2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

PICaSo: A Collaborative Robotics System for Inpainting on Physical Canvas using Marker and Eraser

Shady Nasrat, Jae-Bong Yi, Minseong Jo, and Seung-joon Yi*

***Abstract*— Robotics collaborative drawing involves the interaction between humans and robots to create of visual art using a variety of tools and materials, serving various functions such as communication, narration, and emotional representation. A creative technique within the human natural drawing process is known as inpainting, which involves reconstructing or editing elements in a drawing. This paper introduces PICaSo (Physical Inpainting on Canvas Solution), a robotic drawing system that enables multiple users to collaboratively create artwork on a canvas by integrating the inpainting process. PICaSo utilizes a fine-tuned text-to-image model to interpret natural language prompts into artistic renderings on canvas. Users guide the process by simple descriptive text and specifying desired drawing placement, empowering the robotic arm to autonomously translate these instructions into physical artworks. Our system's innovation lies in its effective translation of digital inpainting processes into physical actions. By leveraging our erasing capability that enables selective removal of specific parts on the canvas without impacting neighboring areas, facilitating the creation of sequential drawings. This paper comprehensively outlines the capabilities of the proposed system, explores potential applications across various domains, and addresses technical challenges encountered during its development.**
Project website: shadynasrat.github.io/PICaSo

I. Introduction

Recent advancements in text-to-image technologies have led to a significant increase in digital content production. However, the integration of these technologies into creating art with robots is still at an early stage, primarily due to the substantial disparity between digital and real-world settings. One notable advancement is the inpainting process, where text-to-image models produce or reinterpret sections of an image based on textual input. PICaSo introduces a robotic system designed to support the inpainting process on a physical canvas, enabling multiple users to cooperate and improve their creative ideas using natural language shown in Fig. 1. This unique approach brings together human creativity and technological precision in visual art, allowing for collaborative brainstorming by adding or refining elements of the canvas. The potential benefits of a collaborative assistant in therapeutic art are evident [1]–[3], as it enables group drawing sessions that can enhance mental health and overall well-being. PICaSo's collaborative inpainting process serves as a supportive partner, offering guidance throughout the artistic process involving multiple users, making it particularly beneficial for those who may have physical or skill-related limitations with drawing.

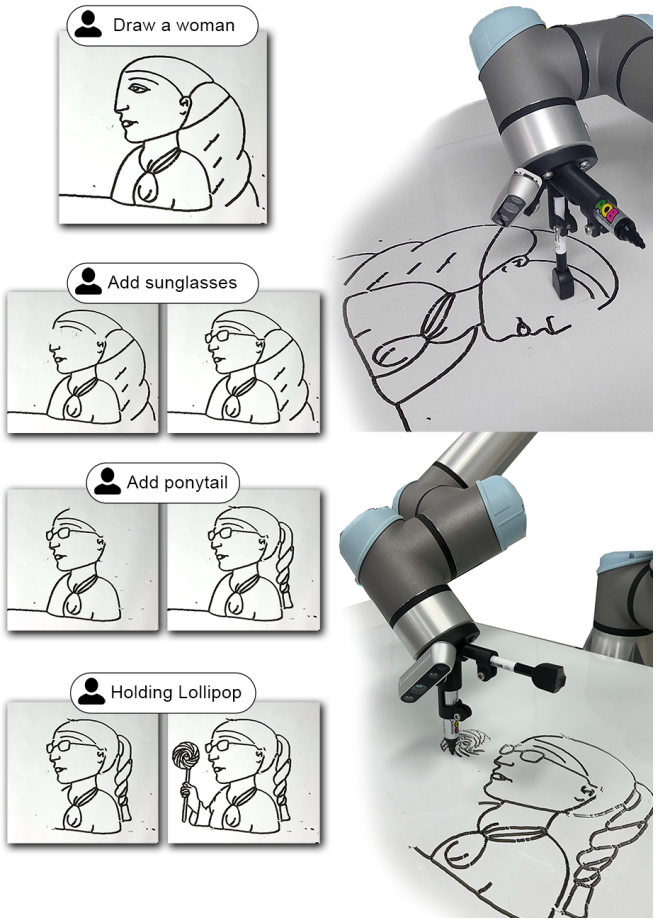

Fig. 1. **PICaSo Inpainting Process**: Illustration demonstrate the collaborative process of drawing by multiple users. The process commences with a user initiating the canvas with an idea, followed by other participants making various edits on the physical canvas. This was made possible through our PICaSo inpainting technique, enabling the robot arm to selectively remove and redraw necessary parts without impacting the entire canvas.

