

ARtVista: Gateway To Empower Anyone Into Artist

Trong-Vu Hoang*

20120025@student.hcmus.edu.vn
University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam

Khanh-Duy Le

lkduy@selab.hcmus.edu.vn
University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam

Quang-Binh Nguyen*

20120412@student.hcmus.edu.vn
University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam

Tam V. Nguyen

tamnguyen@udayton.edu
Department of Computer Science
University of Dayton
Ohio, United States

Duy-Nam Ly

ldnam@selab.hcmus.edu.vn
University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam

Minh-Triet Tran

tmtriet@fit.hcmus.edu.vn
University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam

Trung-Nghia Le†

ltnghia@fit.hcmus.edu.vn
University of Science, VNU-HCM, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam

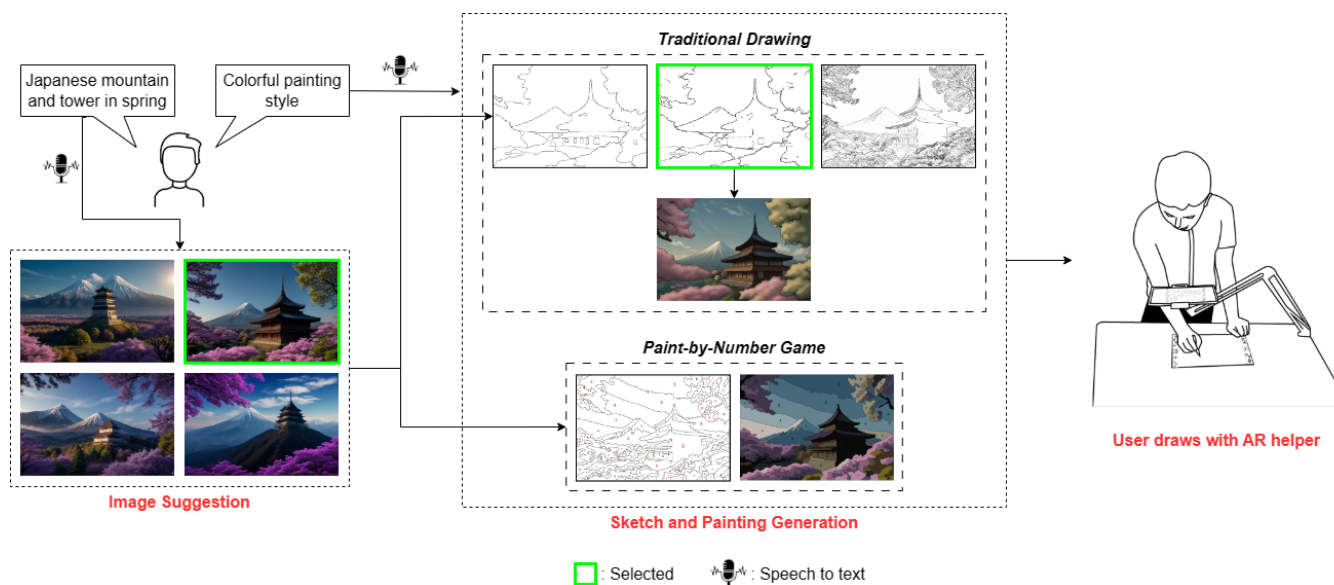


Figure 1: Through AR, the proposed ARtVista system elevates the artistic process, delivering a seamless and immersive experience. Users first input prompts via speech recognition to generate realistic images and then select one to transform it into a specific painting style. After that, the users choose between Traditional Drawing or Paint-by-Number modes to create their preferred painting. Finally, the users draw paintings via an interactive AR interface.

*Equal contribution

†Corresponding author. Email address: ltnghia@fit.hcmus.edu.vn

ABSTRACT

Drawing is an art that enables people to express their imagination and emotions. However, individuals usually face challenges in drawing, especially when translating conceptual ideas into visually coherent representations and bridging the gap between mental visualization and practical execution. In response, we propose ARTVista - a novel system integrating AR and generative AI technologies. ARTVista not only recommends reference images aligned with users' abstract ideas and generates sketches for users to draw but also goes beyond, crafting vibrant paintings in various painting styles. ARTVista also offers users an alternative approach to create striking paintings by simulating the paint-by-number concept on reference images, empowering users to create visually stunning artwork devoid of the necessity for advanced drawing skills. We perform a pilot study and reveal positive feedback on its usability, emphasizing its effectiveness in visualizing user ideas and aiding the painting process to achieve stunning pictures without requiring advanced drawing skills. The source code will be available at <https://github.com/htrvu/ARTVista>.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Interactive systems and tools**; **Mixed / augmented reality**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Text-to-image, AI-generated visual art, AR Painting

ACM Reference Format:

Trong-Vu Hoang, Quang-Binh Nguyen, Duy-Nam Ly, Khanh-Duy Le, Tam V. Nguyen, Minh-Triet Tran, and Trung-Nghia Le. 2024. ARTVista: Gateway To Empower Anyone Into Artist. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3613905.3650760>

1 INTRODUCTION

In the quiet strokes of a pencil or the vibrant hues of a digital canvas, drawing has long been a conduit for human imagination and emotions. Drawing not only enables us to express moments of inspiration or convey meaningful stories, but also, through rhythmic strokes, has been demonstrated to have a therapeutic effect, aiding in the release of stress and anxiety [32].

