The Joy of Co-Painting: Creative Human-AI Collaboration for Traceable Image-Generation Workflows

Jena Satkunarajan* Steffen Koch† Kuno Kurzhals‡

VISUS, University of Stuttgart, Germany

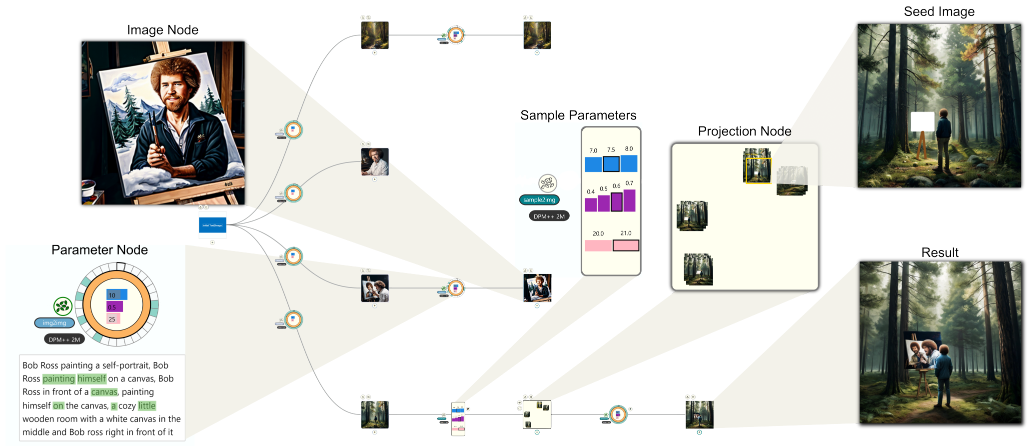

Figure 1: Our approach shows the generation process of an image in a Provenance Graph. The example shows the process of creating an image of the artist Bob Ross drawing himself painting a self-portrait. This type of recursion is typically hard to achieve by prompts alone. By combining prompting, parameter sampling, and manual editing in a comprehensive Co-Painting workflow, we support users in creating content according to their intent and less dependent on the prompt-specific results of an AI model.

Abstract

Image-generative models have gained popularity over the last years with their ability to create realistic artwork. Realizing complex artworks with specific creative ideas often requires iterative optimization of specialized prompts, but may still result in inadequate images. The inclusion of reference images and adapting model-specific parameters can help in steering the model and fostering the creative intent of the user. But by providing text prompts, initial images, and adapting model parameters, users face a vast design space for creating images. To navigate through this space, we propose a visualization approach that combines an interactive Provenance Graph, parameter visualizations, and high-dimensional embeddings. Our approach helps pursue multiple parallel creation paths, makes workflows traceable and parameter changes transparent, and facilitates the reporting of image editing steps. In addition to prompt formulation, we focus on targeted generation by probing parameters, image compositions, and editing details. We integrate the generative process into existing image editing software, enabling users to compose artwork in collaboration with the model. The presented approach is evaluated in a user experiment (n=9) for generating artwork. The results show that users with different levels of experience can create targeted artwork but use different strategies when working with the Provenance Graph.

Index Terms: Human-centered computing—Visualization—Visualization techniques—;

*e-mail: jena.satkunarajan@vis.uni-stuttgart.de

†e-mail: steffen.koch@vis.uni-stuttgart.de

‡e-mail: kuno.kurzhals@visus.uni-stuttgart.de

1 Introduction

The increasing popularity of prompt-based image-generative models led to a multitude of new machine-learning models capable of generating images that cannot be distinguished from digital artwork [15]. Examples of such creative user emancipation comprise artwork for board games [14], book illustrations [46], and movie projects [37]. sers approach creative tasks with a goal in mind and the *intent* to produce a result that best matches their overarching idea, where the details are often developed and adapted iteratively during the creative process. Despite their success, targeted image creation remains difficult when realizing concrete creative intents. *What you prompt is not what you get* - it is often a game of wrangling with text-based image models to create exactly what we envision.

Recently published work [20,57] describes techniques for prompt refinement to steer the image generation. Current research further combines language and image models to address targeted changes in an image [6]. The goal of such procedures is to translate the user's intent into an image result that best fits the user's creative vision. However, it is still an effort to integrate specific objects at certain locations just by prompting alone. Image2image techniques and inpainting [58] enable the targeted variations of locations within images resulting from one specific prompt, but once again boil down to prompt refinement.

To make the image generation process more steerable and foster creativity, digital artists already started to incorporate generative models into their working routines with image editing software such as *Photoshop* [2, 61], *Gimp* [36] or *Krita* [1, 9]. This leads to problems when creating numerous samples because a history of results is difficult to maintain.

Results can also be improved by adapting parameter configurations, which is usually not possible if only the results of specific prompts are saved. Further, the reproducibility of results is important for documentation and instructions. Tracing the origin of generated images can help to argue the originality of artificial artworks, since the length and the depth of the process may prove a creative effort.

We address these problems with a visualization approach that records each step in a creative-generative process. Our approach provides tools for the exploration of parameter configurations and their respective results, and is embedded into the workflow of image editing software. This allows users to explore the influence of hyperparameters on intermediate results. If the model strays away from the user's intent, they can manually edit, remove, and compose images for complex compositions. Figure 1 shows our approach: the visualization is based on a Provenance Graph (see 4.1) to record processing steps [60] and facilitate the navigation between steps. Each step may produce multiple result images from *parameter sampling* (see 4.2), where the user varies the results for a selected prompt by probing different parameter combinations. The image batches produced for each combination are represented as thumbnails in a 2D embedding based on their similarity. An image and parameter configuration can be selected to continue the editing process until results are sent back to the layers of the image editing software.

Our contributions are: **(1)** A visualization approach that enables the targeted image creation by combining prompting, parameter space exploration, and manual editing. We achieve this by combining a graph-based representation of processing steps as an overview, and a 2D projection of result images for individual steps. **(2)** An integration of our visualization approach into existing workflows. We propose the combination of provenance visualization with generative AI and image editing to allow the detailed documentation and data provenance of generation processes. **(3)** We report the findings of an experiment (n = 9) on how users create artwork and assess the workflow and utility of our approach. We identified different strategies during the use of our approach, supporting the individual creative processes of the participants. The visualization provides a step-by-step overview of the creative process important for reproducibility and navigation during the image creation process. We do not target the explainability of models (XAI) but aim to support users in translating their creative intents into digital images. Novice users are given an intuitive framework for creation and tracing of steps, while experienced users can utilize advanced features to employ AI in their sophisticated workflows. We see our approach as a first step away from exclusively prompt-based design into the direction of human-computer co-design.

2 Related Work

Our approach constitutes a combination of visual support for data provenance and visual parameter space exploration. Subsequently, we discuss related work on these aspects as well as research on the visualization of and interaction with image-generative models.

Model fine-tuning Zhang et al [63] propose a generative vision model dedicated to social media text classification and interpretation. Hyung et al. [27] improve the personalized and style-specific generation of portrait photos with image models, designing a novel loss and prompt integration strategy for training. Specialized models perform exceptionally well within their domain, while generalized open-source models such as Stable Diffusion can be more easily embedded into general-purpose tools without the need to train a model beforehand.

Visualization for Human-AI-Collaboration Verheijden and Funk [56] propose the use of image-generative models for creative co-design to help people communicate ideas without requiring artistic skills. In the context of visualization, work on this topic is often categorized as either providing Vis4ML or ML4Vis [48], describing if visualization is used to interpret machine learning models or using machine learning for visualization purposes, respectively. We see our approach in the category of Vis4ML techniques because it provides a means to systematically explore the results of the underlying model. While the explainability of the model is not the focus in this case, we aim to provide control over the image generation processs by parameter exploration and steering, as well as navigation support for previous steps. Hoque et al. [26] describe the value of visualization in *human-centered AI-infused tools (HCAI tools)* to interactively support human users in a transparent and explainable manner. Our approach *empowers* users and *amplifies* their capabilities in creative arts. Within the taxonomy for visual analytics systems for machine learning by Yuan et al. [62], we locate our approach as one to visually explore and understand concrete outputs of existing generative models. Shi et al. [51] provide a systematic review of collaborative approaches, categorizing how AI can assist users in creative tasks. Kim et al. [29] incorporate editing tools in their *DG Comic* system for authoring graph comics to enable flexible content creation, designing editable panels. Our approach aims to help users with the iterative refinement of drafts.

