CICADA: Interface for Concept Sketches Using CLIP

Tomas Lawton
The University of Sydney
Sydney, Australia
tlaw8603@uni.sydney.edu.au

Francisco Ibarrola
The University of Sydney
Sydney, Australia
francisco.ibarrola@sydney.edu.au

Kazjon Grace
The University of Sydney
Sydney, Australia
kazjon.grace@sydney.edu.au

1 Introduction

From Stable Diffusion to DALL·E2, state of the art models for high-resolution text-to-image generation seem to arrive nearly every week [1, 2], along with the promise to cause significant disruption in the creative industries. However, professional designers – from illustrators to architects to engineers – use low-fidelity representations like sketches to refine their understanding of the problem, rather than for developing completed solutions [3, 4]. Conceptual stages of design have been operationalised as the co-evolution of problem and solution "spaces" [5]. We introduce the Collaborative, Interactive, Context-Aware Design Agent (CICADA) [6], which uses CLIP-guided [7] synthesis-by-optimisation to support conceptual designing. Building on previous approaches [8] we optimize a set of Bézier curves to match a given text prompt. In CICADA, users sketch collaboratively with the system in real-time. Users maintain editorial control, although additions to both the optimiser and interaction model enable designers and CICADA to influence one another by engaging with the sketch. CICADA provides an instrument to explore how text-to-image generative systems can assist designers, so we conducted a qualitative user study to explore its impact on designing. Figure 1 shows CICADA's interface and several sample sketches.

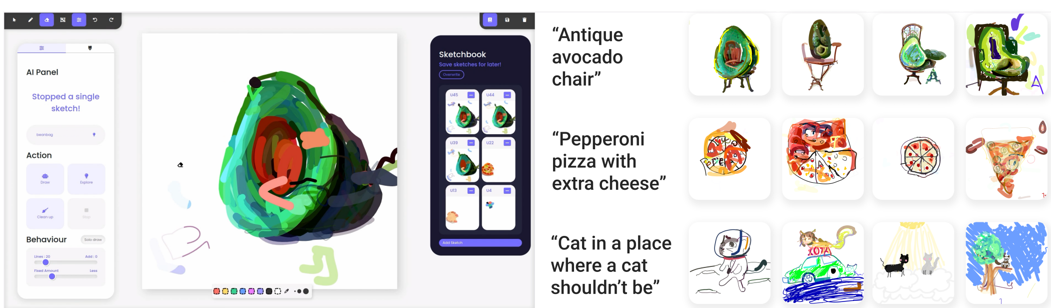

Figure 1: The CICADA drawing interface (left) and several human-CICADA collaborative sketches with the task description provided to users (right).

2 Approach and Findings

We build on the loss function from [8] by adding penalisation terms that encourage fidelity with user-drawn geometry. A new initialisation method is integrated into the model, along with pruning (based on a greedy search for strokes with minimal contribution to the loss) [6]. The resulting CICADA model permits the creation of sketches aligned to both the text prompt and the user's sketch. We implement a client-server architecture with FastAPI and WebSockets: user additions become model inputs, and optimiser outputs become sketch updates (drawn on HTML canvas as vectors using Paper.js). Changes to the prompt, the vector sketch, and the model parameters can all be made smoothly, so CICADA provides a fluid co-drawing experience with the user. Since we cannot

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

backpropagate across user changes, we re-initialise the optimiser with every user event. Multiple CICADA instances can be previewed in the "explorer", suggesting potential design directions. The pruning algorithm is available for “cleaning up” the sketch. A history slider appears when CICADA is inactive to explore model iterations. Assessing the potential directions for designing multi-modal interactions with CICADA, we conducted a qualitative study of user perceptions, expectations and reflections with the system. 12 students at an internationally-recognised school of design completed three tasks with CICADA and were interviewed. We conducted a thematic analysis [9, 10], organising quotes into codes, sub-themes and themes, two of which are described below.

Theme 1: Dialogic Interaction Loop. As illustrated by Figure 2, we found users entered into a feedback loop, where constant interaction with CICADA promoted collaboration. Consistent with sketch studies, traditional visual principles (Gestalts) and abstract forms were adopted to create meaning by “playing around” with the sketch [11]. CICADA changes resulted in unique stroke combinations, so users reported the system building off their ideas through bisociation [12]. As interaction continued, CICADA provided substantial inspiration from unexpected ideas, which challenged users to rethink the sketch. Users therefore strongly valued how CICADA’s “perspective”, and “prompting” provided guidance and a form of creative critique. As their confidence increased, users also noticed their thinking was stimulated because reinterpreting CICADA changes could transform their internal design concepts. This sometimes lead to improvements in the sketch, but more often resulted in reframing conceptual shifts, which expanded the scope of design possibilities [13]. CICADA therefore seems to have the capacity to use relatively simple physical changes to negotiate significant conceptual developments. This may also indicate the interaction loop supported a collaborative-flow experience that benefited designing [14], although users reported that flow being disrupted by the system. Various potential applications were suggested: sparking creativity, increasing design speed, guiding novice designers, and creating original artefacts. Across many participants, CICADA was seen as especially promising for the early ideation-focussed phases of design, as well as for avoiding creative blocks.

