# Creativity in Automatic Sketch-Based Image Creation

**Jannik Hofmann**
Creatividad Computacional 2020-II
Posgrado en Ciencia e Ingeniería de la Computación
Universidad Nacional Autónoma de México

Medieninformatik Master, Fachbereich Informatik
Eberhard Karls Universität Tübingen, Germany
jannik.hofmann@student.uni-tuebingen.de

**Abstract**

This paper analyzes the possibilities of automatic image creation based on sketches submitted by a user, utilizing a generative adversarial neural network. It also discusses the philosophical perspective of computational creativity that is exhibited by this type of system. Furthermore, it is outlined how this coach system could be changed into a co-creative colleague.

## Introduction

In human history, humans seem to always have had new ideas and concepts in their minds. They changed and shaped the world around them, guided by their imagination of what possibilities lie ahead. A central part in the creation and execution of these ideas is the concept of creativity, that humans inherently possess. This level of creativity they exhibit is widely thought to be unmatched (Boden, 2007).

In recent years and decades, humans have tried to teach creativity to computer systems, tried to emulate this inherent characteristic of human intelligence with artificial intelligence systems (Boden, 1998). One kind of these systems is the technology of generative adversarial neural networks, who can generate images after being trained on thousands or even millions of examples of the same type, in order to learn how such a generation works instinctively, without fixed rules that humans need to specify.

In this paper, I will analyze one of these systems, SketchyGAN (Chen & Hays, 2018). This algorithm receives a sketch (preferably drawn by a human) and tries to generate a realistic image out of the sketch, while considering that human sketches represent an intention of an object, rather than an exact representation.

**SketchyGAN and creativity** In the first part of this paper, SketchyGAN will be compared against similar state-of-the-art systems. Then it will be analyzed as to how it fulfills the properties of an autonomous computationally creative system, followed by a discussion of its creative limitations and an evaluation of the products and the system itself regarding the concept of creativity.

**Co-creative colleague version of SketchyGAN** After having discussed the system, its products and its relation to creativity, I outline how SketchyGAN could be converted into a computer colleague, that allows improvisation in a real-time collaboration between a human and the system. The objective of this paper's second part is to paint a detailed picture of what changes would need to be made in order to create a new version that satisfies the standards of co-creativity, rather than just acting as a coach system. This elaboration includes not only concepts, but specific instructions for changes and suggestions for a graphical user interface that enables this co-creative work.

These ideas are not limited to SketchyGAN and possibly could be applied to various generative adversarial networks to produce computer colleague systems that provide opportunity for co-creative interactions with human users.

This part will be followed by a conclusion in the end.

## Related previous work

In recent years, the advent of generative adversarial neural networks has sparked impressive research in the area of automatic image generation. A small part of this research focusses on generating realistic images from crude human sketches.[1] However, most of those systems work with category-specific sketches. This always depends on the datasets that were used to train the algorithms.

### Category-specific constraints

For example, some only work on faces (Osahor et al., 2020), other on faces, birds and cars (Lu et al., 2018).

Liu et al. (2019) trained separately on shoes and chairs. Cross-domain applications are limited. For example, using a sketch of a car with a network that was trained on shoes, yields a result that display a mixture of the two objects.

[1] No research in that area was published by the ICCC yet.

Furthermore, SketchyCoco (Gao et al., 2020) recognizes multiple drawn objects in a sketch, categorizes them using pre-trained categories and places a generated image for each object onto a generated background, that can be indicated by pre-defined elements in the sketch (clouds, trees or tufts of grass).

In contrast to those systems, SketchyGan (Chen & Hays, 2018) and MindReader (Guo at al., 2019) both do not require the user to specify a category for their sketch, however, to improve the quality of resulting images they are still only trained on a limited number of given categories.

## Properties of human sketches

Another problem of sketch-to-image translation is the distortion of human sketches, which often requires a semantic understanding and domain-specific correction of the shapes into more realistic compositions. Human imagination does this automatically, as humans understand that sketches only inaccurately conceptualize the most significant parts of an image. But computers need to be specifically trained on this for a good performance on human-drawn sketches. To combat this problem, Liu et al. (2019) propose a system to distort and simplify the automatically detected edges from object photos, in order to simulate human perception (and drawing skills) in their training data.

## Creativity research

**Artificial Intelligence as a coach system** All systems that were discussed in this section can be classified as creative coach systems, as they help humans flesh out creative ideas. The user always provides the sketch, so they significantly take part in the creative process. For that reason, the systems cannot be counted as self-employed creative entities.

In some systems, the human also further defines the category of that sketch. Then, the system generates an image[2] out of this drawing autonomously, without further input from the user. Therefore, those systems should be regarded as a coach, and not as a colleague, as there is no real-time interaction between human and computer. After the product is created, the human can then freely use it in their creative work, for example to aid with prototyping various different ideas.

**Type of creativity** Boden describes three distinct types of creativity a creator can demonstrate (Boden, 1998): combinational, explorational and transformational creativity.

Transformational creativity happens when the concept space is transformed, and the system generates an artifact that was previously outside of the bounds of possible or expected solutions. By design, this type of creativity can hardly happen with a generative adversarial network since it strives to emulate a similar type of results as found in the knowledge base, especially since it is trained on one domain and does not have any inter-domain knowledge it could utilize.

