ThematicPlane: Bridging Tacit User Intent and Latent Spaces for Image Generation

Daniel Lee
Adobe Inc.
San Jose, CA, USA
dlee1@adobe.com

Nikhil Sharma
Johns Hopkins University
Baltimore, MD, USA
nsharm27@jhu.edu

Donghoon Shin
University of Washington
Seattle, WA, USA
dhoon@uw.edu

DaEun Choi
KAIST
Daejeon, Republic of Korea
daeun.choi@kaist.ac.kr

Harsh Sharma
University of Colorado
Boulder, CO, USA
harsh.sharma@colorado.edu

Jeonghwan Kim
University of Illinois at
Urbana-Champaign
Champaign, IL, USA
jk100@illinois.edu

Heng Ji
University of Illinois at
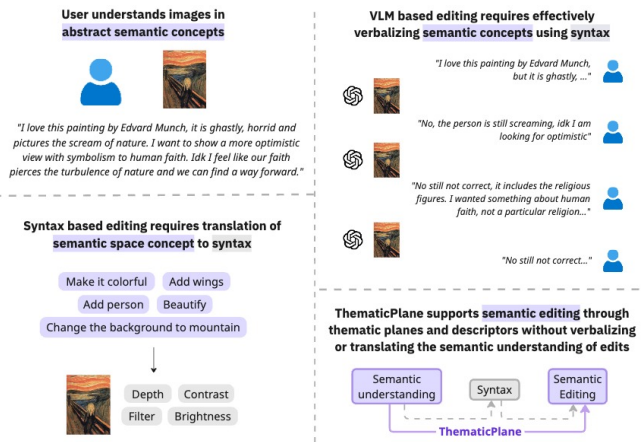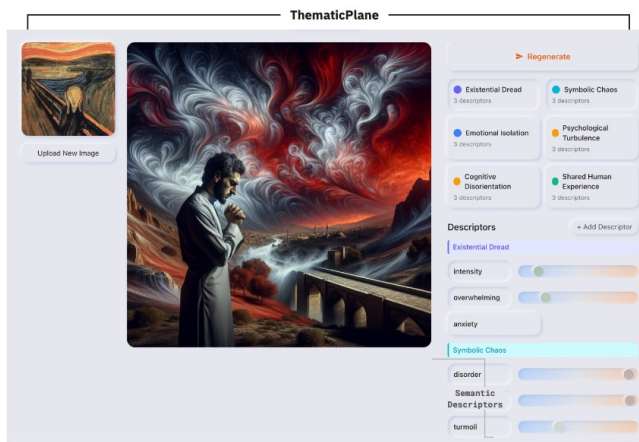Urbana-Champaign
Champaign, IL, USA
hengji@illinois.edu

Figure 1: Overview of THEMATICPLANE. Unlike conventional prompting, THEMATICPLANE supports the navigation and manipulation of high-level semantic concepts within a thematic plan, allowing for tacit editing of the image aligned with the user's understanding of the semantic concept of the image.

Abstract

Generative AI has made image creation more accessible, yet aligning outputs with nuanced creative intent remains challenging, particularly for non-experts. Existing tools often require users to externalize ideas through prompts or references, limiting fluid exploration. We introduce THEMATICPLANE, a system that enables users to navigate and manipulate high-level semantic concepts (*e.g.*, mood, style, or narrative tone) within an interactive thematic design plane. This interface bridges the gap between tacit creative intent and system control. In our exploratory study ($N = 6$), participants engaged in divergent and convergent creative modes, often embracing unexpected results as inspiration or iteration cues. While they grounded their exploration in familiar themes, differing expectations of how themes mapped to outputs revealed a need for more explainable controls. Overall, THEMATICPLANE fosters expressive, iterative workflows and highlights new directions for intuitive, semantics-driven interaction in generative design tools.