We prepared for the inpainting process by selecting the appropriate tools, such as a marker and an eraser, and using a whiteboard as our canvas. To enhance efficiency, we also designed a multi-gripper capable of holding markers, erasers, and a camera for added convenience. Our waypoint generation uses the pixel-to-pixel line extraction algorithm introduced in [4]. This algorithm efficiently extracts points and lines from an image while maintaining quality and reducing processing time. Although drawing proved straight-

*This project was funded by Police-Lab 2.0 Program(www.kipot.or.kr) funded by the Ministry of Science and ICT(MSIT, Korea) & Korean National Police Agency(KNPA, Korea) (No. 082021D48000000) and Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE)(P0008473, HRD Program for Industrial Innovation)

Authors are with Faculty of Electrical Engineering, Pusan National University, Busan, South Korea seungjoon.yi@ pusan.ac.kr

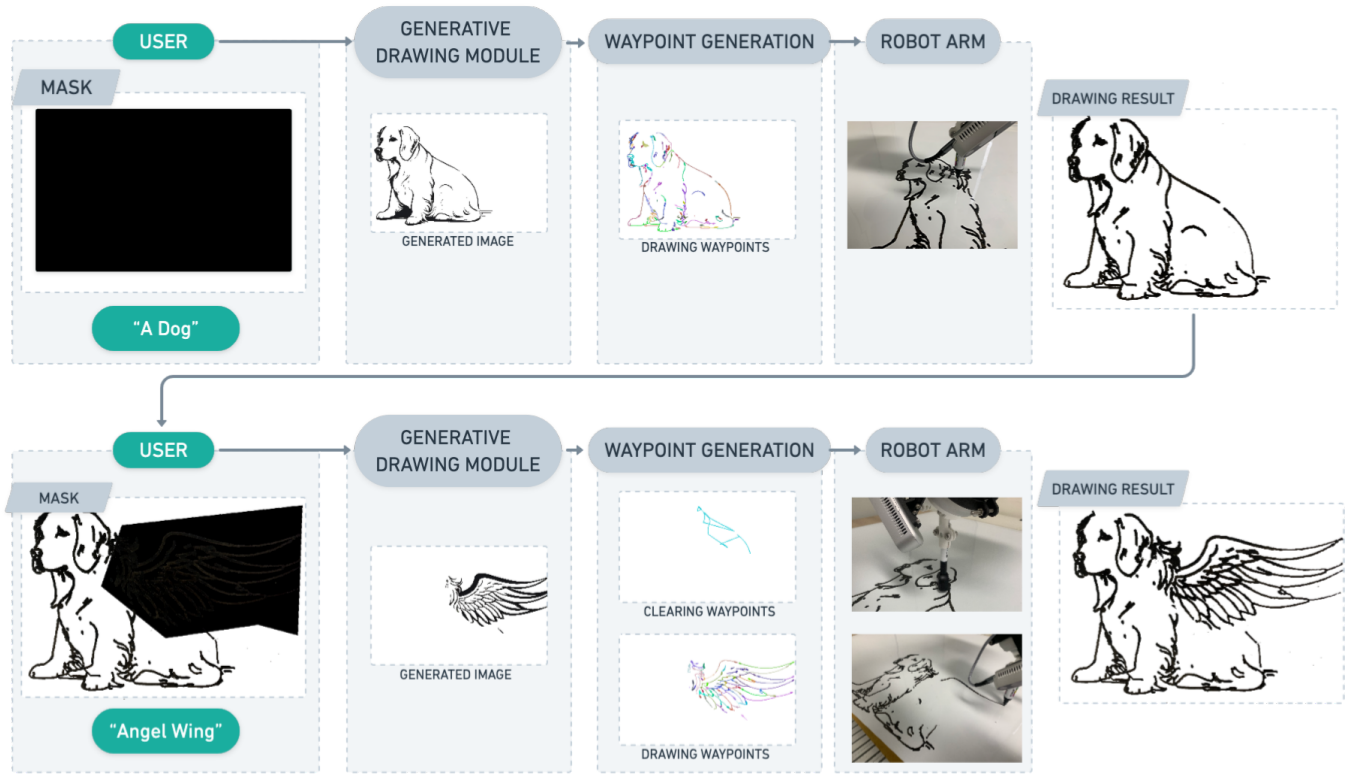

Fig. 2. **System Overview**: The diagram illustrates the step-by-step procedure, beginning with user interface input for mask and text prompt. Then, a generative drawing module creates a new image, followed by the generation of waypoints to direct the robotic arm in the drawing process. The end outcome is presented as the robotic arm transfers the generated drawing onto the canvas.

forward, the challenge arose in clearing process. Prior to our method's development, clearing was an unexplored territory, with no established techniques. Consequently, we innovated a new method to trace back drawn waypoints within a selected mask, enabling their removal from the canvas without impacting the surrounding artwork.

Since our waypoint generation algorithm is based on a morphological transformation for image skeletonization, it required the input image to adhere to a specific style that excludes bold lines, shading, or colors. The pre-trained text-to-image model faced challenges in creating images with this specific style, often resulting in a generic random cartoon styles where regular prompt tuning couldn't yielding the desired outcome style. To address this issue the style gap between the text-to-image style and the pixel-to-pixel line extraction algorithm's desired style needs to be minimized. This requires learning a new style that is better suited for precise drawing and inpainting.