Prompting for image generation Mishra et al. [38] propose *PromptAid* to interactively guide users through prompt adaption. Liu and Hilton [33] propose a guideline for precise prompts for text2image generative models. The guideline can be utilized to influence generative models via specialized textual prompts with the help of proven keyword combinations and by avoiding clashing keywords. Feng et al. [20] combine an image model with a prompt recommendation model to enable interactive, user-steered prompt refinement. Brade et al. [4] propose *Promptify*, a system to ideate and refine personalized prompts with the help of LLMs. Both of these works employ a projection-based visualization, spatially organizing image results based on clustering results with the joint image and prompt embeddings. While they encourage an iterative, prompt-based workflow to refine user prompts, our approach incorporates parameter tuning to achieve specific targets. Wang et al. [57] investigate the usage of a trained proxy model for targeted emotion expression in text-based image generative models. In our approach, we shift the focus from prompt refinement to hyperparameter refinement for better steerability, coupled with an interface for direct image editing.

Multi-modal generation Qiu and Legrady [44] explore a multimodal approach for creating personalized and human-centered art, using human motion data to steer a Stable Diffusion model to create chronophotographs. Sun et al. [54] also highlight the potential of multi-modal systems for *HCAI*, providing a chatbot for fictional story creation complemented by Google Maps locations and AI-generated artworks. Liu et al. [34] create illustrations for articles through a structured search of suitable visual concepts based on keywords or artistic styles. In our approach, we provide interactive tools to explore the parameter space, extending on the idea of in-depth hyperparameter analysis. We visualize the differences in image results due to specific parameter configurations, enabling control of the image generation besides highly specified prompts.

Interactive co-design Hong et al. [25] provide an interactive framework centered around the iterative co-design of AI-generated art. Users provide images for manipulation, where the authors utilize a segmentation model to detect regions of the image for modification and a generative model to redraw these regions. Ko et al. [30] bridge the gap between writers and illustrators for webtoon images, providing an application utilizing a GAN model to generate and synthesize accurate reference images from descriptions. Evirgen and Chen propose *GANzilla* [18] and *GANravel* [19] to steer the generation of images with specific characteristics using GAN-based models. Davis et al. [13] propose an interactive framework to support design space exploration, providing interfaces to navigate the latent space of StyleGAN-models [28] to explore and find new fashion design ideas. Shen et al. [50] integrate multiple generative models in a creativity support system for 3D-sketching called *NeuralCanvas*, which turns 2D sketches into 3D scenes and assists the exploration and iteration of design ideas. Chung and Adar [8] introduce a paint-mixing metaphor in *PromptPaint*, allowing users to mix and match prompts to address specific areas of a generated image. We approach steering of the model to augment specific image areas by embedding parameter space exploration and image editing into our workflow.

Visualization of Provenance Recording provenance information is important to understand analysis processes [7, 40, 45] and often addressed with visualization [60]. In image generation, parameter changes lead to new results that can improve previous images. We visualize this process with a node-link graph representation because it is familiar to most people and a common visualization for workflow tools in general [7, 16]. In the context of our work, the provenance information contributes to reproducibility [21] of achieved results, enables users to branch off during the creation and complements undo/redo functions of image editing tools. Close to our approach are the works by Guo et al. [22] (*PrompTHis*) and Angert et al. [3] (*Spellburst*). *PrompTHis* is a system for prompt provenance analysis, while *Spellburst* provides a node-based creative coding tool for art generation, supporting exploratory workflows. Comparable to our Provenance Graph (see 4.1), *PrompTHis* introduces a graph-based visualization of prompt-image pairs for multiple-generation results. They visualize image distributions depending on word-level modifications of prompts and highlight the impact of specific words. *Spellburst* is focused on the provenance of generative steps and tracking creative paths at any time, providing a node-based overview of the current workflow for the exploration of the design space for sketches. Both approaches encourage an experiment-driven workflow by providing a history of changes and features for the analysis of past steps, but focusing on prompts mostly. In contrast, we combine visual parameter analysis and provenance visualization for targeted art creation, with less attention paid to prompt refinement.

Parameter Space Exploration Machine learning is tightly coupled with a high-dimensional parameter space that is impossible to sample in its entirety. Different exploration techniques have been proposed [49], such as representing the high-dimensional parameter space by a 2D embedding [11, 42]. We use this approach by projecting created instances by the similarity of resulting images. We utilize 2D embeddings [39, 59] of the images in our Projection Nodes (see 4.2), where we project image results from multiple generation steps (sampling) to enable the exploration of the results. With the visual-interactive exploration of subspaces of the parameter space, we provide users with more tools to control the generation process. This aligns with the guidelines proposed by Ko et al. [31] for designing collaborative interfaces with AI models, where they highlight the importance of the controllability of models.

3 Design Concept

We propose a new workflow for Co-Painting. Working with image-generative models involves first finding a promising prompt and then refining it by adjusting prompt and model parameters, or by editing the resulting image manually. Multiple tools have emerged to interact with the models, such as *Midjourney* [24], *Dall-E 3* [41] or the *ComfyUI* web UI [9] for Stable Diffusion. The latter work was further integrated as a plugin into the image editing tool *Krita* [1], incorporating graph-based workflows to create different generation steps. These graphs only consider what happens between one input and the output, while representing results mostly as image matrices, showing many pictures compactly and without overlap. Such representations lack information about similarities between the images and revisiting previous steps of an image generation process can be tedious, as users need to save individual images manually to reuse them. Based on these approaches, we identified the following requirements to improve and unify parts of existing workflows:

R1 Overview: An overview [52] of the results has to be provided on two different levels: (1) for all processing steps and (2) for the images of an individual processing step. Investigating all steps directly contributes to reproducibility (**R2**). We address this requirement (Section 4) by introducing a Provenance Graph as an overview of abstracted nodes. As mentioned, the overview of images of one step should be further supported by ordering elements of similar properties. To emphasize clusters of similar elements, we represent images with 2D embeddings based on image similarity and display their corresponding parameter configuration. This allows users to investigate parameters that lead to similar results or parameter changes that significantly alter the results.

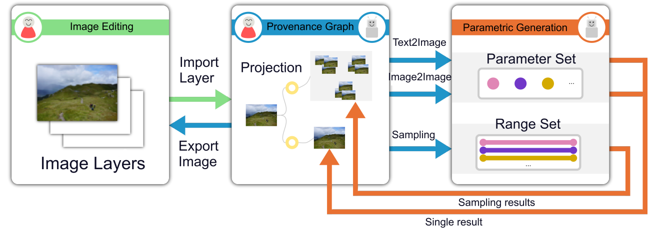

Figure 2: Our proposed Co-Painting workflow brings together manual image editing for **targeted creation** (left) and AI-based parametric generation for **exploration** (right). Image inputs drawn or selected by the user are shown in a Provenance Graph and can be refined by parameter sampling.

R2 Reproducibility: As stated by Hoque and colleagues [26], transparency and provenance are important human concerns that should be addressed for human-AI collaboration. One important aspect of a creative AI collaboration is probing the parameter space in a certain direction while being able to return to previous results and start a new exploration from there. When many steps of Image2Image transformations are performed, current approaches make it difficult to follow or undo changes without massive interaction overhead. If others want to comprehend the creation workflow of an image, a provenance-focused overview helps them learn how to apply these models in a controlled way and achieve specific results.