However, achieving the creation of a beautiful picture involves a challenging process. Firstly, the difficulty lies in translating a conceptual idea into a vivid and coherent visual representation. Even with a preconceived idea in mind, not everyone possesses the ability to envision the complete visual outcome of their drawing, including aspects such as the picture's layout and color. For instance, if a person intends to draw a scene of "a Japanese tower and mountain in spring", where should be the placement of the tower and mountain, and what specific details of the surrounding landscape are appropriate? Secondly, the gap between mental visualization and the practical execution of a drawing can be profound, underscoring the importance of artistic skill and proficiency. Not everyone has many whorls on their fingertips, i.e., the capacity to bring their ideas or even reproduce reference images onto paper

or any artistic medium. Lacking the essential drawing skills can result in a significant investment of time to achieve the desired picture. Finally, coloring a sketch to create a beautiful picture is not a simple process due to varying levels of individuals' aesthetic feelings. It becomes even trickier when we try to color the sketch in a specific painting style, like watercolor painting.

To address drawing challenges, we introduce a novel system, namely ARTVista, to facilitate the creation of beautiful drawings for everyone by synthesizing pictures based on ideas of users and then guiding them to draw those pictures through step-by-step sketching and coloring to follow their desired painting style (See Fig. 1). Users first input their ideas via speech-to-text translation to receive textual prompts. By harnessing the cutting-edge generative AI technology [27], ARTVista can bring textual descriptions to compelling visuals of remarkable quality and realism. Our system also recommends the users in finding reference images that align with their initial concepts, offering them a comprehensive preview of their upcoming drawing. ARTVista then converts users' selected reference images into intricate sketches by employing precise contour extraction of main objects [11, 34]. Taking the creative journey further, our system evolves these sketches into vibrant paintings, tailored to users' preferred styles in our pre-defined style list via generative models [21, 36], resulting in a harmonious and visually stunning output. Finally, inspired by the widely known paint-by-number game, our system offers users an alternative avenue to create impressive paintings based on their ideas through an automatic number painting region process. By integrating AR technology, which overlays the composition draft on the drawing paper through a camera video feed, users can unlock their artistic potential and produce visually appealing artwork without needing advanced drawing skills. The source code will be available at <https://github.com/htrvu/ARTVista>.

We conducted a pilot user study to gather preliminary qualitative insights on the usability of ARTVista. The results of the pilot study yielded the benefits of the proposed system, emphasizing its effectiveness in visualizing users' ideas and facilitating the painting process. Additionally, participants provided valuable feedback highlighting areas for improvement for our system, such as enhancing the drawing experience with AR technology, incorporating a tool to locate the selected color and providing flexibility in adjusting difficulty levels within the Paint-by-Number mode, offering guidance on mixing colors based on the available palette, and supporting additional drawing modes for individuals with varying levels of familiarity or proficiency in drawing.

In summary, our main contributions are as follows:

- We introduced ARTVista, a novel system that incorporates AR technology and generative AI models to enable the creation of beautiful drawings for everyone, without the need for advanced drawing skills.
- We proposed a novel solution that translates users' conceptual ideas into vibrant paintings, adopting a step-by-step guidance process inspired by the paint-by-number game.
- We conducted a pilot study to assess ARTVista's usability, showcasing its effectiveness in visualizing user ideas and simplifying the painting process.

The structure of this paper is as follows. Section 2 discusses previous works. Our proposed ARtVista system is presented in Section 3. Section 4 contains the pilot study and insight discussions. Finally, in Section 5, we summarize the key points of the paper and discuss the implications for future research.

2 RELATED WORK

2.1 AR Drawing Tool

In the past, efforts were made to aid the process of drawing and painting using traditional media and tools. Many individuals struggled with positioning and sizing elements on canvas and executing brushstrokes to achieve desired textures. Prior methods [4, 26] utilized projectors as guides for painting. While effective, these methods lacked flexibility and lightweight execution, often requiring a variety of media tools.

Recently, the development of mobile applications incorporating Augmented Reality (AR) has emerged as a promising tool to enhance physical artistic creation quickly and without the need for specific devices beyond a smartphone. Various works have explored the potential of AR in drawing and painting. Laviole *et al.* [12] proposed AR tools that improve drawing speed and ease, blending physical drawing with projected outlines, and even venturing into stereoscopic drawing with its unique challenges and possibilities. Fischer *et al.* [3] showcased a system to enable interactive manipulation of observed scenes and rendering virtual objects seamlessly integrated with the real environment, offering a good experience for both artists and observers. Beyond drawing, Oh *et al.* [23] presented a mobile AR system that visualizes 3D models from design drawings, significantly reducing recognition time and improving accuracy compared to traditional methods. Ryffel *et al.* [28] demonstrated an interactive application for painting recoloring, achieving realistic color edits with layered alpha channels and an automatic segmentation algorithm for efficient deployment.

