Before A Computer Can Draw, It Must First Learn To See

Derrall Heath and **Dan Ventura**
Computer Science Department
Brigham Young University
Provo, UT 84602 USA
dheath@byu.edu, ventura@cs.byu.edu

Abstract

Most computationally creative systems lack adequate means of perceptually evaluating the artifacts they produce and are therefore not fully grounded in real world understanding. We argue that perceptually grounding such systems will increase their creative potential. Having adequate perceptual abilities can enable computational systems to be more autonomous, learn better internal models, evaluate their own artifacts, and create artifacts with intention. We draw from the fields of cognitive psychology, neuroscience, and art history to gain insights into the role that perception plays in the creative process. We use examples and methods from deep learning on the task of image generation and pareidolia to show the creative potential of systems with advanced perceptual abilities. We also discuss several issues and philosophical questions related to perception and creativity.

Introduction

Some people seem to have a natural talent for drawing, while others only wish they could draw well. Many of these people have turned to books and teachers to help them develop their drawing skills. One of the most widely used and consistently successful books for teaching people how to draw is titled *Drawing on the Right Side of the Brain* (Edwards 1989). This book uses insights from neuroscience to help potential artists improve their drawing skills. One of the main premises in the book is that drawing is not a skill of hand, paper or pencil, but a skill of perception. To quote from the book:

> "The magic mystery of drawing ability seems to be, in part at least, an ability to make a shift in brain state to a different mode of seeing/perceiving. When you see in the special way in which experienced artists see, then you can draw... Drawing is not very difficult. Seeing is the problem, or, to be more specific, shifting to a particular way of seeing."

This idea can extend to any kind of creative ability. Before one can create visual art, compose music, or invent recipes, one must first learn to see, hear/listen, or taste, respectively. Even creative tasks like writing poetry must ultimately be grounded to what has been experienced through perception. Our ability to perceive influences how and what we create. Just as drawing is really about perceptual skills, our ability to think creatively and do creative things heavily depends on how we perceive and understand the world.

In his book, *The Anthropologist On Mars*, Oliver Sacks recounts the story of Shirley Jennings, who had been blind since early childhood and had surgically regained his sight at the age of 50 (Sacks 1995). After the operation, he could not immediately see, could not recognize his family, could not pick out objects, and struggled with depth perception. Over several months, his brain had to learn how to see and make sense of an incredible amount of new information. It was a slow and difficult process reconciling his non-visual mental model of the world with this new form of perception. As he learned to make use of his new sense, things that were aesthetically beautiful to him differed from those others found pleasing. He eventually took painting lessons and created paintings that demonstrate his unique taste in visual art[1].

Shirley's case, along with several other vision disorders and anomalies like blindsight (Weiskrantz 1996), Capgras syndrome (Ellis and Lewis 2001), and agnosia (Farah 2004), has helped to uncover the work and learning that our brain undertakes in order for us to perceive and understand the world. In this paper we argue that the ability to perceive is a necessary and influential piece of the creative process. It enables us to learn a mental model of the world, to understand and continually evaluate our own creations, and to infuse what we produce with meaning. Indeed, even perception itself is a creative act that our brains regularly perform, although often subconsciously. The necessity of perception applies to the field of computational creativity, in which one of the goals is to build computational systems that can autonomously create art. Before a system can learn to create art, we argue that it must first learn to perceive.

We proceed by exploring the relationship between perception and creativity and then discuss the role of perception in computational systems. We then consider how state-of-the-art computer vision methods can enhance the creative potential of systems designed for visual art. We demonstrate, using deep neural networks, how perceptual skills facilitate imagination and can lead directly to generating novel images. We then elaborate further on why perception itself is a creative process and demonstrate a form of creative perception, called pareidolia, using deep neural networks. Finally, we discuss philosophical issues and the implications of our ideas and elaborate on what more advanced perceptual abilities could mean for the future of computational creativity.

[1] http://www.atfirstsightthebook.com/shirls-paintings.html

Perception and Creativity

When talking about visual art, Csíkzentmihályi says, "...the aesthetic experience occurs when information coming from the artwork interacts with information already stored in the viewer's mind..." (Csíkzentmihályi and Robinson 1990). In other words, the viewer's appreciation (perception) of art is determined by his current mental model of the world. Likewise, the artist has her own mental model of the world and created the artwork to convey meaning according to that mental model. How was that mental model established? It's reasonable to say that it was learned through a lifetime of experiences, and people experience the world through perception. Everything we know and understand about the world has come through our senses. Every memory and every thought we have is in terms of what we have experienced in the past (Barsalou 1999).

