Georg-August-Universität Göttingen
Human in the Age of Artificial Intelligence

REPORT ON ARTIFICIAL CREATIVITY

Benjamin Eckhardt (2022-03-15)
benjamin.eckhardt@stud.uni-goettingen.de

ABSTRACT

With the advent of artificial intelligence common intuitions about creativity are being challenged. Creativity may no longer be considered as a distinctively human capability, neither as something divine defying explanation. This report aims to provide a philosophically guided introduction to the field of computational creativity. Thus we explore the concept of creativity it self, show how it decomposes into various different notions and discuss their fundamental applicability to machines. In order to understand the current developments we give an high-level overview of deep learning, a main branch of artificial intelligence. Through the variational autoencoder we consider an example of autonomous creative computers. Thereby we gain an intuition for artificial intelligence's as well as our own means of understanding.

I. INTRODUCTION

In this report we investigate different notions that make up creativity and consider fundamental concerns regarding creative machines. But we will focus mainly on one sort of notions: The measurable means, by which a product of creativity may have impact on its social context (section II). There are other notions that are inevitably tied to controversial concepts, especially when applied to machine intelligence. Human creativity is surely to a great extent self-expression which is deeply tied to feelings and personally lived experiences. As such if creativity should be tied to consciousness, the discussion about machine creativity has to be deferred for the discussion about machine consciousness. The same applies for self-awareness, emotional sensation, empathy and maybe more. Those questions are to some extent purely philosophical, as they involve the definitions of those concepts as well as to define machines in contrast to humans – all of which are highly controversial by themselves alone (cf. Boden 2014). We do not aim here to delve into these discussions; for a comprehensive introduction to each of those the reader may refer to the great SEP[1]. What we aim is to understand mechanical processes that yield a (in the senses of section II) measurably creative artifact. To this end section III starts with the attempt to shortly introduce the inner workings of common machine intelligence – deep neural networks – in a digestible way. Building on this the variational autoencoder (VAE) is proposed as a specific architecture in section IV. It exposes us to some valuable intuition about the means of machine understanding: The yielded concept of latent space shall serve us well in understanding the basic creative capabilities as well as the expressiveness of artificially intelligent systems in section V. We shortly indulge in remarking the philosophical significance we see in the associated empirical findings: Machines means of understanding may serve as a model for our own; and the process by which it's yielded may expand to an inductive proof of the usefulness of our own. We thereafter consider in section VI a simple example of an autonomous creative machine: A random sampler tied to a latent decoder. We investigate how the concepts discussed in section II apply to this. After finally concluding our findings, we give an outlook on hybrid human-computer creativity.

II. MEASURING CREATIVITY

In this section I mainly reflect (Boden 2014).

There is no one general and unifying concept of creativity as it is taken to apply in a multitude of ways in different contexts. Rather there are many different aspects contributing to it and perspectives to take. In this paragraph we shall try to un-

[1] The Stanford Encyclopedia of Philosophy, available at plato.stanford.edu

derstand how to measure creativity in an artifact. This will reveal some corresponding cognitive processes, which may be suitable to port to machines.

Creativity may be measured by the functions a presumably creative product fulfills in some context. Several such functions may be captured by saying that an artifact or an idea is creative iff it is *novel*, *surprising* and *valuable* to someone. Whereby each of these three concepts may have multiple meanings. As such a thing may be valuable, because it is useful for reaching some specific goal, or because it is appealing to some individuals sense of beauty. Novelty may as well either mean novelty to the person yielding the product, novelty to some other recipient, or even novelty to the whole of human thought. Surprising in one way may be something that is unusual, but known to be well possible (like random brush strokes to appeal more than others). Surprising may also be something unexpected and unconsidered, but which still obeys some recognizable regularity (like a new painting that is following a familiar style). And finally it may also mean the surprise of something so completely new, that it breaks with what was before considered to be even possible (typically founding a new genre).