However, a notable limitation of these early works lies in their dependence on pre-existing objects and outlines, potentially constraining the user's creative freedom and limiting the expressive range and originality of the final artwork. Meanwhile, our ARtVista system assists users in materializing diverse imagery from their conceptual musings via generative AI.

2.2 Image Generation

Image generation has evolved through diverse paradigms, including Generative Adversarial Networks [9, 15, 31], autoregressive models [10, 13], and diffusion models [1, 22, 27, 29]. Notably, diffusion models [25, 27, 30] have pushed the boundaries of image generation, showcasing a remarkable ability to synthesize high-fidelity images guided by text. To enhance control in generating output of diffusion models, ControlNet [36] has emerged as an approach for generation based on specified conditions, such as canny edge, sketching, depth, and more. By harnessing the capabilities of these models, our ARtVista system can recommend users in finding reference images that align with their initial textual concepts, offering them a comprehensive preview of their upcoming drawing.

However, advancements in diffusion models have presented challenges in balancing generation efficiency and effectiveness, as these

models often require numerous diffusion steps to produce high-quality images, making it impractical for real-time applications. Currently, a notable advancement is latent consistency distillation [19], which distills pre-trained diffusion models to fewer than 4 steps. However, this approach is still limited by the need for separate distillation for each model. To address this limitation, we employed LCM-LoRA [20] in our proposed ARtVista system to accelerate the generation speed for practical application. In this promising approach, LoRA [7] parameters obtained through LCM distillation serve as a universal acceleration module that can seamlessly integrate into any Stable Diffusion pipeline without the need for separate distillation. Moreover, LoRA can be applied to various tasks such as image-to-image, ControlNet, inpainting, and more.

2.3 Image to Sketch

Traditional image-to-sketch methods have relied heavily on classic computer vision techniques, particularly edge detection algorithms [2, 34]. These methods analyze image information to identify significant edges and translate them into lines, often resulting in sharp, detailed sketches. However, they can struggle with complex textures and lighting, leading to overly simplistic or noisy outputs.

The advent of neural style transfer [5] offered a more holistic approach, allowing the transfer of a desired artistic style, including sketching styles, onto an image. Early methods [6, 33] relied on convolutional neural networks to capture style features and apply them to the content image. More recent work has seen the integration of generative adversarial networks (GANs) [8, 17, 35] into style transfer frameworks, further enhancing the quality of the generated sketches. Through adversarial training, GANs learn to capture the artistic essence of sketch styles and apply them to input images, often producing more artistic and expressive sketches than previous methods. Our ARtVista system utilizes various deep learning based image-to-sketch cutting-edge technologies to offer users different that sketching levels progressively enhance in detail.

3 ARTVISTA - MULTIMODAL AR GENERATIVE DRAWING SYSTEM

3.1 Conceptualization

In light of the aforementioned challenges of drawing that users encounter, we introduce ARtVista, a portable system aimed at facilitating the creation of beautiful pictures by users. We conceptualize ARtVista as a mobile application designed to aid users in easily creating their artwork wherever they may be.

ARtVista Interface is an AR application that leverages a user interface for generating reference pictures and supports users in the painting process. Fueled by the potent capabilities of generative AI, ARtVista empowers users to materialize their imaginative concepts into tangible artistic expressions by providing a user-friendly prompt entry system. To enhance the user experience and streamline the input process, ARtVista incorporates an advanced speech recognition feature, the app incorporates a speech recognition feature. This provides users with the ability to input queries hands-free, addressing the need for hands-on engagement in drawing activities where users may be occupied with other tasks. Verbal input-based prompt entering also aligns with the proposed physical setup of ARtVista, as mentioned in the following part.