To reduce the style gap between generated images and the line extraction algorithm, we propose a fine-tuned model, trained on a style in which it is efficient to extract lines from. First we collected our dataset from cartoon images, which were characterized by uniform black thin lines devoid of colors and shading. Both *SDXL1.0* [5] and *SDXL1.0-inpainting-0.1* models were fine-tuned on this dataset to understand the unique style and features of these images, enabling more accurate line extraction. In order to successfully complete the inpainting process, we selected the suitable tools for the task. Our chosen canvas was a whiteboard, and we utilized a marker as well as an eraser. To optimize our use of these tools, we developed a multi-gripper capable of holding markers, eraser and a camera for added convenience during the process.

We have outlined our primary contributions by introducing physical inpainting, a novel form of canvas manipulation necessary for collaborative drawing tasks involving humans and robots. We introduced clearing algorithm allowing editing canvas. Our system is designed to be scalable, enabling multiple users to work together on a single canvas. Additionally, we have presented a comprehensive approach to bridging the gap between pre-trained models and line extraction algorithms. Both the hardware and software code for PICaSo are available as open-source resources.

II. Related Work

A. Drawing Tools

Previous research papers addressing drawing robotics tasks have primarily focused on aspects such as trajectory planning, stroke generation, and optimization techniques to improve drawing accuracy, documented in [6]–[16]. These techniques include the utilization of various tools such as pens [17], pencils [9-13-15-16], markers [10], paint brushes [8-18] and spray paint [19]. However, a notable gap in these approaches has been the absence of a dedicated clearing process, which is crucial for facilitating the inpainting process.

The ability to clear and redraw areas enables the correction of errors and the refinement of details in the artwork. As a result, PICaSo stands out as the first system to address this by introducing efforts in the area of clearing processes.

B. Generative AI-Driven Drawing

Recent progress has showcased the potential of AI-based solutions in this area. In a study by [20], a generative adversarial network and a collaborative robotic arm worked with a 5-year-old child to create drawings on canvas. Moreover, various research works have introduced neuroadaptive learning algorithms for effective control in constrained nonlinear systems and robotic painting setups such as such as [12-21]–[26]. For instance, Tianying et al. [10] utilized GAN-based style transfer to transform facial images into simplified cartoon representations, while S. Nasrat [6] employed similar techniques to generate high-quality portrait drawings and Gao et al. [13] employed GAN-based style transfer to minimize the number of strokes needed for sketching artworks. Building upon these developments is CoFRIDA [27], which utilizes an adapted Instruct-Pix2Pix model to narrow down the semantic gap between simulated and real-world drawing allowing collaborative drawings between humans and robots. However, with these developments, there is still potential for enhancement in ensuring smooth collaboration and improving the organic nature of the drawing process. Our proposed system utilizes Natural Language Processing to simplify collaboration in drawings. Additionally, other methods lack scalability to facilitate multiple users working together on a shared canvas, which is a fundamental aspect of our system's design.

III. Method

The System Architecture begins with user input through a simple interface as shown in Fig. 2, allowing for the input of both masks and text prompts. Following this, the generative drawing module generates a new image within the specified masked area. The waypoint generation module comes into play, creating drawing and clearing waypoints. Finally, the waypoint data is transmitted to the robotic arm for the drawing process on the canvas.

A. Generative Drawing Module

The primary aim of this module is to produce visuals for text-to-drawing and inpainting tasks, in a style that aligns well with the requirements of the waypoint generation algorithm.

1) Text-To-Drawing: Before inpainting, PICaSo needs to be capable of generating and drawing complete images from text. To achieve this, we employed SDXL1.0, a pre-trained model for text-to-image generation which builds on earlier Stable Diffusion models.

2) Inpainting: For the process of inpainting, we made use of SDXL-1.0-inpainting-0.1 - an enhanced inpainting model initialized with the weights from SDXL1.0 and designed to continue masked images effectively.

The waypoint generation required style involves excluding colored, bold lines, or shaded regions. To accomplish this, we curated a dataset comprising 40 cartoon images characterized by uniform thin black lines without color or shading – a stylistic prerequisite for the efficient operation of the waypoint algorithm. These images were obtained from cartoons in which colors and shadings were manually eliminated and textual annotations were added by hand. When working with pre-trained models, instead of training all network parameters totaling 6.6 billion in each model and to optimize resource usage and time efficiency, we employed LoRA [28] (Low Rank Adaptation) training technique by freezing base model and only training a subset of parameters with a network rank (dimension) of 128. The training process took four hours on a NVIDIA RTX 3090 graphic card. Additionally, we conducted comparative analysis among generated images produced from different models including our own in order evaluate their performance as shown in Fig. 3.