Exploratory creativity takes places when ideas are generated by exploring the concept space in a structured manner. This is arguably possible with artificial intelligence, but also difficult to achieve when the network only produces a single output image per input. However, modern research into exploration of the latent space (Shen et al., 2020) shows that generative adversarial networks have a semantic intuitive understanding of the images they create. This means, in their search for an optimal result, neural networks explore the latent space of possible plausible results. This could be seen as a representative mapping of the relevant conceptual space in their domain. Therefore it could be said, that the SketchyGAN system demonstrates exploratory creativity, too.

Finally, combinational creativity happens when familiar ideas are combined in unfamiliar ways. This definitely happens with neural networks, as they inherently combine elements in a fascinating manner that humans cannot fully comprehend.

# SketchyGAN as an autonomous computationally creative system

In this section, I will analyze the SketchyGAN system (Chen & Hays, 2018) in regard to its architecture as an autonomous computationally creative agent, as described by Dan Ventura in his instruction on how to build a computationally creative system (Ventura, 2017).

## Domain

SketchyGAN belongs to the domain of digital art. As it translates basic human sketches into colored images with more details and textures, the human can use its output in their own creative process or as its own artwork. Therefore, the system mostly supplements an artist's work.

## Representation

**Phenotype** The phenotype in the system of SketchyGAN represents the produced colored image. It is represented by a 64x64 matrix of pixels, each with RGB color values from 0 to 256. This digital image can then be stored as a bitmap file for display on a computer screen.

**Genotype** The genotype has a similar format as the phenotype. It represents a given sketch, that the user submits to the system.

However, rather than just storing the pixels RGB-values the genotype stores a 64x64 matrix of an unsigned Euclidian distance field for a sketches edge map. This improves the networks ability to utilize the genotypes data, as every pixel has some information about its nearest edge – instead of just a sparse number of pixels that contain the edges themselves, which would leave the majority of the matrix empty, mostly void of useful information for the network.

[2] In some cases the system directly generates multiple images, like with MindReader (Guo et al., 2019).

The genotypes information will then be used as an input to generate a corresponding phenotype. During this process, the generative adversarial network will condense the information from the genotype into more high-level information about the input sketch. This data will then be further used to construct the phenotype in a semantical fashion (see subsection “Translation”).

## Knowledge base

As a knowledge base, the system SketchyGAN is using an augmented version of the Sketchy dataset (Sangkloy et al., 2016), which contains pairs of human-drawn sketches and corresponding photographs. This dataset is divided into 125 distinct categories, which contain a total of 75,471 sketches from 12,500 distinct objects.

Chen and Hays (2018) then used image collection methods to collect 61,365 images per category, after excluding a few categories, that often have a human as main object. With these images and an edge detection algorithm to draw the edge maps, they converted the images to sketch-like versions of that image.

The generative part of the network is then trained with this enlarged dataset in order to learn how to convert a sketch into one of those images. First, mainly with the original (human-drawn) sketch pairs and later increasingly with the augmented dataset, as it provides a larger number of differing training examples. This means, the system learns a conceptual model of the image-sketch domain in the previously selected categories, so it can understand the inherent simplifications humans make when they draw a simple sketch.

## Aesthetic

The aesthetic of this system is inherently defined by the knowledge that is used to train it. When creating SketchyGAN, Chen & Hays (2018) did not define certain preferred qualities, that an image converted from a sketch needs to have. Instead, they chose (and augmented) the knowledge in such a way, that the discriminator part of the network can define its own qualities for the output images while trying to distinct between real training images and output images, that the generator part produced by itself.

Moreover, the researchers do not know what qualities the network chose in order to measure the quality of a created image. Nevertheless, the aesthetic of the system directly correlates with the knowledge base, as it has been trained with that data.

## Conceptualization

The conceptualization of SketchyGAN is realized by the generative adversarial network. It is trained on the knowledge base and therefore learns the ability to generate phenotypes from a given genotype.

The network itself has a fixed architecture, containing countless neurons, each of which has weights. These weights are constantly adjusted during the training period to give its network the mentioned abilities. After training is complete, these weights represent a conceptualization that was created from the knowledge base itself.

## Generation

The generation process of a SketchyGAN is directly ingrained with its conceptualization. The generative adversarial network will be run in a cycle, where the generator part tries to create improved output images that seem real to the discriminator, and the discriminator part intents to improve at detecting those images and being able to distinct between real training images and generated ones. This cycle repeats until the output becomes stable and the error between the two systems is minimized. That means, the generator part can generate quite realistic images from sketches, and the discriminator part is good at detecting the quality of such an image.

## Genotypic Evaluator

The generative adversarial network does not have a distinct genotypic evaluator. Genotypic evaluation does not take place explicitly in SketchyGAN. Instead it relies on the fact that the aesthetical preferences of the domain will be assimilated from the knowledge base during the training process. Therefore, the generation process includes some form of genotypic evaluating and filtering.

This form of genotypic evaluation and constrained generation does however limit the ability of the system, to fully explore the conceptual space of that domain.[3]

## Translation

The translation process of SketchyGAN is the typical generation function of a generative adversarial network. In this process, the genotype is given as input data for the generator part of the network.

