Differentiable Quality-Diversity for Co-Creative Sketching AI

Francisco Ibarrola and **Kazjon Grace**
School of Architecture, Design and Planning
The University of Sydney
Sydney, Australia
[francisco.ibarrola,kazjon.grace]@sydney.edu.au

Abstract

Co-creative artificial intelligence systems go beyond one-click generative AI solutions and enable users to participate in the generative process. A key component of co-creative interfaces is the ability to suggest multiple options to the user, to avoid constraining the process and help overcome creative blocks. We explore this diversification problem in a vector drawing synthesis-by-optimisation setting and propose algorithms for generating diversity among user-defined characteristics. Experimental results show improvement in terms of behaviour coverage and image diversity.

Introduction

Co-creative systems built on generative AI models expand upon the latter's increasingly impressive expressive capacity by offering their users additional control and creative agency. This additional agency is critical in the early stages of creative tasks across a wide variety of domains. In creative professions such as art, interface design, engineering, and architecture, sketching is an important aspect of this early conceptual exploration (Goldschmidt 1991; Gero 1998). At the outset of the creative process, artists and designers typically lack a clear idea of what they are looking for, or accept that their current ideas may change entirely as the process progresses. In the context of design, those who approach creative tasks without this level of ideational flexibility often fail to achieve their goals (Dorst 2015), while those who embrace flexbility have been shown to produce more-creative output (Suwa, Gero, and Purcell 2000). This is because designing, especially in its early stages, is not a mere process of synthesising given requirements, but rather an iterative process of discovering and refining both those requirements and how they might be fulfilled. This has been described as a co-evolution: a concurrent emergence of both the problem (the requirements) and its solution (the design) (Poon and Maher 1997). Current research in co-creative systems is exploring how this co-evolution can be supported by AI tools (Lawton et al. 2023; Gero, Liu, and Chilton 2022; Williford et al. 2023).

While the notion of "diversity" depends on context, this paper operationalises co-creative diversity with reference to some observable characteristics quantifiable by numerical values, that can be chosen depending on the context. The values of any given output can then be thought of as a multidimensional point in what is usually called a behaviour space. Subsequently, the breadth of the distribution of a set of points in the behaviour space provides a domain-appropriate measure of diversity. This is related to the notion of a generator's expressive range (Smith and Whitehead 2010), although there user-chosen characteristics are used to assess a generator's diversity in a space, while the term "behaviour space" derives from the quality-diversity (QD) literature, where generators actively seek to cover their behaviour spaces (Pugh, Soros, and Stanley 2016).

In this paper we introduce this QD approach to the CICADA model (Ibarrola, Lawton, and Grace 2022), a drawing agent designed to work cooperatively with a human designer. Previous analysis of user experience with a CICADA-based co-creative system suggests that the capacity to select from between different drawing options would be of value to users (Grace, Lawton, and Ibarrola 2023). Given that CICADA consists of an end-to-end differentiable generation-by-optimisation process, we build on previous approaches for enforcing diversity in a differentiable setting, such as OMG-MEGA (Fontaine and Nikolaidis 2021). This algorithm stochastically explores behaviour space, but we show it to be ill-suited for our context on account of the properties of CICADA's parameter space, and propose an alternative for this kind of settings. In this paper we describe OMG-MEGA along with its shortcomings in our context and propose two better-suited variants.

Method

The general setting of our approach consists of a parameter space Y (or the genome space in evolutionary computation parlance), associated to a differentiable generative model $h : Y \to X$. Additionally, two differentiable functions $f : X \to \mathbb{R}$ and $g : X \to B \subset \mathbb{R}^N$ map the elements of X into an objective score (or fitness) in $\mathbb{R}$ and behaviour dimensions $b \in B$, associated to some characteristics of $x \in X$ in which we want diversity.