Keywords

creativity support tool, visual exploration, generative AI

1 Introduction

The proliferation of generative AI has fundamentally transformed creative expression in various modalities (*e.g.*, text, images, video, code, etc.), enabling users to synthesize artifacts with unprecedented ease (*e.g.*, via tools like ChatGPT or RunwayML). For instance, individuals without expertise in visual editing software

can now generate or modify images through simple prompts, enabling lay users to construct sophisticated outputs [16]. Nonetheless, achieving precise, intent-aligned results still demands technical knowledge (*e.g.*, manually crafting detailed prompts, understanding the underlying model behaviors), hindering non-expert users [20].

Previous works have proposed systems where users can change low-level image parameters (*e.g.* with augmented prompt engineering scaffoldings [9, 24]). However, users seldom conceptualize desired changes in terms of such low-level technical parameters, layers, or explicit instructions. Instead, they reason in high-level semantic terms, like aspiring to make an image 'warmer,' 'more dramatic,' or 'friendlier' [4, 7]. This semantic gap as visualized in Figure 2–between users' tacit, thematic intent and systems' affordances–has spurred recent HCI research into intuitive editing paradigms, including keyword recombinations [5], reference image-based modifications [21], and interactive latent space explorations [1, 17].

Despite these, existing tools predominantly require users to externalize their vision through prescribed mediums, such as textual descriptions or reference images. This 'externalization assumption' presumes users can readily translate abstract ideas into concrete formats, which may not align with natural cognitive processes of creative ideation and refinement [6, 7, 23]. As a result, workflows can become rigid, limiting iterative discovery in generative tasks.

To address this, we introduce THEMATICPLANE, a system that supports interactions within thematic design planes, which scaffolds the representation of multidimensional space for exploring and lets users manipulate thematic attributes of visual content. It enables users to 'ground' and navigate high-level semantic concepts, such as mood, style, narrative, or emotional tone, in an exploratory fashion. This plane serves as a scaffolding layer, connecting users' tacit intent with system capabilities with a 'plane of existence' for themes, drawing on semantic interaction paradigms to support creativity without externalization barriers.
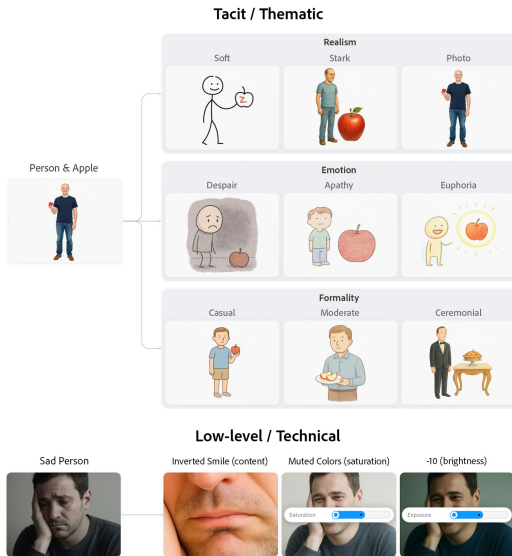

Figure 2: Visualization of tacit/thematic and low-level/technical edits in images.

1.1 System Prototype

THEMATICPLANE implements the thematic design plane as an interactive framework adapted from the Circumplex Model [19] for semantic image editing, enabling users to explore and manipulate high-level thematic attributes without explicit low-level inputs. Given an input image by the user, GPT-4o extracts keywords, which are then refined by removing object-based descriptors and retaining only thematic elements like mood or style. Next, we generate 12 thematic perturbations for each theme, assigning them to left or right directions along a semantic axis. Embeddings are computed using DINOv2 [18] for the input image and its aligned text encoder [15] for textual themes, ranked by cosine similarity. Ranked descriptors are then injected into prompts for the image generation model (*i.e.*, Imagen 3 [3]) to produce transformed images. After each regeneration, users can choose to make the new image the primary reference to continue editing.

