Bridging Generative Deep Learning and Computational Creativity

Sebastian Berns [1] and **Simon Colton** [1,2]
[1] Game AI Group, EECS, Queen Mary University of London, United Kingdom
[2] SensiLab, Faculty of IT, Monash University, Australia

Abstract

We aim to help bridge the research fields of generative deep learning and computational creativity by way of the creative AI community, and to advocate the common objective of more creatively autonomous generative learning systems. We argue here that generative deep learning models are inherently limited in their creative abilities because of a focus on learning for perfection. To highlight this, we present a series of techniques which actively diverge from standard usage of deep learning, with the specific intention of producing novel and interesting artefacts. We sketch out some avenues for improvement of the training and application of generative models and discuss how previous work on the evaluation of novelty in a computational creativity setting could be harnessed for such improvements. We end by describing how a two-way bridge between the research fields could be built.

Introduction

Methods in generative deep learning have become very good at producing high quality artefacts with much cultural value. In particular, Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) provide many exciting opportunities for image generation. Over the course of only a few years, research contributions have pushed models from generating crude low-res images to ones evoking visual indeterminacy, i.e., images which "appear to depict real scenes, but, on closer examination, defy coherent spatial interpretation" (Hertzmann 2019) and even further, to generate images of human faces indistinguishable from digital photographs (Karras, Laine, and Aila 2019). These are just a few of the many applications of generative deep learning around which the notion of 'creative AI' has emerged.

Closely following the fast-paced research on neural networks and generative models in particular, an online community has formed under the hashtag #CreativeAI (Cook and Colton 2018), that has been particularly eager and successful in exploring unconventional applications and has established its place in workshops at major conferences with a focus on artificial neural networks. With extensive knowledge and experience in the development, application and evaluation of machine creativity, the computational creativity community can contribute to this progress by laying out potential ways forward. We aim here to build a bridge between generative deep learning and computational creativity by way of the creative AI community, and we propose avenues for improvements and cross-community engagement.

In the section below, we make a case for generative models as a successful and powerful technology which is inherently limited in its creative abilities by its uni-dimensional objective of perfection. The following section discusses how, in spite of its limitations, GANs have been used and abused as artwork production engines. We then explore how computational creativity research can contribute to further evolve such models into more autonomous creative systems, looking specifically at novelty measures as a first step towards this goal. We conclude by returning to the notion of bridging the two fields and describing future work.

Learning for Perfection

While the purpose of GANs, like all generative models, is to accurately capture the patterns in a data set and model its underlying distribution, guaranteeing convergence for this particular method remains a challenge (Lucic et al. 2018). Theoretical analyses of the GAN training objective suggest that the models fall significantly short of learning the target distribution and may not have good generalisation properties (Arora et al. 2017). It is further suggested that GANs in particular might be better suited for other purposes than distribution learning. Given their high-quality output and wide artistic acceptance, we argue for the adaptation of this generative approach for computational creativity purposes.

Generative models are currently only good at producing 'more of the same': their objective is to approximate the data distribution of a given data set as closely as possible. This highlights two sides of the same fundamental issue. First, in practice it remains unclear whether models with millions of parameters simply memorise and re-produce training examples. Performance monitoring through a hold-out test set is rarely applied and overfitting in generative models is not widely studied. Second, conceptually, such models are only of limited interest for creative applications if they produce artefacts that are insignificantly different from the examples used in training. Hence we further argue for an adaptation such that generative capabilities align with the objectives of computational creativity: to take on creative responsibilities, to formulate their own intentions, and to assess their output independently (Colton and Wiggins 2012).

Proceedings of the 11th International Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

Figure 1: Example results from cross-domain training, i.e., fine-tuning StyleGAN with the Flickr-Faces-HQ data set (Karras, Laine, and Aila 2019) and a custom beetle data set. Reproduced with permission from M. Mariansky.[1]

Figure 2: Samples from Broad, Leymarie, and Grierson (2020) of StyleGAN fine-tuned with a negated loss function. In its state of 'peak uncanny' the model started to diverge but has not yet collapsed into a single unrecognisable output.

Active Divergence

In order to produce artefacts in a creative setting, GANs still require expert knowledge and major interventions. Artists use a variety of techniques to explore, break and tweak, or otherwise intervene in the generative process. The following is a brief overview of some of those techniques. From a purely machine learning perspective, these exploits and accidents would be considered abuses and produce only sub-optimal results. Actively diverging from local likelihood maxima in a generator's internal representation is necessary to find those regions that hold sub-optimal, but interesting and novel encodings.