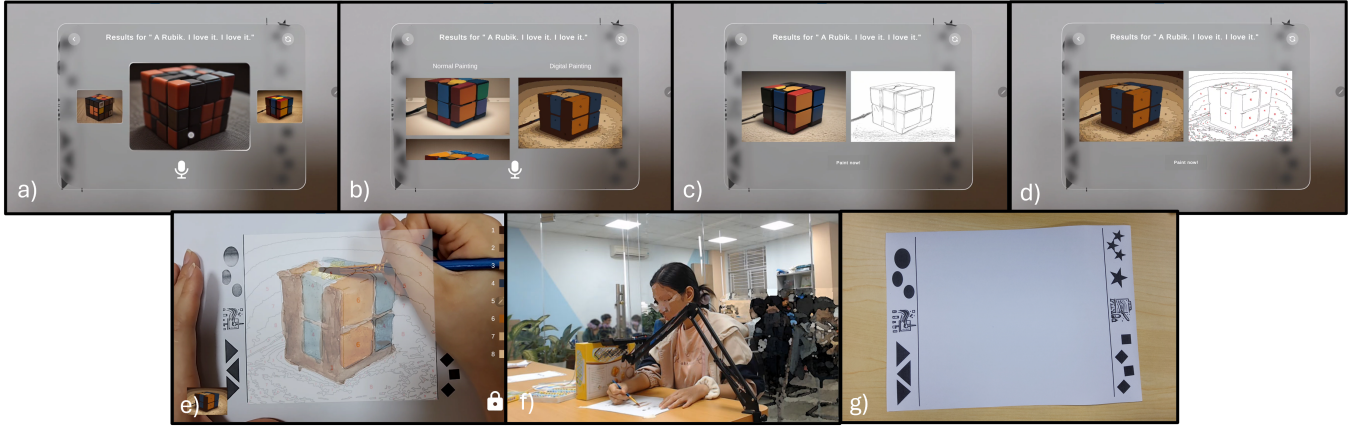


Figure 2: ARTvista interface. a) Realistic reference image suggestion. b) Painting-style picture generating. c-d) Composition of painting and sketch. e) AR view. f) Physical setup for ARTvista. g) Designed drawing paper.

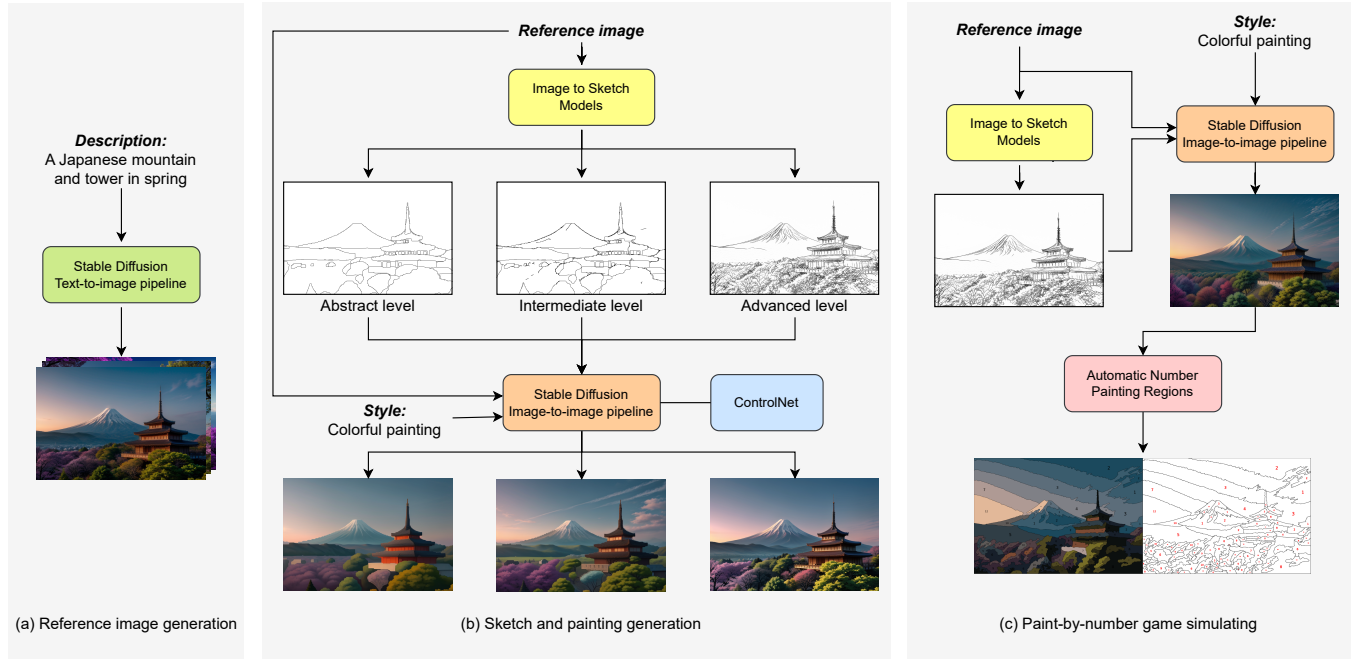


Figure 3: Overview of image generation workflow with three main capabilities: reference image generation, sketch and painting generation, and paint-by-number game simulating.

ARTvista’s generative AI module allows users to craft lifelike pictures through prompt-driven interactions. This stems from the need to artistically create the scenes or things of a number of users instead of capturing them using a camera. In addition, ARTvista provides users with the capability to convert the generated realistic-style pictures into a painting-style format, enhancing the ease and convenience of the subsequent painting process.

However, painting from a reference picture is a big challenge for users who have poor drawing skills. This issue leads them to spend more time calculating the proportion ratio in the drawing.

To help users address this problem, ARTvista incorporate a pass-through AR feature which augments the composition sketch of the painting-style picture they selected before. Thanks to this feature, they can easily complete their artwork with less effort compared to traditional drawing from the reference method. While ARTvista can function with AR/MR headsets, our primary focus is on the integration of the application for widely utilized personal devices like smartphones and tablets. When users engage with the app, the drawing interaction involves looking through the screen of their tablet or mobile device to execute the painting process.

