advances, current robotic painting systems still face fundamental challenges. The painting pipeline inherently involves complex action spaces, intricate physical interactions, and multi-stage procedures such as color mixing, stroke sequencing, and layered rendering. As a result, many existing systems rely on multi-stage heuristics or simplified approximations to remain computationally feasible, leading to insufficient modeling of the full complexity of painting processes. These limitations underscore the need for more integrated, adaptive, and semantically grounded robotic painting frameworks. Motivated by these, the presented work introduces VLA models into the Parallel Art framework to enable more comprehensive multimodal modeling of robotic artistic creation.

III. Evolution of Parallel Art

Parallel Art represents the application of parallel systems theory to the domain of machine artistic creation. Fig. 1 illustrates the origins and evolution of Parallel Art, with origins including Shadow Systems, parallel systems, and Parallel Robots, and evolutionary paths spanning from early virtual-real interactive creation to large model-driven art creation.

A. Origin of Parallel Art

The foundational idea of Parallel Art originates from the Shadow Systems framework proposed by Fei-Yue Wang in 1994 [44]. Building upon this foundation, Wang [26] further

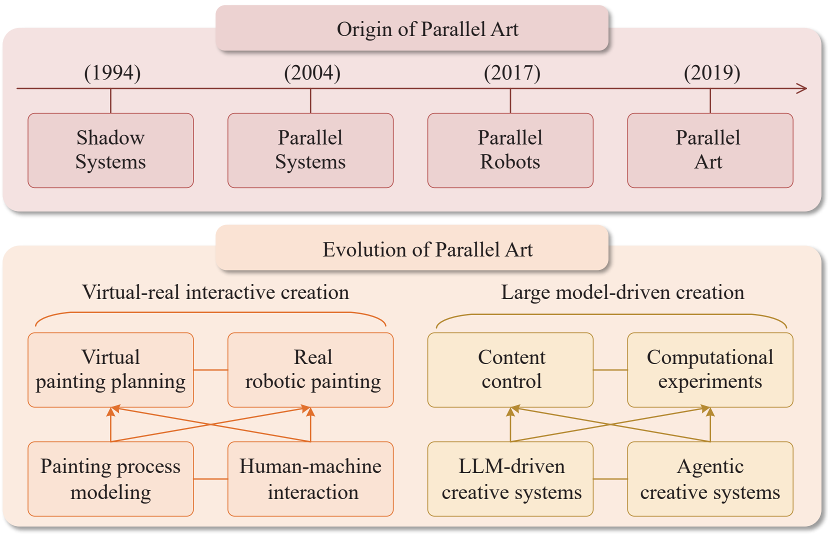

Fig. 1 Origin and evolution of Parallel Art.

introduced the parallel systems theory for the management and control of complex systems, from which the framework of Parallel Robots [45] was derived to address embodied intelligence in robotic systems. Together, these methods provide the theoretical groundwork for the establishment of Parallel Art [23].

Shadow Systems Introduced by Fei-Yue Wang in 1994 [44], Shadow Systems address data scarcity in complex and high-autonomy environments where physical experimentation is costly or constrained. A Shadow System is an artificial counterpart of the actual system, constructed with analytical models, heuristic knowledge, and simulations, and operated in parallel with the real system. This setup allows researchers to explore, test, and optimize strategies without interfering with physical operations. Initially applied in space engineering, Shadow Systems have since been adopted in robotics, intelligent transportation, logistics, and industrial automation, enabling safe and efficient experimentation. By facilitating continuous feedback and comparison between artificial and actual systems, Shadow Systems support iterative design, data-driven optimization, and intelligent decision-making.

Parallel systems To manage and control complex systems, Wang [26] proposed the parallel systems framework, centered on the ACP approach: artificial systems, computational experiments, and parallel execution. Artificial systems replicate the operating principles of the actual system, providing a low-cost experimental environment for analysis and validation when real-world experimentation is expensive or impractical. Computational experiments within these systems simulate diverse scenarios, generating large-scale synthetic data for model development and system optimization. Finally, parallel execution continuously mirrors the actual system in the artificial counterpart, allowing real-time simulations to derive and update optimal control strategies. Parallel systems recently have demonstrated significant progress across a diverse range of domains. These include intelligent transportation and autonomous mobility [46, 47], embodied intelligence and AI agents [48, 49], social and economic systems [50–52], artistic perception and creative intelligence [53–55], as well as scenario-based engineering and virtual experimentation [56, 57].

Parallel Robots With the increasing integration of intelligent and unmanned systems into production and daily life, there is a growing demand for autonomous, intelligent, and systematic robotic platforms [58]. To address challenges such as complex operational scenarios and human-robot collaboration, the Parallel Robots framework [45, 59] applies parallel systems theory and the ACP approach, providing a framework for various unmanned robotic applications. It constructs artificial robot systems for rapid training and validation, enabling virtual experiments with multi-source data and optimization algorithms in a closed-loop simulation environment. Through parallel execution, strategies optimized in the virtual system are deployed to real robots with continuous feedback, supporting iterative refinement. Overall, the framework achieves virtual-real integration, data-driven intelligence, and hierarchical collaboration, overcoming limitations of the traditional robotic systems.

Parallel Art Inspired by parallel systems and Parallel Robots, the Parallel Art paradigm [23] proposes a human-machine collaborative framework comprising artificial artistic systems, computational experiments, and parallel execution [60–62]. First, realistic artificial systems emulate real-world artistic processes, providing platforms for large-scale and long-horizon creative experiments. Second, diverse creation experiments within these systems generate rich models spanning multiple artists and styles. Finally, parallel execution between artificial and real systems enables high-fidelity artistic creation in the real-world scenarios. By integrating human-in-the-loop interaction throughout, Parallel Art supports the development of creative strategies, facilitates human-robot collaboration, and enables intelligent artistic creation.

B. *Evolution of Parallel Art*

Building upon the parallel systems paradigm, Parallel Art has evolved from early virtual-real interactive creation to contemporary large model-driven methodologies. First, virtual-real interactive creation enables large-scale acquisition of creative data and structural modeling of artistic activities. Then, with the emergence of advanced large models, large model-driven methodologies are developed to achieve autonomous and high-fidelity robotic painting.

a. Virtual-Real Interactive Creation

To establish parallel systems for art creation, methods such as painting modeling and human-machine interaction have been developed to support virtual-painting planning and real robotic painting. Together, these efforts form a virtual-real interactive creation framework that enables advanced perception and active learning for robotic painting.

Painting process modeling From the perspective of modeling the painting process, researchers formalize artistic creation as a reinforcement learning framework grounded in Markov decision processes [23], where painting skills are represented as policy functions and aesthetic judgments are modeled as value functions. Through multi-stage learning, the system progressively approximates the real creative process of human artists. In parallel, recent work has built end-to-end robotic painting pipelines that accurately reproduce artist-grade brushstroke colors through data-driven paint mixing and precise dispensing mechanisms [39]. These frameworks continuously calibrates both policy and value models across simulation and reality, providing a scalable formal structure for capturing the complexity of artistic creation.

Human-machine interaction To facilitate coherent and executable artistic creation, researchers have developed systematic architectures that tightly integrate human input, virtual modeling, computational experiments, and physical execution into a closed-loop feedback framework [55]. Recent studies further analyze machine-involved artistic creation through layered human-machine co-creation modes, showing that different levels of machine participation, including tool-level assistance, experiential augmentation, and service-level support, substantially shape collaborative creative processes [63]. These human-machine interaction paradigms ensure that artificial creative processes remain continuously aligned with robotic painting, supporting consistent style transfer, coherent content composition, and feasible brushstroke planning.

Virtual painting planning Building upon the above painting modeling and human-machine interaction paradigm, researchers have developed virtual-system painting simulations [24] within an integrated cyber-physical-social framework to capture human-like artistic creation. Related works extend virtual planning by introducing a multi-scale spray-painting strategy that iteratively refines coarse-to-fine stroke diameters to achieve efficient reproduction of non-uniform and photorealistic grayscale imagery on large surfaces [64]. Therefore, these virtual planning strategies provide a scalable simulation-driven foundation that links human artistic techniques with robotic execution, enabling machines to learn, optimize, and reproduce complex creative behaviors before physical deployment.