Latent space search is a common practice among GAN artists, in which a neural network's internal representation is explored for interesting artefacts. Traversing from one point to another produces morphing animations, so-called 'latent space walks'. The space is often manually surveyed. Wherever precise evaluation criteria are available, evolutionary algorithms can be employed to automate the search for artefacts that satisfy a given set of constraints (Fernandes, Correia, and Machado 2020).

Cross-domain training forcefully mixes two (or more) training sets of the same type but different depictions, such that a model is first fit to the images from one domain (e.g. human faces) and then fine-tuned to another (e.g. beetles). The resulting output combines features of both into cross-over images (fig. 1). Finding the right moment to stop fine-tuning is crucial and human supervision in this process is currently indispensable.

Loss hacking intervenes at the training stage of a model where the generator's loss function is manipulated in a way that diverts it towards sub-optimal (w.r.t. the traditional GAN training objective) but interesting results. Given a model that generates human faces, for example, the loss function can be negated in a fine-tuning process such that it produces faces that the discriminator believes are fake (fig. 2; Broad, Leymarie, and Grierson 2020). Again, human supervision and curation of the results is just as important as devising the initial loss manipulation.

Early stopping and rollbacks are necessary whenever a model becomes too good at the task it is being optimised for. Akin to the pruning of decision trees as a regularisation method or focusing on sub-optimal (in terms of fitness functions) artefacts produced by evolutionary methods, rollbacks can improve generalisation, resulting in artefacts that are unexpected rather than perfect.

Summary

All of the above techniques require manual interventions that rely on human action and personal judgement. There are no well-defined general criteria for how much to intervene, at which point and by how much, or when to stop. It is central to an artistic practice to develop such standards, nurture their individuality and the difference to other practices. A major theme in GAN art, however, and a commonality in the above non-standard usages, is the active divergence from the original objective of the tool of the trade, in pursuit of novelty and surprise. This dynamic appears to be, in contrast to other artistic disciplines, exceptionally pronounced due to the use of state-of-the-art technology that has yet to find its definite place and purpose and whose capabilities are open to be explored. We celebrate and support this endeavour and argue that computational creativity can help by pushing generative models further, towards new objectives.

New Objectives for Generative Models

Two avenues of improvements for generative deep learning come to mind in a creative setting. First, we can consider creativity support tools, where a person is in charge of the creative process and the technology takes on an assistive role. For this, generative models need to be more accessible, as active divergence techniques still require highly

[1] Tweet by @mmariansky
https://twitter.com/mmariansky/status/1226756838613491713

Figure 3: Series of image edits applied to three different GANs with the method from Härkönen et al. (2020)

technical knowledge. They further need to be more controllable in their generative process. Active research on disentangled representation learning has recently proposed interpretable controls for global image manipulation (Härkönen et al. 2020). Common dimensions of variance in the data are first identified by the model and later manually sighted and named. Interpretable controls allow for the manipulation of images in a single specific aspect, such as a person's age, the exposure of a photograph or the depicted time of day, while maintaining the others (fig. 3). Similarly, localised semantic image edits (Collins et al. 2020) transfer the appearance of a specific object part from a reference image to a target image, e.g. one person's nose onto another person's face.

Second, generative models are far from completely autonomous creative systems that are able to formulate an intention and assess their own output. As a start, these models require readjustments and extensions to be pushed from mere imitation to genuine creation. While creativity is arguably an essentially contested concept (Jordanous and Keller 2016) and there exist a variety of individual definitions, many of those include the notions of novelty, surprise and some form of value (e.g., usefulness or significance) (Jordanous 2013). In our analysis of GAN artists' practices, a very clear commonality was the abusing of the standard practice in order to produce novel, perhaps surprising, outputs. Hence we will here focus on the aspect of novelty and on how the output of generative models could be assessed in regards to novelty.