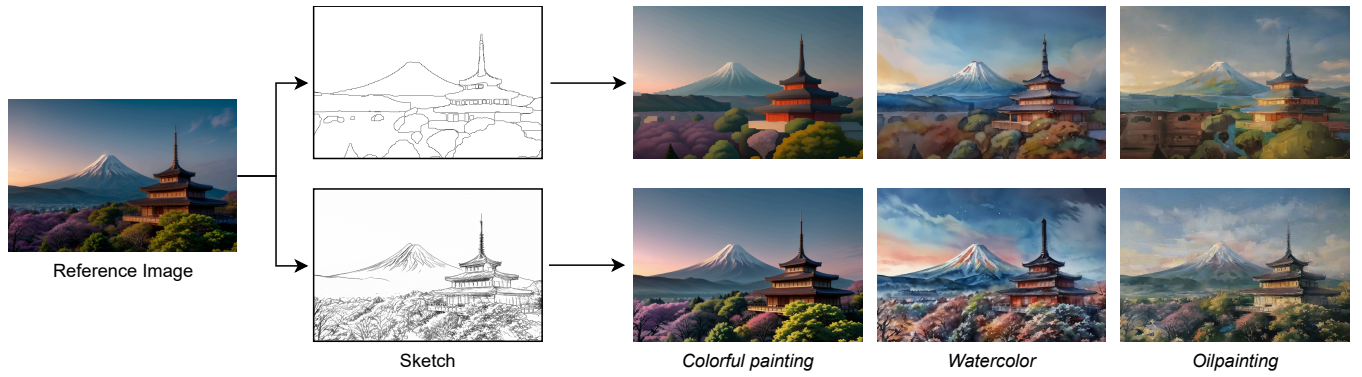


Figure 4: The illustration for the output of some painting styles of our framework: Colorful painting, watercolor, and oil painting.

Drawing setup: In addition to utilizing ARtVista for painting, we strongly recommend users enhance their creative workspace by incorporating a desk-mount arm to elevate the tablet. This helps users stabilize the tablet, making it easier for viewing the screen while drawing. Thus, utilizing this setup, users employing the voice recognition feature will find it more convenient and quicker than entering prompts manually through the keyboard.

Interaction Flow of painting a creative image using ARtVista is as follow (Fig. 1). Users first input the prompt to get realistic images generated by ARtVista using the speech recognition function. After that, users choose one of these pictures to transform it into a specific painting style. Next, users select the mode to start the drawing process. Currently, ARtVista supports two modes for users to create their paintings, which are normal painting and painting by number. Normal painting mode assists users in drawing with the augmentation of the composition draft via a camera video feed with three levels of difficulty. Meanwhile, the Paint-by-Number mode displays a paint-by-number template and a color palette on the UI, users can fill in colors by matching them with corresponding boxes in the template.

3.2 Prototype Implementation

3.2.1 AR Interface: For the development of the ARtVista interface, we utilize Unity as the primary platform. This interface establishes the communication with Image Generation Framework (Part. 3.2.2), which was pre-deployed in our local server, via HTTP APIs to use its functionality. Additionally, to enable users to express their creative input through natural speech interactions, we incorporate OpenAI’s Whisper [24] model provided by the Hugging Face service for recognizing and transforming the user’s speech into textual input. The language we use in our system is English. The illustration for ARtVista’s interactive AR interface with the interaction flow mentioned in Sec. 3.1 is depicted in Fig. 2.

After users select the desired painting-style picture for drawing, ARtVista overlays the composition sketch onto the live video feed captured from the device’s camera. To facilitate users in drawing pictures seamlessly without being influenced by the shifting of the drawing paper with augmented information, we design a special texture for the drawing paper, as seen in Fig. 2-g, and employ

Vuforia to precisely track that paper location. By effectively mitigating the impact of paper movement, users can concentrate on their artistic expressions without spending much time aligning the unfinished painting to match the augmented sketch.

3.2.2 Image Generation Framework: To empower users to translate their mental concepts into tangible artistic expressions regardless of their artistic skill, we design a comprehensive image generation framework as seen in Fig. 3. This framework takes user prompts to generate reference lifelike images, transforming selected images into various sketch versions corresponding with desired painting style pictures. This process aids users in effortlessly obtaining drafts for their paintings, assisting in shaping composition, especially for those unfamiliar with proportional drawing.

Our framework utilizes Stable Diffusion [27] with the Realistic Vision checkpoint, a cutting-edge text-to-image model, to produce reference real images based on the description of users about the thing they want to draw. Additionally, we employ techniques like CPU offloading and attention layer acceleration [14] to optimize the memory consumption of the generation process. By integrating LCM-LoRA [20], we further enhance the generation speed, enabling us to present users with numerous choices within a suitable time-frame.