The generator network of SketchyGAN then sends this information through countless neurons in seven masked residual unit blocks, utilizing three skip connections between the layers. In the first half of this translation, the original input data, which represents an image of a sketch, is compressed into a smaller footprint, with the intention to create a higher-level representation of the data with more meaning than the initial pixel values. In the second half of the translation process, the more semantical information is then converted back into an image of the same size, using the compressed information in combination with original sketch information from skip connections to fill in the picture in a meaningful way that fits the aesthetic of the domain.

Finally, the generator part of the network outputs the produced phenotype, which can be displayed as a colored image

[3] This is an inherent property of the design of generative adversarial networks and always should be considered when choosing such an architecture for an autonomous creative system.

that should display a more realistic version of the original sketch.

## Phenotypic Evaluator

The phenotypic evaluation process of SketchyGAN takes place in the discriminator part of its network. During the training process, the network learns the aesthetic of its domain. While the generator is trained to produce realistic images from given sketches, the discriminator learns to distinguish generated images from real, realistic training data of the knowledge base. This means, after training, the discriminator can predict with a certain accuracy, if a given image is realistic or not, if it fits the aesthetics of the domain that correlate to the knowledge base.

Although the generator part of the network will also be trained in a way to satisfy those qualitative requirements, the output image can be checked against the discriminator to verify if it is recognized as a realistic representation of the sketch, if it fits the aesthetic requirements of the domain. Should it not satisfy the discriminators standards, it is possible to tweak the genotype in absolutely miniscule ways, until the genotype, translated into its corresponding phenotype passes the phenotypic evaluation. This way, it might be possible to implement a phenotypic evaluation and improve the quality of the resulting phenotype a bit more than with just the use of the generator part of the network after training is complete.

# Creative Limitations of SketchyGAN

## Social aspect of creativity

Margaret A. Boden (1998) defines a creative idea as one that is novel, surprising and valuable. This definition is based on the social perception of human creativity is regarded as one of the most widely accepted in the research field of computational creativity. Therefore, the SketchyGAN system (Chen & Hays, 2018) will also be measured against this description.

**Novelty** An idea can either be novel to its creator (classically a human, but in this case the algorithm of artificial intelligence) or novel in the whole known history of mankind.

Basing the definition of creativity on the first description, when the idea was created by an artificial system, would be arguably not very useful, as it is remarkably easy to produce such a system, that creates ideas it in itself has never encountered before. Only because those artifacts are novel to a computer system, does not make them more creative. This description can be seen as a too broad definition of creativity for computationally creative systems.

The second option, novelty based on the history of all ideas that have ever been had (and documented), definitely represents a narrower, more specific idea of creativity. However, it is hard to imagine how SketchyGAN would fulfill this concept, as it is trained specifically on producing realistic images from sketches. While an interesting task for machines, this is not a novel concept for humans, who have sketched and drawn for millennia. And since SketchyGAN is optimally trained on sketches and realistic images from those sketches, the better it becomes, the more difficult it is to create something that is novel.

Therefore, according to Boden's definition, this system might not be considered novel and therefore not creative.

**Surprise** A creative idea should also be surprising, as a predefined or deterministic following of rules, that leads to predictable results, hardly counts as creative. Here, a generative adversarial network again has difficulties fulfilling that requirement, as the network is trained to produce artifacts that fulfil the aesthetic of the knowledge base, imitating the style of the artifacts that have been used to train it. Therefore, most of the results of the system will be expected, if it has been trained well enough.

However, there is still potential in the system to produce surprising products. Especially in the results, that do not look like realistic images or renderings of the given sketch.

**Value** An idea should also be valuable in order to count as creative. This aspect of creativity is highly social and hinges on the ideas of humans, their cultural norms, customs and experiences. An idea might be completely useless to one culture of people, while highly appreciated in another. An idea that is novel and surprising might be discarded as useless, if it does not have any perceived value.[4] Then it is also rarely appreciated as creative by other people in the domain.

Ignoring the problem, that an autonomous phenotypic evaluator has trouble representing all of this nuance while aiding the creation of such a system, it is also difficult to train a system in a way that satisfies these ideas. Using a generative adversarial network, this is only possible via selecting the knowledge base appropriately, preferably by experts in the same domain. While it is clearly not easy to satisfy this description, it is definitely possible for SketchyGAN to provide valuable results.

## Autonomy

Kyle E. Jennings (2010) defines three criteria a system must fulfil in order to be seen as having creative autonomy.

The first is autonomous evaluation. This means, that system has the ability to evaluate the quality of its own creations to help itself create better artifacts. This requires a certain level of self-awareness. With the discriminator part of the network, this is fulfilled.

The second part is autonomous change. This criterion is fulfilled, when a system can change its evaluation function on its own. While artificial intelligence is going into that direction, neural networks can only change their weights, but

[4] Note that perceived beauty or appreciation in art also count as values. A given idea does not necessarily need to serve a tangibly practical value in the classical sense of usefulness in order to be creative.

not the way, in which they operate. That means the system does not have the ability to decide for itself how it wants to create things and does not have the creative freedom to decide what to create, a property which is ingrained into the human understanding of creative freedom.