Real robotic painting Building on the above methods, researchers developed ShadowPainter [61], a learning-from-human robotic painting framework designed to reproduce fine art with high fidelity. This system addresses several limitations of prior robotic painters, including monotonous brushwork, limited state awareness, and the absence of objective evaluation criteria. Researches in collaborative drawing show that process-driven methods allow robots to engage meaningfully in co-creative artistic workflows through custom interfaces and tangible tools [65]. These process-driven methods demonstrate that robotic painting systems can achieve both high-fidelity artistic reproduction and meaningful engagement in co-creative workflows with human artists.

b. Large Model-Driven Art Creation

Building on the developments of virtual-real interactive creation, the emergence of large models has further catalyzed the evolution toward large model-driven creation. Large language model (LLM)-driven creative systems and agentic creative systems have been established to enable content control and computational experiments. Collectively, these efforts form a comprehensive large model-driven creation framework that supports end-to-end and high-level artistic production.

LLM-driven creative systems Leveraging the reasoning and generative capabilities of LLMs, researchers position LLMs as the central driver of artistic workflows [12]. LLMs coordinate human artists, digital artists, and robotic artists by enabling large-scale virtual experimentation and precise physical execution, thereby fostering a unified multi-modal creative process. Specifically, LLMs simulate artistic reasoning, organize conceptual abstractions, and guide downstream generative models in transforming high-level creative intents into the visual outputs. Complementary works further demonstrate that fine-tuned LLMs can translate interactive user input into precise robotic drawing commands, enabling real-world AI-assisted artistic creation with generative models [66]. Together, these works illustrate how LLMs can both orchestrate multi-agent virtual and physical artistic workflows and translate user interactions into actionable robotic commands, bridging high-level creative intent with precise embodied execution.

Agentic creative systems To establish collaborative art production, researchers introduce multiple AI agents that form an intelligent and automated ecosystem [67]. They combine foundation models with agent technologies to create a self-organizing creative environment in which digital humans, AI agents, biological artists, and robotic systems collaborate across cyber-physical-social spaces. Related agentic architectures can facilitate cross-domain creative reasoning, allowing specialized agents to translate artistic concepts across modalities while preserving semantic and emotional content [68]. These agentic frameworks demonstrate how multi-agent and foundation-model-driven systems can enable both autonomous and adaptive creative workflows, as well as cross-domain semantic reasoning, offering a scalable and interpretable paradigm for art production.

Content control Building on the LLM-driven creative systems and agentic creative systems, researchers have explored content-control mechanisms that expand and refine conceptual inputs for robotic painting [25]. This line of work typically adopts a two-stage assistant generation pipeline: the LLM first conducts chain-of-thought reasoning and interactive dialogue to structure, clarify, and enrich the conceptual description; subsequently, the refined prompts are fed into text-to-image generators to produce artistic outputs. Complementary works demonstrate that robotic artists can iteratively refine paintings using visual feedback, combining underpainting and successive adjustments to realize the intended content and composition [18]. The above approaches enable robotic painting systems to interpret and refine conceptual guidance while iteratively adjusting visual execution and enhancing compositional coherence, stylistic consistency, and fidelity to the artist's intent in the human-machine collaborative creation.

Computational experiments With the above techniques, researchers conduct systematic computational experiments on robotic painting to explore diverse aesthetic expressions in the virtual environments in the proposed ArtVerse framework [53]. By leveraging large-model-driven techniques, the system decomposes existing artworks into content and style representations, recombines these features, and generates stylized variants that convey different visual emotions while preserving the underlying semantic structure. Complementary works further demonstrate that capturing and modeling an individual artist's brushstrokes enables generative models to reproduce personalized stylistic nuances, which then can guide robotic painting for high-fidelity stylistic replication [69]. Therefore, current approaches demonstrate that computational experiments can both explore diverse stylistic variations in the virtual environments and capture artists' brushstroke characteristics, providing high-fidelity guidance for subsequent robotic painting.

IV. Overall Framework of VLA-Driven Robotic Painting

Based on the analysis of the Parallel Art paradigm, VLA models can be seamlessly integrated into this framework, enabling the system to meet the growing creative demands of both professional artists and the general public. Therefore, we develop a VLA-driven robotic painting framework that integrates the multimodal reasoning capabilities of VLA models with the structural foundations of Parallel Art.

To facilitate VLA-based learning of robotic painting, we model the entire painting process as a multimodal state-language-action tuple

$$P = (S, L, A) \tag{1}$$

which jointly captures the creation states, language guidance, and action sequence involved in artistic creation. The three components are defined as follows.

(1) The creation state is expressed as $S = (s_1, s_2, \ldots, s_T)$, where each state s_t is defined as $s_t = (s_t^{\text{canvas}}, s_t^{\text{aux}})$. Here, s_t^{canvas} denotes the RGB observations of the canvas at step t, while s_t^{aux} represents auxiliary workspace information, such as palette color distribution, brush condition, and environmental context.

(2) The language guidance is represented as $L = (l_1, l_2, \ldots, l_T)$, where l_t denotes a language token, caption, or high-level instruction at time step t, encoding creative goals, style descriptors, affective states, semantic constraints, or step-wise commands. The language stream grounds the painting behavior in the human-aligned creative intent. Future extensions can incorporate physiological signals reflecting internal human states to further enrich the modeling of creative intent.

(3) The action sequence is defined as $A = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_T)$, where $\boldsymbol{a}_t$ represents a high-level execution step (e.g., one continuous stroke), containing a variable-length sequence of low-level actions: $\boldsymbol{a}_t = (a_{t,1}, a_{t,2}, \ldots, a_{t,K_t})$. Each atomic action is parameterized as $a_{t,k} = (\boldsymbol{p}_{t,k}, \boldsymbol{r}_{t,k}, c_{t,k})$, where $\boldsymbol{p}_{t,k} \in \mathbb{R}^3$ is the brush-tip position in the canvas coordinate system, $\boldsymbol{r}_{t,k} \in \mathrm{SO}(3)$ is the orientation, and $c_{t,k}$ encodes tool attributes such as brush type, color selection, or paint loading. This formulation allows the model to express continuous and stroke-level motor control with high temporal resolution.

Given the above multimodal representation, we employ a VLA model to learn a unified generative policy for robotic painting. The VLA model jointly encodes creation states and language guidance to generate continuous robotic actions. The learning objective is to model the conditional distribution

$$\pi_\theta(A \mid S, L) \tag{2}$$

where θ denotes the parameter of the VLA model.

Action generation At each time step, the VLA model directly predicts the robotic action from multimodal inputs through a unified perception-action mapping

$$\boldsymbol{a}_t = f_\theta(s_t, l_t) \tag{3}$$

where f_θ denotes the VLA model, which jointly interprets the canvas state and the linguistic instruction to produce the corresponding action. More explicitly, the model parameterizes the distribution over the end-effector pose in six degrees of freedom (6DoF) and tool-related parameters

$$p(\boldsymbol{a}_t \mid s_t, l_t) = p(\boldsymbol{p}_t \mid s_t, l_t)\, p(\boldsymbol{r}_t \mid s_t, l_t)\, p(c_t \mid s_t, l_t) \tag{4}$$

Training objective To train the VLA model, paired datasets of multimodal trajectories $\{(s_i, l_i, \boldsymbol{a}_i)\}$ are employed. The training objective is formulated as minimizing the distributional discrepancy between the predicted action chunk $\hat{\boldsymbol{a}}_i$ and the ground-truth chunk $\boldsymbol{a}_i$. Instead of treating actions as isolated steps, we optimize chunk-level consistency by aligning their trajectory distributions.

$$\mathcal{L}_{\text{action}} = \ell(p_\theta(\hat{\boldsymbol{a}}_i \mid s_i, l_i),\, p_{\text{data}}(\boldsymbol{a}_i \mid s_i, l_i)) \tag{5}$$

where $\ell(\cdot)$ is a trajectory-level divergence measure.