R3 Editing/Steering: Abstract semantic concepts are still hard to grasp by current AI models [32]. For example, concepts like *morality* have no concrete, physical entity associated with them. Also, recursive ideas, as depicted in Figure 1, are difficult for the models. Hence, creating images by prompts only has limitations regarding targeted content. Image2Image processing provides means to influence the results of a model more targeted than prompt adjustment alone. For this purpose, rudimentary editing possibilities with simple brushes are provided in many interfaces. However, more advanced tools for masking, lighting, and composing are typically not available and require people to switch to other software and later continue processing. The integration of generative models into image editing software is a good way to create an intuitive interaction with the model, i.e., by prompting and drawing. To address these requirements, we developed a framework that works as a plugin for open-source image editing software. We propose a new workflow incorporating layer-based image editing with traceable, intermediate generation, and refinement steps.

3.1 Proposed Workflow

The new proposed workflow (Figure 2) introduces a framework using a Provenance Graph (see Section 4.1) with encodings for processing steps (**R1**, **R2**) and the communication with image editing software (**R3**) and an AI model. Each layer in the editing software can be selected individually and imported into the framework, allowing for the composition of different images. Users may create an initial image by prompting the model (Text2Image), manual drawing, or starting the generation process with an existing image (Image2Image). The image generation and editing can be continued in different ways: (1) Users can linearly refine and edit previous image results, performing Image2Image generation. In this case, one parameter set is used to control the model output. A single image is added to the Provenance Graph. (2) The results can be varied by sampling multiple generations based on parameter ranges. This results in multiple images that are added to the graph as a Projection Node. Intermediate results can be revisited at any time, including sampling results, where the generation process can be restarted with new parameter settings. Images from the Provenance Graph can be exported to *Krita* for manual refinement. Each image is an individual layer, allowing the combination of the images by editing and merging the layers. Individual and combined layers can then be imported into the Provenance Graph for generative steps again.

The presented workflow aims to assist users in the iterative and often repetitive process of generating artwork. Reproducible results enable users to fine-tune configurations, that is, the prompt, reference image, or model parameters, without the need to begin anew. Editing can be a necessary tool where the model fails at complex compositions. Different branches allow for testing diverse strategies, where parameter sampling helps to understand and tweak the model in a way that prompting alone does not allow.

3.2 Parameter Space Considerations

Image-generative models provide a multitude of options to tweak the results, including many different parameters to experiment with. Users can specifically change and adapt these hyperparameters before each generation step. Depicting and exploring the entire parameter space is impossible. Hence, we restricted our visual parameter space exploration to a subset of the model parameters. For Stable Diffusion, this subset includes the Guidance Scale, the Denoising Strength, the Sampling steps, and the Sampler. The Guidance Scale is a discrete parameter $\in [1,30]$ and determines how closely the model follows the provided prompt. The Denoising Strength takes discrete values in the range of $[0,1]$ and instructs the model how much to deviate from the provided seed image. The number of Sampling steps is provided as a discrete integer value starting from 1 to control the amount of image refinement. The Sampler is a categorical parameter that defines the underlying function that maps from the latent space of the model to the image space.

Exposing control over the selected set of parameters to the user provides more opportunities to steer the generative process. Certain fine-grained changes are challenging to express via language, such as specifying that the resulting image should remain arbitrarily close to the reference image. This is achievable by adapting the Denoising Strength. Combining prompting, experimentation, and parameter analysis enables users to translate difficult aspects of ideas without repetitive prompt changes. While we focus on the aforementioned parameters when using Stable Diffusion, our approach is not confined to a particular model. Our visualization approach provides an overview of the combination of quantitative parameters and their impact on the image results. Any image model exposing quantitative parameters to steer the generative model can thus be utilized in our system. There are more parameters that can be varied for different impacts [33] or to achieve highly specific results. Image generative models take in a random seed, which acts as an initialization parameter for the generation. Varying the seed changes the starting point of the generative process, thus allowing for the variation of results. Users can vary the seed as they like, but we decided against including the seed in the parameter sampling for better reproducibility of results. The random seed remains fixed until the user changes it directly. Since the value itself has no concrete meaning, we provide the function to randomize the seed. We also provide interfaces to specify the text prompt and the negative prompt. Users can revise the prompt at any time resulting in a new step in the Provenance Graph, but prompt variations are not considered for sampling.

After comparing approaches from existing literature with our own experiences from working with image models, we believe that the included set of parameters allows for an extensive amount of customization. As we couple image generation with image editing functionalities, our approach provides a large feature set for beginner and expert users alike to realize their ideas in a goal-oriented way.

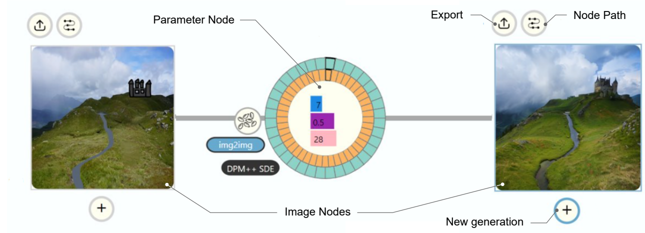

Figure 3: Image Nodes are connected by an edge with a Parameter Node that displays model parameters. Each Image Node can be utilized as a new seed image for subsequent generation steps or exported as an image layer for manual editing.

4 VISUALIZATION FRAMEWORK

Our implemented framework to support the proposed workflow consists of a Provenance Graph as the main visual representation of the current image generation process. The graph is visualized as a node-link diagram consisting of Image Nodes and Parameter Nodes. Additional views such as the Node Path (see Figure 6) enable the user to view the image results of one particular generation path as a sequence of images. Dedicated Projection Nodes display the results of parameter sampling steps. The following section elaborates on the details of the visual representations.

4.1 Provenance Graph

We designed a Provenance Graph that provides an overview of processing steps (**R1**), enables users to navigate and reproduce all steps taken during the generation process (**R2**), and makes the image creation process more transparent and communicable. A Provenance Graph denotes a graph-like structure that provides an overview of temporarily ordered events in a process. Our Provenance Graph represents all steps of the current image generation process and consists of two types of nodes: (1) Rectangular nodes representing image results and (2) glyph-based nodes between Image Nodes representing the parameter configuration that led to the resulting image. The graph is designed as a tree from left to right, i.e., with a dedicated, unique root node, without any cycles and such that a unique path exists between any two nodes of a generation process.

Image Nodes: Figure 3 shows one image generation step consisting of two Image Nodes connected by a Parameter Node. The input picture shows a photo of a mountain landscape with a sketched river and a castle in the background. The resulting image shows a realistic depiction of an actual water stream and a medieval castle on the mountain. At each Image Node users can spawn new child nodes to continue generating from this specific step and be sent back to image editing for detailed changes. The path in the graph to the selected node can also be highlighted on demand, showing all relevant steps to achieve a specific result.

Parameter Nodes: The other node type provides more information about the particular generation step. They are composed of multiple visual elements representing a different attribute of the generation step resulting in the target image. The exemplary Parameter Node in Figure 4 consists of two major visual components: the outer prompt rings and the inner parameter visualization. The prompt rings provide a visual abstraction of both the prompt and the negative prompt that were used for the text2image or image2image generation. Each ring is constructed in the same way, where the outer ring represents the prompt and the inner ring the negative prompt. The rings are segmented based on the number of words in a prompt, where the prompt itself is entered by the user as a string and then split by whitespace, resulting in an array of words. The prompt representation starts at the top-most segment, the first word in the prompt, and continues in a clockwise direction. New elements in a prompt are highlighted by color. For instance, Figure 3 shows a

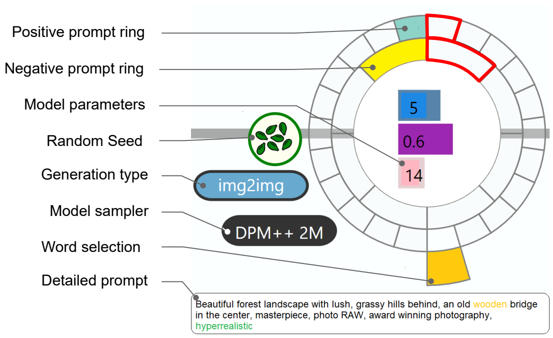

Figure 4: A parameter node shows prompts as rings segmented by individual words. New words are highlighted in the segments and the detailed prompt. The chart in the center displays current and previous (less saturated) values of the parameters.

