Visual Hallucination For Computational Creation

Leonid Berov and **Kai-Uwe Kühnberger**
Institute for Cognitive Science
University of Osnabrück
49076 Osnabrück, Germany
{lberov, kkuehnbe}@uos.de

Abstract

Research on computational painters usually focuses on simulating rational parts of the generative process. From an art-historic perspective it is plausible to assume that also an arational process, namely visual hallucination, played an important role in modern fine art movements like Surrealism. The present work investigates this connection between creativity and hallucination.
Using psychological findings, a three-step process of perception-based creativity is derived to connect the two phenomena. Insights on the neurological correlates of hallucination are used to define properties necessary for modelling them. Based on these properties a recent technique for feature visualisation in Convolutional Neural Networks is identified as a computational model of hallucination. Contrasting the thus enabled perception-based approach with the Painting Fool allows to introduce a distinction between two distinct creative acts, sketch composition and rendering.
The contribution of this work is threefold: First, a computational model of hallucination is presented and discussed in the context of a computational painter. Second, a theoretic distinction is introduced that aligns research on different strands of computational creativity and captures the differences to current computational painters. Third, the case is made that computational methods can be used to simulate abnormal mental patterns, thus investigating the role that "madness" might play in creativity – instead of simply renouncing the myth of the mad artist.

Introduction

Computational creativity research often stresses that the creative act is a rational process instead of a divine gift or the byproduct of madness. But while it is certainly true that "one does not need to be [...] an ear-lopping manic-depressive to be creative" (Veale 2012, p. 16), it is also the case that some creative artefacts owe their uniqueness precisely to the workings of a deranged mind. Self-reports indicate that visual hallucinations were an important source of inspiration for many artists. Some, like van Gogh, were involuntarily influenced by the changes of perception inherent in their psychological disorders (van Gogh 1889). Others, like Joan Miró, willingly induced hallucinations to draw creativity from the arational (Phillips 1948). And while self-reports are not necessarily reliable evidence, recent findings will be introduced that also establish quantitative evidence for this connection. Such findings can be seen as contradictory to a view that rejects the arational as corroborating the myth of the "mad genius". By deriving a computational model of hallucination, and explaining how it can be used in a computational painter, the present paper will illustrate how arational processes can be employed as generative computational models of creativity.

For this we will first present the mentioned art-historic findings on the role of hallucination in creativity. Thus motivated, we will investigate from a psychological perspective how a creative process based on perception (be it normal, or aberrant) can be formulated. After having identified the role of hallucination in such a process, we will descend one level of abstraction to outline how hallucination is implemented in the human brain. This will allow us to derive functional properties that a computational model of hallucination must posses. These properties will be used to argue that a recently introduced technique for visualising features of Deep Convolutional Neural Networks (ConvNets) can be used to simulate hallucination. This argument will be partially validated by demonstrating how three phenomena associated with hallucination can be modelled with this technique. Coming back to the introduced psychological process of perception-based creativity we will present how such a model of hallucination can be used to implement a computational painter, and discuss it in the context of the current state of the art.

Taking all together, our goal is not just to present a computational model of hallucination as a potential source of inspiration, but also to broaden the scope of computational creativity. By providing a case study on how to model a part of creativity that is arational, we argue that accepting abnormal mental patterns as potential sources of creativity does not imply yielding to the myth of the "mad genius".

Hallucinations in the Fine Arts

The role hallucination plays in the fine arts is most apparent in modern art movements due to their departure from the primacy of naturalistic depiction. Post-impressionist artwork, for instance, is characterised by depictions of the artists' subjective impression of a scene – something that can be influenced by perceptual disorders. Most notable in this context is painter Vincent van Gogh, who increasingly suffered from psychotic episodes including visual hallucinations (Blumer 2002) and had to move to mental asylum in 1889. This development was accompanied by a noticeable, qualitative change in his style, tending to wavy lines and thick, intensive colouring. Van Gogh himself noted this connection in a letter where he states that "some of my pictures certainly show traces of having been painted by a sick man" (van Gogh 1889). This can be backed by quantitative evidence, as artwork from van Gogh's psychotic phases appears to capture mathematical properties of light-turbulences in a way that artwork from healthy phases does not (Aragón et al. 2008).