VLA-based painting execution During deployment, the trained VLA model performs autoregressive action generation starting from the initial canvas state s_0 and a human-provided linguistic instruction l.

$$\begin{cases} \hat{\boldsymbol{a}}_t = f_\theta(s_t, l), \\ s_{t+1} = F(s_t, \hat{\boldsymbol{a}}_t), \end{cases} \quad t = 0, 1, \ldots \tag{6}$$

where the predicted action is executed in either the real world or the virtual painting environment and F denotes the underlying environment dynamics. Through this iterative perception-action loop, the system achieves fully autonomous and step-by-step painting driven by the VLA model.

The above multimodal representation establishes a unified formulation for robotic painting that aligns with the requirements of VLA models. It enables the system to jointly learn visual scene dynamics, language-grounded creative reasoning, and high-fidelity action execution. Widely adopted vision-language-action models can be used as foundational backbones to support robotic artistic painting tasks. Consequently, the model provides a stronger foundation for learning human-like artistic behaviors.

Grounded in parallel systems theory and built upon the above multimodal formulation, the overall framework of VLA-driven robotic painting is composed of three interdependent components: artistic artificial systems, artistic computational experiments, and artistic parallel execution, as illustrated in Fig. 2. The artistic artificial systems establish high-fidelity digital counterparts of actual artistic creation processes, capable of replicating diverse creative behaviors and stylistic expressions. The artistic computational experiments utilize artificial systems to perform large-scale creative simulations. The artistic parallel execution transfers these strategy models into actual creative scenarios. Together, these components form a closed-loop framework that integrates artificial and actual artistic processes, realizing real-data-efficient and adaptive art creation under the paradigm of Parallel Art and the cross-modal modeling ability of VLA models.

A. *Artistic Artificial Systems*

The artificial systems for art creation serve as realistic experimental platforms for modeling and analyzing artistic activities. Their central goal is to reconstruct the essential elements involved in the actual artistic creation, including virtual data such as stroke sequences, action trajectories, and painting states; virtual elements such as virtual tools, digital paintings, and virtual pigments; and virtual environments such as rendering engines, canvas simulation, and stroke modeling. By integrating these components, the artificial systems establish a comprehensive virtual environment for training VLA-driven creation models and simulating the entire creative process. This virtual platform provides a controlled foundation for conducting art-related experiments, reducing experimental costs, and improving the fidelity of artistic creation.

In terms of data acquisition, constructing the artificial systems requires both precise morphological and dynamic modeling of key entities used in artistic creation, such as canvases, brushes, palettes, and other artistic tools. High-precision data capture is also essential for documenting artistic processes and outcomes, including painting gestures, color-mixing dynamics, design principles, and the compositional and color-related attributes of finished artworks. Motion capturing and high-resolution video recording can be used to obtain the full-chain work process of artistic creation.

Beyond physical modeling, the system integrates art historical, theoretical, and aesthetic knowledge to characterize mainstream art styles and the distinctive creative signatures of representative artists. Through this multi-layered integration, ranging from physical and behavioral modeling to stylistic and genre-level abstraction, the artificial systems provide comprehensive simulation platforms for artistic creation, enabling dynamic interactions and vivid simulations within the virtual environments.

B. *Artistic Computational Experiments*

The artistic computational experiment serves as a core mechanism for improving VLA-based artistic creation models. By leveraging the artificial system's virtual components, it effectively integrates scarce real-world artistic data with abundant synthetic data generated through the autonomous virtual exploration. This hybrid-data paradigm enables the model to learn both realistic artistic rules and diverse creative behaviors, as outlined in Algorithm 1.

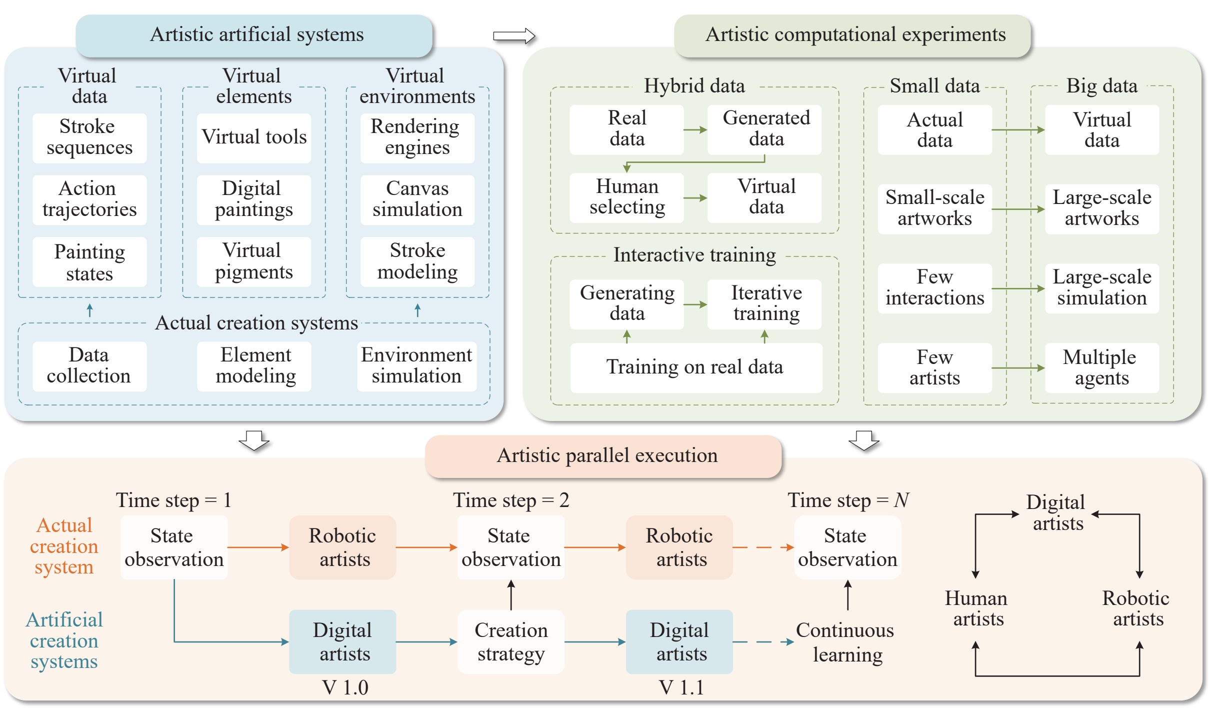

Fig. 2 Overall framework of VLA-driven robotic painting, comprising artistic artificial systems, artistic computational experiments, and artistic parallel execution.

Algorithm 1 VLA-driven computational experiments

Input:

(1) Real-world artistic creation data (state-language-action tuples) $\mathcal{D}_{\text{real}} = \{(s_i, l_i, \boldsymbol{a}_i)\}$;

(2) VLA creation model f_θ;

(3) Training epochs N_{e}, batch size N_{b}, unified action chunk length T_{uni}, and exploration steps T_{exp}.

Output:

(1) Trained artistic creation model f_θ^*;

(2) Complete set of selected generated samples $\mathcal{D}_{\text{gen}}^{\text{all}}$.

$\mathcal{D}_{\text{train}} = \mathcal{D}_{\text{real}}$, $\mathcal{D}_{\text{gen}} = \emptyset$, $\mathcal{D}_{\text{gen}}^{\text{all}} = \emptyset$;

for $t_{\text{e}} = 1$ to N_{e} **do**

  Shuffle $\mathcal{D}_{\text{train}}$ and divide into mini-batches $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{N_{\text{b}}}\}$;

  for $t_{\text{b}} = 1$ to N_{b} **do**

    Load batch $\mathcal{B}_{t_{\text{b}}} = \{(s_i, l_i, \boldsymbol{a}_i)\}$ and normalize all action trajectories to obtain $\{(s_i, l_i, \tilde{\boldsymbol{a}}_i)\}$, where each $\tilde{\boldsymbol{a}}_i$ is resampled to length T_{uni} via temporal interpolation or trajectory resampling;

    Predict robotic action chunk

$$\hat{\boldsymbol{a}}_i = f_\theta(s_i, l_i);$$

    Compute action loss

$$\mathcal{L}_{\text{action}} = \frac{1}{|\mathcal{B}_{t_{\text{b}}}|} \sum_{i \in \mathcal{B}_{t_{\text{b}}}} \ell(\hat{\boldsymbol{a}}_i, \tilde{\boldsymbol{a}}_i);$$