These three different modes of surprising may correspond to a psychological mechanism in the creative process each (Boden 2004). By mimicking such a mechanism we can create machines that may be considered creative in the respective sense. Surprise in the weakest sense is facilitated by *combinational* creativity, which is to recombine known ideas in new ways. The second mode of surprise is yielded by *exploratory* creativity, the search for new ideas in a known conceptual space, thereby orienting on known regularities therein. And there is *transformative* creativity, which is the venture beyond the understood and the discovery of in some part completely new ways of thinking, and which facilitates the strongest mode of surprise. The variational autoencoder we present in section IV shall be our example of creativity in the different senses.

III. DEEP LEARNING

For more comprehensive explanations of the concepts introduced in this and the next section as well as many related ones the reader may refer to (Akmen 2021).

A *neural network* (NN) is a network of many simple units, called *neurons* or *perceptrons* (Rosenblatt 1958). A neuron combines the signals of multiple other neurons into a new signal, thereby performing a simple nonlinear computation involving multiple tweakable *parameters*. The neurons are typically organized in layers (see Figure 1).

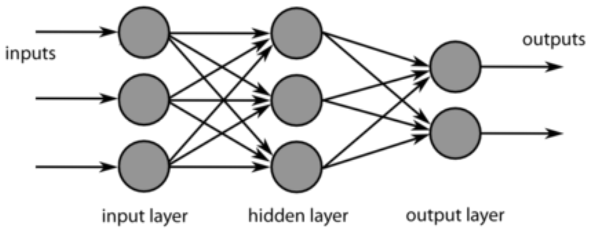

Figure 1: A basic deep neural network. The balls depict neurons. Arrows between neurons depict that the value the one yields is involved in the computation of the other. For simplicity each neuron only receives values from the layer directly behind and feeds only to the layer directly ahead. Source: de.wikipedia.org/wiki/Deep_Learning

Trough the non-linearity NNs of a certain depth (having minimum three layers of neurons and called *deep*) are *universal function approximators*, meaning that (roughly) any real-world relation between two domains of data (*input* and *output*) may be modeled arbitrarily accurate just by setting the parameters right – a certain NN is thus called a *model* of the relation. The process of setting those parameters right, called *learning,* is iterative and approximative (and thereby highly computation and energy intensive). Most importantly it can be automated (though gradient descent on a loss-function evaluating the NNs performance – typically by comparing the models output to an expectation). A neural network thus is a computation from some arbitrarily choosable input domain to some arbitrarily choosable output domain, with the distinctive property of being able to have this mapping improved automatically (besides being made up of neurons). In such systems complexity arises from the interaction of many simple pieces – hence the fascinating capabilities of deep learning. But in the end, it is just a computation.

IV. VARIATIONAL AUTOENCODERS

Variational Autoencoders (VAE) (introduced in (Kingma 2014)) are networks trained to represent complex high dimensional data optimally in very few dimensions. Consider images as an example. A common medium sized digital image contains hundred-thousands of pixels, each varying per color-channel. Thus to represent all the information an image contains one could need some millions of numbers. The most time though not all this information is needed: The content of a natural image (like a photography) may be described to great extent using much fewer than a million words (though the opposite is claimed often). Each individual pixels color is not the relevant way in which we process images and fluctuations therein don't make meaningful differences. Rather pixel colors are very much interrelated in ways we capture by talking about shapes and whole objects.

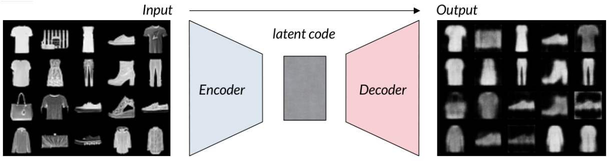

Figure 2: Our variational autoencoder learned to represent 32x32 Pixel gray-scale images of fashion articles using just 4 numbers.

The job of a VAE is to exploit this well structuredness of natural images and to condense them into much fewer numbers called *latent code*. VAE are mainly constituted of two separate neural networks (NNs): The *encoder* and the *decoder*. Both can be imagined as computational black-boxes in the shapes of a trapeze, one mirroring the other (see Figure 2.). The encoders large side is faced with the raw numbers of a data sample (like an image of a fashion article). The multiple layers it consists of decrease in width towards the small end. Thereby the computation stepwisely reduces the dimensionality (fancy word for size or amount of numbers) of the sample. This implies an interesting process: With each layer more and more information gets condensed into each number. The small side finally outputs the much fewer numbers constituting the latent code of the respective data sample. The decoder tries to reverse this computation: layer by layer information from multiple dimensions is combined to yield a greater sample size, eventually arriving at the original dimensions of the data. The training objective is to maximize the similarity between the input and it's reconstruction, the output. This is done by minimizing the euclidean distance between both and/or their intermediate (between the layers) representations.