Parameter Node of a new process, so all words in the positive and negative prompt are new. Figure 4 shows an example where only one new word, "hyperrealistic", was added at the end of the positive prompt. Users can select segments on demand (orange) and look up the respective prompt. We display the prompt as segmented rings to prevent visual clutter and overdraw from the explicit display of the text. The ring representation allows us to use the available space efficiently, such that the prompts can be visualized alongside the quantitative parameters. Our design of the prompt arcs allows us to highlight changes between consecutive prompts. This can be valuable in complex generation processes with multiple branches, where the explicit display of the prompt or keywords becomes difficult. If the user wants to view the prompt explicitly, a textbox can be opened with the prompt.

The inner part of the glyph encodes the model parameters by bars. Figure 4 shows the Guidance Scale, Denoising Strength, and the number of Sampling Steps. We scale the Guidance Scale and the Denoising Strength between 0 and 1, allowing for a better comparison of the parameters. We emphasize value changes of specific parameters between image generation steps by displaying both the previous value and the current value as horizontally stacked bars with varying color saturation. For example, Figure 4 shows a change in the Guidance Scale (blue): the current step uses a value of 5, while the previous iteration utilized a higher value. A visual representation of the quantitative parameters is needed for the comparison of different generation steps and their results. The bar chart representation is intuitive to interpret and compare, allowing for a quick and informed comparison of parameter choices.

Each Parameter Node further contains a symbol with information about the type of processing (Image2Image, Text2Image, Sampling), the applied sampler of the model, and if the seed remained constant (same color). The icons are colored differently using a color-blind categorical color scheme to ease the distinction between the different types of details. The display of these categorical parameters is kept simple to highlight the more relevant, quantitative parameters and the prompts, while still providing necessary information to compare or replicate a generation step. The Provenance Graph is zoomable and pannable. It is possible to work on multiple generative processes simultaneously by switching between browser tabs. This facilitates the dedicated, fine-granular exploration of different generative paths in isolation while also providing an overview of all concurrent generative processes (**R1**). If users have difficulties making progress in the current process, returning to previous steps and branching for a different approach enables them to adapt their strategy and to try again, without having to restart the entire process.

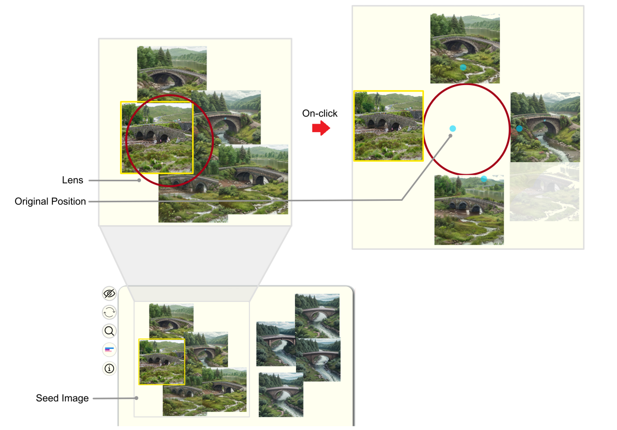

Figure 5: A Projection Node displays the results of a parameter sampling step as a 2D projection that shows similar images in spatial proximity. An interactive lens helps explore the images and remove overlaps by dynamic reordering.

4.2 Parameter Sampling

Open-source models such as Stable Diffusion enable users to change and adapt hyperparameters to steer the generation process. Users can specify ranges of values for numerical parameters and a selection of samplers to facilitate the exploration of the parameter space. The prompt and the negative prompt are not varied in this sampling process. A detailed overview of the generation dialog can be found in the supplemental material (section B). An overview of resulting images (**R1**) can be achieved by ordering elements according to their similarity in a matrix. However, visually identifying clusters of similar elements is impaired due to the lack of spatial distance between less similar elements. Hence, we use a 2D embedding of high-dimensional feature vectors which keeps similar elements spatially together and separates those clusters better. For this purpose, we apply UMAP [35] based on feature vectors derived from the images using the Img2Vec library [47].

Projection Node: The image results are presented in a dedicated Projection Node. It is embedded as part of the Provenance Graph and can be readily accessed within the graph. Figure 5 exemplifies how an expanded Projection Node displays the image results of a parameter sampling step as a 2D projection of the corresponding image vectors. This allows for the similarity-based overview of the sampling results and the observation of patterns and outliers within the results (**R1**). The emphasis on finding parameter configurations producing similar images motivated our decision for a projection-based visualization, instead of utilizing a grid-based display. Parameters for individual results are depicted on demand by glyphs similar to those applied in the Parameter Nodes. For larger sampling quantities, many of the image samples may overlap due to a high degree of image similarity. To reduce overdraw and clutter, we gridify the images using the method proposed by Cutura et al. [12]. For remaining overlaps, we implemented a lens-based interactive function (Figure 5) for the controlled displacement of images that enables the comparison of the results within the spatial context of the projection. With the proposed strategy we can display the origin of the images as small dots within the lens, allowing users to compare the original placement and proximity of images they displaced for further analysis.

Modal View: Users can select individual images to open a Modal View that presents the images with parameters (Figure 6). Users can switch between images to explore and compare the results of the projection. If an image from the projection is selected to continue processing, it is added to the Provenance Graph as a new Image Node. To enable revisiting sampling results, Image Nodes originating from a sampling step always display a Projection icon (P) beside them.

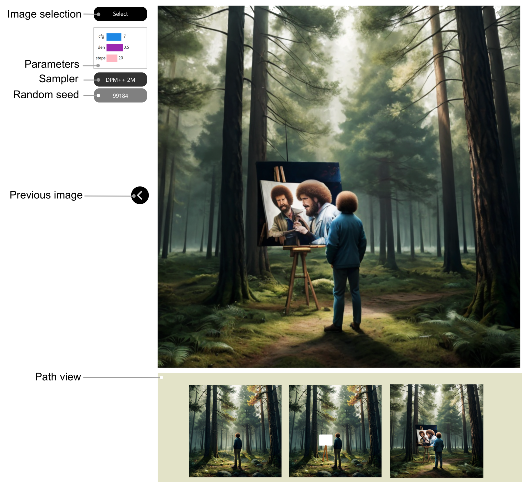

Figure 6: The Modal View shows selected image nodes and sampling results in detail. The Path View displays all steps along the path in the Provenance Graph leading to the selected image.

The combination of the described interactive mechanisms simplifies the search for a fitting image. The focus on parameter changes helps especially in the later stages of the editing and generation process, where users may desire very few changes to the existing image, producing many samples with few overall variations.

4.3 Implementation Details

Our implemented prototype runs on mid-range hardware to enable creativity support without the need for large computing power. Communication with image editing software is implemented as a plugin for *Krita* [55] (**R3**) because it is open-source and provides numerous drawing tools that are helpful for potential editing and guiding image generation. The co-painting visualization framework is written as a web frontend using JavaScript (*Svelte* [10]). Finally, we run a *Flask* [43] backend for the communication between the *AUTOMATIC1111* Stable Diffusion backend and our framework. By this design, our framework can potentially also be included in other software such as *GIMP* or *Photoshop* by re-implementing just the communication interface.

5 Evaluation

We invited n=9 participants to investigate aspects of usability and the emerging strategies when using our approach. We put particular emphasis on understanding how the provenance aspect of our system encourages the iterative generation and refinement of images, and how the parameter visualizations enable the fine-granular adjustment for targeted generation. Our task design reflects this with **1)** a more goal-oriented task to understand whether our approach is intuitive and quick to use and **2)** a creative task to observe and analyze the complexity of the workflows participants develop. We prepared a desktop PC with a mid-range graphics card (RTX 3070) to run the framework. The participants were provided a keyboard and a mouse but no dedicated drawing tablet for image editing. Participants were actively encouraged to think aloud [17] and explain their thought processes and intermediate steps. The screen was recorded during the study and participants' statements were logged. The study was approved by our university's ethics board.