Even more relevant is surrealist artwork, which is characterised by the drive to capture the subconscious. This can be taken quite literally, since one group of surrealists intended their art to be an "exact transcription of personal hallucination" (Frey 1936). In fact, Frey emphasizes that in this approach hallucination is to be considered "antecedent" to the painting, which, in turn, is just a means for "immediate fixation of the violent [. . .] images that haunt the brain". Consequently, healthy surrealists resorted to artificial means for inducing hallucinations, like Joan Miró, who painted from hunger hallucinations: "I began gradually to work away from the realism I had practiced [...] until, in 1925, I was drawing almost entirely from hallucinations" (Miró, qtd. in Phillips 1948). A qualitative analysis suggests that also Max Ernst was influenced by visual hallucination, as all spatial properties of hallucinatory phenomenology were identified in his artwork (Keeler 1970).

A thorough analysis of art-historic material is outside of the scope of this paper. However, what we have shown is that (1) hallucination can be systematically related to modern art agendas, that (2) artistic self-reports support such theoretic conceptions and that (3) qualitative and quantitative evidence corroborate artists' claims.

Perception-based Creativity

Psychological inquiry on the artistic use of hallucination can be found in a discussion of the role of perception in creativity, which identified two relevant types of mental processes (Flowers and Garbin 1989): The first type are executively controlled perceptual processes like mental imagery or selective attention. These processes can be used to generate novel mental representations by effortful construction. Individuals with superior control of such faculties derive their creative abilities from the scope and complexity of available mental operations. Processes of the second type, on the other hand, are involuntary because they are based on the perceptual organisation of input data, which is performed automatically by the visual system. Individuals whose perceptual organisation operates less deterministically, or fundamentally divergent from what is typical, can derive novel mental representations straight from their percept. Their creativity stems literally from seeing things in an unusual way. Creative behaviour usually results from a combination of both types, with the emphasis shifting from one individual to the other. The role of hallucinations can be identified as one possible source of "loose" perceptual organisation.

Because "common mental resources are used in executive control of mental representations and processing of corresponding forms of sensory data" (Flowers and Garbin 1989) the authors state that interference effects can occur when processes of different types happen to coincide temporally. This will become relevant later, when we show that such behaviour can actually be observed in the proposed computational model. Flowers and Garbin furthermore point out that conceiving a creative artefact includes selection processes in order to identify if a mental representation is novel and valuable. This can be especially hard for individuals with a loose perception, since it involves hypothesizing about the judgement of non-aberrant perceivers.

A widely accepted psychological model of creativity (Csikszentmihalyi 1997) postulates a five steps process:

1. preparation: gathering knowledge and values of the relevant domain,
2. incubation: subconscious combination, consolidation and re-organisation of knowledge,
3. insight: unexpected event, the surfacing of an idea,
4. evaluation: deciding weather the idea is novel and valuable,
5. elaboration: detailed concretisation and implementation of the idea.

The process is not linear bur rather recursive in nature, and especially the last three steps can reoccur iteratively thereby informing each other.

The first two steps of this process can not be linked to the account of Flowers and Garbin directly, however, preparation and incubation could be mapped to the maturation and knowledge acquisition of the visual system, most prominently during infancy, which must be an implicit precondition for any perception-based account. An actual creative process would thus start with the insight phase. In the context of Flowers and Garbin this can be taken to be the generation of an unconventional percept by a loose perception-process; the characteristic phenomenology of an insight-event is attributable to the involuntariness of sensory organisation. The evaluation step can be connected to Flower's selection processes, while elaboration, taken to be the most effortful step, maps well to Flower's description of executively controlled construction. We thus arrive at an iterative, perception-based process of creativity: (1) loose perceptual organisation, (2) selection and (3) executively controlled construction.

This synthesis could in principle serve as the basis of a computational model of hallucination based creativity. However, while there is work on computational accounts that might be dubbed effortful construction (Cohen-Or et al. 2006; Bhattacharya, Sukthankar, and Shah 2010) and aesthetic selection (Li and Chen 2009; Luo, Wang, and Tang 2011; Yao et al. 2012), to the best of our knowledge, no work has been done on computational models of the sensation of visual hallucination-based loose perception. In order to devise such a model we first need to understand how visual hallucinations are implemented in the brain.

Neurological Correlates of Hallucination

Visual sensory information is ambiguous. Thus in order to generate a stable, unambiguous percept, and perform higher-order tasks like object recognition, a processing of the input data has to take place in the visual cortex (Teufel et al. 2015). The primate visual cortex is comprised by a hierarchical system of specialised brain areas. Lower areas are responsive to primitive visual features like oriented gratings, while higher areas use information from lower layers, and are responsive to complex features like e.g. faces, houses or landscapes (Zeki et al. 1991; Felleman and Van Essen 1991). The visual system thus combines bottom-up sensory input processing with top-down predictions based on prior-knowledge of the environment.