Summing up, a variational autoencoder is a chain of two networks, each optimized for providing one way of a bidirectional mapping between data space and latent space. There is a twist to consider: In contrast to plain autoencoders the variational ones don't directly encode data into latent space, rather they encode into parameters of a probabilistic distribution residing in latent space. This distribution is sampled from to generate the latent code that is then forwarded to the decoder. While this procedure may seam abstruse, it will provide a useful property for our example of autonomous creativity in section VI.

V. LATENT SPACE

Due to the training objective of a VAE the whole space of the data in consideration has to be exhaustively covered by latent representations. In our example from Fig. 2 this means that every fashion article shall have its individual approximative code. Because the whole diversity of the data has to be captured in the much fewer dimensions of the latent code, each of them inevitably has to be very expressive. Given the latent encoding of some data when we modify a number at a time in it, we observe something interesting: Some directions in it correspond to very meaningful human semantic categories. For example by manipulating a single value in it, we may observe a reconstructed images transition along the axis of one specific concept, like the degree of smile in an image of a face (Kingma 2014). We thereby see some concepts, humans use to understand the world emerging out of data using fully autonomous systems. Those systems are mechanically optimized for the aim to represent some data in the most efficient way. Fulfilling this aim may be taken as a deductive account that some of the emerging means to categorize data (like the degree of smile) are indeed useful. This may seam like a no-op, but expanding on this we have found, that the corresponding notions we use as the basis of our understanding and communication are not just arbitrarily made up of thin air. There is empirical evidence now that they indeed reflect structure inherent to the world. Of course it is important to consider the bias induced by subjective data aggregation. The provided training data may be strongly influenced by the notions we use in the first place. The emergent notions are only of particular use relative to that data – unable to capture dissimilar or underrepresented varieties. Also we observe another very interesting property of latent space: Doing simple arithmetic with latent codes is reflected in the reconstructions in meaningful ways. Take e.g. the code of a man wearing glasses, subtract the code of the man without glasses and add this to the code of a woman without glasses; decoding may yield the woman wearing the mans glasses (Radford 2015). This is what we would expect from doing the same arithmetic with our notions of woman, man and glasses: The difference between a man with and one without glasses is the glasses, that together with a woman is a woman with glasses. We thereby see that by operating in the conceptual space of artificial intelligence we have intuitive means to control their enormous expressiveness. This may prove very useful for enhancing human creativity with computational powers.

If anything, latent space may be considered as the means by which artificial intelligence facilitates human-like understanding. Abstraction and information condensation is the key by which a word may say more than a thousand pixels.