    Update parameters

$$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}_{\text{action}};$$

  end for

  Autonomous exploration: Initialize the painting environment with an incomplete canvas state s_0 and a human-provided linguistic instruction l. Then generate a multi-step rollout by iteratively applying

$$\begin{cases} \hat{\boldsymbol{a}}_t = f_\theta(s_t, l), \\ s_{t+1} = F(s_t, \hat{\boldsymbol{a}}_t), \end{cases} \quad t = 0, 1, \ldots, T_{\text{exp}} - 1,$$

  where F models the dynamics of the environment;

  Store the synthesized transition sequence

$$\mathcal{D}_{\text{gen}} \leftarrow \{(s_t, l, \hat{\boldsymbol{a}}_t)\}_{t=0}^{T_{\text{exp}}-1};$$

  Human experts filter $\mathcal{D}_{\text{gen}}$ and update $\mathcal{D}_{\text{gen}}^{\text{all}}$

$$\mathcal{D}_{\text{gen}}^{\text{selected}} \subseteq \mathcal{D}_{\text{gen}}, \quad \mathcal{D}_{\text{gen}}^{\text{all}} \leftarrow \mathcal{D}_{\text{gen}}^{\text{all}} \cup \mathcal{D}_{\text{gen}}^{\text{selected}};$$

  Augment the training set

$$\mathcal{D}_{\text{train}} \leftarrow \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{gen}}^{\text{all}};$$

end for

Return f_θ^* and $\mathcal{D}_{\text{gen}}^{\text{all}}$.

We first construct an initial training dataset $\mathcal{D}_{\text{train}}$ using the real-world artistic creation samples. A pretrained VLA model acts as the backbone, whose multimodal encoder and action decoder are further adapted to robotic painting through supervised training. This step allows the model to align its general vision-language-action representations with the domain-specific characteristics of artistic creation tasks.

After each training epoch, the updated model is deployed in a high-fidelity rendering environment to perform virtual painting. During virtual interaction, the model encodes multimodal observations and predicts painting actions. This virtual execution generates diverse synthetic trajectories. Compared with the purely real data, these synthetic samples enrich stylistic diversity, increase coverage of rare states, and expose the model to exploration-driven variations difficult to obtain from physical painting processes.

To ensure the data quality, a human-guided sample selection stage is introduced. Instead of unselectively accepting all synthetic outputs, human evaluators filter the generated samples based on artistic quality, action rationality, and creative relevance. This produces a curated subset $\mathcal{D}_{\text{gen}}^{\text{selected}}$ to be accumulated into a long-term synthetic dataset $\mathcal{D}_{\text{gen}}^{\text{all}}$. Such selective incorporation prevents error accumulation and maintains a stable supply of high-quality supervision.

The long-term synthetic samples are then incorporated into the training data.

$$\mathcal{D}_{\text{train}} \leftarrow \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{gen}}^{\text{all}} \tag{7}$$

Through repeated cycles of model updating, virtual exploration, human-in-the-loop curation, and dataset augmentation, the VLA model gradually enhances its perceptual grounding and creative expressiveness. Future extensions can explore integrating additional feedback channels, potentially capturing subtle or non-conscious human influences, to further refine the alignment between artificial and human artistic decision-making within the parallel execution framework.

C. Artistic Parallel Execution

The parallel execution mechanism governs the interaction between the artificial and actual systems during artistic creation. Its core principle lies in enabling the artificial systems to dynamically mirror and guide the actual creative process through continuous synchronization and adaptive control. Through continuous observation of the actual system and synchronized simulation within the artificial system, it enables the selection of optimal execution policies and the dynamic adjustment of creative behavior in real time.

Parallel execution addresses two fundamental challenges: the domain discrepancy between artificial and actual systems and the occurrence of unforeseen events during the real-world creation. To mitigate these issues, the artificial systems continuously track the state of the actual system in real time and conduct multiple virtual trials exploring alternative creative strategies. By analyzing the simulated outcomes, it identifies the optimal control policies for real-world implementation, thereby reducing the creative risk, enhancing the decision-making accuracy, and improving the overall efficiency in creative process.

Real-time synchronization further enables continuous learning and adaptive refinement of the artificial system. When human feedback or unexpected conditions arise during actual creation, the artificial system accordingly incorporates such feedback and environmental changes, updates its internal models, and re-optimizes creative strategies. This process establishes a collaborative workflow among three roles: digital artists, human artists, and robotic artists. Consequently, the system achieves a human-centered, context-aware, and adaptive creative process that seamlessly integrates human intuition with artificial intelligence, realizing the vision of intelligent artistic collaboration.

V. Key Features of VLA-Driven Robotic Painting

Building upon the three fundamental components, including artistic artificial systems, artistic computational experiments, and artistic parallel execution, the VLA-driven robotic painting framework enables intelligent artistic creation. These components embody the descriptive, predictive, and prescriptive processes in the context of artistic creation, respectively, as shown in Fig. 3.

A. Descriptive Process

Descriptive process serves as the foundation for simulation and modeling in artistic creation, providing a structured representation of artistic activities through virtual entities, creative processes, and artistic outcomes. The VLA model plays a central role in guiding this process, jointly modeling the three key modalities: vision-language-action, thereby informing the organization, collection, and utilization of multi-source artistic data.

Guided by the requirements of VLA models, the descriptive process must simulate artistic creation from both the environmental and mechanistic perspectives. On the environmental side, it models the key elements such as lighting simulation, pigment texture, material models, and brush dynamics, providing a realistic and creative setting for simulation. On the mechanistic side, it reconstructs detailed processes including stroke interaction, palette management, artist profile, and mixing simulation, enabling the high-fidelity reproduction of artistic mechanisms.

Through the multi-level modeling of artistic creation and encompassing physical interactions, stylistic evolution, and aesthetic outcomes, descriptive process supports large-scale experimental inference using limited real-world artistic data. This establishes an informational, intelligent, and structural foundation for subsequent computational training and strategy derivation.

B. Predictive Process

Predictive process constitutes the analytical and inferential core of the VLA-driven robotic painting framework. It enables large-scale computational experiments that simulate and predict the future states of artistic creation process based on the current conditions and decision-making actions. By integrating VLA models, the system can efficiently perform extensive creative simulations, supporting diverse explorations of artistic strategies and outcomes.

Relying on the high-fidelity computer graphics and emerging the world model technologies, the VLA-driven robotic painting framework performs hierarchical simulations of artistic creation. At a global level, state-prediction models are constructed and trained using data from the actual system and high-quality synthetic data from artificial systems. These models capture artistic structure, style, and compositional logic, ensuring coherence in creative planning. At a fine-grained level, rendering engines and fluid interaction models simulate pigment dynamics and brushstroke behaviors with high precision, achieving realistic visual effects.

Through these layered simulations, predictive process provides reliable support for data generation, large-scale computational inference, and construction of high-precision artistic strategy models. It thus forms the computational backbone that bridges descriptive modeling and prescriptive control in the proposed framework.

C. Prescriptive Process

Prescriptive process represents the decision-making and control capability of the VLA-driven robotic painting framework, responsible for guiding and regulating the creative processes of actual system. It completes the closed-loop cycle of data collection, model training, and real-world execution by continuously regulating and refining the interaction between artificial systems and actual creative environment, ensuring seamless and adaptive integration.