B. Waypoints Generation

We utilized the approach described by S. Nasrat et al. [4]. Initially, a morphological transformation is applied to simplify the sketch, smoothing pixel edges and removing isolated pixels. Subsequently, an efficient pixel-to-pixel algorithm extracts lines from the sketch, preserving important details while optimizing processing time. Line clustering reduces the number of lines by merging closely located ones to ensure precision and guide the robotic arm along the desired path information. In this study, a scaling algorithm was developed to convert pixel dimensions of a canvas image into metric sizes in Cartesian coordinates. The process involves converting each individual pixel into a meter-based Cartesian system using the pixel's x and y coordinates shown in Eq. 1. This allows users to easily adapt to changes in pixel or physical canvas sizes according to their preferences. The vectors $\boldsymbol{C}$ and $\boldsymbol{P}$ represent the position on the canvas and the pixel position in the image, with $\boldsymbol{C_0}$ and $\boldsymbol{C_f}$ as start and end points on the canvas, while $\boldsymbol{P_0}$ and $\boldsymbol{P_f}$ correspond to start and end points of the image.

$$\boldsymbol{C} = \boldsymbol{C_0} + ((\boldsymbol{C_f} - \boldsymbol{C_0})/(\boldsymbol{P_f} - \boldsymbol{P_0})) * (\boldsymbol{P} - \boldsymbol{P_0}) \quad (1)$$

Both drawing and clearing uses an interpolation movement approach through the generated waypoints showing visually appealing drawings. Utilizing the scaling algorithm with the interpolation movement made it was possible to generalize our system and be able to be used on different robotics arms as we tested on UR5 and RB-180 documented at Sec. IV.

C. Clearing Algorithm

During the clearing phase, each waypoint is revisited on the current canvas. Utilizing the user-defined mask, it precisely discerns which waypoints require erasure, facilitating the preparation for subsequent creative iterations. The process of tracing back waypoints falling within the mask is represented by Eq. 2, where W_{erase}, $W_{current}$, and M_{mask} respectively signify the set of waypoints to erase, the current waypoints on the canvas, and the user-defined mask area.

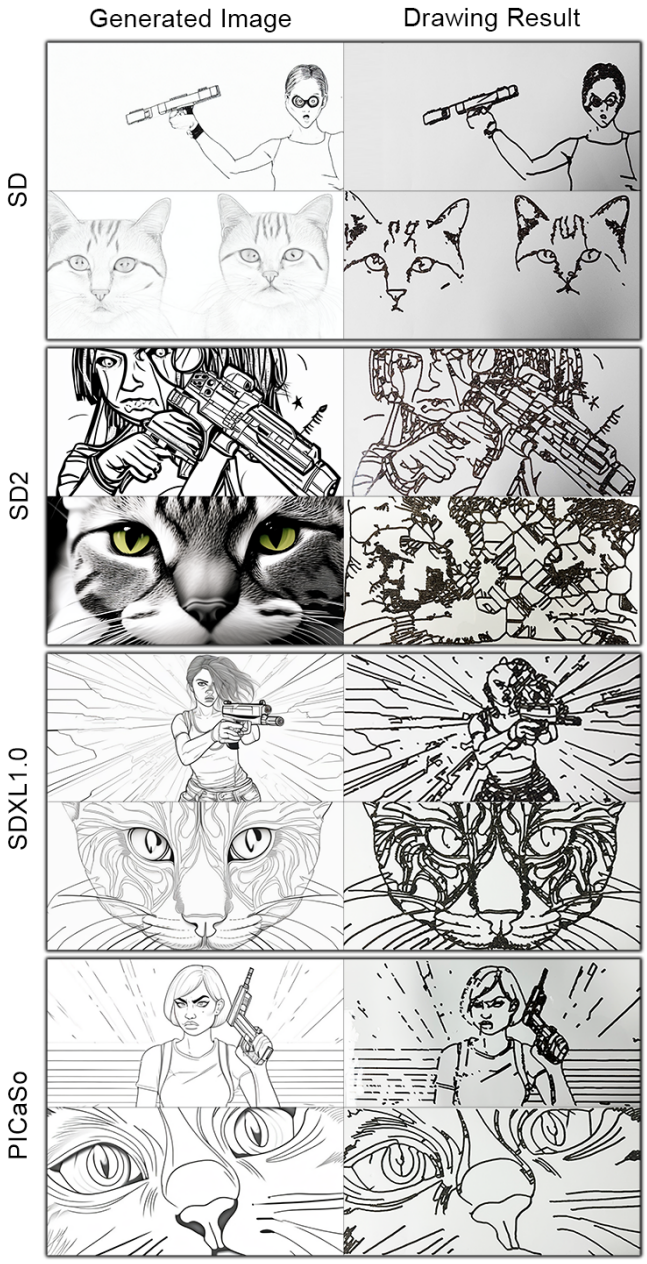

Fig. 3. **Text-To-Drawing Comparison**: This illustration shows a comparison between multiple text-to-drawings results using different models. Image size: 1024x512, Canvas size: 600x300 mm, Prompts: *"angry girl with a laser gun"*, and *"a close up shot of a cat face"*

The updated waypoints $W_{updated}$ after performing the erase waypoints W_{erase} onto the canvas is represented by Eq. 3:

$$W_{erase} = \{\boldsymbol{C} | \boldsymbol{C} \in (W_{current} \cap M_{mask})\} \tag{2}$$

$$W_{updated} = W_{current} - W_{erase} \tag{3}$$

D. Gripper Design

Enabling a seamless transition between the different functionalities was essential to this system, to achieve this we implemented a spring loaded multi-tool gripper mechanism.