The final part is non-randomness, also called aleatoricism. This means, that the system should not act in a completely random fashion, but rather that it should have a reason for its decisions and an intent behind its own actions. This philosophical idea of creativity is hard to match to deterministic machines, especially neural networks, that operate in a manner that is opaque to human understanding. However, it can clearly be said, that these networks do not operate completely randomly, but that they instead have a directive they intend to follow as they try to match the learned aesthetic of the given domain.

Moreover, SketchyGAN needs a (preferably human-drawn) sketch in order to create an artifact. Therefore it might not be considered to create with autonomous creativity, but rather as a coach system to human creativity.

In conclusion, the system does only have a limited amount of self-determination and autonomy over its creative process. While it is difficult to define, at what point an algorithm can reflect on its decisions and process, it can arguably be said, that neural networks do not yet possess the capability for that kind of advanced thinking, that humans are used to in their own creative processes. Therefore, the autonomy in decision-making of SketchyGAN is not comparable to that of human creativity.

## Understanding of domain

SketchyGAN was trained on a knowledge base, that contains only limited categories. Therefore it only has information about those distinct categories of objects and can only assimilate the aesthetic of the domain in the area of those limited categories. This limits the system to create only images of objects in those categories. Were the system to be used in the digital art domain for example, it could also be trained on works of art itself (and their corresponding sketches) instead of limiting it to realistic objects from the real world.

Furthermore, it does not seem like the system would be able to transfer that knowledge into a new domain, which would enable it to demonstrate transformational creativity. While research into domain transfer for neural networks is making advancements, for now it seems that SketchyGAN lacks this adaptivity and could mostly just be used for the categories it was trained in, as it has no understanding of other objects.

Also, it is hard to determine if SketchyGAN really obtains an intelligent, intuitive understanding of the domain, or if it just imitates the style and images it sees without making creative decisions about the composition. Because of years of experience in the real world, humans have an intrinsic understanding of how different domains relate, a commonsense of reality. Even if a human sees a sketch of an unknown object, they could make educated guesses about its three-dimensional shape, use, size, texture and environment. On the other hand, artificial intelligence systems would not be able to understand the sketch in such depth and just fill in random colors that it learned from other objects. For example, the researchers conclude, that SketchyGAN has difficulty recognizing the human intent, the sketch artist had when creating the sketch. While probably useful in the domain of digital art, this system might not be as helpful in prototyping or rendering applications. And even in digital art, it might only prove useful to the user in the categories it has been trained in.

# Evaluation

Before continuing with the evaluation, one remark about the analysis that will be conducted in this and the following sections: While the original code of the artificial intelligence system was made available online by its creators, unfortunately the dataset that was used is no longer available (Chen, 2020). This means, in order to prepare and train the SketchyGAN system, one would have to recollect, process and augment the training data set before configuring and training the neural network. And even then, it is not guaranteed, that it would produce the same or similar results, when it was trained with different original images. Rather, in this and the following sections I will analyze the results presented in the original paper (Chen & Hays, 2018), taking especial care of which products are representative of the system and which ones are the best products, handpicked by the researchers. Furthermore, the analysis and discussion in the following sections about potential changes to the system will take place in a theoretical fashion.

## Evaluation of the generated products

**General observations** The products SketchyGAN produces seem to be in the digital art domain, as such a system would mostly be used by artists to quickly develop an idea for a more detailed image from a rough sketch. That means, the proposed audience for such a system would be that digital artists who could use the products in their further work. Outside of the digital art community, the results might not find appreciating, as they lack a certain quality and photorealism, as the researchers concluded: “Ideally, we want our results to be both realistic and faithful to the intent of the input sketch. For many sketches, we fail to meet one or both of these goals.” (Chen & Hays, 2018). The problem itself of synthesizing realistic images is part of the computer vision research category.

**Modified Turing test** The products created by the SketchyGAN system would not pass the modified Turing test. Apart from the fact, that the output images have a resolution of only 64x64 pixels, the results lack a certain realism, that is definitely detectable by humans (from the domain of digital art, as well as of the general population). When compared to realistic photographs or paintings that are scaled down to the same resolution, even the best results of SketchyGAN would be recognizable as coming from a machine, due to digital artifacts and distortions in textures and shapes. However, that does not mean, that the products could not provide any value or be appreciated in its own form.

**Appropriateness** The images SketchyGAN creates are definitely appropriate for the digital art domain. Not only, that the art created by artificial intelligence in itself is a category that people greatly appreciate, the products can be used in their own right outside of this field. Not necessarily in photorealistic rendering of objects, but rather in a context where distorted images of real objects are appreciated and desired (for example in an art piece of dreamscapes).

**Typicality** As in many domains of art, typicality is not easy to define, especially in areas that do not strive for photorealism and have no definite rules for its products. Arguably, the products are appropriate pieces of digital art and can be used and by some appreciated as such, therefore could easily be regarded as typical for its field. However, introducing an objective measure for such a wide field of art is almost impossible, therefore in order to give a definite response to this question, surveys would need to be conducted to objectively determine if the product is regarded as typical for its domain by the subjective evaluation of people (either among the normal population or experts from the digital art community).

**Novelty** First of all, in order to discuss the novelty of the product, a few rules need to be satisfied by its class, the domain of digital art: Art is unarguably defined by societal norms, cultural interpretation and appreciation of the people living in that culture. Consequently, humans can assess the quality of artifacts in this domain. The domain of digital art has existed for many decades, long before neural networks produced anything that could be regarded as art. Therefore it was not created or significantly shaped by the system, nor defined by its operation.[5] Also, the domain is basically infinitely large, countless digital artworks are created and recognized every day.