2 Exploratory Study

We recruited 6 participants (4 women, 2 men), aged 19–31, through a corporate Slack channel. Participants had prior experience with image generation tools (*e.g.*, Midjourney [13]). They were first provided 10 minutes to explore the system, using their own images or study provided examples. We then designed three image transformation tasks requiring participants to modify source images to match target images. Each participant completed tasks across three conditions: (1) baseline ChatGPT constrained to image generation only, (2) ChatGPT with full features, and (3) our system—with 5 minutes per condition, followed by a semi-structured interview and survey. Each participant received $20 for their participation.

3 Discussion and Future Work

Aligning interfaces to the human creativity process. Our study revealed two emergent behaviors—open-ended discovery (*i.e.*, divergent exploration) and goal-based editing (*i.e.*, convergent refinement) [10, 25]. Although participants using THEMATICPLANE did not always anticipate the resulting images ($M = 3.43$, $SD = 1.27$), they reported high satisfaction with the outcomes ($M = 5.86$, $SD = 0.90$). Participants noted that non-linear generations either spurred intentional shifts in design direction during divergent tasks (P3) or enabled iteration in a space where themes amalgamated into cohesive affordances not characterized by a single plane (P6).

Different perceptions of thematic planes. We observed that participants grounded themselves in themes based on their own understanding, enabling them to interpolate and navigate the rest of the thematic design plane; however, such understanding varied across participants, leading to differing expectations during exploration. Specifically, some participants (P1, P3, P7) anticipated linear mappings between themes and outputs, while others (P2, P4, P5) encountered unclear connections between the plane and initial themes, prompting a desire for explainable methods to comprehend steering mechanisms, rather than merely executing them [2, 12].

Interactive visualizations of themes. Despite the steerability it offers, our interaction method currently provides a single axis for users to steer generative outputs based on an initial thematic attribute, which may restrict the exploration of interconnected

semantic dimensions [12]. Future work could advance this to multidimensional representations, such as interactive planes or graphs, enabling users to simultaneously manipulate multiple attributes (*e.g.*, mood and style) through gestures like dragging or clustering for more nuanced control over latent space trajectories [22].

Steerable model generations for images. Building on the visual representation of steerable dimensions, these should reflect the technical methods for navigating these spaces, extending text-based approaches like LM-Steer [11] to incorporate vector-based steering in multimodal generative systems [8, 14] to create hybrid navigation mechanisms that dynamically adjust trajectories across semantic attributes during inference.

Acknowledgments

References

[1] Krzysztof Adamkiewicz, Paweł W Woźniak, Julia Dominiak, Andrzej Romanowski, Jakob Karolus, and Stanislav Frolov. 2025. PromptMap: An Alternative Interaction Style for AI-Based Image Generation. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 1162–1176. doi:10.1145/3708359.3712150

[2] Goonmeet Bajaj, Valerie L Shalin, Srinivasan Parthasarathy, and Amit Sheth. 2024. Grounding from an AI and cognitive science lens. *IEEE Intelligent Systems* 39, 2 (2024), 66–71. doi:10.1109/mis.2024.3366669

[3] Jason Baldridge, Jakob Bauer, Mukul Bhutani, Nicole Brichtova, Andrew Bunner, Lluis Castrejon, Kelvin Chan, Yichang Chen, Sander Dieleman, Yuqing Du, et al. 2024. Imagen 3. *arXiv preprint arXiv:2408.07009* (2024). https://arxiv.org/abs/2408.07009

[4] Yali Bian and Chris North. 2021. Deepsi: Interactive deep learning for semantic interaction. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 197–207. https://doi.org/10.1145/3397481.3450670

[5] DaEun Choi, Sumin Hong, Jeongeon Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–25. doi:10.1145/3613904.3642794

[6] DaEun Choi, Kihoon Son, HyunJoon Jung, and Juho Kim. 2025. Expandora: Broadening Design Exploration with Text-to-Image Model. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–10. doi:10.1145/3706599.3720189

[7] Alex Endert. 2014. Semantic Interaction for Visual Analytics: Toward Coupling Cognition and Computation. *IEEE Computer Graphics and Applications* 34, 4 (2014), 8–15. doi:10.1109/MCG.2014.73