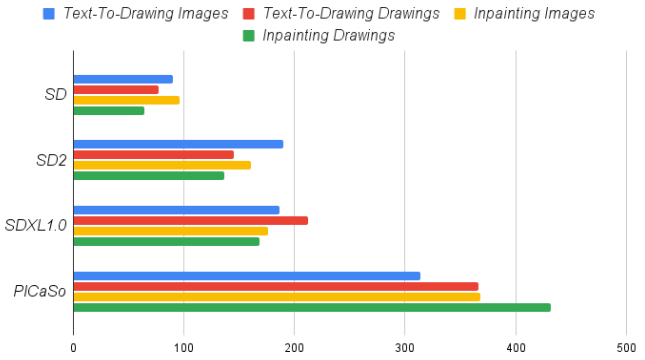

Fig. 4. **Human Evaluation Study**: A survey involving selecting from most appealing image according to description out of four images generated by different models. The survey consists of the following sets: text-to-drawing images, text-to-drawing drawings, inpainting images and inpainting drawings.

This gripper was designed to facilitate quick switches between the marker, eraser and camera modes offering a flexible and responsive approach to the process. Each tool in the gripper was configured with a 45-degree offset from the Tool Center Point (TCP) origin angle, allowing versatility and reach while maintaining precise control over the execution. Additionally, we included a spring mechanism in the marker tool to maintain consistent lines.

IV. Experiments

A. Baselines Comparison

We examine how the fine-tuning procedure affects the generation of consistent waypoints in both the SDXL1.0 and SDXL1.0-inpainting models by contrasting them with alternative models in PICaSo's Generative Drawing module, encompassing text-to-drawing and inpainting tasks.

B. Multiple User Inpainting

We test PICaSo's inpainting scalability in multiple iterations of human-robot interactions, by allowing multiple participants to introduce new creative element onto the canvas, such as adding new components or modifying existing ones. and giving the freedom to change drawing canvas sizes to testing our scaling algorithm. Additionally we got participants satisfaction feedback while using the inpainting process.

C. Evaluation

1) Text-To-Drawing: We utilized CLIPScore [29] to gauge the similarity between images and prompts using pre-trained image-text encoders. Other benchmarks are incorporated such as BLIPScore [30], Aesthetics [31] and ImageReward [32] into our analysis. Additionally, we calculated the loss in the semantic meaning $\triangle_{id}$ In Eq. 4. We measure the $\triangle_{id}$ by calculating the absolute mean-error between the CLIPScore values of both the generated image $\boldsymbol{I_{img}}$ and its corresponding drawing $\boldsymbol{I_{draw}}$.. Due to potential bias related to CLIPScore, human evaluation was also conducted.

TABLE I

TEXT-TO-DRAW BENCHMARK SCORES COMPUTED ON BOTH GENERATED IMAGES AND DRAWINGS RESULT

	CLIPSCORE↑		BLIPSCORE↑		AESTHETICS↑		IMAGEREWARD↑		$\triangle_{id}\downarrow$	HUMAN EVAL.↑	
	DRAW	IMAGE	DRAW	IMAGE	DRAW	IMAGE	DRAW	IMAGE		DRAW	IMAGE
CoFRIDA	0.6240	——-	0.1920	——-	–	–	–	–	–	–	–
PICaSo (SD)	0.6378	0.6595	0.3336	0.3227	4.0551	4.8051	-2.2119	-2.2363	0.0216	0.087	0.109
PICaSo (SD2)	0.6055	0.6277	0.3622	0.3242	4.0997	4.9585	-2.2330	-2.2187	0.0221	0.168	0.233
PICaSo (SDXL1.0)	0.6891	0.6619	0.3354	0.3255	4.2808	5.0909	-2.1630	-2.1433	0.0272	0.252	0.233
PICaSo (fine-tuned)	**0.7023**	**0.6942**	0.3389	0.3218	**4.3448**	5.0425	**-2.1575**	**-2.1272**	**0.0080**	**0.491**	**0.425**

TABLE II

INPAINTING BENCHMARK SCORES COMPUTED ON BOTH GENERATED IMAGES AND DRAWINGS RESULT

	CLIPSCORE↑		BLIPSCORE↑		AESTHETICS↑		IMAGEREWARD↑		$\triangle_{id}\downarrow$	HUMAN EVAL.↑	
	DRAW	IMAGE	DRAW	IMAGE	DRAW	IMAGE	DRAW	IMAGE		DRAW	IMAGE
PICaSo (SD-inpaint)	0.6943	0.6908	0.5692	0.5490	4.4933	4.9165	0.3035	0.6667	0.0035	0.124	0.082
PICaSo (SD2-inpaint)	0.6582	0.6994	0.5606	0.5415	4.6311	4.8423	0.5580	0.5592	0.0412	0.202	0.168
PICaSo (SDXL1.0-inpaint)	0.6689	0.7064	0.5271	0.5471	4.6290	4.9966	-0.2514	0.9001	0.0374	0.216	0.213
PICaSo (fine-tuned)	**0.7275**	**0.7298**	**0.5763**	**0.5712**	**4.7390**	4.8771	**0.6149**	**1.3088**	**0.0022**	**0.458**	**0.537**

$$\triangle_{id} = 1/n \sum_{i=1}^{n} |CLIP(\boldsymbol{I_{draw_i}}) - CLIP(\boldsymbol{I_{img_i}})| \quad (4)$$

difference

2) Inpainting: We evaluated the data collected from the multi-user inpainting sessions we conducted, evaluation is done using CLIPScore and other benchmarks between inpainting models.