The products created by SketchyGAN's neural network arguably display a certain dissimilarity to most other works in the digital art domain. And there definitely exists a disparity between the produced images and the human interpretation of a sketch, which expects a more realistic result. Therefore, it can be said that SketchyGAN produce some amount of novelty in its results, due to the process that the sketch is interpreted by a trained algorithm instead of by a human who has a more complete understanding of the world, the displayed objects and the abstract concept of sketching.

The novelty of these products can be mostly attributed to random processes. It is hard to recognize the neural network having an intent on producing certain novel features that deviate from the expected photorealistic result. Rather, those features usually origin in distortions of shape and texture, that can be seen as non-intentional errors in the produced image, as the network normally strives towards producing realistic output instead of novel, creative interpretations of the original sketch.

Regarding the types of novelty: I think the products satisfy the definition of novelty by a set of knowledge, as the images appear neither too predictable and boring, nor too weird and crazy to be valuable. Novelty relative to complexity is hard to define, as the domain of digital art is basically infinitely big, a lot of which is therefore completely unexplored. While an uncountable number of digital artworks exist, this type of dreamlike products does not seem to be very common. Therefore this type of digital art has a certain novelty, even though it is nothing completely new and therefore not displaying transformational creativity. As already shown in the typicality section, the system sometimes produces results, that are surprising. This will be further elaborated in the section "Factor of surprise." And finally, for the perceived novelty, a definitive answer could only be given after an extensive study that analyses if experts of the digital art domain subjectively classify the products as novel. While it can be said with certainty, that they do not demonstrate transformational creativity, maybe some artists would find the results novel, surprising and even useful.

By the definition, that creativity lies between the input and the output, that the level of creativity is exhibited with the information that is added to this process, the system does demonstrate novelty. Since the output is only a simple and mostly deformed human sketch, the system has to show a lot of interpretation skill, understanding of the domain aesthetic and recognize the intent of the sketch author. In this way, SketchyGAN does demonstrate a level of understanding that could be interpreted as creativity.

**Value** The value and quality of such a product is difficult to determine, especially in the highly subjective art domain, where human appreciation mostly defines the value of a piece. Here as well, a survey between experts of the digital art domain would be the most objective measure in order to definitely say if the product has value. Of course, it can be argued that probably some artists would find the products valuable and would like to use them in their work, as well as some people who could appreciate the products on their own merit. In the same survey, the participants' emotional response to the products could be measured, when they see a product for the first time. This would give a very clear indication of the product's value, as especially for art, the

[5] This could not be said for the category of digital art produced by artificial intelligence, which is shaped by what modern research produces. But in this analysis, we will focus on the whole domain of digital art, as the product can be used and understood in this form, as any other artifact from this domain.

emotional response is crucial for its value. Boring artworks that do not have any effect on the observer usually are valued less than pieces that generate enormous emotions when viewed.

Unfortunately without such data, it is hard to determine how valuable this product could be for its domain and for society in general. I could only give my subjective opinion here[6], but the assessment of a singular person could only initiate a conversation about creativity in digital art, it would not add much to an objective evaluation of the system. A consensual assessment by survey therefore would be needed and should be explored in further research on the creativity of SketchyGAN.

**Factor of surprise** Due to the opaque nature in which neural networks operate, it is hard to predict what exact result they will produce from a given input. This is also the case with SketchyGAN. While the produced images strive for a realistic representation of the given sketch, the exact details are only known after the process. And because the network also incorporates error, distortions and digital artifacts into the product, they always contain a certain level of surprise, compared to the more realistic representation that the human imagination would expect from one of those original sketches.

It is hard to say if this level of surprise would be enough to consider the products a result of a creative system. Especially since this kind of distortion and visual errors is somewhat typical for generative adversarial networks. But this further depends on the environment, the audience and the desired aesthetic of the system. So here too, a survey of human emotional response would help to measure the level of surprise that people (or domain experts) exhibit when viewing the products for the first time.

## Process assessment

**Randomness** When generating its products, SketchyGAN's neural network appears to have some randomness in its process. And although the same input should yield the same result, an input sketch, that has been only slightly changed (like a minimally different shade of black in one pixel) might produce a completely different result. Overall, the products do orient themselves on the input sketch and often appear faithful to the intent of the sketch artist. Therefore, it could be said that the level of randomness in the process is at a healthy medium.

This randomness especially appears in the way in which the network deviates from the normal realistic representation of a sketch, that a human would expect, for example in distortion of the object, unrealistic textures or digital artifacts that are created during the generation process.

**Intention** Due to the opaque nature of a neural network, artificial intelligence research cannot yet explain, how exactly these systems make their decisions and produce the results. Therefore, it is hard to say where the intent lies in such a system. It could be, that all of the product is the intentional result of how the network understands the world given their knowledge base. Or on the other hand, it could be proposed that neural networks do not have any intent and simply imitate their training material in a more structured way, that considers the complete details and composition for imitation. This question also has a philosophical aspect as it begs for a definition of when and how machines can exhibit intent in a creative process.