[8] Patrick Esser, Robin Rombach, and Bjorn Ommer. 2021. Taming Transformers for High-Resolution Image Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12873–12883. doi:10.1109/CVPR46437.2021.01268

[9] Rohit Gandikota, Joanna Materzyńska, Tingrui Zhou, Antonio Torralba, and David Bau. 2024. Concept Sliders: LoRA Adaptors for Precise Control in Diffusion Models. In *European Conference on Computer Vision*. Springer, 172–188. doi:10.1007/978-3-031-73661-2_10

[10] J.P. Guilford. 1959. Three Faces of Intellect. *American Psychologist* 14, 8 (1959), 469–479. doi:10.1037/h0046827

[11] Chi Han, Jialiang Xu, Manling Li, Yi Fung, Chenkai Sun, Nan Jiang, Tarek Abdelzaher, and Heng Ji. 2024. Word Embeddings Are Steers for Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. 16410–16430. doi:10.18653/v1/2024.acl-long.864

[12] Yuanning Han, Ziyi Qiu, Jiale Cheng, and Ray Lc. 2024. When Teams Embrace AI: Human Collaboration Strategies in Generative Prompting in a Creative Design Task. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–14. doi:10.1145/3613904.3642133

[13] Midjourney Inc. 2025. Midjourney. https://www.midjourney.com/.

[14] Ali Jahanian, Lucy Chai, and Phillip Isola. 2019. On the "steerability" of generative adversarial networks. *arXiv preprint arXiv:1907.07171* (2019). https://arxiv.org/abs/1907.07171

[15] Cijo Jose, Théo Moutakanni, Dahyun Kang, Federico Baldassarre, Timothée Darcet, Hu Xu, Daniel Li, Marc Szafraniec, Michaël Ramamonjisoa, Maxime Oquab, et al. 2025. DINOv2 Meets Text: A Unified Framework for Image- and Pixel-Level Vision-Language Alignment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 24905–24916.

[16] Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. 2022. Locating and Editing Factual Associations in GPT. *Advances in Neural Information Processing Systems* 35 (2022), 17359–17372.

[17] Jonas Oppenlaender, Rhema Linder, and Johanna Silvennoinen. 2024. Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering. *International Journal of Human–Computer Interaction* (2024), 1–23. doi:10.1080/10447318.2024.2431761

[18] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. 2023. DINOv2: Learning Robust Visual Features without Supervision. *arXiv preprint arXiv:2304.07193* (2023). https://arxiv.org/abs/2304.07193

[19] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[20] Jingyu Shi, Rahul Jain, Hyungjun Doh, Ryo Suzuki, and Karthik Ramani. 2023. An HCI-Centric Survey and Taxonomy of Human-Generative-AI Interactions. *arXiv preprint arXiv:2310.07127* (2023). https://arxiv.org/abs/2310.07127

[21] Kihoon Son, DaEun Choi, Tae Soo Kim, Young-Ho Kim, and Juho Kim. 2024. GenQuery: Supporting Expressive Visual Search with Generative Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–19. doi:10.1145/3613904.3642847

[22] Sangho Suh, Meng Chen, Bryan Min, Toby Jia-Jun Li, and Haijun Xia. 2024. Luminate: Structured Generation and Exploration of Design Space with Large Language Models for Human-AI Co-Creation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–26. doi:10.1145/3613904.3642400

[23] Michael Terry, Elizabeth D Mynatt, Kumiyo Nakakoji, and Yasuhiro Yamamoto. 2004. Variation in Element and Action: Supporting Simultaneous Development of Alternative Solutions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 711–718. doi:10.1145/985692.985782

[24] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21. doi:10.1145/3613904.3642803

[25] Robin H Willemsen, Isabelle C de Vink, Evelyn H Kroesbergen, and Ard W Lazonder. 2023. The role of creative thinking in children's scientific reasoning. *Thinking Skills and Creativity* 49 (2023), 101375. doi:10.1016/j.tsc.2023.101375