V. RESULTS

A. Text-To-Drawing

In order to assess the effectiveness of our refined approach in the text-to-drawings method, we conducted a comparison of image and drawing outcomes between runwayML(SD) [33], stable-diffusion-2(SD2) [34] and SDXL1.0 as base models, including our fine-tuned model. Our investigation focused on examining the impact of our fine-tuned model's ability in text-to-drawing transformation. We produced 50 pairs of images and drawings pairs per model while maintaining consistent settings for prompt, size, and generation seed, where some of the results are shown in Fig 3. CLIPScore, BLIPScore, Aesthetics and ImageReward measures are reported in Table I. PICaSo (fine-tuned) demonstrates the highest scores followed by the SDXL1.0 a base model, this was expected given that our fine-tuned model is based on SDXL1.0. in terms of the absolute mean-error ($\triangle_{id}$) between images and drawings CLIPScore, our fine-tuned model outperforms the baselines highlighting the efficiency in producing the required style for efficient waypoint generation resulting in visually appealing drawings.

We adequately evaluate the dataset through human evaluation, as utilizing AI-based evaluation may lead to biased results, a survey was carried out as depicted in Fig. 4. 20 distinct participants were involved and presented with a language description along with four images generated by different base models as well as our fine-tuned model, resulting in a total of 100 drawing-image generated pairs. Participants were instructed to select the most aesthetically pleasing image and drawing based on the textual description provided. For both images and drawings, the majority favored the output of our fine-tuned model for its simple yet descriptive style. The particular preference for our drawings emphasizes PICaSo's effectiveness in translating the refined style onto the canvas while maintaining visually appealing results.

B. Inpainting

Similarly to the text-to-drawing benchmark, we conducted a comparison and survey to evaluate the inpainting dataset curated from several multi-user inpainting sessions. In this comparison we use stable-diffusion-inpainting(SD), stable-diffusion-2-inpainting(SD2), and SDXL1.0-inpainting as baseline models along with our fine-tuned inpainting model. Our dataset consists of 50 image-drawing pairs generated from each model while maintaining consistent settings and input masks. Some examples of the dataset are shown in Fig. 5. The inpainting dataset consists of different tasks involving removing objects, redrawing character pose and drawing objects from an empty background. Table II reports CLIPScore, BLIPScore, Aesthetics, ImageReward and absolute mean-error ($\triangle_{id}$) measures. PICaSo (fine-tuned) out performs other models. On the CLIPScore metric the base model SDXL1.0-inpainting ranks second after our model, which aligns with expectations as our fine-tuned model is built upon it. The decrease in absolute mean error ($\triangle_{id}$) demonstrates the quality of image translation onto the canvas using our trained style. In addition, we carried out a survey using the inpainting dataset under the same conditions as the text-to-drawing survey. The results of this survey are depicted in Fig.4.

Fig. 5. **Inpainting Comparison.** This illustration shows a comparison between multiple inpainting image-drawing pairs generated using different models. Image Size: 512x512 pixels, Canvas size: 300x300 mm, Prompts: *"eyes"*, *"holding a balloon"* and *"a dog"*.

C. Inpainting Scalability

To evaluate the performance and collaborative scalability of the PICaSo system's inpainting process in multiple iterations of human-robot interactions, several multi-user inpainting sessions were conducted, wherein each participant contributed new creative element to the canvas, Examples of these sessions are depicted in Fig. 6. Notably, one session involved 10 participants, as detailed in Table III. Within this session, a participant initiated the process by adding a character which was subsequently edited by others through the addition of various features such as glasses, a suit, a smile, a full body, and the transformation of a balloon into an umbrella etc... . This collective effort resulted in a single artwork collaboratively crafted by 10 individuals, utilizing our system to articulate their ideas into a physical canvas through Natural Language. Based on the participants feedback PICaSo scores a 86.7% satisfaction rate in using the system in collaborative inpainting.

VI. Discussion

The exploration of new artistic styles and techniques stands as a promising avenue for the advancement of the PICaSo system. Despite the current emphasis on traditional

Fig. 6. **Inpainting Results.** We showcase PICaSo inpainting outcomes obtained from several multi-user collaborative sessions. Each session begins with a unique concept from the participant and their desired canvas dimensions, leading to a diverse array of drawings on different canvas sizes, using our scaling algorithm.

artistic forms, numerous unexplored avenues beckon for exploration. For instance, the mimicking of an artist style or incorporating drawing techniques as minimalist one-line drawing, which offer rich potential for experimentation and innovation within the PICaSo framework. Leveraging the capability to merge different LoRA adapters further expands the system's capacity to generate diverse artistic outputs by blending various LoRA image styles.