It is difficult to comprehend what life would be like without perception because it is so fundamental to how we think. Would it even be possible to think, imagine, or create anything at all without some kind of input? There is no definite answer to that question; however, studies of long term sensory deprivation and solitary confinement suggest significant mental deterioration (Grassian and Friedman 1986; Allen, Celikel, and Feldman 2003). Perception directly influences our ability to think and understand, and the better and more varied our perceptual abilities are, the more we are able to think about, imagine, and ultimately, create. We can take this idea further and say that, with our current senses, there are thoughts we cannot think simply because we lack additional (or adequate) senses to know how to think them. To quote Richard Hamming (Hamming 1980):

> "Just as there are odors that dogs can smell and we cannot, as well as sounds that dogs can hear and we cannot, so too there are wavelengths of light we cannot see and flavors we cannot taste. Why then, given our brains wired the way they are, does the remark 'Perhaps there are thoughts we cannot think,' surprise you? Evolution, so far, may possibly have blocked us from being able to think in some directions; there could be unthinkable thoughts."

We need perception (i.e., input) in order to build a mental model that can facilitate thinking, which can then facilitate creativity (Flowers and Garbin 1989). Indeed, as noted earlier in the case of Shirley Jennings (Sacks 1995), the mental model itself is what does the perceiving. Our eyes merely translate light into brain signals, but it is our brain that must learn to make sense of that information, which then allows us to think in those terms.

Imagination is clearly tied to this idea and is closely linked with creativity in cognitive psychology literature (Gaut 2003). Imagination is typically generalized as thinking of something (real or not) that is not present to the senses. Most psychologists agree that our perceptions (senses), our conceptual knowledge, and our memories make up our mental model and form the bases of imagination (Barsalou 1999; Beaney 2005). As we perceive the world and have experiences, our mental model is formed by establishing and strengthening connections in our mind. These connections form concepts, which are in turn interconnected. Creative imagination is achieved by combining these connections and experiences in different ways that produce novel results.

Thinking Beyond Natural Perception

It is possible for us to indirectly experience things outside of our perceptual abilities by translating other modalities into our range of senses. For example, we visualize infrared light by shifting it into the visible spectrum. We create charts and graphs that represent data we cannot observe directly, like barometric pressure, or electromagnetic fields. In this way we can vicariously think in terms of other modalities and perhaps even be creative in those modalities.

This idea is applied explicitly in the case of *sensory substitution* (Bach-y-Rita and Kercel 2003), where one sense can take the place of another that has been lost. For example, devices have been made that can allow blind people to literally "see" with their tongue. They work by mounting a video camera on the blind person's forehead, which sends video data to a plate that sits on the person's tongue. This plate contains a grid of "pixels" consisting of pressure points. These pressure pixels correspond to grayscale video by pressing harder where the image is brighter and pressing lighter where the image is darker. The tongue can then "feel" the video information and, after several months of training, a blind person's brain starts to see images in their mind. It's certainly not high resolution, but it's enough to allow a blind person to read large print text and navigate new terrain.

Another way that we humans can communicate and understand things that we ourselves have not perceived directly is through language. In other words, through verbal/written communication we can experience by proxy what others have directly experienced (Zwaan and Kaschak 2008). In this case, language acts as an analogy between two people's experiences. Our interpretation of a described experience must still be grounded by our personal perceptions and experiences (Barsalou 1999). For example, it is very difficult to describe colors to a congenitally blind person because colors are inherently visual and the blind person has no visual grounding at all. This is why even creative literature and poetry also require perception—the writer must have experiences to write about and the reader must have experiences with which to interpret the writing.

Art, whether it be visual, written, musical, etc, acts as a metaphor between the experiences of the artist and the experiences of the receiver. Successful artists are creative because they have a unique perspective on the world that they are trying to communicate through their art, and people appreciate art that helps them gain new perspectives. In other words, having unique experiences and perceiving the world differently plays a role in the creative process. It has been postulated in cognitive psychology that creative people literally see the world differently (Flowers and Garbin 1989; Berns 2008), which is in turn why they tend to think differently and can produce novel things and ideas.

Quality of Perception Affecting Visual Art

There have been studies analyzing several famous artists with documented visual impairments (Marmor and Ravin 2009). For example, Claude Monet developed cataracts, while Edgar Degas began to suffer from retinal disease. These studies point out that the earlier works of these painters (when they had good eyesight) were better formed and detailed, while later works (made with poorer eyesight) became more and more abstract. These studies generally conclude that the failing eyesight of the artists *did* have a large impact on the quality and style of their work. Although some researchers say that this was not necessarily a bad thing, and some artists would use their visual impairments to their advantage by removing their corrective lenses for certain paintings.

These artists had issues with just their eyes, but what about *cognitive* impairments involving vision? How do different cognitive disorders of the brain affect artists' work? This question was explored by Anjan Chatterjee, where he reported on multiple studies analyzing the drawing ability of several artists with various cognitive disorders, including spacial neglect, visual agnosia, epilepsy, TBI, etc (Chatterjee 2004). The results for many of the disorders, like epilepsy were mixed, but artists with disorders more specific to vision, like agnosia, had some notable peculiarities to their drawing ability.