5.1 Participants and Tasks

Overall, 9 people (avg. age = 27) participated in the study; 3 identified as female and 6 as male. All participants have a Computer Science background; four are active researchers at our university, and two have working experiences in UI/UX design at different research faculties. **(P1)** had never worked with image editing in their life and had little knowledge about generative models (<6 months). **(P2–P5)** had little experience with image editing and casually used generative models (1–2 years). **(P6–P8)** had reasonable experience with editing images using free and commercial software and had little to medium experience using generative models (<6 months–2 years). **(P9)** had over 20 years of experience with image editing, using it in their everyday life, and had worked with generative models in the past (1.5 years). After a restricted tutorial to learn the basic functionalities, the tasks consisted of (1) targeted photo editing and (2) a creative, open-ended task. The participants received no formal training for the tasks before the study.

Tutorial (20 min) As an introduction, the participants were provided a photo of a flower. The goal of the task was to complement the scene by adding a realistic honey bee. The participants were encouraged to keep the overall style realistic and retain the flower as it is in the original.

Task 1: Editing (25 min) The participants were provided a photo of a landscape. The photo was taken by one of the authors and contains many distracting features such as random bypassers and street signs. The participants were tasked with removing all such features. They were provided a *Krita* project with two layers, one containing the landscape image and the other one empty. The suggested workflow was to roughly hide or draw over distracting features using *Krita* and then utilize our features to compose and generate a coherent image.

Task 2: Creative (35 min) The participants were free to edit the landscape from the previous task or start with a new one. Their task was to add new elements to compose a fantasy scene. They were encouraged to remain close to the original composition of the landscape image but were given the freedom to change certain elements and the style of the image if necessary. The task started with an open *Krita* project with one layer containing the background image. The participants were encouraged to experiment with different approaches.

5.2 Procedure and Measurements

The participants were informed about the study procedure and signed a consent form. They then performed the tutorial and were supported by one of the authors in case of questions about functionalities. When the participants finished a task, they were asked to confirm that they wanted to proceed and the resulting image was saved. After task 2, we asked them to complete a questionnaire before the experiment ended. The questionnaire consisted of demographic questions and a set of 18 statements and text boxes for general and system-related comments. The first 9 questions were adapted from the System Usability Scale [5], while the remaining 9 questions were dedicated towards creativity support. The participants were asked to indicate their agreement or disagreement with the statement on a Likert scale ranging from -2 (strong disagreement) to 2 (strong agreement). The questionnaire ended with specific questions about the visual elements of the system and offered text boxes to share general thoughts and points of improvement.

6 Results

Our results comprise the evaluation of the questionnaire and a detailed investigation of strategies for solving the given tasks. The results indicate that our approach supports all levels of expertise and that experience might influence the applied strategies.

6.1 Questionnaire

Figure 7 presents the answers of all 9 participants. Participants found the overall system simple to use (Q1–Q3, Q8), with a reasonable learning curve (Q10). After a brief introduction to the technical aspects, participants could effectively use our prototype for the creative image generation process (Q4, Q7, Q11).

Like to use frequently (Q1)	P4	P5	P8	P9	P2	P1	P6	P7	P3
System too complex (Q2)	P9	P2	P3	P5	P1	P7	P4	P6	P8
Easy to use (Q3)	P2	P4	P5	P9	P3	P1	P6	P7	P8
Technical help needed (Q4)	P2	P7	P9	P3	P6	P8	P4	P5	P1
Workflow intuitive (Q5)	P4	P5	P6	P2	P9	P3	P7	P8	P1
Workflow inconsistent (Q6)	P9	P2	P4	P1	P6	P7	P3	P5	P8
Quick to learn (Q7)	P5	P8	P2	P4	P6	P7	P3	P9	P1
Tedious to use (Q8)	P2	P8	P9	P3	P1	P6	P7	P4	P5
Penalizes Mistakes (Q9)	P3	P8	P2	P4	P5	P6	P7	P9	P1

Steep learning curve (Q10)	P5	P1	P6	P7	P9	P2	P3	P4	P8
Fosters creativity (Q11)	P5	P4	P7	P9	P2	P1	P6	P8	P3
Encourages undo/redo (Q12)	P4	P6	P7	P8	P9	P2	P3	P5	P1
Confident in Art (Q13)	P4	P5	P9	P2	P1	P6	P7	P8	P3
Create pleasant images (Q14)	P5	P4	P6	P7	P9	P2	P1	P8	P3
Satisfied with results (Q15)	P5	P8	P4	P7	P9	P2	P3	P1	P6
Images not my work (Q16)	P6	P9	P2	P4	P7	P3	P8	P5	P1
Parameters uninteresting (Q17)	P9	P2	P3	P1	P4	P5	P6	P7	P8
Editing software not beneficial (Q18)	P9	P5	P3	P7	P4	P6	P8	P2	P1

Strongly disagree	-2	-1	0	1	2	Strongly agree

Figure 7: The participants generally enjoyed using our prototype for AI-assisted art creation. The proposed workflow incorporating the provenance of generation steps and image editing was considered intuitive and useful. The participants found the system to be a helpful creativity support tool.

The participants stated that the system was "*overall helpful*", "*was good for creative tasks*", "*made the usage of generative AI intuitive*" and was "*quick to learn*". They developed workflows for their preferred strategies using the (visual) tools provided in our system (Q5, Q6). In particular, participants considered the ability to adapt and retry failed steps as helpful (Q12), and the graph-based workflow was "*good for trying out and jumping around (creatively)*".

The Projection Nodes were seen as a central tool for parameter experimentation, as it was "*intuitive to use*" and "*useful as an overview and helped make sense of the parameter space*". Participants also proposed features to ease the navigation in the Projection Nodes, such as "*a grid-display or something similar, to make better use of the space*" and more tools to handle overlap for large samplings. Since they "*need time to find good ranges*", participants suggested "*more help for selecting good parameters*". The depiction of the parameters involved in each generation step was considered insightful ("*helps with orientation and replication of results*", "*helps with reproducibility*") and the integration of an image editing software as providing necessary tools (Q17, Q18). At the same time, participants noted that "*the parameters were sometimes hard to compare*" and that they "*needed some time to learn*", proposing a more abstract representation such as "*something like icons for fast comparison*". Some of the participants highlighted that they "*did not use [the parameter nodes] much*".

Being able to undo and redo any step at any time within the Provenance Graph, participants did not consider the tool to penalize mistakes when creating art with AI models (Q9). The resulting images were mostly considered successful (Q15). Two participants attributed the results more to the generative model and less to their capabilities (Q16). Participants reported difficulties in steering the model at times, for example, the model "*sometimes randomly adding elements I don't want or omitting very specific elements I wanted to have*". The majority of the participants considered our approach to increase their confidence in creating art (Q13), with some exceptions among the less experienced users (P4 and P5). Comments such as "*waiting for sampling results was sometimes cumbersome*" stem from the increased waiting time for the results of large sampling steps. This may have deterred them from further experimenting.

6.2 Tasks and Strategies

We analyzed the different workflow patterns depending on each task and the participants' proficiency in image editing. For each resulting workflow, we paid special attention to the following aspects:

Degree of exploration: How often do participants vary their approach? Is their process mostly linear, with consecutive generations based on the most recent result? Do they branch out often or return to previous results and begin new attempts with variations in the parameters, the prompt, or the seed image?

Editing behavior: When and why do participants attempt to edit seed images or intermediate image results? Is it mostly a deliberate choice or out of necessity to add or remove elements?

Parameter experimentation: How willing are the participants to experiment with different parameter configurations? When and how do they utilize the parameter sampling?