Deep learning models are employed to generate sketch versions of those images. We offer users three distinct sketching levels that progressively enhance in detail, as seen in Fig. 3-b. *Abstract-level sketches* are obtained by leveraging SAM model [11] to semantically segment the given image, followed by region mining, resulting in visually engaging abstract sketch suitable for beginners and children. *Intermediate-level sketches*, a baseline for everyone to start with our system, are created by fusing edge detection algorithms [11, 34] to enhance the sketch’s detail and intricacy. *Advanced-level sketches* are achieved through Line Art style transfer model¹ for capturing intricate details and nuances, providing users with advanced sketching options. Stable Diffusion [27] with DreamShaperV8 checkpoint, which is known for the proficiency in generating artistic images, is employed to transform users’ sketches into colored paintings according to their preferred painting style. ControlNet [36] and

¹<https://huggingface.co/spaces/awacke1/Image-to-Line-Drawings>



Figure 5: The workflow of automatic number painting regions process.

LCM-LoRA [20] also are integrated to guarantee that the resulting painted images align with the strokes of the original sketches. The output of different painting styles produced by our framework is illustrated in Fig. 4.

To create the template for the Paint-by-Number mode of ArtVista, we perform an *automatic number painting regions* procedure. This involves utilizing the K-Means algorithm [18] along with a graph traversal algorithm to compress and cluster color bits. Subsequently, we extract edge details to outline the boundaries of each region and allocate the cluster’s index to these delineated areas. Figure 5 illustrates the workflow of this process.

4 PILOT STUDY

In our preliminary investigation, we conducted a pilot study aimed at obtaining initial feedback on ArtVista. The primary objective was to explore the efficacy of ArtVista in facilitating users to transform their mental concepts into tangible artistic expressions and offer assistance throughout the drawing process, as well as shed light on potential improvements for our system.

4.1 Study Setup

We invited four participants to experience our system. All participants were students at our faculty with ages ranged between 19 and 20 and non-native English speaker. Among them, two girls exhibit basic painting skills, whereas the two boys rarely engage in drawing activities.

In this study, all participants were equipped with a Samsung Tab S7 device for the execution of ArtVista. Additionally, we also supplied wax color crayons, watercolors, and pre-designed drawing paper to enable users to engage in painting while utilizing the support of the app. To optimize the viewing experience of the augmented reality (AR) screen during drawing sessions, a desk-mount arm was employed to elevate the tablet, aligning its screen orientation parallel to the table surface.

The assigned task for participants involved utilizing ARPainting to create desired images and employing the Painting-by-number mode for guidance. Due to the app’s language configuration, participants are required to use English while engaging with the application. Throughout the process, we closely monitored user behaviors, conducting post-painting interviews immediately after each participant completed their artwork. The drawing session for each participant averaged around 30 minutes, followed by a 10-minute post-interview period. The whole study was video-recorded for our analysis purposes.

4.2 Initial Insights

During the observation of users interacting with ArtVista, a primary challenge emerged in their ability to generate images based on verbal input. This difficulty predominantly stemmed from issues related to the performance of the speech recognition feature and variations in users’ pronunciation. As a temporary solution during the study, participants were provided with an additional smartphone equipped with the Google Translate application’s voice feature to aid in articulation. However, this was acknowledged as a provisional measure for study completion.

In the Paint-by-Number mode, P1 expressed difficulty with the initial 16-color setup, finding it time-consuming. In response, we adjusted the IGF module to require only 8 colors, aiming to streamline the drawing session. P1 also suggested a customization feature, allowing users to choose their preferred number of colors, thus personalizing the difficulty level of this mode.

Meanwhile, P4 tried to tap on a color-1 button in the color palette panel to expect the system to highlight the areas he needed to fill color-1 but did not work. P4 also conveyed the difficulty he faced in locating the colors recommended by the app and keeping track of the available colors. This is evident in his result painting, where the color tone did not truly align with the output of the Paint-by-Number mode (Fig. 6). He wished for a mechanism within the system that would allow him to register existing colors and provide guidance to mix a new color to enhance his painting experience.

Additionally, during the time experienced our system, P2 recommended that it would be more engaging if our system incorporated a paint-by-layer mode, tailored for individuals with some familiarity with watercolor painting or oil painting, as it aligns with their preferred learning approach for those styles and they frequently struggle with blending colors accurately. P2 also suggested a support for mimicking the artistic styles of popular artists like Van Gogh and Leonardo da Vinci, specifically catering to individuals with proficient drawing techniques.

In terms of the drawing experience, users expressed a slight discomfort when drawing through the screen compared to the tactile sensation of traditional drawing. Notably, participants occasionally diverted their gaze from the tablet screen, opting to look directly at the paper while drawing. This behavior was attributed to the perceived ease of drawing without augmented reality (AR) support. Participants mentioned concerns such as color smudging, particularly noticeable when using wax color crayons, and the difficulty in estimating the distance between the pen nibs and the drawing paper. These issues may result in unexpected strokes and hinder the pace of their painting process.