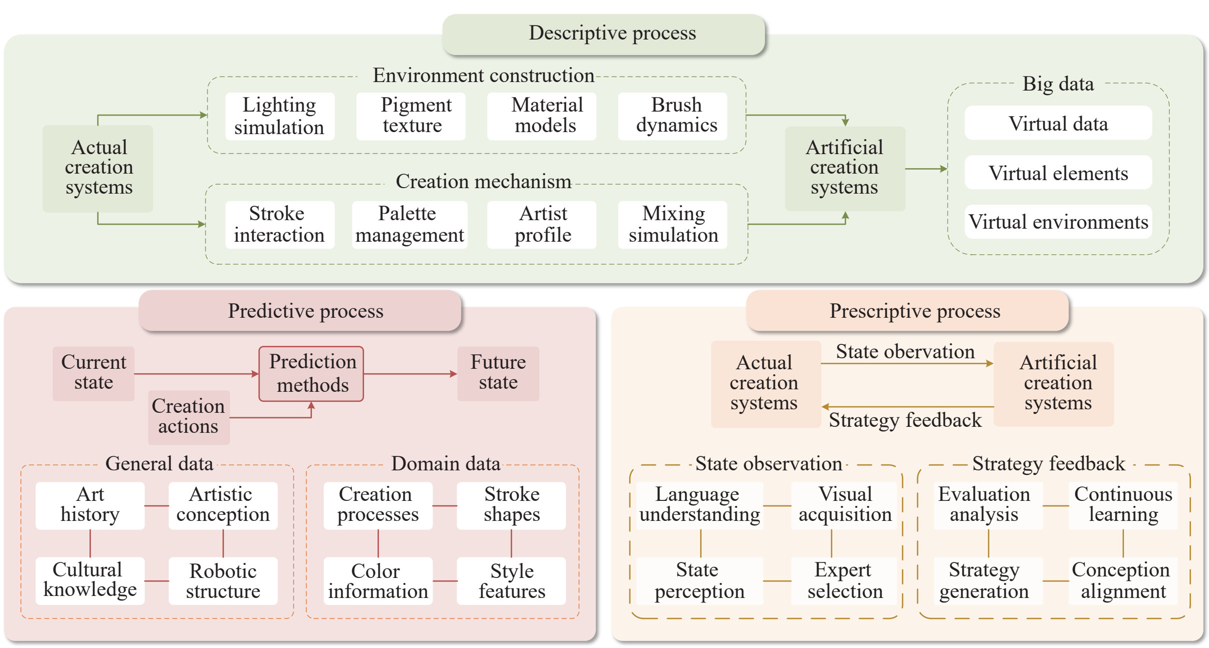

Fig. 3 The descriptive, predictive, and prescriptive processes of VLA-driven robotic painting.

The essence of prescriptive process lies in real-time interaction between artificial and actual systems. The artificial systems provide real-time decision support for artistic execution, generating creation strategies based on the current state of the actual system while simultaneously updating its internal representations. Through continuous simulation and iterative refinement, it identifies high-quality control actions that significantly enhance both the creative quality and operational efficiency, driving more effective artistic outcomes.

The actual system provides feedback reflecting environmental variations and human creative intentions. The prescriptive process integrates this feedback into a structured control loop with two key components: state observation and strategy feedback. In the state observation module, the system acquires comprehensive context information through language understanding, visual acquisition, state perception, and expert selection, capturing both environmental conditions and human creative intent. In the strategy feedback module, the system conducts evaluation and analysis, performs continuous learning, generates refined strategies, and aligns conceptual plans with human aesthetic goals. When unexpected events or new artistic requirements arise, the artificial systems incorporate this feedback, rapidly adapt their models, and refine decision strategies to maintain coherence with human preferences and contextual dynamics. By unifying decision-making, simulation, and iterative feedback, the prescriptive process enables precise, adaptive, and human-centered management of artistic creation, empowering the VLA-driven robotic painting framework to achieve high-quality and context-aware capabilities in the complex creative environments.

VI. Case Study

A. Implementation Details

Leveraging multimodal joint modeling capabilities, the VLA model is seamlessly integrated into the processes of artificial systems, computational experiments, and parallel execution. Within this structure, the proposed VLA-driven robotic painting framework enables high-fidelity, human-aligned, and context-aware artistic creation, as illustrated in Fig. 4. The framework provides a conceptual and methodological demonstration of how multimodal perception, creative reasoning, and embodied action can be cohesively unified within a single and dynamically evolving artistic workflow. In doing so, it highlights the expanding potential of intelligent systems to engage in complex, adaptive, and iterative forms of creative expression.

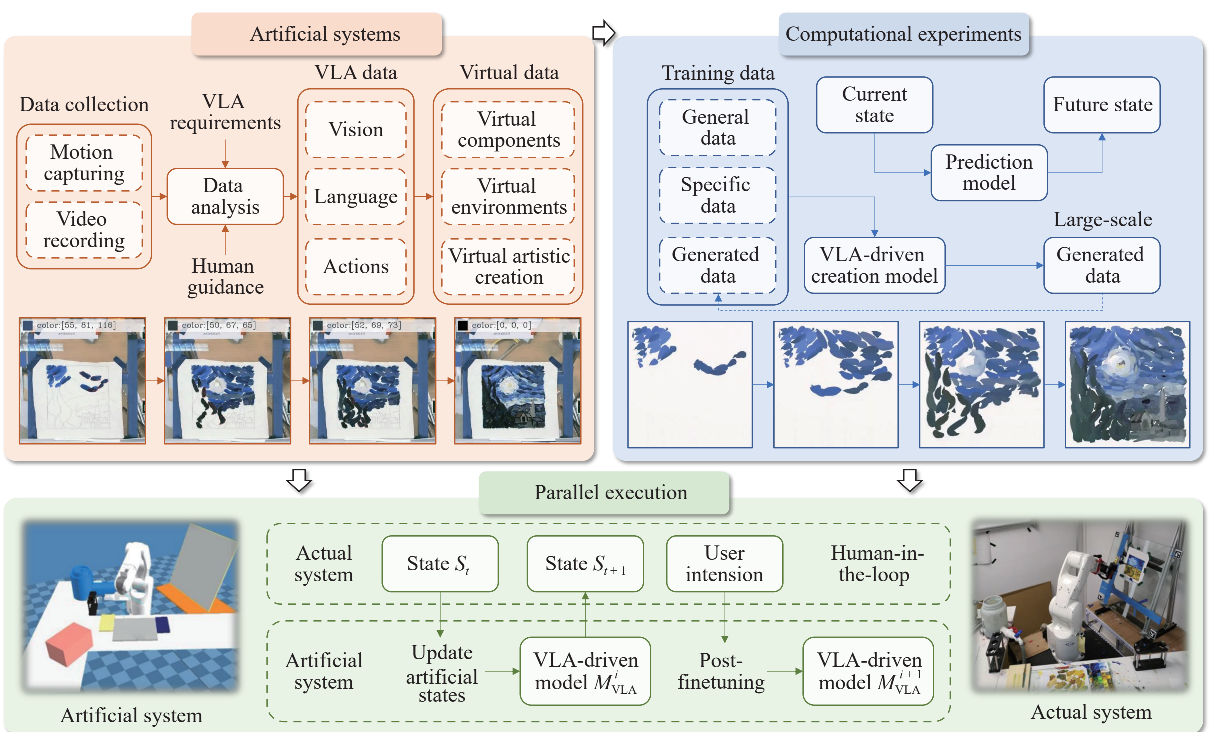

Fig. 4 Example case of the VLA-driven robotic painting framework, illustrating the workflow from computational experiments to real-world painting.

In terms of artificial systems, the VLA model allows for capturing coupled visual, linguistic, and action data, forming long-term multimodal sequences that provide high-precision inputs for training decision models. The system collects high-precision multimodal data during the human painting process, including brushstroke color, stroke region, and motion trajectories, thereby constructing a comprehensive and fine-grained art creation database. This database not only captures the stylistic characteristics and dynamic gestures of human artists, but also provides essential inputs for the subsequent training of multimodal creation models. In data modeling, the VLA model offers a general interface that supports pre-training on related domains, such as robotic manipulation, to acquire fundamental capabilities. After fine-tuning on extensive artistic data, the model can generate large-scale and high-quality creative datasets to support the construction and optimization of artificial systems. These mechanisms significantly enhance both the fidelity of artificial systems and the accuracy and depth of data modeling.

During computational experiments, the VLA model functions as the core of creative decision-making process. It directly outputs sequences of robotic actions, enabling seamless integration with artistic workflows and precise robotic control. Large-scale computational experiments are performed within the artificial systems. Brushstroke rendering, policy training, and model construction produce an initial art creation model. Generated outputs are evaluated by humans to retain high-quality creations to iteratively refine both the model and the virtual dataset. Additionally, its flexible language output allows human artists to interact naturally with the system, controlling stylistic aspects or guiding creative interventions. This combination of precise action control and flexible human interaction enables the generation of rich datasets and robust creative models. For the network architectures, widely adopting vision-language-action models can be used as foundational backbones to support painting computational experiments.