We pursue the issue of finding diverse alternatives in a generative process as an exploration of the behaviour space B. That is, we want to generate a set of solutions that "behave" differently in terms of the outputs of g. We explore this problem in the setting of co-creative design using CICADA (Ibarrola, Lawton, and Grace 2022), where the pa-

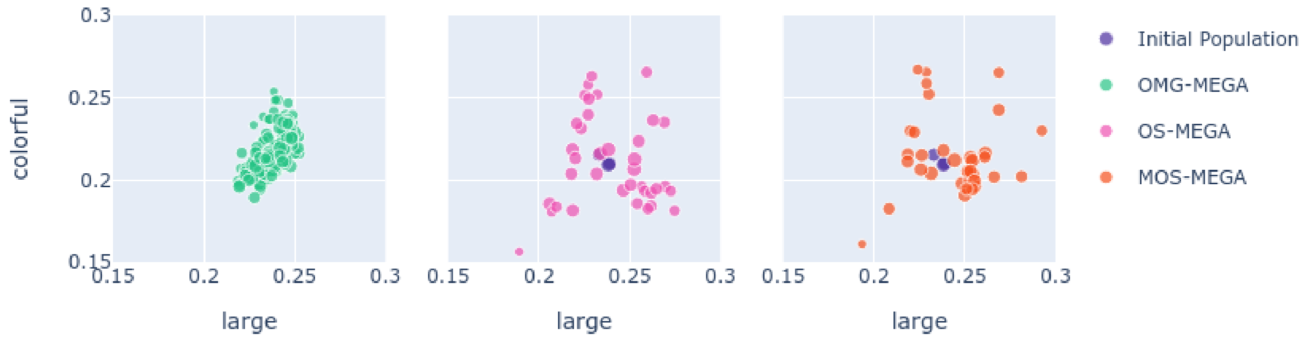

Figure 1: Distribution of populations generated with different algorithms in the (colorful, large) behaviour space.

rameter space $Y \subset \mathbb{R}^M$ is the set of arrays of parameters that determine the width, color and spatial locations of the set of Bézier curves that constitute a sketch or drawing. Also, h is a differentiable rasterizer (Li et al. 2020) that generates an RGB image x from an array of parameters y. Additionally, we define the objective function

$$f(x;t) \doteq \langle c_{img}(x), c_{txt}(t)\rangle,$$

as the normalised inner product between the CLIP (Radford et al. 2021) encoded latents of the image x and a text prompt t, provided by the user as the drawing's description.

In a similar fashion, we can define behaviours in terms of CLIP losses by addressing how well x matches additional characteristics. In our experiments below, we define

$$g(x) \doteq (\,\langle c_{img}(x), c_{txt}(\mathit{large})\rangle\,,\,\langle c_{img}(x), c_{txt}(\mathit{colorful})\rangle\,)$$

to set CLIP interpretations of size and chromatic variance as the behaviour dimensions.

We shall then consider the problem of producing a population of CICADA-drawn images $\{x_1, \ldots, x_K\} \subset X$ such that the elements x_k are maximally behaviourally diverse (i.e. in this case small-to-large and colourless-to-colourful).

OMG-MEGA

A recently developed and end-to-end differentiable approach to the quality-diversity problem is the Objective and Measure Gradient MAP-Elites via Gradient Arborescence (OMG-MEGA) algorithm (Fontaine and Nikolaidis 2021) ('measure' is the authors' term for what we call a behaviour). In brief, this consists of iteratively picking an existing element x_k from an existing population, generating a new individual by modifying y_k according to

$$y_{K+1} = y_k + |\alpha_0|\nabla_y f(y_k) + \sum_{n=1}^{2} \alpha_n \nabla_y g \circ h(y_k),$$

where $\alpha_n \sim \mathcal{N}(0, \sigma I), \forall n = 0, \ldots, N$. In other words, by using gradient descent over both the objective and the behaviours, but with random weights (positive-only in the case of the objective).

While this model is really good for thoroughly exploring the behaviour space, it is not very well suited for the CICADA problem, where drawings typically spend hundreds of iterations progressing towards recognisable shapes. We contend that for synthesis-by-optimisation tasks such as ours OMG-MEGA does not allow for significant enough per-iteration changes to each individual before randomly varying the objective and behaviour weightings, which jeopardises the algorithm's capacity to converge on images that are recognisable as a representation of the prompt.

OS-MEGA

To address the lack of convergence in OMG-MEGA on CICADA tasks, we propose to optimise for longer between selecting new coefficients. We also wish to avoid duplicating work by re-searching areas that have already been well-traversed, so we additionally want to enforce the random coefficients to be biased towards directions that purposefully lead them away from explored regions of behaviour space.