Evaluating Novelty

As many evaluation schemes for creativity include notions of novelty, an exhaustive review of the literature is beyond the scope of this paper, as is the relationship and subtle differences between novelty and surprise (Macedo and Cardoso 2017). We focus here on explicit measures of novelty, in particular in the context of generative models. Currently, novelty can be achieved by tuning the stochasticity of a generative process whenever it is conditioned on a distribution of probabilities. In GANs, the latent code truncation trick clips values drawn from a normal distribution to fall within a limited range (Brock, Donahue, and Simonyan 2019). On the other end, a temperature parameter can be applied to scale a network's softmax output (Feinman and Lake 2020). Both improve the quality of individual artefacts at the cost of sample diversity. While the original intention is to decrease randomness in order to obtain artefacts closer to the mean, they may also be able to achieve the inverse. Neural network-based methods have been proposed for the generation of novel artefacts, e.g., CAN (Elgammal et al. 2017), Combinets (Guzdial and Riedl 2019), as well as a number of metrics for the evaluation of GANs, e.g., the Inception Score (Salimans et al. 2016) and FID (Heusel et al. 2017). However, none of these can be used to measure novelty or to compare the extent to which deep learning methods are capable of producing it. For a measure that might fill this gap, we can draw from work in computational creativity.

Ritchie (2007) proposes a formal framework of empirical criteria for the evaluation of a computer program's creativity, advocating for a post-hoc assessment based on a system's output and independent of its process. A definition of creativity focuses on novelty, quality and typicality, where the latter refers to whether an artefact matches the intended class (e.g., when generating jokes, whether it has the formal structure of a joke). Quality (also denoted as value) and typicality are expressed as ratings, novelty is seen as the relationship between the input and output of a program and formalised in a collection of proportions in set-theoretic terms.

Most interesting for our purposes is Ritchie's concept of an 'inspiring set', which could be treated as the knowledge base but, in the context of learning algorithms, does not have to be equivalent to the training set. Representing the examples that the author of a generative system hopes to achieve, it would be too trivial to allow a learning algorithm a glimpse at such examples. Rather, an inspiring set can inform about the necessary choices in the design process of a generative system that might evoke the desired output. Current discussions around the inductive biases of the fundamental building blocks in deep learning pose similar questions. Recent work has tried to leverage the specific choice of structure in hybrid neuro-symbolic models (Feinman and Lake 2020). This idea leaves room for the question of how the concept of an inspiring set could be integrated into the training and sampling schemes of a generative model.

In work on curious agents, Saunders et al. (2004; 2010) use Self Organising Maps (SOM) (Kohonen 1988), to measure the novelty of an input through a distance metric in vector space and in comparison to all other examples stored in the SOM. 'Interestingness' is estimated through an approximation of the Wundt curve (Berlyne 1960) (the sum of two sigmoids), to the effect that the score peaks at moderate values of novelty and rapidly drops thereafter. This model is based on the understanding that for new stimuli to be arousing, they have to be sufficiently different but not too dissimilar from known observations.

Pease, Winterstein, and Colton (2001) discuss novelty in relation to complexity, archetypes and surprise, and propose specific metrics for these aspects. First, an item is deemed more novel the more complex it is. Complexity is defined in terms of the size of a given domain and how unusual and

complicated the generation of an item is, which attempts to capture how many rules and how much knowledge was necessary in the process. Second, responding to Ritchie's typicality, novelty is defined as the distance of an item to a domain's given archetypes. This approach is similar to Saunders et al. (2004; 2010) in that it compares items to a knowledge base and computes distances in vector space. Third, the authors argue that 'fundamental' novelty evokes surprise as a reaction. However, a metric for surprise cannot be used to prove novelty, it only shows the absence of 'fundamental' novelty through the lack of surprise.

Conclusions and Future Work

The bridge between computational creativity (CC) and generative deep learning is currently one-way only. That is, CC researchers regularly draw on deep learning techniques in their projects, but the artificial neural network community rarely draws from the philosophy, evaluation or techniques developed in CC research, even for generative projects. The methodology for reversing the traffic presented here seems sound: survey ways in which artists use and abuse deep learning for creative purposes, identify how current practice limits this, and draw from CC research to address the shortcomings. For the bridge to be truly successful, any flow of information from CC into deep learning must respect the culture of the latter field. In particular, we will be aiming to develop concrete numerical evaluation methods for important aspects such as novelty, against which different models can be compared and progress shown, perhaps framing novel generation of artefacts as solving a problem of generating surprising results. This could lead to test suite data sets and potentially a kaggle.com competition, etc.