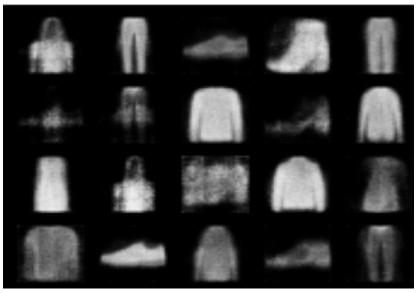

Figure 3: Random samples from the latent space of the VAE in Figure 2

VI. AUTONOMOUS CREATIVITY: LATENT RANDOM SAMPLE DECODING

Our example for a autonomously creative machine is of simple construction: Take a random generator to produce random latent codes. Decode them into images (or any other domain) by applying a (trained) decoder to it. Sample produced by our example VAE in Fig. 2 are depicted in Fig. 3. Let's analyze this automaton with our tools form section II and discuss in which ways it may yield creative artifacts. The produced images are to some extent novel – quite likely images are generated that didn't exist before as with any random generator of sufficient dimensionality. The generated images valuability draws from the valuability of the training data. This is due to two facts: A variational autoencoder does learn the distribution of data in latent space, by sampling form it, we gain plausible latent codes. Whereby plausible means, that natural data could have yielded it. Secondly, latent codes are optimized for representing the data they are trained with, thus decoding will almost surely result in reconstructions that resemble some structure of the data. As such, if the training data exposes any valuable qualities, it probably will reflect in the randomly generated images. Some portion of the them will however be invaluable: Like the first one in the second row in Fig. 3, sometimes the sampled code lies far outside of the space covered by the training data, thus there are no means by which the decoder could have been trained to make something of value form it. What counts as surprise? Our example suggests a process corresponding to combinational creativity, yielding the weakest sense of surprise. As each dimension in latent space corresponds to a distinct feature(set), choosing any new instantiation (like with random sampling) is to be considered as combining known ideas in new ways. Our VAE apparently recombined some known features in new ways, yielding some blurry new pants and shirts, but also some fashion-chimeras. The second degree of surprise, the one corresponding to exploratory creativity, is questionable though. The shirt-bag in row one, column one in Fig. 3 may be considered as a genuine creation; nothing like this has existed before (not in the training data at least) or was to be expected; plus it superficially resembles the style of the other items. Randomly sampling from a distribution that is informed by the training data (the special feature of VEAs) may be considered as a means to orient in the conceptual space with the aim to stay close to what is known. It is though the simplest and most disputable way to 'orient' in conceptual space. Transformative creativity seams far out of the possibilities of such a system: It could never produce anything that wasn't partly contained in the training data. Portions of latent space that are not involved in the training have no means to be learned as there is per construction nothing to compare a decoding against. An offbeat portion of latent space will yield an image that contains *any* structure at most by chance due to random initialization or accidental complex interplay of parts of the computation.

VII. CONCLUSION AND OUTLOOK

We found that autonomous machines may be in a limited manner creative. We therefore constructed and studied a random sampler tied to a latent decoder. The means by which such systems create are inherently combinational: Each latent dimension associates with some high-level concept(s), thus random sampling is like combining them in new ways. There is no way for such a system to produce something completely unseen other than by rare accident, as it is only able to make sense of data similar to that it was trained with. Thus transformative creativity is far out of scope. Exploratory creativity tough may be present: Random sampling from a known distribution in latent space may be considered as a borderline-case of exploration of conceptual space that is oriented on known regularities. We glimpsed on the tremendous potential of deep neural networks capabilities to abstract, thereby facilitating human-like understanding of image contents. A much more impressive demonstration of this has been made in (Ramesh 2021)[2]. In (Akten 2021) these capabilities are explored with the aim to provide ways for humans to meaningfully express through them, pioneering deep visual instruments. A different approach for exploring this is in (Park 2020), showcasing impressive content alteration of images by high-level and precise human guidline. A great reading to get into many concepts of machine intelligence as well as very recent research on computational creativity is the dissertation by Memo Akten (2021). The authors filled with computer aided art website[3] is also to be recommended.

REFERENCES

Memo Akten (2021): *Deep Visual Instruments: Real-time Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression* (https://research.gold.ac.uk/id/eprint/30191/)

Margaret Boden (2014): *Creativity and Artificial Intelligence, A Contradiction in Terms?* (https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780199836963.001.0001/acprof-9780199836963-chapter-12)

Taesung Park et al. (2020): *Swapping Autoencoder for Deep Image Manipulation* (https://arxiv.org/pdf/2007.00653v2.pdf)

Radford et al. (2015): *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks* (https://arxiv.org/pdf/1511.06434v2.pdf)

Aditya Ramesh et al. (2021): *Zero-Shot Text-to-Image Generation* (https://arxiv.org/abs/2102.12092v2)

Frank Rosenblatt (1985): *The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain*

Diederik Kingma and Max Welling (2014): *Auto-Encoding Variational Bayes* (https://arxiv.org/abs/1312.6114v10)

Deep Learning lecture by Alexander Ecker, Winter 2021/2022, Göttingen

[2] See also their interactive website at openai.com/blog/dall-e.

[3] www.memo.tv