SketchyGAN is incapable of evaluating its own process. However, it can evaluate its own products, as already discussed in the "Autonomy" section. Its discriminator network has the ability to evaluate and rate products for their realism and other qualities of the aesthetic it deems important, it is trained in a generation-evaluation cycle that aims for constant improvement of the system. This could be seen as an intent, to improve itself and generate results, that fulfill the domain's aesthetic.

**Level of creativity** As discussed further above, SketchyGAN demonstrates a certain novelty in its products, that only arguably might have a small amount of value. And it is hard to prove that the network generates its products with intent. Therefore the level of creativity of this system is in the "Generalization" stage.

SketchyGAN therefore has autonomy to create whatever "it wants" (with intent) or at least autonomy to change its process to create any possible combination of images from a source, some only more probable than others (without intent).

The system can generalize over all its training data that it was given and does not just memorize them. This would definitely be impossible anyways, as a neural network inherently lacks the capacity to memorize so many different combinations. Rather, the network learns a little bit from each example and generalizes over all parts of the knowledge base until it learned some general rules for sketch-to-image synthesis. It can create its products with variation in their creation. This is even proven by an example in the paper, where the same input with different noise vectors produced different images, that have a comparable composition, similar foreground, but completely different background, as the background was not indicated in the sketch and can be generated arbitrarily, as long as it still fits as environment for the object. In this case the system produced images of a bee with different flowers or leaf-like structures in the

[6] My own subjective opinion about the value of SketchyGAN's products: I find them interesting, see how they could be useful for digital artists (e.g. by incorporating them into a distorted dreamscape). But I do not intend to use the products on my own and do not know anyone who would probably utilize these artifacts for their creation or appreciate the artistic value of this type of product.

background, which are definitely plausible environments for a bee.

Furthermore, it could be argued that the system even reaches the level of “filtration” in the creative process. This would be the case if the process of image generation does have an intent behind it. And in this case, the filtration part of the system would be intrinsic in the network structure of the generative adversarial network. It is hard to prove that the system belongs to that category and it rather seems like it is part of the “generalization” level. However, this assessment could be the initialization of a further conversation about the topic of creativity in generative adversarial networks, as it seems difficult to objectively determine the level of intent and filtering such a system shows.

## A co-creative colleague version of SketchyGAN

In its current version, SketchyGAN imports an image of a sketch, calculates the corresponding product and output one singular product, which is a realistic representation of an input sketch. This process does not support any co-creative collaboration between human and system, as the human user needs to create the sketch in a different program (or by hand, then photograph it and increase the contrast, so that the background is pure white). This process is far from immediate and could be streamlined to not stand in the way of the users creative ideas.

**Real-time sketch manipulation** Instead of the user having to create the sketch in a separate program every time they want to make changes, the system itself should have a built-in application to create and edit sketches. Every time the user adds or removes a line, the system will update its product according to the new input. The user would then only need to open the program, create a sketch or edit an already existing one and could see their result immediately. This real-time interaction would enable the user to think less about the process of how they will feed their information to the algorithm and can easily have a real-time interaction with the system, making many creative changes along the way and seeing how the affect the resulting product.

**Transparent sketch overlay** To further move SketchyGAN from a coach to a colleague system, the distinction between input and output should be blurred. Users experience more interactive co-creativity, when they can directly interact with the system on the same canvas. So instead of separating input and output canvas, the input area (on the image the left canvas) will have a transparent overlay of the current version of the product. That means, the user can see the product directly under their own sketch and can interact more easily with it. For example, if some lines in the sketch are distorted, the user could utilize the system’s interpretation of the object to correct their sketch and create a more realistic representation of that object, which further enables the network to create better results. Of course, still a full-colored version, can be displayed on another canvas (right on the image), but the main focus lies on the interactive canvas that holds sketch and result at the same time. With this implementation, a user study about the networks delay after editing the sketch would be suggested, as it might confuse the user if the network has a slight delay on a slower computer, as this might hinder the user’s real-time interactive experience.

**Multiple product suggestions** When humans exhibit co-creative collaboration, they like to brainstorm for many different ideas and suggestions before agreeing on one. This fuels their creativity, as it helps them compare various directives, instead of only chasing one specific goal. This principle can also be applied to SketchyGAN. In their paper (Chen & Hays, 2020), the system’s researchers showed an example of different products created from the same input image, with only slightly different noise vectors (see section “Level of creativity” of the process assessment). This variation could be utilized by displaying multiple suggestions of the given sketch to the user. The user could then select the sketch, that they like best or that fits their directive the most instead of only being able to deal with one product. These suggestions could be displayed on top of the main result and the user could select a different suggestion by clicking on it.

**Filtering by color points** One problem that arises with the interactive design of SketchyGAN is that the user might have selected one product they like, for example where the background has a certain property (e.g. color of a flower behind the bee, castle next to a river). Due to the nature of sketches, these properties, especially ones of the background, are usually not expressed within the sketch, as this input only is a very crude and abstract representation of an object. Therefore, if the user would add that property to their sketch, the system would most probably not understand it as such a desired background feature. Therefore, after selecting a different product suggestion, the user could not make any more changes to the sketch, as this would prompt the network to produce completely new results, discarding the preference, the user had for one specific type of image.