In the digital arts, deep generative models have found wide application as avant-garde tools, continuously demonstrating their potential. However, as these tools emerged from the discipline of machine learning, the objective of perfectly modelling patterns in data stands in the way of generative models further evolving towards autonomous creative systems. Active divergence is the common theme of a number of techniques we have explored, illustrating how GAN artists strive for sub-optimal solutions rather than perfect reproductions in the pursuit of novelty and surprise. These techniques, however, require much human intervention, supervision and highly technical knowledge which further limits their accessibility. We believe that computational creativity methods and methodologies, evaluation criteria and philosophical discourses can help progress deep generative learning so that non-standard creative usages become standard and the machine learning community embraces currently (seemingly) fuzzy ideas such as novelty and surprise. In the process, CC researchers will have access to increasingly powerful, autonomous and possibly creative techniques for exciting and ground-breaking projects.

Acknowledgements

We thank the anonymous reviewers. This work was supported by the EPSRC Centre for Doctoral Training in Intelligent Games & Games Intelligence (IGGI) [EP/S022325/1].

References

Arora, S.; Ge, R.; Liang, Y.; Ma, T.; and Zhang, Y. 2017. Generalization and Equilibrium in Generative Adversarial Nets (GANs). In *Proceedings of ICML*.

Berlyne, D. 1960. *Conflict, Arousal & Curiosity*. McGraw-Hill.

Broad, T.; Leymarie, F. F.; and Grierson, M. 2020. Amplifying The Uncanny. In *Proceedings of xCoAx*.

Brock, A.; Donahue, J.; and Simonyan, K. 2019. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In *ICLR*.

Collins, E.; Bala, R.; Price, B.; and Süsstrunk, S. 2020. Editing in Style: Uncovering the Local Semantics of GANs. In *Proc. CVPR*.

Colton, S., and Wiggins, G. A. 2012. Computational Creativity: The Final Frontier? In *Proceedings of ECAI*.

Cook, M., and Colton, S. 2018. Neighbouring Communities: Interaction, Lessons and Opportunities. In *Proceedings of ICCC*.

Elgammal, A.; Liu, B.; Elhoseiny, M.; and Mazzone, M. 2017. CAN: Creative Adversarial Networks, Generating Art by Learning About Styles and Deviating from Style Norms. In *Proc. ICCC*.

Feinman, R., and Lake, B. 2020. Generating new concepts with hybrid neuro-symbolic models. In *Proc. CogSci*.

Fernandes, P.; Correia, J.; and Machado, P. 2020. Evolutionary latent space exploration of generative adversarial networks. In *Proc. Int. Conf. Applications of Evolutionary Computation (EvoStar)*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In *Advances in Neur. Inf. Proc. Sys.*

Guzdial, M., and Riedl, M. O. 2019. Combinets: Creativity via Recombination of Neural Networks. In *Proc. ICCC*.

Härkönen, E.; Hertzmann, A.; Lehtinen, J.; and Paris, S. 2020. GANSpace: Discovering Interpretable GAN Controls. *arXiv preprint 2004.02546*.

Hertzmann, A. 2019. Visual Indeterminacy in Generative Neural Art. In *NeurIPS Workshop on Creativity for Learning & Design*.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Adv. NIPS*.

Jordanous, A., and Keller, B. 2016. Modelling creativity: Identifying key components through a corpus-based approach. *PloS one* vol. 11(10).

Jordanous, A. K. 2013. *Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and Its Application*. PhD Thesis, University of Sussex.

Karras, T.; Laine, S.; and Aila, T. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. In *Proc. CVPR*.

Kohonen, T. 1988. *Self-Organization & Assoc. Memory*. Springer.

Lucic, M.; Kurach, K.; Michalski, M.; Gelly, S.; and Bousquet, O. 2018. Are GANs Created Equal? A Large-Scale Study. In *Advances in Neural Information Processing Systems*.

Macedo, L.; and Cardoso, A. 2017 A Contrast-Based Computational Model of Surprise and its Applications *Topics in Cognitive Science*, 11(1).

Pease, A.; Winterstein, D.; and Colton, S. 2001. Evaluating Machine Creativity. In *Proc. ICCBR Workshop on Creative Systems*.

Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, vol. 17.

Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; and Chen, X. 2016. Improved Techniques for Training GANs. In *Advances in Neural Information Processing Systems*.

Saunders, R., and Gero, J. S. 2004. Curious Agents and Situated Design Evaluations. *AI EDAM* vol. 18(2).

Saunders, R.; Gemeinboeck, P.; Lombard, A.; Bourke, D.; and Kocaballi, A. B. 2010. Curious Whispers: An Embodied Artificial Creative System. In *Proc. ICCC*.