For example, one artist with a type of visual agnosia could create beautiful drawings as long as the item he was drawing was continuously right in front of him. However, if he was asked to draw from memory (e.g., draw a 'bus'), then his drawings were simplistic and often unrecognizable. Another artist, with a traumatic brain injury, produced drawings that were more abstract and "expressive" than drawings produced before the accident. Although these studies appear anecdotal due to the rarity of many of these disorders, it is apparent that how the brain sees and understands the world does affect the ability to draw.

There are several cases of successful artists who are blind, and one artist in particular has received a lot of attention because he is *congenitally* blind (Amedi et al. 2008). This blind man has a remarkable ability to paint and draw pictures that are consistently meaningful to sighted people. He uses special paper and pencils that form ridges that he can feel as he draws. He first explores an object with his hands and then, remarkably, can draw it from different perspectives. He's never been able to see, yet can understand perspective. His case provides insight into how the brain perceives and builds invariant mental representations of the world.

The blind artist's case is related to sensory substitution, where a blind person can "see" through touch, and further supports the idea that the brain is what processes and makes sense of perceptual input. Researchers who study blindness and visual art have indicated that vision and touch are linked and make use of similar processes and similar features in the brain (Kennedy 1993; Kamel and Landay 2000). The brain can do remarkable things even when the quality of the input signal is disrupted or re-routed. Perception is really about being able to build these mental models and using them to interact with the world. It's not that *visual* art requires *vision*, but that creating visual art requires *some* form of perception that establishes and continually informs the artist's mental model.

Perception and Computational Creativity

We've discussed the role of perception in *human* creativity, but what about computational creativity? Certainly, there's no requirement that computers can only be creative in the same way as humans. However, we are positing that perception is *fundamentally* a necessary component of the creative process. So, just as perception is important for human creativity, perceptual ability is also important for computational creativity. The exact methods of perceiving and creating may be different than those of humans, but some form of perceptual grounding is requisite for a truly creative system.

Colton proposed the creative tripod as necessary criteria for a creative system (Colton 2008). A creative system must have imagination, which is analogous to producing novel artifacts; it must have skill, which corresponds to generating quality artifacts; and it must have appreciation, which is the ability to recognize the novelty and quality of its own artifacts (i.e., self-evaluation). In other words, there must be a perceptual component that directs the creative process by helping the system explore new ideas (imagination), and understanding which ideas are worth pursuing (appreciation).

Many creative systems exist across several domains that can generate novel artifacts. Most of these systems, however, are merely mimicking example human-created artifacts without understanding or appreciating what they are producing, like a parrot mimicking human speech. For example, the PIERRE system generates new crockpot recipes according to a model trained on user ratings of existing recipes, but it has no sense of what the recipes actually taste like, only that humans have liked similar recipes (Morris et al. 2012). In music, there are several systems that analyze patterns and n-grams from existing melodies, then probabilistically draw from those distributions or construct grammars when producing music (Cope 1996; Pachet and Roy 2011). Likewise, poetry systems are also often based on corpora and n-gram distributions, without much understanding of what the words actually represent (Colton, Goodwin, and Veale 2012; Netzer et al. 2009).

Other existing creative systems produce artifacts according to hand engineered metrics and databases, where the ability to appreciate and perceive what is produced is limited to those explicit metrics. For example, some musical systems rely on rules and metrics based on musical theory in order to produce and evaluate melodies (Ebcioğlu 1988; Melo and Wiggins 2003). Visual art systems often use some form of evolutionary algorithm for producing art, which involves a fitness function by which the art is evaluated at each iteration. The fitness functions in these systems are usually based on models trained using extracted image features in order to evaluate aesthetic quality or novelty (Machado, Romero, and Manaris 2007; DiPaola and Gabora 2009). In these cases the perceptual ability is ultimately limited to those specific features.

There are some creative systems that do attempt to incorporate a sophisticated model of perceptual ability. For example, there is a system that invents recipes based on actual chemical properties of the individual ingredients (Varshney et al. 2013). It at least has some understanding of what would actually taste good in a recipe and isn't limited to just producing something that is mimicking human examples. The DARCI system extracts various image features and trains neural networks to evaluate how well the images convey the meaning of particular adjectives (Heath and Ventura 2016). Although DARCI still relies on extracting specific low level features, it at least attempts to learn the semantic qualities related to those features (in the form of adjectives). In this way DARCI, more than other visual art systems, is able to at least partially perceive *meaning* in the art that is produced.