To solve this problem, it is suggested, that the user should have the ability to add color points to their sketch, in order to express a preference for a certain color in this part of the image. The results of the system could then be filtered to only allow product with this feature (a similar color in this area of the image) and the user could continue editing their sketch without losing the desired preference for a certain feature. In this way, they would also have influence over the background of the image and other properties of the foreground object without just having to accept one of the system’s suggested products. This would give the user more creative freedom and the system the ability to make better suggestions that fit the user’s preferred aesthetic.

Since too many color points in one image could make it difficult to find an appropriate result when filtering, the system could also be trained on recognizing these color points as input and expressing them in the output image. For this

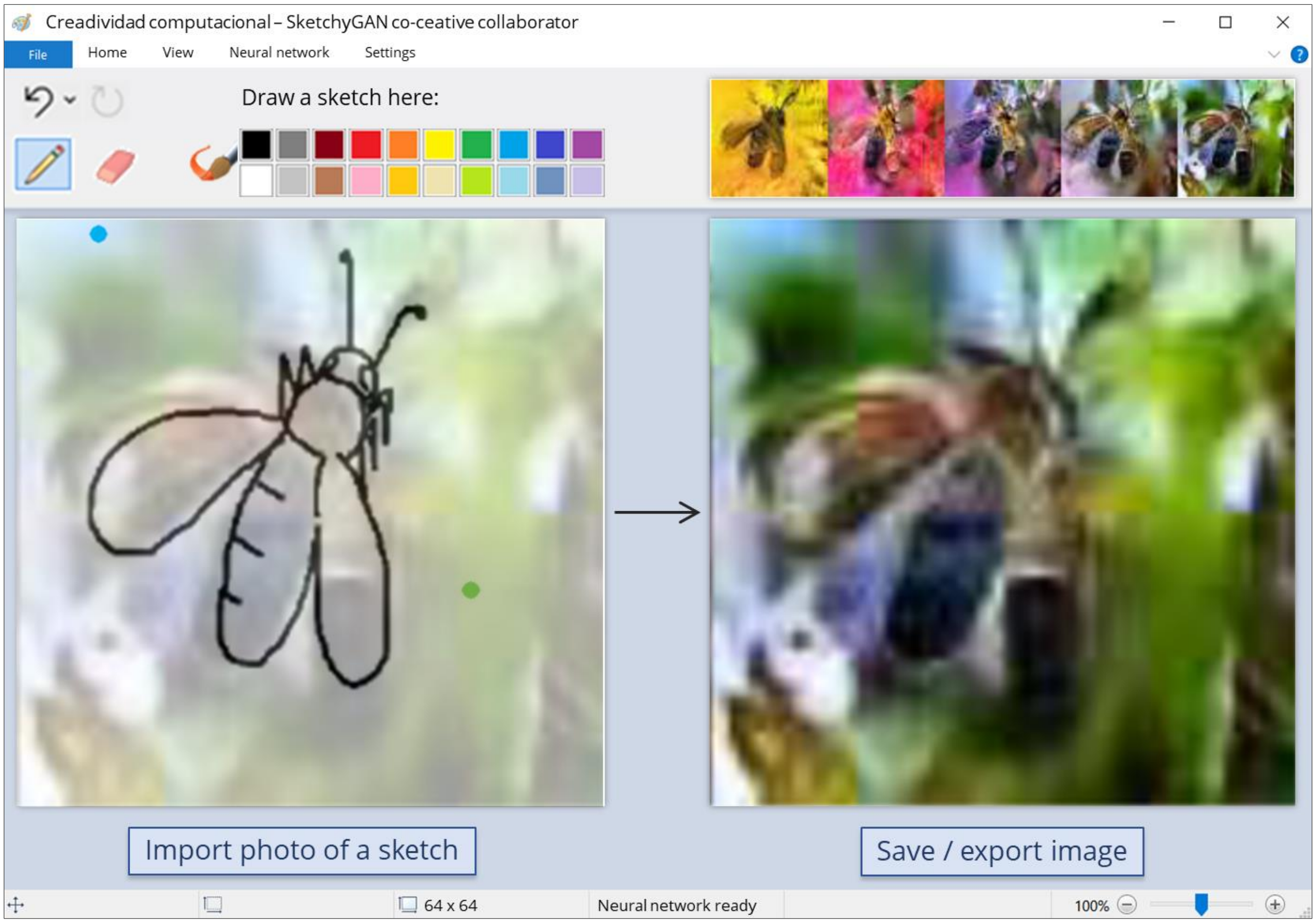


Figure 1: Suggestion for the graphical user interface of a modified version of SketchyGAN, which acts as a co-creative colleague system.

process, during the later training phase of the neural network, the input sketch should include data about a random number of randomly chosen color points (with specified deviations in the color value, as the network does not need to match the exact color shade). This expanded training could then be used to train the network further. A quantitative evaluation of the products according to their color points is then trivial, as it only needs to be checked, that the color in that part of the image matches the specified color point (with some allowed deviation). If these criteria are added to the neural net's fitness function, the training process can include color information and the network can learn how to incorporate these features into its output, while still striving to provide a realistic product.