We start by picking an element at random from the current population, and build a new individual by moving away from the population centroid in the direction of least variance. That is, let $\{b_1, \ldots, b_K\}$ be the set of two-dimensional behaviour scores of the population, such that $g(x_k) = b_k \in \mathbb{R}^2$, and let $\bar{b}$ and C be the associated empirical mean and covariance matrix, respectively. Also, let v be the eigenvector of C with the smallest eigenvalue, i.e. the direction the population is least diverse in. Then, we can build a new individual by starting with $y = y_k$ and iteratively running

$$y' = y + \nabla_y f(y) - \lambda \nabla_y \|g \circ h(y) - b_k - \sigma v\|^2,$$

where $\lambda > 0$ and σ are weighting parameters, and the sign of σ is the sign of $\langle v, b_k - \bar{b}\rangle$, meaning the optimisation process is directed "outwards" from the explored area along the direction that has been least explored thus far. We refer as this algorithm as Outbound Scattering MEGA or OS-MEGA.

MOS-MEGA

In addition to directing the search in the behaviour space "outwards", the characteristics of our problem space suggest we may be able to improve on the diversity of the search by directly introducing noise through smart "mutation" strategies. OMG-MEGA replaces the traditional "mutation" genetic operator designed to introduce genetic diversity with the random coefficients on the objective and behaviours, which have the effect of adding noise to the search. However, the resulting changes are local and small-scale, and in our vector image context larger changes may be more effective. We take advantage of our vector representation to modify paths that are not significantly contributing to the objective, replacing the least-contributing with new, random traces. This operation is not differentiable, so we perform it after choosing an individual from the population but before conducting gradient descent.

Let us assume we have chosen an individual k, and let $P = \{p_1, \ldots, p_J\}$ be a partition of the set of parameters y_k, such that every p_j contains the parameters of a single trace. Then, we can compute a set of "irrelevance scores" $\{s_1, \ldots, s_J\}$ where s_j is the objective score of the image generated from y_k after subtracting (i.e. not drawing) the j-th trace. A low value of s_j means that discarding the j-th trace undermines the quality of the drawing. Consequently, we may improve diversity by adding Gaussian white noise with variance s_j to every p_j, obtaining a drawing which maintains the relevant traces of the original but differs in those that do not significantly contribute to the objective.

From here on, we can proceed with the gradient descent iterations as in the OS-MEGA algorithm. The full process (which we call Mutated Outbound Scattering MEGA or MOS-MEGA) is outlined in Algorithm 1.

Algorithm 1 MOS-MEGA

Initialization
  Build starting population $\{x_1, \ldots, x_K\}$
  $B = [g(x_1), \ldots, g(x_K)]$
  for $t = 1, \ldots, T$
    Mutation Phase
    Choose a random $k \in \{1, \ldots, K\}$
    Compute the irrelevance scores $\{s_1, \ldots, s_J\}$
    for $j = 1, \ldots, J$
      $p_j \leftarrow p_j + \eta,\ \eta \sim \mathcal{N}(0, s_j^2)$
    $y = [p_1, \ldots, p_J]$
    Optimisation Phase
    $\bar{b} = \text{mean}(B)$
    $C = (B - \bar{b})^T (B - \bar{b})$
    $v =$ eigenvector with min eigenvalue of C
    $\sigma = \text{sign}(\langle v, b_k - \bar{b} \rangle)$
    for $i = 1, \ldots, I$
      $y \leftarrow y + \nabla_y f(y) - \lambda \nabla_y \|g \circ h(y) - b_k - \sigma v\|^2$
      $K \leftarrow K + 1$

Results

In order to test how the proposed algorithms work, we run a few examples using CICADA, starting from a partial sketch of "a red chair" and generating three random completions, whose behaviour scores are depicted in blue in Figure 1. From there, we run OMG-MEGA, OS-MEGA and MOS-MEGA. We run all algorithms for 1,000 seconds, to be able to fairly compare their performance, as they are intended to use in a co-creative setting, where time is a relevant factor.

The results are illustrated in Figure 1, where it can be seen that while OMG-MEGA produces more individuals, both of our proposed variants, OS-MEGA and MOS-MEGA cover a larger area of the behaviour space.