The example systems just described can produce interesting and novel artifacts. However, without advanced perceptual abilities, the systems lack any notion of understanding and intentionality. The systems can produce something, but can't necessarily tell us why, or what it means. They are instances of Searle's Chinese room (Searle 1980), that simply follow rules and algorithms, with no comprehension of what is taking place. Just as humans cannot think beyond our perceptions, computational systems cannot think beyond theirs. Some have argued that even human thought and creativity is subject to the Chinese room analogy at the biological (cellular) level. This may be true, but if we aim to build systems that can be creative at a human level, then they must at least have human-level perception.

Somewhat surprisingly, in the case of visual art, current creative systems rarely use state-of-the-art computer vision techniques, like deep neural networks. Certainly having more advanced perceptual abilities would improve the quality of their art by enabling these systems to understand more concrete things. For example, a system could conceivably create an original image of a dog, if it knew how to see and recognize dogs. It seems, then, that incorporating advanced computer vision techniques, especially ones tied to semantic understanding, should be a high priority in the field of computational creativity.

Visual Art and Deep Learning

The last few years have shown a resurgence of *deep neural networks* (DNNs), especially for computer vision tasks, where they hold current records for several vision benchmarks (Farabet et al. 2013; Szegedy et al. 2015). Deep learning has the potential to significantly improve visually creative systems as well. A key advantage of DNNs is that they are capable of learning their own image features, while the visual art systems described above all rely on manually engineered features. Thus, deep learning models can provide more advanced perceptual abilities by building better "mental" models of the world.

Some of these deep learning models can already be used directly to improve current artistic systems. In recent work on DARCI, we built a sophisticated semantic model that uses a shallow neural network to associate image features with a vector space model (Heath and Ventura 2016). Here we can show significant improvement by replacing the shallow neural network and extracted features, with a DNN (and the raw image pixels as input). Specifically, we used a deep learning framework, called Caffe (Jia et al. 2014), and started with the CaffeNet model, which was first trained to recognize 1000 different items using the ImageNet 2012 competition data (Russakovsky et al. 2015). We then further trained and fine-tuned the model on DARCI's image-adjective dataset (with a vector space model).

	Random	*DARCI*	*Deep Network*
Coverage	0.709	0.444	**0.202**
Ranking Loss	0.502	0.199	**0.102**

Table 1: Zero-shot image ranking results comparing the DARCI system with our modification of DARCI that uses a deep neural network (lower scores are better). We used the same test set from the original DARCI paper (Heath and Ventura 2016). The use of a DNN improves the system's ability to perceive and understand adjectives in images.

The DARCI system is capable of zero-shot prediction (using the vector space model), meaning it can successfully evaluate images for adjectives that it was not explicitly trained on. We compare DARCI's original results (Heath and Ventura 2016) with our deep neural network version in Table 1. The results show significant improvement using the DNN to evaluate images, and fully incorporating a deep model into the DARCI system will likely help it to produce more semantically relevant images.

In fact, DNNs have already been used to generate images directly (Denton et al. 2015; Gregor et al. 2015; Leon A. Gatys and Bethge 2015). One particular method, called *gradient ascent* (Simonyan, Vedaldi, and Zisserman 2013), works by essentially using the DNN in reverse. The trained network starts with a random noise image and tries to maximize the activation of the output node corresponding to the desired class to generate. The network then backpropagates the error into the image itself (keeping the network weights unchanged) and the image is slightly modified at each iteration to look more and more like the desired class.

We demonstrate gradient ascent using the same deep model that we trained with the DARCI image-adjective data set, and the resulting images can be seen in Figure 1. These images can be thought of as visualizations of the features learned by the model for each adjective. Each adjective's features seem fairly general, except in the case of 'peaceful', where the visualized features are consistent with the fact that most of the training images depict calm beaches. It is theorized that imagination in humans can be partially thought of as running our vision processing systems in reverse (Barsalou 1999), in which case our deep neural network is analogously demonstrating its own kind of imagination.

The generated images seem fairly abstract, which is expected for adjectives, especially since the DARCI data set contains a wide variety of scenes, objects, genres, and styles for each adjective label. Deep neural networks are becoming powerful enough to render actual recognizable objects using the gradient ascent method. The ImageNet 2012 competition consists of classifying 1000 different categories of objects ranging from various animals, to clothing, to household items. We took the CaffeNet model (used as the base for the DARCI model), as well as another successful model called GoogleNet (Szegedy et al. 2015), and generated several images depicting objects from the 1000 possible cate-

Figure 1: Four images generated using gradient ascent from the deep neural network trained on the DARCI dataset. From left to right the images were generated for the adjectives 'vibrant', 'cold', 'fiery', and 'peaceful'. These images are essentially visualizations of the features that the model has learned and demonstrate a form of imagination.

gories. The resulting images for several objects can be seen in Figure 2.

