PICaSo: Physical Inpainting on Canvas Solution

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

***Abstract*— The integration of robotics and artificial intelligence has paved the way for innovative solutions in various domains. This paper introduces PICaSo (Physical Inpainting on Canvas Solution), a pioneering robotic painting system driven by fine-tuned text-to-image models. Departing from conventional approaches, PICaSo harnesses the power of tailored text-to-image algorithms to interpret natural language prompts and execute precise artistic renderings on canvas. Users guide the process by furnishing descriptive text, specifying desired imagery and 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. Leveraging our multi-tools gripper, capable of seamlessly switching between eraser and pen functions, enables to seamlessly execute detailed drawings on canvas while also providing the flexibility to erase and redo sections as required. 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

Representing a novel approach to artistic creation by combining robotics precision with generative AI capabilities to generate visual artworks based on textual prompts. Unlike traditional drawing systems, our approach allows the robot to interpret and create visual compositions from textual descriptions, bridging the gap between human creativity and technological precision in visual art.

In addition to its generative capabilities, our system includes an eraser feature, allowing users to modify and refine their creations. This functionality facilitates the selective removal of elements from the artwork, followed by reinterpretation in alternative styles, as shown in Figure 1, the robotic arm employs its eraser to edit the drawing.

At its core, leveraging a custom trained state-of-the-art text-to-image model which is trained on a specific style in which the waypoint generation algorithm can articulate on canvas, Users can articulate ideas through text using our system, not only one use, multi users can collaborate on single canvas refining and collaborating their ideas realtime on the canvas.

Paragraph about the experiment

*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`

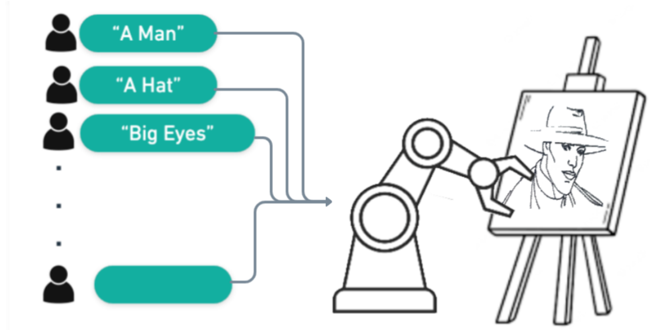


Fig. 1. Illustration showing the collaborative drawing between multiple users on the canvas through a robotic arm.

II. Related Work

The intersection of artificial intelligence and robotics in the realm of drawing tasks has become a focal point of research in academic and digital art circles. Previous endeavors in robotic drawing, as evidenced by studies such as [2]–[17] have encountered notable challenges.

Recent advancements, however, have demonstrated the potential of AI-driven solutions in this context. Tianying et al. [9] utilized GAN-based style transfer to convert facial images into simplified cartoon representations, while Shady N. et al. [1] employed similar techniques to create portrait drawings of high quality. These studies exemplify the capacity of AI to replicate artistic styles with precision, suggesting a promising synergy between AI and robotics in the artistic domain.

Beyond robotics, the digital art world has witnessed a significant movement that combines AI technologies to produce artworks that challenge conventional boundaries. These initiatives encompass a wide range of applications, from neural networks generating generative art to interactive installations driven by machine learning algorithms. While distinct from robotics-centered approaches, these digital endeavors collectively reflect the broader trend of leveraging AI technologies to push the boundaries of artistic expression. Some notable examples include [18]–[20].

However, prior research had not explored the domain of text-to-drawing, making this approach a pioneering effort in this unexplored domain. With the introduction of our proposed artistic system, we address this critical gap, offering a unique and transformative approach to art generation with potential implications for both the artistic and technological communities. PICASO not only enriches the landscape of AI-driven art but also represents a significant milestone in human-machine collaboration, setting the stage for exciting developments at the intersection of AI, robotics, and artistic