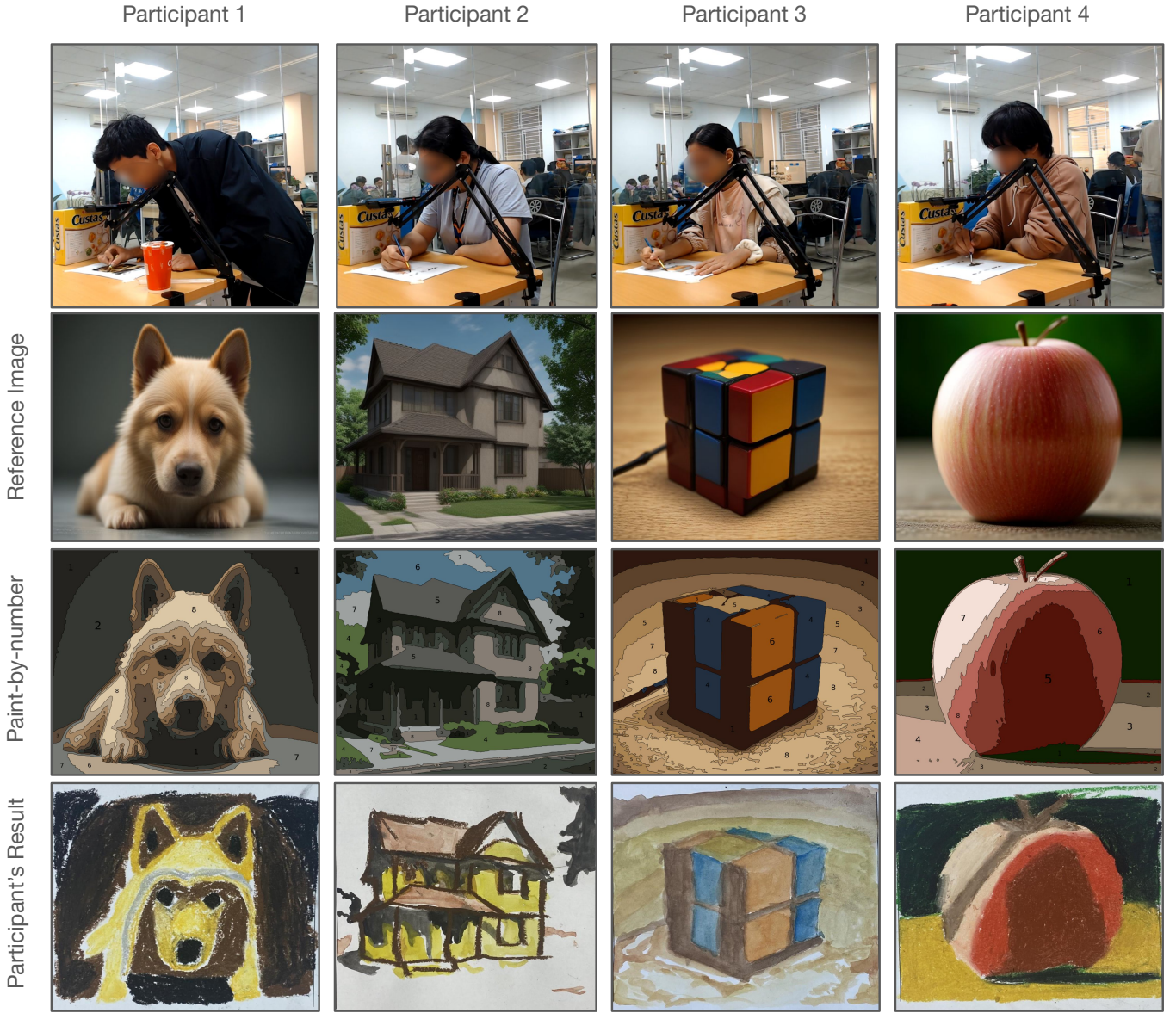


Figure 6: Paintings created by participants using ARtVista.

5 CONCLUSION AND FUTURE WORK

In this paper, we introduced ARtVista, a novel system that enables the creation of beautiful drawings for everyone regardless of advanced drawing skills. By incorporating cutting-edge generative AI models and AR technology, ARtVista can synthesize images based on users' conceptual ideas and guide them through step-by-step sketching and coloring processes to achieve visually captivating artworks. We conducted a pilot study to explore the effectiveness of ARtVista in supporting users to visualize and picture their ideas.

Based on the feedback gathered in this study, we will tackle the existing issues in the current prototype and enhance both the user experience and the variety of drawing modes in the future. Firstly, we intend to improve the drawing experience through the screen

of the device by providing more accurate color visualization on the screen and integrating the depth estimation feature of the device's camera to simulate the interaction between the pen and the drawing paper. Next, we will consider the possibility of specific painting styles (e.g. Impressionist and Baroque styles) and drawing modes tailored for individuals with some familiarity with drawing skills. To replicate popular artistic painting styles more effectively, we intend to utilize more robust checkpoints for the Stable Diffusion model. The upcoming drawing mode that we plan is paint-by-layer, which can be achieved by integrating deep learning models capable of stroke prediction, like PaintTransformer [16].

ACKNOWLEDGMENTS

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant Number 102.05-2023.31.

Additionally, Dr. Tam V. Nguyen is supported by National Science Foundation (NSF) under Grant Number 2025234.