While the images are not photo-realistic, they are original and do resemble the intended item. Notice how the two models generated images with different styles as each model learned different features. The generation of images using DNNs is currently an active area of computer vision and machine learning research, and several researchers have produced impressive results (Denton et al. 2015; Leon A. Gatys and Bethge 2015). The field of computational creativity has yet to significantly leverage the potential of deep learning, although some have alluded to it (Heath, Dennis, and Ventura 2015). However, some researchers have already begun incorporating deep learning into evolutionary art systems that are capable of rendering images that resemble concrete objects, with interesting results (Nguyen, Yosinski, and Clune 2015).

To See Is To Create

We have argued that perception is an important aspect of creativity and that more advanced perceptual abilities can lead to more sophisticated creative systems. We also argue that perception is a creative act in its own right. When light hits a person's eyes, it is converted into signals, which travel to the visual cortex via the optic nerve. The brain itself does not receive any light, only information about the light. The brain must then learn to make sense of that information, and an image in the mind is fabricated, and that is what a person "sees". Our brain over our lifetime has built a mental model of the world through the various signals it has received from our senses. This mental model is what determines our personal reality, and it is an impressively creative act (Hoffman 2000; Peterson 2006).

We do not think of perception itself as a creative act because it happens instantly, constantly, and seemingly without effort. We take for granted how difficult perception is because it is an ordinary part of life, and we have become desensitized to it. However, even the most advanced state-of-the-art computer intelligence cannot process visual information as well as a child can almost instantaneously. The case of Shirley Jennings (Sacks 1995), in which he spent months learning how to see for the first time at age 50, and other cognitive visual disorders, shed light on the tremendous amount of work that goes into vision.

Figure 2: Images generated using gradient ascent from the CaffeNet model and the GoogleNet model, both trained on the 2012 ImageNet challenge data. The first two rows of images are from CaffeNet and, from left to right, were generated for 'pool table', 'broccoli', 'flamingo', 'goldfish', 'bald eagle', 'lampshade', 'starfish', and 'volcano'. The last row of images are from GoogleNet and were generated for 'bald eagle', 'tarantula', 'starfish', and 'ski mask'. These original images are certainly not photo realistic, but it is still fairly easy to identify each image's subject. Notice that the two models have different styles because they have learned different features.

Optical illusions also provide insights on how the brain understands visual input and constructs images in the mind (Hoffman 2000). Different people given the same input, experience it differently. A person's subjective experience is unique to them, an act of novelty by their creative brain. This idea became even more evident when a particular image of a dress sparked huge debate on social media over the color of the dress (Lafer-Sousa, Hermann, and Conway 2015). Some saw white/gold, others saw blue/black, because our brains construct differing realities based on our mental models.

If we accept the idea that our brains are doing the actual creating of the images we see, then what is the artist doing when she paints a picture? The artist is providing a set of constraints, in the form of a painting, that viewers use to create an image in their minds. The more realistic a painting is, the more it constrains the viewer to mentally create it a certain way. The more abstract or ambiguous the painting is, the less it constrains the viewer, and the more variety and novelty in the individual aesthetic experiences.

Pareidolia

Attributing creativity to a system just because it has some perceptual abilities does not appear very compelling. However, there are some perceptual tasks that seem more creative than others. Pareidolia is the phenomenon of perceiving a familiar pattern where none actually exists. For example, seeing constellations in the stars, faces in ordinary things, objects in blotches of ink, or shapes in the clouds. Sometimes these are considered mistakes or optical illusions, but they can actually be a deliberate act of creativity. When a child says a cloud looks like a particular animal, we admire her imagination, especially when we can then see the shape too. We obviously know it's a cloud, but we have chosen to see it as something else.

Pareidolia is a creative act because it is is not about seeing things for what they are but seeing things for what they could be. Creative systems capable of pareidolia may have applications in visual communication, advertising, story telling, illustration, and non-photo-realistic art. As it stands, there are few computational systems developed for automatically performing pareidolia. One group of researchers developed a system for recognizing "faces" in ordinary pictures and then automatically determining the emotion expressed by the "face" (Hong et al. 2013). Here we demonstrate how deep learning can be used for pareidolia and argue that it is a form of creativity because the model is interpreting images in novel ways.

Finding Faces Seeing faces in objects is by far the most common type of pareidolia and provides a simplified version of the problem to begin with. The initial task is to use a deep neural network (DNN) to identify what aspects of an image could make up a face. We then have the DNN iteratively emphasize those features, using gradient ascent, until the "face" that the network sees emerges in the image. We use two different DNN models trained on faces. The first is called VGG-Face and was trained to recognize the faces of over 2500 different celebrities (Parkhi, Vedaldi, and Zisserman 2015). The second model, which we'll call AGE-Face, was trained to determine the age of a person (one of eight age ranges) in a provided image (Levi and Hassner 2015).