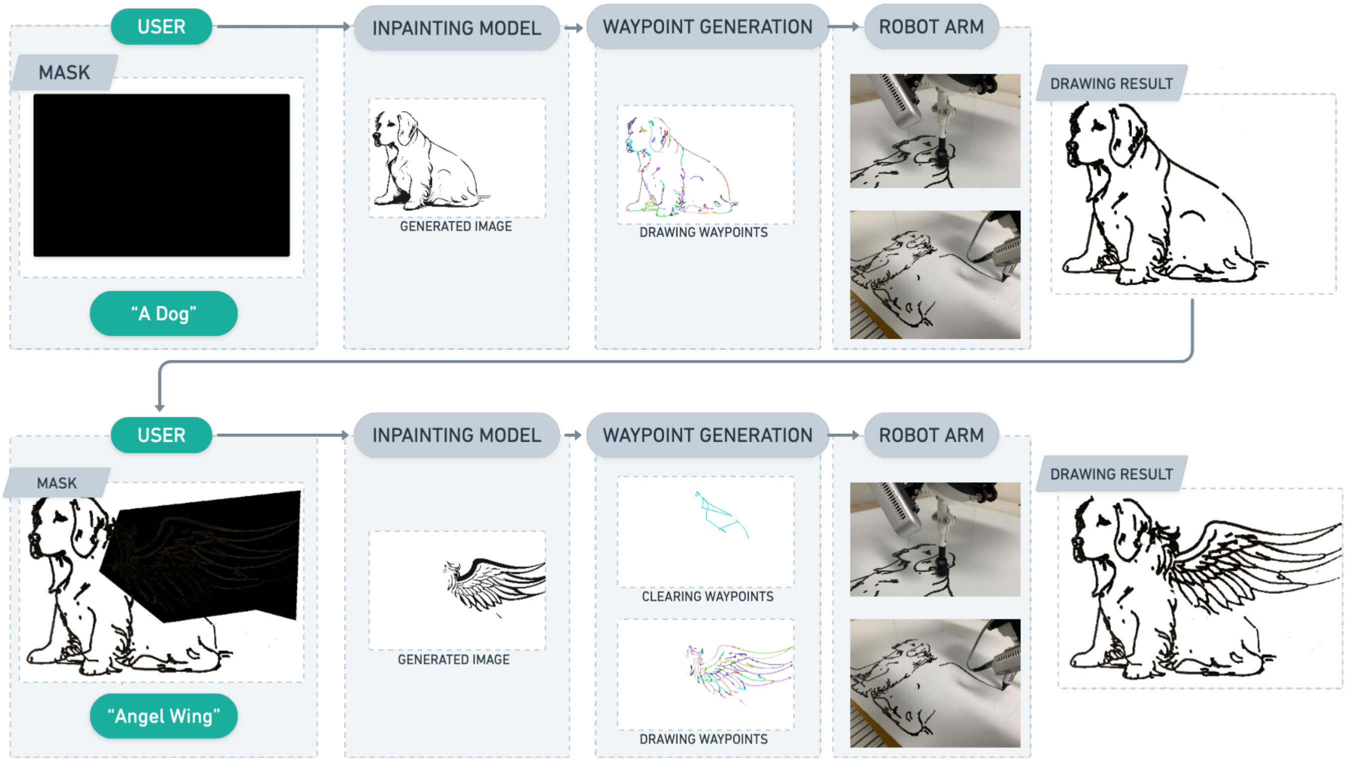


Fig. 2. System architecture and output flow overview: This diagram visualizes the sequential process, starting with user interface input for both mask and text prompt. Subsequently, the inpainting model generates a new image, and waypoints are generated to guide the robotic arm in the drawing process. The final results are displayed as the robotic arm translates the generated drawing onto the canvas.

expression.

III. Method

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

A. Fine-Tuning Inpainting Model

It is a challenging task for the waypoint generation algorithm to extract waypoints from an image that contains colored or shaded regions. Also, the use of a marker on the physical canvas necessitated the drawing of lines in consistent thickness. Therefore, our text-to-image generation model required learning a new style in order to enable the robotic arm to draw the image precisely on canvas.

We utilized the SDXL1.0 model for image generation, rather than training all 6.6 billion network parameters. To conserve resources and time, we employed the LoRA (Low Rank Adaptation) training technique, which freezes the base model and trains only a few parameters with the network rank (dimension) of 128. The training process took four hours on an NVIDIA RTX 3090. Our dataset consisted of 40 images of cartoon drawings, characterized by consistent black thin lines devoid of colors and shading, a crucial style requirement for the waypoint algorithm, which translates each line onto the canvas using a marker. The training dataset was sourced from cartoons, wherein colors were removed and annotations were manually added. We also conducted a comparison between generated images from various models alongside our own model as shown at Table I.

B. Waypoints Generation and Tracing

We utilized the approach described by Shady N. et al. [1]. 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 equation 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.