**Co-creativity and collaboration** This modified version of SketchyGAN would satisfy the criteria of a co-creative colleague system:

The user and system contribute at the same time on the same canvas and shape their directive in reaction to suggestions of the other agent. When the user changes their sketch or expresses a color preference in one part of the image, the system reacts accordingly and suggests modified results. And from these suggestions, the user might consider new ideas, change their directive or confirm suggestions the system has made. This direct interaction enables more ideas and leads to a co-creative system. The creative product that emerges in the end cannot anymore be attributed to only one agent, as both have worked together in the process to create that product.

While the system generates the images and the user only modifies the sketch (with given color preferences), the tasks are intertwined and influence each other, a complete division of work disappears.

This interaction happens in real time, as the system reacts to the users input quickly. However, the user has time to think about what changes they want to make next, so one could say that this interaction is turn-based, since the network does not change the products (or the other product suggestions) while the input stays the same. So while the system only reacts once per input change, the user can keep modifying their sketch while the system is still processing and suggesting new ideas.[7]

**Graphical user interface design** The design should mainly focus on the human-computer interaction. Its purpose is to enable co-creativity, in whatever way the user prefers to use it and not stand in the way of the creation. In Figure 1, I suggest a design that would fulfil these criteria and incorporate the changes and features that have been discussed in this section. A focus is put on usability, while advanced options for configuring the network, exporting sketches and changing view or size are hidden away behind menus, so the user can focus on their creation and the interaction with the system.

In the top bar, menus for configuration, neural network training, settings, file import/export and various other preferences can be found. The user can either import a sketch that has already been created and saved or start drawing. For drawing, a pencil and eraser work to create the sketch itself, the brush can be used to express color preferences at certain points in the image (as discussed above). Of course, undo and redo are available to give the user full creative control of their process and minimize doubts about making changes and trying out ideas. On the right half of the application, the system suggests various different results and displays the main results (that best fit the aesthetical preferences) in the right canvas and as an overlay on top of the left canvas for improved interaction. When finished, this product can easily be saved and exported as a digital image file.

## Conclusions

As demonstrated in this paper, it is hard to define whether a generative adversarial network can be counted as creative or not. Depending on the definition of creativity and philosophical ideas about autonomy and intent of such an autonomous system, SketchyGAN and similar systems could be seen as creative in its own right. But it is almost impossible to determine with certainty if neural networks demonstrate an intuitive understanding of the domain and its aesthetic. However, there are still obstacles and limitations that need to be overcome to inarguably grant these systems the description of having creativity.

The products they produce can have value and definitely display the potential, that artificial intelligence has to create creative artifacts. While novelty still is hard to find in their results, a certain level of surprise can be observed in the results of neural networks. The performance of these systems and therefore the value and quality of their products will only improve in the next years, until it might one day be obvious that computers can exhibit the same level of creativity as humans or even more.

Further studies have been suggested in this paper, in order to determine how humans of the general population and experts in the domain of digital art react to SketchyGAN's products, if they appreciate them and if they consider them creative, novel, valuable, surprising and typical for its domain.

There definitely exists potential in a system like SketchyGAN to change it into a co-creative colleague system that can be used in real-time together with a human to create creative products. This would enable a more fluent human-computer-interaction in which both agents could create the product together, with ideas being inspired by the other agent. A concept for these changes, including a draft for a graphical user interface, is outlined. These concepts could easily be transferred to many other generative adversarial networks to give more potential for creative contributions.

## References

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial intelligence*, *103*(1-2), 347-356.

Boden, M. A. (2007). Creativity in a nutshell. Think, 5(15), 83-96.

Chen, W., & Hays, J. (2018). Sketchygan: Towards diverse and realistic sketch to image synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 9416-9425).

Chen, W. (2020, July 14). SketchyGAN. Retrieved August 06, 2020, from https://git.droidware.info/wchen342/SketchyGAN

Gao, C., Liu, Q., Xu, Q., Liu, J., Wang, L., & Zou, C. (2020). Image Generation from Freehand Scene Sketches. arXiv preprint arXiv:2003.02683.

Guo, L., Liu, J., Wang, Y., Luo, Z., Wen, W., & Lu, H. (2017, October). Sketch-based image retrieval using generative adversarial networks. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 1267-1268).

Jennings, K. E. (2010). Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines*, *20*(4), 489-501.

Liu, R., Yu, Q., & Yu, S. (2019). An unpaired sketch-to-photo translation model. *arXiv preprint arXiv:1909.08313*.

[7] If the user is fast enough. This depends on the performance of the system and the hardware that is used to execute the system. Also, the implementation should be done with process threads, so that the user can keep editing the sketch while the system is processing changes and calculating new results. If the user had to wait for the computer to "finish their turn," this would only hinder the creative flow of the user and not help a co-creative, immersive interaction.

Lu, Y., Wu, S., Tai, Y. W., & Tang, C. K. (2018). Image generation from sketch constraint using contextual GAN. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 205-220).

Osahor, U., Kazemi, H., Dabouei, A., & Nasrabadi, N. (2020). Quality Guided Sketch-to-Photo Image Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 820-821).

Sangkloy, P., Burnell, N., Ham, C., & Hays, J. (2016). The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (TOG)*, *35*(4), 1-12.

Shen, Y., Gu, J., Tang, X., & Zhou, B. (2020). Interpreting the latent space of gans for semantic face editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9243-9252).

Ventura, D. (2017). How to build a CC system paper type: System description paper. ICCC.