For **Task 1** we observed two prominent approaches to editing unwanted objects in the image. (1) Five participants (P1, P2, P3, P6, P7) began by seeding the given image directly into the generative model. They experimented with the prompt, negative prompt, and the parameters to generate an image close to the original. (2) The other four (P4, P5, P8, P9) started by (roughly) editing the image and proceeding with image generation. Subsequently, all participants followed a mostly linear exploration pattern. Figure 8 shows the original image of Task 1 and two exemplary results from participants P3 and P5. As the images showcase, both described strategies lead to similar results. Both groups utilized parameter sampling in all but one case to employ small modifications in consecutive steps. P1 saw parameter sampling as a tool for parameter space exploration and did not attempt to explore and fine-tune specific generation paths. The remaining participants used parameter sampling to investigate and understand the interaction of different parameters while observing and adapting to the changes induced by parameter tuning. Overall, applying targeted changes with small adaptations of the prompts and parameters helped avoid strong deviations and unwanted results, leading to simple generative paths. The main differences lay in the willingness to edit images, which was greater among the experienced participants (P6–P9). The graph exploration and parameter experimentation were kept light for applying smaller targeted changes.

In **Task 2** all participants utilized results from the previous task, and we identified three major strategies: (1) P1–P3 and P6 began with image generation right away. The goal was to provide an accurate and detailed prompt with all fantastical elements at once and try to steer the generative model via parameter adaptation to modify the image as required. Some of the targeted compositions of the participants were quite complex, thus requiring multiple attempts with iterative prompts and parameter experimentation and refinement. The resulting Provenance Graphs thus had numerous branches close to the root node, with a small subset of the branches developed further. This strategy was employed mostly by the participants with little to no image editing experience. Participants with this strategy often commented loudly that they were unsure how to begin and lacked confidence in their editing skills, so they started by experimenting. The resulting images indicate that this type of strategy might lead to high variance in the results: the experiment-driven, iterative generation aims at finding nice results by chance, thus deviating strongly from the original image. (2) The second strategy put strong emphasis on the generative model as well, albeit utilizing the model to generate fantastical elements one by one. We saw variations of this strategy being employed by participants P5, P7, and P8. Each element was then integrated into the current seed image using very light image editing. Subsequently, the edited image was seeded into image generation to let the model seamlessly integrate the new elements into the overall composition. This last step required some fine-tuning of parameters using parameter sampling to integrate new elements without unwanted modifications, such as sudden style changes. Participants purposefully experimented with parameters to carefully implement individual elements into the image. The resulting Provenance Graphs thus exhibited more branches,

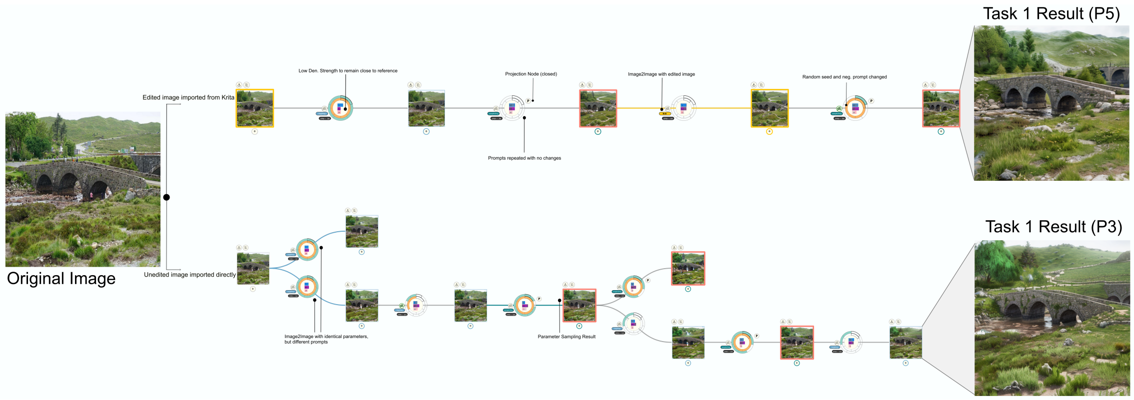

Figure 8: Results of task 1 showcase that an emphasis on image editing (P5) and an emphasis on multiple image generation steps (P3) both lead to good results. Both Provenance Graphs emphasize the mostly linear workflow with multiple parameter sampling steps. The annotations show examples of Parameter Nodes and edited images that were imported into the Provenance Graph and highlight source and target nodes.

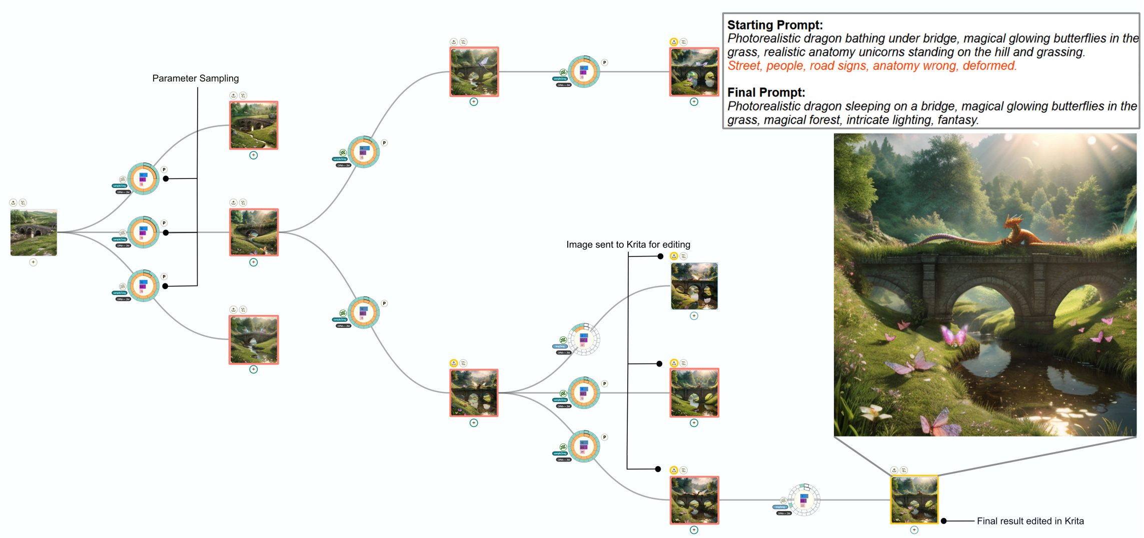

Figure 9: The Provenance Graph of participant P9 for task 2 provides insight into the strategy employed by them. P9 made extensive use of parameter sampling to generate multiple fantastical elements, combining them with advanced image editing techniques into the final result. The beginning (left) shows three initial sampling steps, each one continued with an image chosen from the results. The annotations highlight some examples where the user exported intermediary results to *Krita*.

but only certain paths were developed further. The dedication to one *working path* aligns well with the participant's intent and experience: participants employing the described strategy had sufficient experience with generative models and were aware of the capabilities of Stable Diffusion. Each new element required an initial generation process, where the dedicated path was abandoned as soon as a good image resulted from it. Intermediate results were then integrated into the composition using image editing and different image layers. Examples of the first two strategies can be found in the supplemental material (section A). (3) The third strategy was employed by the most experienced participant (P9). They integrated image editing into their workflow, combining multiple image generation results to produce the final result. Figure 9 presents the Provenance Graph of the participant for Task 2, demonstrating the utility of the graph-based visualization for ideation and experimentation by exploring multiple paths. P9 utilized multiple paths dedicated to different elements of the overall image composition, targeting a multi-layer approach where they combine the different generation results into one composition. They made extensive use of parameter sampling, aiming to identify a parameter configuration that produces images with the same or very similar perspectives, as this allowed for the combination of different images by masking. The result deviates strongly from the original image: the participant preferred the aesthetics of some of the intermediary results, despite those introducing new perspectives or elements not found in the original image.

6.3 Workflow Takeaways

Our approach supports all levels of expertise for image generation, allowing even laypeople to create targeted images. However, we observed indicators that proficiency may influence user behavior and our approach reaches its full potential with experienced users.

Users with little experience in image editing and generation preferred linear exploration when aiming for specific image results. Traceability of results mattered less while learning the ropes of image generation. This motivates such users to experiment with incremental parameter adaptions rather than explore different parameter configurations or manual image manipulation and, thus, different paths in the Provenance Graph. Our proposed approach supports inexperienced users in exploring the capabilities of image-generative models. Visual tracing of individual parameter alterations in the Provenance Graph and the sampling-based projections help understand how the model can be steered to produce specific results.