Furthermore, advancements in text-to-image models hold significant promise for enhancing the capabilities of the PICaSo system. Future iterations of these models may facilitate a more nuanced understanding of images, enabling the system to interpret and render complex visual concepts with greater fidelity. This could greatly enrich the system's ability to translate intricate textual descriptions into dynamic and multifaceted artistic compositions on canvas. As such, ongoing research into text-to-image models presents an exciting opportunity for the continued evolution and enhancement of the PICaSo system's creative capabilities.

VII. Conclusion

The PICaSo system is a flexible platform that allows for the exploration and development of the combination of artificial intelligence, human creativity, and robotics. Our research into fine-tuning and tool selection has shown how adaptable and scalable the system is in the process of robotics inpainting on physical canvas. Additionally, we have implemented a clearing process to enable multi-user editing on a single canvas using natural language input. Pre-trained models are unable to produce the desired style according to the waypoint algorithm. However, our fine-tuned models provides an effective style in terms of both

Prompt	Mask	Clearing	Drawing	Time
"Rick from rick and morty"				Clear: – Draw: 59 sec
"glasses"				Clear: 49 sec Draw: 26 sec
"A suit"				Clear: 65 sec Draw: 60 sec
"smile"				Clear: 3 sec Draw: 6 sec
"full body"				Clear: 25 sec Draw: 30 sec
"holding balloon"				Clear: 28 sec Draw: 60 sec
"umbrella"				Clear: 49 sec Draw: 109 sec
"space rocket"				Clear: – Draw: 52 sec
"smoke out of the rocket"				Clear: 5 sec Draw: 59 sec
"ghost"				Clear: 103 sec Draw: 17 sec

TABLE III

ILLUSTRATION OF COLLABORATIVE DRAWING EXPERIMENT CONSISTING OF 10 VOLUNTEERS EACH ONE CONTRIBUTING NEW IDEA TO THE CANVAS USING PICASO COLLABORATIVE INPAINTING METHOD, CANVAS SIZE 600X300 MM

time and quality, as confirmed by both image-text encoders and human evaluation. Additionally, the innovative method for clearing canvas sections enabled multiple users to engage in inpainting sessions, with one notable session involving 10 participants working together to combine their distinct ideas onto a single canvas. In future work, we aim to explore the potential of integrating different painting styles into our robotic system and introducing colors into the system, allowing users to articulate more complex ideas onto the canvas.

REFERENCES

[1] A. Hansen, C. Duna, C. Sandu, and E. Jochum, "Towards creative applications for socially assistive robots," 2020, p. 3, aCM/IEEE International Conference on Human-Robot Interaction, HRI '20 ; Conference date: 23-03-2020 Through 26-03-2020. [Online]. Available: https://humanrobotinteraction.org/2020/

[2] M. D. Cooney and M. L. R. Menezes, "Design for an art therapy robot: An explorative review of the theoretical foundations for engaging in emotional and creative painting with a robot," *Multimodal Technologies and Interaction*, vol. 2, no. 3, p. 52, 2018.

[3] S. Rasouli, G. Gupta, E. Nilsen, and K. Dautenhahn, "Potential applications of social robots in robot-assisted interventions for social anxiety," *International Journal of Social Robotics*, vol. 14, no. 5, pp. 1–32, 2022.

[4] S. Nasrat, T. Kang, J. Park, J. Kim, and S.-J. Yi, "Artistic robotic arm: Drawing portraits on physical canvas under 80 seconds," *Sensors*, vol. 23, no. 12, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/12/5589

[5] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv preprint arXiv:2307.01952*, 2023.

[6] S. Nasrat, T. Kang, J. Park, J. Kim, and S.-J. Yi, "High-speed, high-quality robotic portrait drawing system," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 1042–1047.

[7] L. Scalera, S. Seriani, A. Gasparetto, and P. Gallina, "Non-photorealistic rendering techniques for artistic robotic painting," *Robotics*, vol. 8, no. 1, p. 10, 2019.

[8] R. C. Luo and Y. J. Liu, "Robot artist performs cartoon style facial portrait painting," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7683–7688, 2018.

[9] L. Dong, W. Li, N. Xin, L. Zhang, and Y. Lu, "Stylized portrait generation and intelligent drawing of portrait rendering robot," *2018 International Conference on Mechanical, Electronic and Information Technology*, 2018.

[10] T. Wang, W. Q. Toh, H. Zhang, X. Sui, S. Li, Y. Liu, and W. Jing, "Robocodraw: Robotic avatar drawing with gan-based style transfer and time-efficient path optimization," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34-06, pp. 10 402–10 409, 2020.

[11] W. P-L, H. Y-C, and S. J-S, "Artistic robotic pencil sketching using closed-loop force control," *Proceedings of the Institution of Mechanical Engineers, Part C: JOURNAL of Mechanical Engineering Science*, vol. 236, no. 17, pp. 9753–9762, 2022.