In parallel execution, the VLA model facilitates real-time interaction between artificial systems and actual systems. Its visual perception, language understanding, and action generation modules can be decoupled, allowing human artists to adjust model behavior during execution. Pretrained perception and control modules support flexible adaptation to different artistic styles and preferences. The model can also be fine-tuned based on execution feedback, dynamically aligning its performance with human intent. These mechanisms improve both the adaptability and effectiveness of parallel artistic execution. Through parallel execution, the trained model guides an actual painting robotic arm to produce artworks according to human creative intent. Real-time interaction between artificial and actual systems enables feedback integration and human-in-the-loop collaboration.

B. Experimental Results and Analysis

Evaluating the performance of robotic painting systems remains a challenge. These systems widely vary in their robotic arm configurations, pigment types, painting styles, and artistic objectives. Moreover, the artistic quality of the generated paintings is inherently subjective. Unlike well-defined tasks such as image classification and segmentation, the evaluation of robotic painting shall be tailored to specific analytical perspectives, with visual comparison playing a central role in the evaluation.

To examine the feasibility of proposed framework, we compare the preliminary results produced by the Parallel Art-based painting system with the representative robotic painting systems. Fig. 5 presents a qualitative comparison of artworks generated by the Parallel Art framework and several notable systems, including (a) Paul [16], (b) ArtCybe [14], (c) Robot artist [18], and (d) E-David [19].

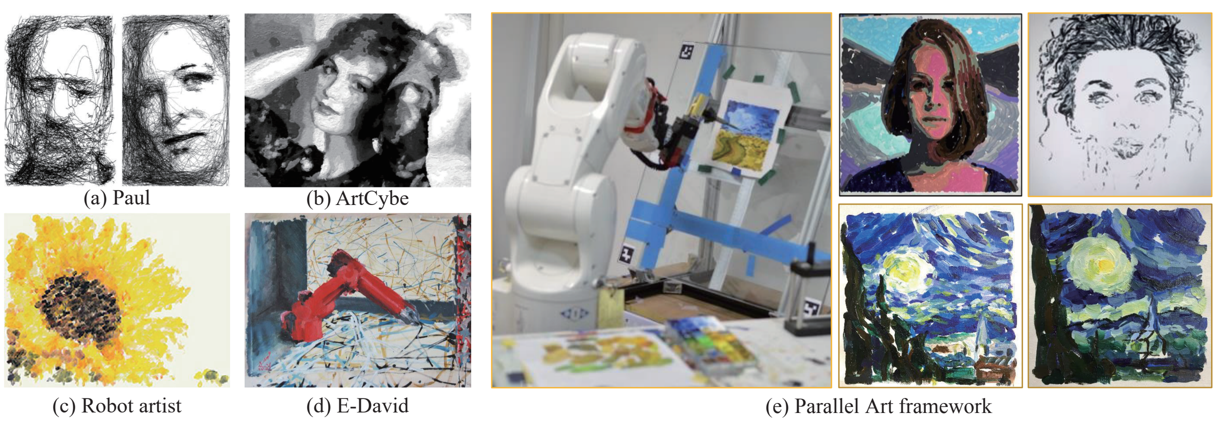

Fig. 5 Qualitative comparison of creation results from representative painting robots and the Parallel Art framework: (a) Paul [16], (b) ArtCybe [14], (c) Robot artist [18], (d) E-David [19], and (e) the Parallel Art framework.

As shown in Fig. 5, current robotic painting systems exhibit notable gaps from the human-level artistic performance in terms of painting form, color richness, and stroke diversity. Paul relies primarily on uniform monochromatic line strokes, which limits expressiveness across different colors and structural forms. ArtCybe produces vivid outputs but remains constrained to monochrome paintings. Robot artist is capable of full-color painting. However, its stroke patterns are dominated by short tapping actions, falling short of the diverse stroke dynamics in the human artmaking. E-David is able to generate longer brushstrokes, yet its control over stroke technique, angle variation, and stylistic consistency still leaves room for performance improvement.

In contrast, the Parallel Art framework performs human-like painting actions, executes a wide range of brushstroke motions, and autonomously selects and mixes pigments to obtain the desired colors. The system successfully produces multiple impressionist-style artworks inspired by Van Gogh's starry night, demonstrating its capability for achieving high-level autonomous artistic creation.

VII. Conclusion

This work addresses the key technical challenges in AI-driven artistic creation, with a particular focus on performance generation and human-machine collaborative processes. A VLA-driven artistic creation framework grounded in the parallel-systems theory and the ACP methodology is developed, its methodological structure and core characteristics are articulated, and its feasibility through practical case studies is discussed. Collectively, these contributions aim to enhance the intelligence, autonomy, and creative adaptability of computational artistic systems.

Looking forward, the continued advancement of large foundation models is expected to further strengthen the framework's capabilities in multimodal artistic data modeling and strategic creative synthesis. Under the guidance of parallel-systems principles, the proposed framework holds promise for integration into both professional artistic workflows and everyday creative practices, offering a forward-looking paradigm for deep human-AI co-creation. Future work can explore incorporating additional channels of human feedback, including subtle physiological or non-conscious signals, to further refine the alignment between artificial and human artistic decision-making. Ultimately, the VLA-driven artistic creation framework is positioned not only to augment technical performance but also to reshape artistic practice itself, enabling new aesthetic forms, broadening access to creative expertise, and advancing the evolution toward more intelligent and human-aligned modes of the artistic expression in future creative work.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 62503280 and 62306319) and the "Digital/Parallel Arts and Artistic Intelligence for Smart Education and Parallel Science" Project (No. DeSCII-0001-0123-2021, PI: Yue Lu) initiated by DeSci International. This work is initiated by and benefited from many valuable discussions with Professor Fei-Yue Wang.

References

[1] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Akhtar, N. Barnes, and A. Mian, A comprehensive overview of large language models, *ACM Trans. Intell. Syst. Technol.*, 2025, 16(5), 106.

[2] B. Li, J. Hao, Y. Sun, Y. Meng, J. Huang, Y. Tian, and Z. He, Vision-language-action models under parallel intelligence paradigm: The state of the art and future perspectives, *Chin. J. Intell. Sci. Technol.*, 2025, 7(3), 290–303, (in Chinese).

[3] H. Zhang, X. Jia, G. Luo, Y. Tian, S. Liang, and D. Hong, Parallel images: Knowledge and data-driven multi-modal information intelligent analysis, *Int. J. Intell. Control Syst.*, 2025, 30(2), 93–107.

[4] Q. Du, X. Wang, S. Yin, L. Li, and H. Ning, Social force embedded mixed graph convolutional network for multi-class trajectory prediction, *IEEE Trans. Intell. Vehicl.*, 2024, 9(9), 5571–5580.

[5] T.-Y. Shen, Z.-R. Tao, Y.-D. Wang, T.-Z. Zhang, Y.-H. Liu, X.-X. Wang, J. Yang, Z.-W. Li, L. Chen, K.-F. Wang et al., Key problems of embodied intelligence research: Autonomous perception, action, and evolution, *Acta Autom. Sin.*, 2025, 51(1), 43–71, (in Chinese).

[6] Q. Du, L. Xu, Q. Wu, H. Ning, X. Wang, L. Lin, and C. Sun, SafeCrossNet: Multi-modal fusion with social-aware for pedestrian crossing intention prediction, *Inform. Fusion*, 2026, 126, 103609.

[7] S. Yang, W. Zhang, R. Song, J. Cheng, H. Wang, and Y. Li, Watch and act: Learning robotic manipulation from visual demonstration, *IEEE Trans. Syst., Man, Cybern.: Syst.*, 2023, 53(7), 4404–4416.

[8] E. S. Mikalonytė and M. Kneer, Can artificial intelligence make art?: Folk intuitions as to whether AI-driven robots can be viewed as artists and produce art, *ACM Trans. Hum.-Robot Int.*, 2022, 11(4), 43.

[9] R. Chamberlain, C. Mullin, B. Scheerlinck, and J. Wagemans, Putting the art in artificial: Aesthetic responses to computer-generated art, *Psychol. Aesthet., Creat., Arts*, 2018, 12(2), 177–192.