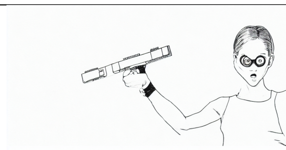
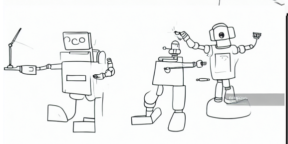
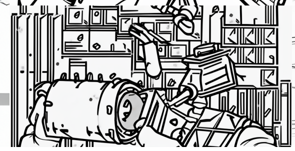
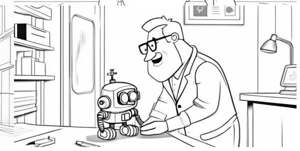
Prompt	RunwayML	SD2	SDXL	OURs
"angry girl with a laser gun"				
"scientist fixing a robot"				

TABLE I

A COMPARISON OF THE GENERATED IMAGE USING DIFFERENT TEXT-TO-IMAGE MODELS AND OUR FINE-TUNED MODEL. IMAGES SIZE: 1024X512 PIXELS, SEED:2935428, FULL PROMPT: BLACK AND WHITE, THIN BLACK LINES, HIGHLY DETAILED, WHITE BACKGROUND, COLORING BOOK STYLE.

Fig. 3. Gripper design 3 tools marker, eraser and camera

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

Both drawing and clearing uses an interpolation movement approach through the generated waypoints showing visually appealing drawings. In the clearing step, the robotic arm retraces its path through all current canvas waypoints. By applying the user's defined mask, it accurately identifies the waypoints necessitating erasure preparing subsequent creative iterations. 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.

C. Gripper Design

Enabling a seamless transition between the different functionalities was essential to this system, to achieve this we implemented a multi-tool gripper mechanism. This gripper was designed to facilitate quick switches between the marker, eraser and camera modes offering a flexible and responsive approach to the process seen at Figure 3. The multi-tool gripper each tool 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.

IV. EXPERIMENTS

We standardized the generation of images to a size of 1024 x 512 pixels for our experiments. Utilizing the PICaSo system, we employed both UR5 and RB-180 robotic arms. In experiments conducted by our team with the RB-180, images were translated onto a physical canvas measuring 600 x 300 cm. For collaborative experiments, we utilized the UR5, translating images onto a canvas sized at 300 x 150 cm. To evaluate the performance and collaborative capabilities of the PICaSo system, we conducted an experiment involving 20 volunteers. Each participant contributed by introducing new creative elements onto the canvas, such as adding new components or modifying existing ones. This collective effort resulted in an artwork collaboratively crafted by 20 individuals, utilizing our system to articulate their ideas into a physical canvas.

Our study demonstrates a 70% success rate in generating in-painted images, as determined by volunteer feedback on our PICaSo system. This finding suggests potential for further enhancement of text-to-image models, particularly in the generation of cartoon images. We have designed our code to be adaptable for future state-of-the-art image-to-text models, enabling our system to advance its drawing capabilities.

V. CONCLUSION

In conclusion, the PICaSo system showcases promising potential in the generation of in-painted images on physical canvases using robotic arms. The use of the SDXL1.0 model, combined with the LoRA training technique, allowed for efficient training and generation of images with consistent black thin lines, meeting the crucial style requirements for the waypoint algorithm. The successful collaborative experiment involving 20 volunteers further demonstrated the system's capabilities in translating creative ideas onto a physical canvas. The 70% success rate, as determined by volunteer feedback, indicates the system's potential for further advancement and enhancement in text-to-image models, particularly in generating cartoon images. With adaptable code designed for future state-of-the-art image-to-text models, the PICaSo system is poised to continue evolving and

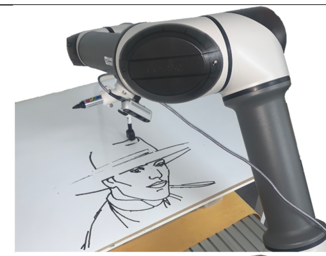
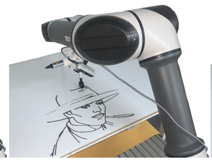
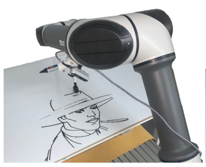
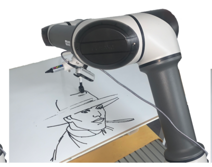
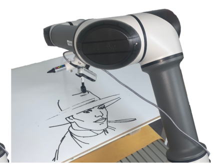
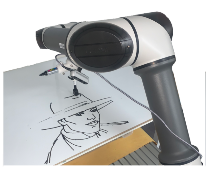
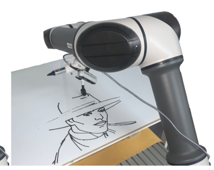
Prompt	Mask	Clearing	Drawing	Time / Attempt No.	Result
"A sheep in a farm"				Draw: 5.6 min Clear: 0.0 min Attempts: 1	
"Big eyes"				Draw: 2.9 min Clear: 1.6 min Attempts: 1	
"A hat"				Draw: 3.2 min Clear: 0.0 min Attempts: 3	
"Stop sign"				Draw: 1.6 min Clear: 1.9 min Attempts: 2	
"Stop sign"				Draw: 0.9 min Clear: 1.2 min Attempts: 1	