[12] G. Lee, M. Kim, M. Lee, and B. T. Zhang, "From scratch to sketch: Deep decoupled hierarchical reinforcement learning for robotic sketching agent," *2022 International Conference on Robotics and Automation (ICRA)*, pp. 5553–5559, 2022.

[13] F. Gao, J. Zhu, Z. Yu, P. Li, and T. Wang, "Making robots draw a vivid portrait in two minutes," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9585–9591, 2020.

[14] L. Chen, A. Swikir, and S. Haddadin, "Drawing elon musk: A robot avatar for remote manipulation," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4244–4251, 2021.

[15] Q. Gao, H. Chen, R. Yu, J. Yang, and X. Duan, "A robot portraits pencil sketching algorithm based on face component and texture segmentation," *2019 IEEE International Conference on Industrial Technology (ICIT)*, pp. 48–53, 2019.

[16] T. Xue and Y. Liu, "Robot portrait rendering based on multi-features fusion method inspired by human painting," *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 2413–2418, 2017.

[17] P. Tresset and F. Fol Leymarie, "Portrait drawing by paul the robot," *Computers & Graphics*, vol. 37, no. 5, pp. 348–363, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0097849313000149

[18] L. Scalera, S. Seriani, A. Gasparetto, and G. P., "Non-photorealistic rendering techniques for artistic robotic painting," *Robotics*, vol. 8, no. 1, p. 10, 2019.

[19] R. Chancharoen, K. Chaiprabha, L. Wuttisittikulkij, W. Asdornwised, M. Saadi, and G. Phanomchoeng, "Digital twin for a collaborative painting robot," *Sensors*, vol. 23, no. 1, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/1/17

[20] R. Twomey, "Three stage drawing transfer: collaborative drawing between a generative adversarial network, co-robotic arm, and five-year-old child," in *ACM SIGGRAPH 2022 Art Gallery*, ser. SIGGRAPH '22. New York, NY, USA: Association for Computing Machinery, 2022. [Online]. Available: https://doi.org/10.1145/3532837.3534954

[21] G. Yang, "Asymptotic tracking with novel integral robust schemes for mismatched uncertain nonlinear systems," *International JOURNAL of Robust and Nonlinear Control*, vol. 33, no. 3, pp. 1988–2002, 2023.

[22] G. Yang, J. Yao, and Z. Dong, "Neuroadaptive learning algorithm for constrained nonlinear systems with disturbance rejection," *International JOURNAL of Robust and Nonlinear Control*, vol. 32, no. 10, pp. 6127–6147, 2022. [Online]. Available: https://onlinelibrary.wiley.com/DOI/abs/10.1002/rnc.6143

[23] A. Beltramello, L. Scalera, S. Seriani, and P. Gallina, "Artistic robotic painting using the palette knife technique," *Robotics*, vol. 9, no. 1, 2020. [Online]. Available: https://www.mdpi.com/2218-6581/9/1/15

[24] A. Karimov, E. Kopets, G. Kolev, S. Leonov, L. Scalera, and D. Butusov, "Image preprocessing for artistic robotic painting," *Inventions*, vol. 6, no. 1, 2021. [Online]. Available: https://www.mdpi.com/2411-5134/6/1/19

[25] A. Karimov, E. Kopets, S. Leonov, L. Scalera, and D. Butusov, "A robot for artistic painting in authentic colors," *JOURNAL of Intelligent & Robotic Systems*, vol. 107, no. 34, 2023. [Online]. Available: https://www.mdpi.com/2411-5134/6/1/19

[26] 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," *JOURNAL of Intelligent & Robotic Systems*, vol. 105, no. 105, 2022. [Online]. Available: https://www.mdpi.com/2411-5134/6/1/19

[27] P. Schaldenbrand, G. Parmar, J.-Y. Zhu, J. McCann, and J. Oh, "Cofrida: Self-supervised fine-tuning for human-robot co-painting," 2024.

[28] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=nZeVKeeFYf9

[29] J. Hessel, A. Holtzman, M. Forbes, R. Le Bras, and Y. Choi, "CLIPScore: A reference-free evaluation metric for image captioning," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 7514–7528. [Online]. Available: https://aclanthology.org/2021.emnlp-main.595

[30] J. Li, D. Li, C. Xiong, and S. Hoi, "Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation," 2022.

[31] Z. J. Wang, E. Montoya, D. Munechika, H. Yang, B. Hoover, and D. H. Chau, "Large-scale prompt gallery dataset for text-to-image generative models," *arXiv:2210.14896 [cs]*, 2022. [Online]. Available: https://arxiv.org/abs/2210.14896

[32] J. Xu, X. Liu, Y. Wu, Y. Tong, Q. Li, M. Ding, J. Tang, and Y. Dong, "Imagereward: Learning and evaluating human preferences for text-to-image generation," 2023.

[33] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[34] ——, "High-resolution image synthesis with latent diffusion models," 2021.