Figure 2 shows some (randomly chosen) examples of the actual images obtained, making it clear that the observed greater coverage in Fig. 1 translates to much more visible variance in the images. As previously stated, OMG-MEGA does not produce significant variations between CICADA images (at least not without prohibitive amounts of compute), whereas OS-MEGA significantly changes the characteristics of the drawings, and MOS-MEGA moreso again.

Visual comparison has a high degree of subjectivity, so we have made use of the Truncated Inception Entropy (TIE), as introduced in (Ibarrola, Lawton, and Grace 2022), to quantify the diversity of each of our resulting sets of drawings. TIE uses the same feature space as the well-known FID image quality measure, but assesses variance rather than comparing two sets of images. This is computed as

$$\text{TIE}(A; K) \doteq \frac{K}{2} \log(2\pi e) + \frac{1}{2} \sum_{k=1}^{K} \log \lambda_k, \tag{1}$$

where A is the population being evaluated, λ_k are the eigenvalues (in descending order) of their covariance matrix, after mapping with an inception network, and K is a truncation parameter. High values of this metric are associated with high population diversity.

Four completion tasks of sketches of common household items ("a chair", "a lamp", "a hat" and "a blue dress") were run starting with the same populations for the three algorithms, and the obtained TIE scores (using $K = 16$) are shown in Figure 3. The larger TIE values obtained with OS-MEGA, and larger still using MOS-MEGA corroborate the effects observed in Figure 2. Some examples of the obtained results can be seen on Figure 4, showcasing variety in size and chromaticity.

Conclusions

In this paper we have proposed new ways to explore behaviour space in the setting of co-creative drawing based on vector image optimisation. Experiments show that the proposed algorithms result in better coverage of behaviour space in the same CPU time, as well as greater diversity as attested by visual inspection and the TIE diversity metric. While our explorations have thus far focused only on the CICADA drawing context, we hope they may generalise to other quality-diversity contexts.

Future work will focus on how well the proposed algorithm works in a real co-creative setting, including both the time taken to usably generate different suggestions and the user ratings of their appropriateness and utility. Furthermore, additional studies are needed to explore the difficulties that may arise when the users are to define the behaviour space on their own (i.e. providing arbitrary prompts for both objective and behaviours).