TABLE II

EXAMPLE OF COLLABORATIVE DRAWING OF 5 VOLUNTEERS EACH ONE CONTRIBUTING NEW IDEA TO THE CANVAS DRAWN USING UR5

expanding its drawing capabilities in the future. The PICaSo system, utilizing both the UR5 and RB-180 robotic arms, has effectively demonstrated its ability to translate images onto physical canvases and facilitate collaborative artistic expression. In future work, we aim to explore the potential of integrating different painting styles into our robotic system, allowing users to imitate specific artists or artistic movements in their collaborative creations.

REFERENCES

[1] Nasrat, S.; Kang, T.; Park, J.; Kim, J.; Yi, S.-J. Artistic Robotic Arm: Drawing Portraits on Physical Canvas under 80 Seconds. Sensors 2023, 23, 5589. https://doi.org/10.3390/s23125589

[2] Scalera, L.; Seriani, S.; Gasparetto, A.; Gallina, P. *Non-Photorealistic Rendering Techniques for Artistic Robotic Painting*. Robotics 2019, 8, 10. https://doi.org/10.3390/robotics8010010

[3] Carlos Aguilar, Hod Lipson. *A robotic system for interpreting images into painted artwork*. GA2008, 11th Generative Art Conference, 2008, pp. 372-387

[4] Lindemeier, Thomas. (2015). *Hardware-Based Non-Photorealistic Rendering Using a Painting Robot. Computer Graphics Forum*. 34. 311-323. 10.1111/cgf.12562.

[5] 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), 2018, pp. 7683-7688, doi: 10.1109/IROS.2018.8594147.

[6] Li Dong, Weijun Li, Ning Xin, Liping Zhang and Yaxuan Lu. (2018). *Stylized Portrait Generation and Intelligent Drawing of Portrait Rendering Robot*, 2018 International Conference on Mechanical,

Electronic and Information Technology

[7] Tresset Patrick, Fol Leymarie Frederic. *Sketches by Paul the robot*, Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and ImagingAt: Annecy, France, 2012

[8] Calinon Sylvain, Epiney J and Billard Aude. (2006). *A Humanoid Robot Drawing Human Portraits*. Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots. 2005. 161 - 166. 10.1109/ICHR.2005.1573562.

[9] Wang, T., Toh, W. Q., Zhang, H., Sui, X., Li, S., Liu, Y., and Jing, W. (2020). *RoboCoDraw: Robotic Avatar Drawing with GAN-Based Style Transfer and Time-Efficient Path Optimization*. Proceedings of the AAAI Conference on Artificial Intelligence, 34(06), 10402-10409. https://doi.org/10.1609/aaai.v34i06.6609

[10] Wu P-L, Hung Y-C, Shaw 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. 2022;236(17):9753-9762. doi:10.1177/09544062221096946

[11] 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), 2022, pp. 5553-5559, doi: 10.1109/ICRA46639.2022.9811858.

[12] 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), 2020, pp. 9585-9591, doi: 10.1109/IROS45743.2020.9340940.

[13] P. McCorduck, *Aaron's code: meta-art artificial intelligence and the work of Harold Cohen.*, Macmillan, 1991.

[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), 2021, pp. 4244-4251, doi: 10.1109/IROS51168.2021.9635879.

[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), 2019, pp. 48-53, doi: 10.1109/ICIT.2019.8755142.

[16] Chyi-Yeu Lin, Li-Wen Chuang and Thi Thoa Mac, *Human portrait generation system for robot arm drawing*, 2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2009, pp. 1757-1762, doi: 10.1109/AIM.2009.5229810.

[17] 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), 2017, pp. 2413-2418, doi: 10.1109/ROBIO.2017.8324781.

[18] Robin Rombach and Andreas Blattmann and Dominik Lorenz and Patrick Esser and Björn Ommer, *High-Resolution Image Synthesis with Latent Diffusion Models*, 2022, pp. 2112-10752

[19] Dustin Podell and Zion English and Kyle Lacey and Andreas Blattmann and Tim Dockhorn and Jonas Müller and Joe Penna and Robin Rombach, *SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis*, 2023, pp 2307-01952

[20] Aditya Ramesh and Prafulla Dhariwal and Alex Nichol and Casey Chu and Mark Chen, *Hierarchical Text-Conditional Image Generation with CLIP Latents*, 2022. pp. 2204-06125