Moderately experienced users in image editing and generation made more use of the provided features. The Provenance Graph was employed mainly in the ideation and exploration of different approaches to create specific images and to analyze variations in results. The more targeted experimentation of moderately experienced users is facilitated by our approach: users can quickly compare image results after varying their approach, either by fine-tuning the model or manual image manipulation, without the need to memorize individual results.

Experts in image editing and generation showed great interest in the sampling-based projections and incorporated those heavily in their exploration-driven generation of images. Branching out quickly and trying different initial ideas, such users aimed to identify good intermediary results to further work with. The integration of image editing software and the graph-based visualization of generative steps assisted skilled users in editing especially. Such users could utilize the tool to add fine-granular detail into different intermediary results and combine them into complex compositions.

Overall, our study showed that participants with different levels of experience can create targeted images with our approach. All participants stated that solving the given tasks with the provided system was a pleasant experience. In the simplest case, people use the framework similar to other established systems, i.e., by playing around with prompts and investigating what happens. However, with more experience, it is possible to use the Provenance Graph in combination with editing capabilities to fully harness the benefits of targeted image generation and parameter steering. In all cases, the resulting graphs provide a detailed summary of the user behavior and strategies for dissemination, or for analyzing results and teaching users how to improve their skills in generating new artwork.

7 Discussion

We received overall much positive feedback from our participants. However, from the feedback as well as our experience in developing and testing the proposed approach, we identified some problematic issues. We discuss these subsequently and indicate how they can be addressed in the future.

Model Control: Our proposed approach provides an intuitive and easy-to-learn way to create workflows to design and refine images. For further improvement, it would be necessary to have even more control over the generation, preventing frustrating processes with subpar image results. One way would be to integrate state-of-the-art prompting methods [8, 23] that guide users in constructing complex prompts for model-specific and concept-specific results. This would also help with the initial trial-and-error stage to find a suitable starting point for further, directed manipulation.

Graph Representation: The graph-based representation of generation processes serves as an overview. In more complex workflows, the graph is likely to grow until scrolling is necessary to view all steps from beginning to end. Such dense and large graphs with potentially numerous branches further become difficult to analyze and understand, necessitating the means to selectively truncate or summarize certain parts. We plan to enable the collapsing of uninteresting paths to improve the use of screen space. This would further help with reducing the computational cost associated with rendering large, complex Provenance Graphs in the web browser.

Projection Alternatives: Projection Nodes provide a quick overview of the samples and their similarity. Temporarily removing or rearranging cluttered regions within the projection handles overdraw but diminishes efficiency as the number of results increases. Participants suggested a grid-based display of preselected groups within the results. We utilize the *hagrid* [11] space-filling technique to arrange images locally, but it does not provide the same level of structure as a global grid. A global grid could display many samples with optimal use of display space but at the cost of information loss.

Parameter Encoding: We observed that participants were most interested in the parameters when comparing sampling results in the Projection Nodes and the Modal View as seen in Figure 6. While it was helpful for comparing current results to previous steps, participants stated that the parameter visualization was mostly an afterthought during the generation process. An alternative to the explicit glyph presentation could be a more abstract one, such as using icons to represent either the absolute or relative value changes between consecutive steps. Placing such an abstract parameter encoding directly alongside the images may put more emphasis and attention on the parameters.

Prompting: As noted by the study participants, there was a lack of prompting support, that is, some form of guideline or assistance on prompting the image-generative model for specific compositions or details. According to the participants, this particularly affected the initial stages of the image generation process, where they tried to realize broad ideas to start working with. A possible extension of our approach to help with his issue could be to incorporate some of the existing techniques to provide prompt support, such as templates [53] or similarity-based prompt-image suggestions [20].

Parameter Sampling: Sampling different parameter ranges allowed users to vary the generation results for a fixed prompt and to observe the influence of each parameter. As a singular generation step rarely took over a second to finish, the time cost remained acceptable for smaller ranges producing approximately 50 to 75 images at once. Larger sampling ranges invoked a blow-up of the number of images to be generated, thus becoming computationally unpractical. To keep the system responsive and interactive, it might be helpful to allow the scheduling of generation jobs that keep running in the background.

Assisting Users: Our approach allows users with creative goals and different levels of expertise to realize their creative intent. We see user intent as an important aspect of a creative process and designed our system to assist users in adapting their process to their skills and needs. More complex features such as parameter sampling require a short learning phase, but provide great value in sophisticated workflows.

Generalizability: Findings from our study show that users can realize their creative goals using our approach. Their expertise in AI and editing translates into interesting workflows with varying levels of complexity. For a more nuanced distinction of workflows, it would be helpful to recruit a broader spectrum of participants with different interests. Our visual representation can be adapted with other image generation models and image editors. Adjustable parameters can differ between models but are often quantitative or categorical and can thus be used for the glyph-based visualization in Parameter Nodes. The image editing further relies on the *layers approach*, which is a common concept in most editors.

8 Conclusion

We presented an approach for the realization of multi-layered artworks by interactive exchange between users and an image-generative model. Our Provenance Graph provides detailed insights into different workflows employed by users with different levels of expertise. In the future, we plan to adjust aspects such as the distance between nodes to reveal insights into the frequency of changes and their impact on the intermediate results. We further plan to revise our Provenance Graph's interface design to be more adaptive, such that users are introduced to more complex features, such as parameter sampling, step-by-step. Overall, we think prompt design alone will not harness the full creative potential when working with image-generative models. Therefore, we see the future for targeted use in integrated components for image editing software. This allows professionals and laypeople to realize ideas in collaboration with AI supported by novel workflows and visual interfaces such as the presented one.

Acknowledgments

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618. We thank Julian Hummel, who developed an early prototype as part of his bachelor thesis.