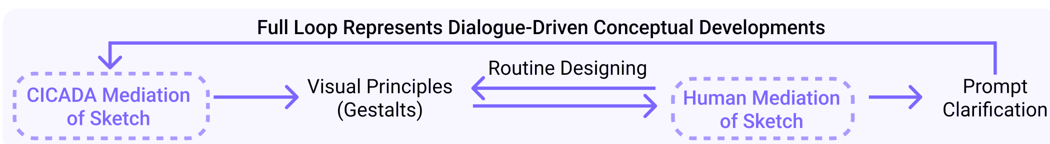

Figure 2: Dialogic Interaction Loop instigating conceptual development through joint mediation.

Theme 2: Strategic Recovery. CICADA has a distinctive drawing style that diverged from user expectations, so “creative differences” disrupted the “dialogic interaction loop.” Users struggled most with the number of strokes, CICADA making “child-like” modifications to their *good* strokes, and with difficulty refining strokes into sketch components. CICADA’s abstraction threshold sometimes assisted creative thinking, but occasionally disrupted the interaction loop when users couldn’t interpret the sketch. When this happened, it was unclear if CICADA could understand what the user had drawn, or if changing the prompt would alter the drawing process. Users therefore felt confused, and frustrated when CICADA was too difficult to control. In response, users felt the need to adapt to CICADA, so they experimented with strategies to get collaboration back on track. Avoiding loss of control by using features like Stop, Undo and Rewind provided the least improvement, as it led to fighting for control over the sketch. By contrast, when users were able to follow CICADA moves, they reported the most fun and experienced beneficial conceptual tangents. Some users focused on drawing individual elements of the composition to gain finer control, for which we could provide improved UI support. Addressing issues related to latent feedback and ambiguous system status could also alleviate some of the confusion users experienced.

3 Conclusion

CICADA users were highly engaged and had fun overall, despite some frustration. Our study indicated CICADA sketches strongly encouraged designers to investigate novel design directions, and interaction flow and drawing control were important design considerations. A quantitative study focusing on these aspects is underway, although CICADA evaluation indicates dynamic inspiration systems could use text-image generative models to support conceptual designing.

4 Ethical Considerations

Text-to-image generation models could exacerbate job disruption in the near future, since reducing the difficulty and cost of creative work creates an economic incentive to reduce creative employment. This may already be happening since platforms like Canva are lowering the skill barriers to do work that previously required extensive training or effort. As a system designed to augment rather than replace human designers, CICADA is less likely to contribute to this effect directly, although it should be acknowledged that boosting designer productivity may have negative impacts on the labour market in the creative industries.

References

[1] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8821–8831.

[2] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 10 684–10 695.

[3] I. M. Verstijnen, C. van Leeuwen, G. Goldschmidt, R. Hamel, and J. Hennessey, "Sketching and creative discovery," *Design studies*, vol. 19, no. 4, pp. 519–546, 1998.

[4] D. A. Schon and G. Wiggins, "Kinds of seeing and their functions in designing," *Design studies*, vol. 13, no. 2, pp. 135–156, 1992.

[5] J. Poon and M. L. Maher, "Co-evolution and emergence in design," *Artificial Intelligence in Engineering*, vol. 11, no. 3, pp. 319–327, 1997.

[6] F. Ibarrola, T. Lawton, and K. Grace, "A collaborative, interactive and context-aware drawing agent for co-creative design," *arXiv preprint arXiv:submit/4511059*, 2022.

[7] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*. PMLR, 2021, pp. 8748–8763.

[8] K. Frans, L. B. Soros, and O. Witkowski, "Clipdraw: Exploring text-to-drawing synthesis through language-image encoders," *arXiv preprint arXiv:2106.14843*, 2021.

[9] V. Braun and V. Clarke, "What can "thematic analysis" offer health and wellbeing researchers?" p. 26152, 2014.

[10] V. Clarke, V. Braun, and N. Hayfield, "Thematic analysis," *Qualitative psychology: A practical guide to research methods*, vol. 222, no. 2015, p. 248, 2015.

[11] B. Tversky, "What do sketches say about thinking," in *2002 AAAI Spring Symposium, Sketch Understanding Workshop, Stanford University, AAAI Technical Report SS-02-08*, vol. 148, 2002, p. 151.

[12] A. Koestler, "The act of creation," in *Brain Function, Volume IV: Brain Function and Learning*. University of California Press, 2020, pp. 327–346.

[13] M. A. Boden, "Computer models of creativity," *AI Magazine*, vol. 30, no. 3, pp. 23–23, 2009.

[14] M. Csikszentmihalyi, S. Abuhamdeh, and J. Nakamura, *Flow*. Springer Netherlands, 2014, pp. 227–238.