We perform pareidolia by having each network, when given an image, determine the output node (corresponding to a class) with the highest activation. The model then performs the gradient ascent algorithm in an attempt to increase that node's activation further, thus emphasizing the strongest features it found initially. Figure 3 shows example pareidolia images generated with both the VGG-Face and AGE-Face networks. The networks generally do a decent job of drawing (cartoony) faces on the source images in ways that make sense, although some are harder to appreciate. The VGG-Face model tends to draw more realistic facial features (i.e., eyes, nose, etc) than the AGE-Face model. However, the VGG-Face model will often highlight isolated facial features (especially when a face in the source image is not apparent to humans), while the AGE-Face model tends to keep the facial features together for a full face.

Finding Objects We now move on to a harder version of pareidolia in which we ask the model to find and highlight any kind of object in an image. We again use the CaffeNet model that was trained on the 1000 category 2012 ImageNet data; thus the model could potentially see any of those 1000 items in an image. We use the same method as just described in the faces version. The model is given a source image, then performs gradient ascent on the source image in order to further maximize the highest activated output node. We

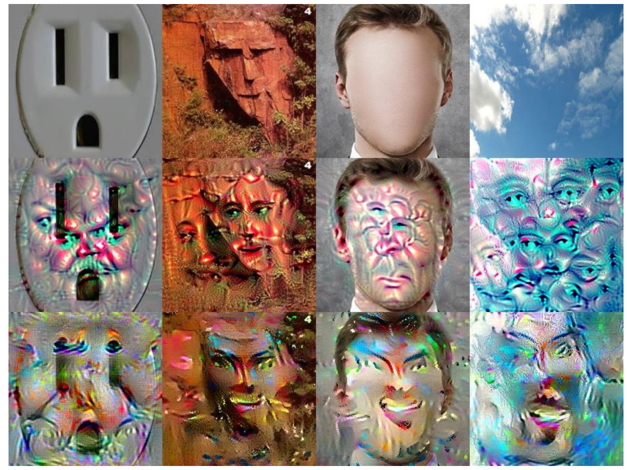

Figure 3: Images created for face pareidolia using deep neural networks. The top row are the source images, the second row are faces highlighted by the VGG-Face model, and the third row are faces highlighted by the AGE-Face model.

Figure 4: Images for object pareidolia using CaffeNet, trained on the 2012 ImageNet data for 1000 object categories. From left to right, the items highlighted in the images (bottom row) from each source image (top row) are 'mask', 'arctic fox', 'scorpion', and 'ringworm'.

applied this method to several source images, and the results can be seen in Figure 4.

For some of the examples it is easy to see why the model did what it did. For instance, it is understandable how the [Figure 4, 1st] source image looks like a mask, and we can see how the modified image came from it. However, it is more difficult to appreciate how the model saw an 'arctic fox' in the [Figure 4, 2nd] source image. Other examples are hard to relate to initially, but on inspection, we can start to see the connection. For example, the [Figure 4, 4th] source image looks, to most humans, like a spider, but the model saw it as a ringworm. After considering the resulting image, we can at least appreciate why the model thought ringworm.

This leads to an interesting discussion about perception and creativity. If a person says that a particular cloud looks like a horse, then *if we can also see it*, we think the person has imagination. However, if we can not see it ourselves, then we do not necessarily praise the person's imagination.

Conversely, if a person says that a photo of a horse looks like a horse, we also do not admire the imagination, and we end up wondering why they bothered to say something so obvious. We appreciate creativity when it is different from the norm, but not so different that we cannot connect.

When it comes to visual art, how a person sees will influence their art; thus, people that see things differently (but not too differently) can potentially be more creative with their art. Using deep learning models for pareidolia helps us to understand how these models are actually seeing, and it helps us to visualize what features are being learned. The features learned by each model are likely different than the features that human brains use when processing visual input. This is why the CaffeNet model sees an arctic fox in the [Figure 4, 2nd] source image, but most humans would say it looks like an elephant.

If a computational system perceives things differently than a human, and accordingly produces different kinds of art, then is the art only viable if we humans can relate to it? It has been suggested that the most creative and influential people are ones that see (and therefore think) differently (Flowers and Garbin 1989; Berns 2008), and Colton argues that computational systems that see differently than humans have enhanced creative potential (Colton et al. 2015), but is that true only to an extent? Could a computational system that perceives differently (even radically differently) than humans actually help us to extend our notions of what constitutes good art?

To go even further, could we build a system capable of understanding and creating art beyond the capabilities of current human perception? For example, could we build a system that creates infrared art? Or electromagnetic field art? Or gravity art? Or some other kind of art? Would there be any purpose in doing so? Or perhaps augmenting computational systems with other forms of perception could help them gain a richer, deeper understanding of the world, and allow them to create visual art that can be even more meaningful to humans.