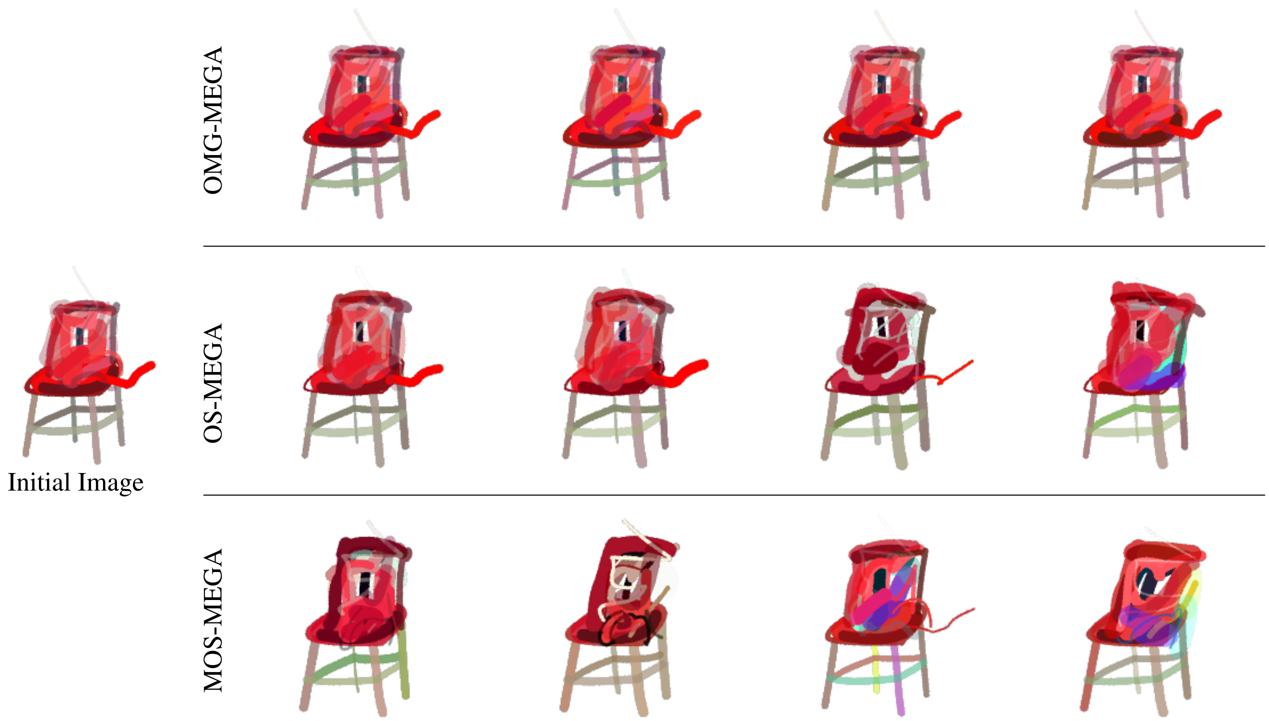

Figure 2: Each row shows four images randomly chosen from the population branched out from the initial image on the left, using one of the three algorithms.

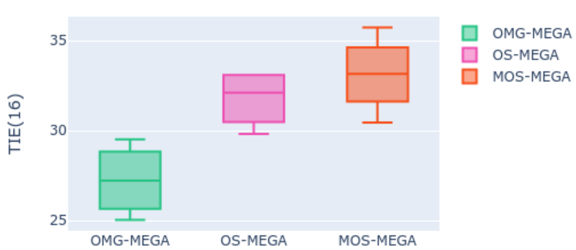

Figure 3: TIE values for the populations obtained with different algorithms. The experiments were carried using four different partial sketches.

Acknowledgments

We would like to acknowledge the Australian Research Council for funding this research (ID #DP200101059).

References

Dorst, K. 2015. *Frame innovation: Create new thinking by design*. MIT press.

Fontaine, M., and Nikolaidis, S. 2021. Differentiable quality diversity. *Advances in Neural Information Processing Systems* 34:10040–10052.

Gero, K. I.; Liu, V.; and Chilton, L. 2022. Sparks: Inspiration for science writing using language models. In *Designing Interactive Systems Conference*, 1002–1019.

Gero, J. S. 1998. Conceptual designing as a sequence of situated acts. *Artificial intelligence in structural engineering* 165–177.

Goldschmidt, G. 1991. The dialectics of sketching. *Creativity research journal* 4(2):123–143.

Grace, K.; Lawton, T.; and Ibarrola, F. 2023. When is a tool a tool? user perceptions of system agency in human-ai co-creative drawing. In *Designing Interactive Systems*.

Ibarrola, F.; Lawton, T.; and Grace, K. 2022. A collaborative, interactive and context-aware drawing agent for co-creative design. *arXiv preprint arXiv:2209.12588*.

Lawton, T.; Ibarrola, F. J.; Ventura, D.; and Grace, K. 2023. Drawing with reframer: Emergence and control in co-creative ai. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 264–277.

Li, T.-M.; Lukáč, M.; Gharbi, M.; and Ragan-Kelley, J. 2020. Differentiable vector graphics rasterization for editing and learning. *ACM Transactions on Graphics* 39(6):1–15.

Poon, J., and Maher, M. L. 1997. Co-evolution and emergence in design. *AI in Engineering* 11(3):319–327.

Pugh, J. K.; Soros, L. B.; and Stanley, K. O. 2016. Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI* 40.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from

Figure 4: Randomly selected samples of sketches obtained with MOS-MEGA sampling for “a red chair”, “a hat”, “a lamp” and “a blue dress”, using “colorful” and “large” as behaviour dimensions.

natural language supervision. In *International Conference on Machine Learning*, 8748–8763. PMLR.

Smith, G., and Whitehead, J. 2010. Analyzing the expressive range of a level generator. In *Proceedings of the 2010 workshop on procedural content generation in games*, 1–7.

Suwa, M.; Gero, J.; and Purcell, T. 2000. Unexpected discoveries and s-invention of design requirements: important vehicles for a design process. *Design studies* 21(6):539–567.

Williford, B.; Ray, S.; Koh, J. I.; Cherian, J.; Taele, P.; and Hammond, T. 2023. Exploring creativity support for concept art ideation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–7.