REFERENCES

- [1] Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, Bryan Catanzaro, et al. 2022. ediffi: Text-to-image diffusion models with an ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324* (2022).
- [2] John Canny. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* 6 (1986), 679–698.
- [3] Jan Fischer, Dirk Bartz, and Wolfgang Straßer. 2006. The augmented painting. In *ACM SIGGRAPH 2006 Emerging technologies*. 2–es.
- [4] Matthew Flagg and James M Rehg. 2006. Projector-guided painting. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*. 235–244.
- [5] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).
- [6] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2414–2423.
- [7] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021).
- [8] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.
- [9] Minguk Kang, Jun-Yan Zhu, Richard Zhang, Jaesik Park, Eli Shechtman, Sylvain Paris, and Taesung Park. 2023. Scaling up gans for text-to-image synthesis. In *CVPR*. 10124–10134.
- [10] Diederik P Kingma, Max Welling, et al. 2019. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning* 12, 4 (2019), 307–392.
- [11] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4015–4026.
- [12] Jeremy Laviolle and Martin Hachet. 2012. Spatial augmented reality to enhance physical artistic creation. In *Adjunct proceedings of the 25th annual ACM symposium on User interface software and technology*. 43–46.
- [13] Doyup Lee, Chiheon Kim, Saehoon Kim, Minsu Cho, and Wook-Shin Han. 2022. Autoregressive image generation using residual quantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11523–11532.
- [14] Benjamin Lefaudeux, Francisco Massa, Diana Liskovich, Wenhan Xiong, Vittorio Caggiano, Sean Naren, Min Xu, Jieru Hu, Marta Tintore, Susan Zhang, Patrick Labatut, and Daniel Haziza. 2022. xFormers: A modular and hackable Transformer modelling library. <https://github.com/facebookresearch/xformers>.
- [15] Wentong Liao, Kai Hu, Michael Ying Yang, and Bodo Rosenhahn. 2022. Text to image generation with semantic-spatial aware gan. In *CVPR*. 18187–18196.
- [16] Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Ruifeng Deng, Xin Li, Errui Ding, and Hao Wang. 2021. Paint transformer: Feed forward neural painting with stroke prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*. 6598–6607.
- [17] Xueting Liu, Wenliang Wu, Huisi Wu, and Zhenkun Wen. 2021. Deep style transfer for line drawings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 353–361.
- [18] Stuart Lloyd. 1982. Least squares quantization in PCM. *IEEE transactions on information theory* 28, 2 (1982), 129–137.
- [19] Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. 2023. Latent consistency models: Synthesizing high-resolution images with few-step inference. *arXiv preprint arXiv:2310.04378* (2023).
- [20] Simian Luo, Yiqin Tan, Suraj Patil, Daniel Gu, Patrick von Platen, Apolinário Passos, Longbo Huang, Jian Li, and Hang Zhao. 2023. Lcm-lora: A universal stable-diffusion acceleration module. *arXiv preprint arXiv:2311.05556* (2023).
- [21] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. 2021. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. In *International Conference on Learning Representations*.
- [22] Alexander Quinn Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2022. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. In *International Conference on Machine Learning*. PMLR, 16784–16804.
- [23] Yeon-Jae Oh, Kyoung-Yook Park, and Eung-Kon Kim. 2014. Mobile augmented reality system for Design Drawing visualization. In *16th International Conference on Advanced Communication Technology*. IEEE, 1296–1300.
- [24] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*. PMLR, 28492–28518.
- [25] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1, 2 (2022), 3.
- [26] Alec Rivers, Andrew Adams, and Frédo Durand. 2012. Sculpting by numbers. *ACM Transactions on Graphics (TOG)* 31, 6 (2012), 1–7.
- [27] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *CVPR*. 10684–10695.
- [28] Mattia Ryffel, Fabio Zünd, Yağız Aksoy, Alessia Marra, Maurizio Nitti, Tunç Ozan Aydın, and Bob Sumner. 2017. AR Museum: A mobile augmented reality application for interactive painting recoloring. *ACM Transactions on Graphics (TOG)* 36, 2 (2017), 19.
- [29] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *NeurIPS* 35 (2022), 36479–36494.
- [30] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *NeurIPS* 35 (2022), 36479–36494.
- [31] Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, and Timo Aila. 2023. Stylegan-t: Unlocking the power of gans for fast large-scale text-to-image synthesis. In *International conference on machine learning*. PMLR, 30105–30118.
- [32] Kayla Smolarski, Kristy Leone, and Steven J Robbins. 2015. Reducing negative mood through drawing: Comparing venting, positive expression, and tracing. *Art Therapy* 32, 4 (2015), 197–201.
- [33] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor Lempitsky. 2016. Texture networks: Feed-forward synthesis of textures and stylized images. *arXiv preprint arXiv:1603.03417* (2016).
- [34] Saining "Xie and Zhuowen" Tu. 2015. Holistically-Nested Edge Detection. In *Proceedings of IEEE International Conference on Computer Vision*.
- [35] Ran Yi, Mengfei Xia, Yong-Jin Liu, Yu-Kun Lai, and Paul L Rosin. 2020. Line drawings for face portraits from photos using global and local structure based GANs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 10 (2020), 3462–3475.
- [36] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.