Conclusion

We have argued that perceptual abilities are fundamental to the creative process. We have discussed the relationship between perception and creativity from a cognitive psychology perspective and also in terms of computational systems. We have even asserted that perception itself is a creative act and that perceiving things differently can facilitate creative thinking. We have demonstrated how state-of-the-art deep neural networks can be used to create images and perform certain types of imagination, and we have also demonstrated how they can see creatively through pareidolia.

As with humans, advanced perceptual abilities can provide a foundation on which computational systems can think, imagine, and create. In the field of artificial general intelligence, current trends and ideas are also advocating the need for perception, and recent general AI systems are learning to perform intelligent tasks exclusively from raw inputs (Hawkins and Blakeslee 2007; Mnih et al. 2015). They argue that having a system learn from the ground up, with raw inputs (e.g., raw pixel values), is essential for general/adaptable intelligence. Perceiving and understanding various raw inputs can act as a basis for a large variety of intelligence tasks, and learning how to perceive and perform for one task should transfer to additional tasks. Furthermore, advanced cognitive ability, such as language and reasoning, could emerge naturally from these perceptual primitives as they form connections and hierarchies of understanding.

The idea of perceptual primitives can also be applied to a general notion of computational creativity. Ideally, we would like to develop a universal creative process, which allows for connections to form across multiple domains, experiences, and knowledge. Perceptual abilities for multiple modalities establish an internal mental model of the world, which can provide a system with freedom and adaptability to be creative in any of its modalities or combination of modalities. For example, a system trying to invent recipes could benefit from visually recognizing ingredients (in addition to understanding how they taste) and could invent new recipes by substituting similar looking ingredients. It is possible that developing and incorporating advanced perceptual abilities in computational systems will not only increase the creative potential of those systems but may also facilitate the abstraction of a domain-independent, general creativity "algorithm".

References

Allen, C. B.; Celikel, T.; and Feldman, D. E. 2003. Long-term depression induced by sensory deprivation during cortical map plasticity in vivo. *Nature Neuroscience* 6(3):291–299.

Amedi, A.; Merabet, L. B.; Camprodon, J.; Bermpohl, F.; Fox, S.; Ronen, I.; Kim, D.-S.; and Pascual-Leone, A. 2008. Neural and behavioral correlates of drawing in an early blind painter: a case study. *Brain Research* 1242:252–262.

Bach-y-Rita, P., and Kercel, S. W. 2003. Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences* 7(12):541–546.

Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22(04):637–660.

Beaney, M. 2005. *Imagination and Creativity*. Open University Milton Keynes, UK.

Berns, G. 2008. *Iconoclast: A neuroscientist reveals how to think differently*. Harvard Business Press.

Chatterjee, A. 2004. The neuropsychology of visual artistic production. *Neuropsychologia* 42(11):1568–1583.

Colton, S.; Halskov, J.; Ventura, D.; Gouldstone, I.; Cook, M.; Blanca, P.; et al. 2015. The Painting Fool sees! new projects with the automated painter. In *Proceedings of the 6th International Conference on Computational Creativity*, 189–196.

Colton, S.; Goodwin, J.; and Veale, T. 2012. Full face poetry generation. In *Proceedings of the Third International Conference on Computational Creativity*, 95–102.

Colton, S. 2008. Creativity versus the perception of creativity in computational systems. *Creative Intelligent Systems: Papers from the AAAI Spring Symposium* 14–20.

Cope, D. 1996. *Experiments in musical intelligence*, volume 12. AR editions Madison, WI.

Csíkzentmihályi, M., and Robinson, R. E. 1990. *The Art of Seeing*. The J. Paul Getty Trust Office of Publications.

Denton, E. L.; Chintala, S.; Fergus, R.; et al. 2015. Deep generative image models using a Laplacian pyramid of adversarial networks. In *Advances in Neural Information Processing Systems*, 1486–1494.

DiPaola, S., and Gabora, L. 2009. Incorporating characteristics of human creativity into an evolutionary art algorithm. *Genetic Programming and Evolvable Machines* 10(2):97–110.

Ebcioğlu, K. 1988. An expert system for harmonizing four-part chorales. *Computer Music Journal* 12(3):43–51.

Edwards, B. 1989. *Drawing on the Right Side of the Brain*. New York: Tarcher.

Ellis, H. D., and Lewis, M. B. 2001. Capgras delusion: a window on face recognition. *Trends in Cognitive Sciences* 5(4):149–156.

Farabet, C.; Couprie, C.; Najman, L.; and LeCun, Y. 2013. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1915–1929.

Farah, M. J. 2004. *Visual Agnosia*. MIT press.