[10] D. Winter, Aesthetic aspects of digital humanism: An aesthetic-philosophical analysis of whether AI can create art, in *Introduction to Digital Humanism*. Cham, Germany: Springer, 2024, 211–224.

[11] E. Cetinic and J. She, Understanding and creating art with AI: Review and outlook, *ACM Trans. Multim. Comput., Commun., Appl.*, 2022, 18(2), 66.

[12] C. Guo, Y. Lu, Y. Dou, and F.-Y. Wang, Can ChatGPT boost artistic creation: The need of imaginative intelligence for Parallel Art, *IEEE/CAA J. Autom. Sin.*, 2023, 10(4), 835–838.

[13] B. Wang, Q. Chen, and Z. Wang, Diffusion-based visual art creation: A survey and new perspectives, *ACM Comput. Surv.*, 2025, 57(10), 268.

[14] A. I. Karimov, E. E. Kopets, V. G. Rybin, S. V. Leonov, A. I. Voroshilova, and D. N. Butusov, Advanced tone rendition technique for a painting robot, *Robot. Auton. Syst.*, 2019, 115, 17–27.

[15] H. Cohen, The further exploits of Aaron, painter, *Stanford Hum. Rev.*, 1995, 4(2), 141–158.

[16] P. Tresset and F. F. Leymarie, Portrait drawing by Paul the robot, *Comput. Graph.*, 2013, 37(5), 348–363.

[17] S. Calinon, J. Epiney, and A. Billard, A humanoid robot drawing human portraits, in *Proc. 5th IEEE-RAS International Conference on Humanoid Robots*, Tsukuba, Japan, 2005, 161–166.

[18] R. Luo, M.-J. Hong, and P.-C. Chung, Robot artist for colorful picture painting with visual control system, in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Daejeon, Korea, 2016, 2998–3003.

[19] J. M. Gülzow, L. Grayver, and O. Deussen, Self-improving robotic brushstroke replication, *Arts*, 2018, 7(4), 84.

[20] A. Beltramello, L. Scalera, S. Seriani, and P. Gallina, Artistic robotic painting using the palette knife technique, *Robotics*, 2020, 9(1), 15.

[21] L. Scalera, S. Seriani, A. Gasparetto, and P. Gallina, Watercolour robotic painting: A novel automatic system for artistic rendering, *J. Intell. Robot. Syst.*, 2019, 95(3), 871–886.

[22] Y. Lu, C. Guo, Y.-L. Lin, F. Zhuo, and F.-Y. Wang, Computational aesthetics of fine art paintings: The state of the art and outlook, *Acta Autom. Sin.*, 2020, 46(11), 2239–2259.

[23] C. Guo, Y. Lu, Y. Lin, F. Zhuo, and F.-Y. Wang, Parallel Art: Artistic creation under human-machine collaboration, *Chin. J. Intell. Sci. Technol.*, 2019, 1(4), 335–341.

[24] C. Guo, T. Bai, Y. Lu, Y. Lin, G. Xiong, X. Wang, and F.-Y. Wang, Skywork-daVinci: A novel CPSS-based painting support system, in *Proc. IEEE 16th International Conference on Automation Science and Engineering*, Hong Kong, China, 2020, 673–678.

[25] Y. Lu, C. Guo, Y. Dou, X. Dai, and F.-Y. Wang, Could ChatGPT imagine: Content control for artistic painting generation via large language models, *J. Intell. Robot. Syst.*, 2023, 109(2), 39.

[26] F.-Y. Wang, Parallel system methods for management and control of complex systems, *Control Dec.*, 2004, 19(5), 485–489, 514, (in Chinese).

[27] X. Wang, J. Yang, Y. Liu, Y. Wang, F.-Y. Wang, M. Kang, Y. Tian, I. Rudas, L. Li, M. P. Fanti et al., Parallel intelligence in three decades: A historical review and future perspective on ACP and cyber-physical-social systems, *Artif. Intell. Rev.*, 2024, 57(9), 255.

[28] L. Chen, P. Schaldenbrand, T. Shankar, L. Coleman, and J. Oh, Spline-FRIDA: Towards diverse, humanlike robot painting styles with a sample-efficient, differentiable brush stroke model, *IEEE Robot. Autom. Lett.*, 2025, 10(6), 5983–5990.

[29] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid et al., RT-2: Vision-language-action models transfer web knowledge to robotic control, in *Proc. 7th Conference on Robot Learning*, Atlanta, GA, USA, 2023, 2165–2183.

[30] H. Zhang, S.-T. Liang, M.-X. Li, Y.-L. Tian, J.-W. Ge, H. Yu, L.-X. Li,

and F.-Y. Wang, Vision-language-action models: From the early foundations to the state-of-the-art, *Acta Autom. Sin.*, 2025, 51(9), 1922–1950.

[31] L. Scalera, S. Seriani, A. Gasparetto, and P. Gallina, Busker robot: A robotic painting system for rendering images into watercolour artworks, in *Mechanism Design for Robotics*. Cham, Germany: Springer, 2018, 1–8.

[32] P. V. Arman, Cloud painter: AI artist on the road [Online], https://www.cloudpainter.com/, 20 November 2025.

[33] L. Scalera, S. Seriani, A. Gasparetto, and P. Gallina, Non-photorealistic rendering techniques for artistic robotic painting, *Robotics*, 2019, 8(1), 10.

[34] L. Scalera, E. Maset, S. Seriani, A. Gasparetto, and P. Gallina, Performance evaluation of a robotic architecture for drawing with eyes, *Int. J. Mech. Control*, 2021, 22(2), 53–60.

[35] L. Scalera, S. Seriani, A. Gasparetto, and P. Gallina, A novel robotic system for painting with eyes, in *Advances in Italian Mechanism Science*. Cham, Germany: Springer, 2020, 191–199.

[36] P. Schaldenbrand, J. McCann, and J. Oh, FRIDA: A collaborative robot painter with a differentiable, Real2Sim2Real planning environment, in *Proc. IEEE International Conference on Robotics and Automation*, London, UK, 2023, 11712–11718.

[37] V. Misra, P. Schaldenbrand, and J. Oh, Robot synesthesia: A sound and emotion guided AI painter, arXiv preprint arXiv: 2302.04850, 2023.

[38] P. Schaldenbrand, G. Parmar, J.-Y. Zhu, J. McCann, and J. Oh, CoFRIDA: Self-supervised fine-tuning for human-robot co-painting, in *Proc. IEEE International Conference on Robotics and Automation*, Yokohama, Japan, 2024, 2296–2302.

[39] A. Karimov, E. Kopets, S. Leonov, L. Scalera, and D. Butusov, A robot for artistic painting in authentic colors, *J. Intell. Robot. Syst.*, 2023, 107(3), 34.

[40] A. Karimov, M. Strelnikov, S. Mazin, D. Goryunov, S. Leonov, and D. Butusov, Physically motivated model of a painting brush for robotic painting and calligraphy, *Robotics*, 2024, 13(6), 94.

[41] A. Karimov, E. Kopets, G. Kolev, S. Leonov, L. Scalera, and D. Butusov, Image preprocessing for artistic robotic painting, *Inventions*, 2021, 6(1), 19.

[42] A. Karimov, E. Kopets, T. Shpilevaya, E. Katser, S. Leonov, and D. Butusov, Comparing neural style transfer and gradient-based algorithms in brushstroke rendering tasks, *Mathematics*, 2023, 11(10), 2255.

[43] P. Schaldenbrand and J. Oh, Content masked loss: Human-like brush stroke planning in a reinforcement learning painting agent, in *Proc. 35th AAAI Conference on Artificial Intelligence*, Virtual Event, 2021, 505–512.

[44] F.-Y. Wang, Shadow Systems: A new concept for nested and embedded co-simulation for intelligent systems, Tech. Rep., University of Arizona, Tucson, AZ, USA, 1994.

[45] T.-X. Bai, S. Wang, Z. Shen, D.-P. Cao, N.-N. Zheng, and F.-Y. Wang, Parallel robotics and parallel unmanned systems: Framework, structure, process, platform and applications, *Acta Autom. Sin.*, 2017, 43(2), 161–175, (in Chinese).