References

[1] Acly. Generative ai for krita. https://github.com/Acly/krita-ai-diffusion, 2023.
[2] A. Alfaraj. Auto-photoshop-stablediffusion-plugin. https://github.com/AbdullahAlfaraj/Auto-Photoshop-StableDiffusion-Plugin, 2023.
[3] T. Angert, M. Suzara, J. Han, C. Pondoc, and H. Subramonyam. Spellburst: A node-based interface for exploratory creative coding with natural language prompts. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023.
[4] S. Brade, B. Wang, M. Sousa, S. Oore, and T. Grossman. Promptify: Text-to-image generation through interactive prompt exploration with large language models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023.
[5] J. Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 1995.
[6] T. Brooks, A. Holynski, and A. A. Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 18392–18402, 2023.
[7] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo. Vistrails: visualization meets data management. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pp. 745–747, 2006.
[8] J. J. Y. Chung and E. Adar. Promptpaint: Steering text-to-image generation through paint medium-like interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23. ACM, 2023.
[9] comfyanonymous. Stable diffusion backend and gui. https://github.com/comfyanonymous/ComfyUI, 2023.
[10] S. contributors. Svelte - cybermetically enhanced web apps. https://github.com/sveltejs/svelte, 2024.
[11] R. Cutura, S. Holzer, M. Aupetit, and M. Sedlmair. Viscoder: A tool for visually comparing dimensionality reduction algorithms. In *Esann*, 2018.
[12] R. Cutura, C. Morariu, Z. Cheng, Y. Wang, D. Weiskopf, and M. Sedlmair. Hagrid — gridify scatterplots with hilbert and gosper curves. In *Proceedings of the 14th International Symposium on Visual Information Communication and Interaction*. Association for Computing Machinery, 2021.
[13] R. L. Davis, T. Wambsganss, W. Jiang, K. G. Kim, T. Käser, and P. Dillenbourg. Fashioning the future: Unlocking the creative potential of deep generative models for design space exploration. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.
[14] Elad. How I proposed with a board game I created. https://bogomolnyelad.substack.com/p/board-game, 2023.
[15] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone. Can: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms. https://arxiv.org/abs/1706.07068, 2017.
[16] N. Elmqvist, J. Stasko, and P. Tsigas. Datameadow: a visual canvas for analysis of large-scale multivariate data. *Information visualization*, 7(1):18–33, 2008.
[17] K. A. Ericsson and H. A. Simon. *Protocol Analysis: Verbal Reports as Data*. The MIT Press, Cambridge, 1993.
[18] N. Evirgen and X. A. Chen. GANzilla: User-driven direction discovery in generative adversarial networks. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022.
[19] N. Evirgen and X. A. Chen. Ganravel: User-driven direction disentanglement in generative adversarial networks. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.
[20] Y. Feng, X. Wang, K. K. Wong, S. Wang, Y. Lu, M. Zhu, B. Wang, and W. Chen. Promptmagician: Interactive prompt engineering for text-to-image creation, 2023.
[21] M. Franke, G. Reina, and S. Koch. Toward reproducible visual analysis results. In *2023 IEEE 16th Pacific Visualization Symposium (PacificVis)*, pp. 102–106, 2023.
[22] Y. Guo, H. Shao, C. Liu, K. Xu, and X. Yuan. Prompthis: Visualizing the process and influence of prompt editing during text-to-image creation. *IEEE Transactions on Visualization and Computer Graphics*, 2024.
[23] Y. Hao, Z. Chi, L. Dong, and F. Wei. Optimizing prompts for text-to-image generation. In *Advances in Neural Information Processing Systems*, vol. 36, pp. 66923–66939, 2023.
[24] D. Holz. Midjourney. https://www.midjourney.com, 2023.
[25] Y.-G. Hong, J.-Y. Jeong, J.-W. Jeong, and E. Lukianova. Spot the difference: Ai, please make this for me! In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.
[26] M. N. Hoque, S. Shin, and N. Elmqvist. Harder, better, faster, stronger: Interactive visualization for human-centered ai tools. https://arxiv.org/abs/2404.02147, 2024.
[27] J. Hyung, J. Shin, and J. Choo. Magicapture: High-resolution multi-concept portrait customization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38:2445–2453, 2024. doi: 10.1609/aaai.v38i3.28020
[28] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 43(12):4217–4228, 2021.
[29] J. Kim, H. Lee, D. M. Nguyen, M. Shin, B. C. Kwon, S. Ko, and N. Elmqvist. Dg comics: Semi-automatically authoring graph comics for dynamic graphs. https://arxiv.org/abs/2408.04874, 2024.
[30] H.-K. Ko, S. An, G. Park, S. K. Kim, D. Kim, B. Kim, J. Jo, and J. Seo. We-toon: A communication support system between writers and artists in collaborative webtoon sketch revision. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022.
[31] H.-K. Ko, G. Park, H. Jeon, J. Jo, J. Kim, and J. Seo. Large-scale text-to-image generation models for visual artists' creative works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, p. 919–933. Association for Computing Machinery, 2023. doi: 10.1145/3581641.3584078
[32] J. Liao, X. Chen, Q. Fu, L. Du, X. He, X. Wang, S. Han, and D. Zhang. Text-to-image generation for abstract concepts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.
[33] V. Liu and L. B. Chilton. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022.
[34] V. Liu, H. Qiao, and L. Chilton. Opal: Multimodal image generation for news illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022.
[35] L. McInnes, J. Healy, and J. Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints*, pp. 1–63, 2018.
[36] R. Metcalfe. Openvino™ ai plugins for gimp. https://github.com/intel/openvino-ai-plugins-gimp, 2023.
[37] Mickmumpitz. A more perfect union. https://www.youtube.com/watch?v=4kBl3VNaht4, jan 2023.
[38] A. Mishra, U. Soni, A. Arunkumar, J. Huang, B. C. Kwon, and C. Bryan. Promptaid: Prompt exploration, perturbation, testing and iteration using visual analytics for large language models. *ArXiv*, abs/2304.01964, 2023.
[39] G. P. Nguyen and M. Worring. Interactive access to large image collections using similarity-based visualization. *Journal of Visual Languages & Computing*, 19(2):203–224, 2008.
[40] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic provenance: process+interaction+insight. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, p. 33–36. Association for Computing Machinery, 2011.
[41] OpenAi. Dalle3. https://openai.com/dall-e-3, 2024.
[42] D. Orban, D. F. Keefe, A. Biswas, J. Ahrens, and D. Rogers. Drag and track: A direct manipulation interface for contextualizing data

instances within a continuous parameter space. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):256–266, 2018.

[43] Pallets. Flask. https://flask.palletsprojects.com/en/2.3.x/, 2023.

[44] W. Qiu and G. Legrady. Fencing hallucination: An interactive installation for fencing with ai and synthesizing chronophotographs. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.

[45] E. D. Ragan, A. Endert, J. Sanyal, and J. Chen. Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):31–40, 2015.

[46] M. Ruskov. Grimm in wonderland: Prompt engineering with midjourney to illustrate fairytales. *arXiv preprint arXiv:2302.08961*, 2023.

[47] C. Safka. Img2vec. https://github.com/christiansafka/img2vec, 2024.

[48] V. Schetinger, S. Di Bartolomeo, M. El-Assady, A. McNutt, M. Miller, J. P. A. Passos, and J. L. Adams. Doom or deliciousness: Challenges and opportunities for visualization in the age of generative models. *Computer Graphics Forum*, 42(3):423–435, 2023.

[49] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller. Visual parameter space analysis: A conceptual framework. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2161–2170, 2014.

[50] Y. Shen, Y. Shen, J. Cheng, C. Jiang, M. Fan, and Z. Wang. Neural canvas: Supporting scenic design prototyping by integrating 3d sketching and generative ai. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2024.

[51] Y. Shi, T. Gao, X. Jiao, and N. Cao. Understanding design collaboration between designers and artificial intelligence: A systematic literature review. *Proc. ACM Hum.-Comput. Interact.*, 7, 2023. doi: 10.1145/3610217

[52] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In B. B. Bederson and B. Shneiderman, eds., *The Craft of Information Visualization*, Interactive Technologies, pp. 364–371. Morgan Kaufmann, San Francisco, 2003.

[53] H. Strobelt, A. Webson, V. Sanh, B. Hoover, J. Beyer, H. Pfister, and A. M. Rush. Interactive and visual prompt engineering for ad-hoc task adaptation with large language models. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1146–1156, 2023.

[54] Y. Sun, Y. Xu, C. Cheng, Y. Li, C. H. Lee, and A. Asadipour. Explore the future earth with wander 2.0: Ai chatbot driven by knowledge-base story generation and text-to-image model. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.

[55] K. Team. Krita — digital painting. creative freedom. https://krita.org/, 2023.

[56] M. P. Verheijden and M. Funk. Collaborative diffusion: Boosting designerly co-creation with generative ai. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2023.

[57] Y. Wang, S. Shen, and B. Y. Lim. Reprompt: Automatic prompt editing to refine ai-generative art towards precise expressions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.

[58] S. Xie, Z. Zhang, Z. Lin, T. Hinz, and K. Zhang. Smartbrush: Text and shape guided object inpainting with diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22428–22437, 2023.

[59] X. Xie, X. Cai, J. Zhou, N. Cao, and Y. Wu. A semantic-based method for visualizing large image collections. *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2362–2377, 2018.

[60] K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch. Survey on the analysis of user interactions and visualization provenance. In *Computer Graphics Forum*, vol. 39, pp. 757–783, 2020.

[61] C. Yan, J. J. Y. Chung, Y. Kiheon, Y. Gingold, E. Adar, and S. R. Hong. Flatmagic: Improving flat colorization through ai-driven design for digital comic professionals. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2022.

[62] J. Yuan, C. Chen, W. Yang, M. Liu, J. Xia, and S. Liu. A survey of visual analytics techniques for machine learning. *Computational Visual Media*, 7:3–36, 2021. doi: 10.1007/s41095-020-0191-7

[63] X. Zhang, H. Kuang, X. Mou, H. Lyu, K. Wu, S. Chen, J. Luo, X. Huang, and Z. Wei. SoMeLVLM: A large vision language model for social media processing. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2366–2389. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.140