Flowers, J. H., and Garbin, C. P. 1989. Creativity and perception. In *Handbook of Creativity*. Springer. 147–162.

Gaut, B. 2003. Creativity and imagination. *The Creation of Art* 148–173.

Grassian, S., and Friedman, N. 1986. Effects of sensory deprivation in psychiatric seclusion and solitary confinement. *International Journal of Law and Psychiatry* 8(1):49–65.

Gregor, K.; Danihelka, I.; Graves, A.; and Wierstra, D. 2015. DRAW: A recurrent neural network for image generation. In *Proceedings of the 32nd International Conference on Machine Learning*, 1462–1471.

Hamming, R. W. 1980. The unreasonable effectiveness of mathematics. *American Mathematical Monthly* 87:81–90.

Hawkins, J., and Blakeslee, S. 2007. *On intelligence*. Macmillan.

Heath, D., and Ventura, D. 2016. Creating images by learning image semantics using vector space models. In *Proceedings of The Thirtieth AAAI Conference on Artificial Intelligence*.

Heath, D.; Dennis, A.; and Ventura, D. 2015. Imagining imagination: A computational framework using associative memory models and vector space models. In *Proceedings of the 6th International Conference on Computational Creativity*, 244–251.

Hoffman, D. D. 2000. *Visual Intelligence: How We Create What We See*. W.W. Norton.

Hong, K.; Chalup, S. K.; King, R.; Ostwald, M. J.; et al. 2013. Scene perception using pareidolia of faces and expressions of emotion. In *IEEE Symposium on Computational Intelligence for Creativity and Affective Computing*, 79–86.

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; and Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, 675–678.

Kamel, H. M., and Landay, J. A. 2000. A study of blind drawing practice: creating graphical information without the visual channel. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies*, 34–41.

Kennedy, J. M. 1993. *Drawing & the Blind: Pictures to Touch*. Yale University Press.

Lafer-Sousa, R.; Hermann, K. L.; and Conway, B. R. 2015. Striking individual differences in color perception uncovered by 'the dress' photograph. *Current Biology* 25(13):R545–R546.

Leon A. Gatys, A. S. E., and Bethge, M. 2015. A neural algorithm of artistic style. *Computing Research Repository*.

Levi, G., and Hassner, T. 2015. Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 34–42.

Machado, P.; Romero, J.; and Manaris, B. 2007. Experiments in computational aesthetics: An iterative approach to stylistic change in evolutionary art. In Romero, J., and Machado, P., eds., *The Art of Artificial Evolution: A Handbook on Evolutionary Art and Music*. Berlin: Springer. 381–415.

Marmor, M. F., and Ravin, J. 2009. *The Artist's Eyes*. Harry N Abrams Incorporated.

Melo, A. F., and Wiggins, G. 2003. A connectionist approach to driving chord progressions using tension. In *Proceedings of the AISB Symposium on Creativity in Arts and Science*.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; et al. 2015. Human-level control through deep reinforcement learning. *Nature* 518(7540):529–533.

Morris, R. G.; Burton, S. H.; Bodily, P. M.; and Ventura, D. 2012. Soup over bean of pure joy: Culinary ruminations of an artificial chef. In *Proceedings of the 3rd International Conference on Computational Creativity*, 119–125.

Netzer, Y.; Gabay, D.; Goldberg, Y.; and Elhadad, M. 2009. Gaiku: Generating haiku with word associations norms. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, 32–39.

Nguyen, A. M.; Yosinski, J.; and Clune, J. 2015. Innovation engines: Automated creativity and improved stochastic optimization via deep learning. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 959–966.

Pachet, F., and Roy, P. 2011. Markov constraints: Steerable generation of Markov sequences. *Constraints* 16(2):148–172.

Parkhi, O. M.; Vedaldi, A.; and Zisserman, A. 2015. Deep face recognition. *British Machine Vision* 1(3):6.

Peterson, E. M. 2006. Creativity in music listening. *Arts Education Policy Review* 107(3):15–21.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; and Fei-Fei, L. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* 115(3):211–252.

Sacks, O. 1995. *An Anthropologist on Mars*. New York: Knopf.

Searle, J. R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3(03):417–424.

Simonyan, K.; Vedaldi, A.; and Zisserman, A. 2013. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. 2015. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9.

Varshney, L. R.; Pinel, F.; Varshney, K. R.; Schörgendorfer, A.; and Chee, Y.-M. 2013. Cognition as a part of computational creativity. In *Procceedings of the 12th IEEE International Conference on Cognitive Informatics and Cognitive Computing*, 36–43.

Weiskrantz, L. 1996. Blindsight revisited. *Current Opinion In Neurobiology* 6(2):215–220.

Zwaan, R. A., and Kaschak, M. P. 2008. Language in the brain, body, and world. *The Cambridge Handbook of Situated Cognition* 368–381.