[46] Z. Li, Z. Li, X. Hu, G. Du, Y. Nie, F. Zhu, L. Bai, R. Zhao, and Y. Lv, VisionTraj: A noise-robust trajectory recovery framework based on large-scale camera network, *IEEE Trans. Intell. Transp. Syst.*, 2025, 26(10), 15826–15837.

[47] Y. Lv, From cold algorithms to warm intelligence: Infusing autonomous driving systems with humanistic care, *IEEE Intell. Transp. Syst. Mag.*, 2025, 17(4), 2–3.

[48] W. Xie, C. Guo, Y. Wang, Y. Lv, and F.-Y. Wang, Plug-and-play dramaturge: A divide-and-conquer approach for iterative narrative script refinement via collaborative LLM agents, arXiv preprint arXiv: 2510.05188, 2025.

[49] X. Xue, D. Zhou, X. Yu, G. Wang, J. Li, X. Xie, L. Cui, and F.-Y. Wang, Computational experiments for complex social systems: Experiment design and generative explanation, *IEEE/CAA J. Autom. Sin.*, 2024, 11(4), 1022–1038.

[50] X. Wang, Y. Wang, J. Yang, X. Jia, L. Li, W. Ding, and F.-Y. Wang, The survey on multi-source data fusion in cyber-physical-social systems: Foundational infrastructure for industrial metaverses and industries 5.0, *Inform. Fusion*, 2024, 107, 102321.

[51] R. Qin, J. Li, and F.-Y. Wang, MetaEconomics and MetaManagement for MetaCities and MetaSocieties in Metaverse, *IEEE Trans. Syst., Man, Cybern.: Syst.*, 2024, 54(11), 6849–6858.

[52] F.-Y. Wang, Parallel management: The DAO to smart ecological technology for complexity management intelligence, *Acta Autom. Sin.*, 2022, 48(11), 2655–2669, (in Chinese).

[53] C. Guo, Y. Dou, T. Bai, X. Dai, C. Wang, and Y. Wen, ArtVerse: A paradigm for parallel human-machine collaborative painting creation in metaverses, *IEEE Trans. Syst., Man, Cybern.: Syst.*, 2023, 53(4), 2200–2208.

[54] Y. Lu, C. Guo, X. Dai, and F.-Y. Wang, ArtCap: A dataset for image captioning of fine art paintings, *IEEE Trans. Comput. Soc. Syst.*, 2024, 11(1), 576–587.

[55] C. Guo, Y. Lu, X. Wang, D. Yi, X. Wang, and F.-Y. Wang, Architecture and key techniques of parallel creation through the fusion of human-cyber-physical intelligence in CPSS, *Chin. J. Intell. Sci. Technol.*, 2022, 4(3), 344–354.

[56] X. Li, Y. Tian, P. Ye, H. Duan, and F.-Y. Wang, A novel scenarios engineering methodology for foundation models in metaverse, *IEEE Trans. Syst., Man, Cybern.: Syst.*, 2023, 53(4), 2148–2159.

[57] C. Chang, J. Zhang, J. Ge, Z. Zhang, J. Wei, L. Li, and F.-Y. Wang, VistaScenario: Interaction scenario engineering for vehicles with intelligent systems for transport automation, *IEEE Trans. Intell. Vehicl.*, to be published.

[58] F.-Y. Wang, From humanoids of embodied intelligence to robotoids of parallel intelligence, *Chin. J. Intell. Sci. Technol.*, 2025, 7(3), 287–289, (in Chinese).

[59] T. Bai, Z. Shen, Y. Liu, and X. Dong, Parallel machine: A framework for the control and management for intelligent machines, *Chin. J. Intell. Sci. Technol.*, 2019, 1(2), 181–191, (in Chinese).

[60] Y. Lu, C. Guo, X. Dai, and F.-Y. Wang, Generating emotion descriptions for fine art paintings via multiple painting representations, *IEEE Intell. Syst.*, 2023, 38(3), 31–40.

[61] C. Guo, T. Bai, X. Wang, X. Zhang, Y. Lu, X. Dai, and F.-Y. Wang, ShadowPainter: Active learning enabled robotic painting through visual measurement and reproduction of the artistic creation process, *J. Intell. Robot. Syst.*, 2022, 105(3), 61.

[62] Y. Lu, C. Guo, X. Dai, and F.-Y. Wang, Data-efficient image captioning of fine art paintings via virtual-real semantic alignment training, *Neurocomputing*, 2022, 490, 163–180.

[63] F. Zhuo, Human-machine co-creation on artistic paintings, in *Proc. IEEE 1st International Conference on Digital Twins and Parallel Intelligence*, Beijing, China, 2021, 316–319.

[64] S. Seriani, A. Cortellessa, S. Belfio, M. Sortino, G. Totis, and P. Gallina, Automatic path-planning algorithm for realistic decorative robotic painting, *Autom. Constr.*, 2015, 56, 67–75.

[65] C. Gomez Cubero, M. Pekarik, V. Rizzo, and E. Jochum, The robot is present: Creative approaches for artistic expression with robots, *Front. Robot. AI*, 2021, 8, 662249.

[66] S. Xie, E. B. Sandoval, K. A. Shaik, and F. Cruz, Embodied generative AI art for enhanced human-robot interaction through a human-centric LLM-guided robotic arm drawing system, in *Proc. 20th ACM/IEEE International Conference on Human-Robot Interaction*, Melbourne, Australia, 2025, 1727–1730.

[67] C. Guo, X. Dai, W. Liao, J. Qi, Y. Lu, Q. Ni, L. Wang, and F.-Y. Wang, Intelligent art factory: Achieving creative automation through Parallel Art agents, *Chin. J. Intell. Sci. Technol.*, 2024, 6(2), 179–188.

[68] V. Matanski, A. Iliev, N. Kyurkchiev, and T. Terzieva, Ontology-driven

multi-agent system for cross-domain art translation, *Future Internet*, 2025, 17(11), 517.

[69] A. Bidgoli, M. L. De Guevara, C. Hsiung, J. Oh, and E. Kang, Artistic style in robotic painting; A machine learning approach to learning brushstroke from human artists, in *Proc. 29th IEEE International Conference on Robot and Human Interactive Communication*, Naples, Italy, 2020, 412–418.

Yue Lu received the PhD degree in pattern recognition and intelligent systems from University of Chinese Academy of Sciences, Beijing, China, in 2023. From 2023 to 2025, he was a postdoctoral researcher at School of Control Science and Engineering, Shandong University, Jinan, China. Since 2025, he has been with School of Information Science and Engineering, Shenyang University of Technology, Shenyang, China. His research interests include AI for art, machine learning, and intelligent robotic systems.

Yong Dou is currently pursuing the PhD degree at Faculty of Innovation Engineering, Macau University of Sciences and Technology, Macao, China. His research interests include robotic art, intelligent robotic systems, and machine learning.

Chao Guo received the PhD degree from University of Chinese Academy of Sciences, Beijing, China, in 2022. He is currently an assistant professor at Institute of Automation, Chinese Academy of Sciences, China. His research interests include AI for creation, large foundation models, autonomous agents, and intelligent robotic systems.

Qinghua Ni is currently pursuing the PhD degree in intelligent science and systems study at Faculty of Innovation Engineering, Macau University of Sciences and Technology, Macao, China. Her research interests include parallel intelligence, decision intelligence, blockchain and DAO, and parallel theater.

Tianxiang Bai received the MS degree in control science and engineering from Institute of Automation, Chinese Academy of Sciences, China, in 2022. He is currently an engineer at Aerospace Science and Industry Group Intelligent Technology Research Institute Co., Ltd., China. His research interests include reinforcement learning, complex systems, and human-machine intelligence.

Tom Gedeon is currently a professor as Human-Centric Advancements chair in AI and was recently the Optus chair in AI at Curtin University, Australia. Prior to this, he was a professor and former deputy dean with College of Engineering and Computer Science, Australian National University, Australia. He has more than 400 publications and has run multiple international conferences. He is a former president of Asia-Pacific Neural Network Assembly and former president of Computing Research and Education Association of Australasia.

Thales S. W. Theseus is currently a senior researcher at DeSci Center of Parallel Intelligence, Obuda University, Hungary. His research interests include Parallel Art and music for education and smart living and working environments.