Hallucinations occur when the balance in information processing shifts to prefer this knowledge over sensory evidence (Teufel et al. 2015; Mocellin, Walterfang, and Velakoulis 2006). Mocellin and colleagues take hallucination to be a “sensory perception that has the compelling sense of reality of a true perception but that occurs without stimulation of the relevant sensory organ”. Functional imaging has shown that visual hallucinations (at least within the Charles Bonnet Syndrome[1]) correlate with increased cerebral activity in specialised visual cortex areas: “colour hallucinations [are] accompanied [by] increased activity in cortex specialized for colour; face hallucinations, increased activity in cortex specialized for faces [...] and so forth.” (Santhouse, Howard, and Ffytche 2000). A recent study by Mégevand et al. (2014) was actually able to induce complex visual hallucinations (CVH) of outdoor scenes in non-psychotic subjects, by applying direct electrical stimulation to the parahippocampal place area. This implies a causative connection between increased activity of specialised areas and hallucinations.

For our purpose we can sum up hallucinations to be the product of the visual cortex where the processing balance between input and prior knowledge shifted towards the latter, for instance due to an artificial increase of activity in a specialised brain area, resulting in a percept that is not rooted in sensory information. Abstracting away from the neurological implementation in humans, this would mean a system that (1) performs visual processing, is (2) comprised by specialised subsystems, and where (3) increasing a subsystem’s activity leads to the generation of a visual representation that has no correlate in the input image but in the knowledge encoded in the respective subsystem. We thus have derived three properties that a system needs to demonstrate in order to be taken to model hallucinations.

A Computational Model of Hallucination

The state-of-the-art approach to many computer vision problems are Deep Convolutional Neural Networks, a specific type of the Multilayer Perceptron (MLP) that is informed by the workings of the mammalian visual cortex (LeCun et al. 1998).

Simple-cells in the primary visual cortex are sensitive to a small part of the retinal image, the so called *receptive field*. Neighbouring cells are processing neighbouring parts of the retinal image and have overlapping receptive fields which results in a topographical map of the input. This is beneficial due to the specific statistical properties of natural images, specifically the strong spatially-local correlations. Analogously, units in the convolutional layer of a ConvNet are only connected to a small subset of neighbouring units from the previous layer, instead of being dependent on all the units of the input, like it is the case in the fully connected layers of conventional MLPs. Because of this, the filters computed by each unit are not responsive to variations outside of their respective receptive field – they are just responsive to spatially local patterns. However, stacking several convolutional layers allows the receptive fields of units from deeper layers to become bigger with respect to the input image, and facilitates the detection of more complicated, global patterns.

Pattern-detectors that are useful in one part of the input-image are likely to be of use in other parts as well. This is exploited by ConvNets by employing parameter sharing. Each convolutional layer is organised in planes, consisting of units whose combined receptive fields cover the complete input-layer. The same parameters are used to compute the activation of each unit of a plane, which results in the same filter being applied on each patch of the input. Since the activation of units from a plane indicates the presence of the encoded pattern in the respective patch of the input, the output of each plane is referred to as a *feature-map*. Mathematically this operation can be described as a convolution of the filter-function with the input image. Usually, each convolutional layer is comprised by several different feature-maps.

The precise positions of a detected pattern is not as important as its position relative to other features. This allows ConvNets to perform a sub-sampling of the feature-maps computed by the convolutional layers by using pooling layers. These layers split the input into non-overlapping regions and compute a function (usually max) over the activation of all the units of each region. Thus the feature-map of a pooling-layer encodes the information whether a pattern was detected in a certain region of the image, without storing its precise position.

These three architectural idiosyncrasies significantly reduce the number of parameters involved in training ConvNets, as compared to conventional MLPs. This allows to create much deeper networks, which is quite beneficial if we consider that with each additional layer more complex, position-invariant features can be detected. Nevertheless, a significant amount of data is required to train a ConvNet, with state of the art approaches leveraging over one million of labeled images (Szegedy et al. 2015).

Investigations have been performed on the nature of the feature-representations learned by ConvNets. As it turns out, the features are not random and even interpretable (Zeiler and Fergus 2014). The first convolutional layer, seems to consistently learn to be responsive to (among others) oriented gratings (Zeiler and Fergus 2014; Krizhevsky,

[1] The Charles Bonnet Syndrome describes visual hallucinations correlating with a partial loss of vision (Burke 2002). Mocellin et al. argue that the distinction between CBS and lesion-based hallucinations is not clear. Thus these findings might generalise.

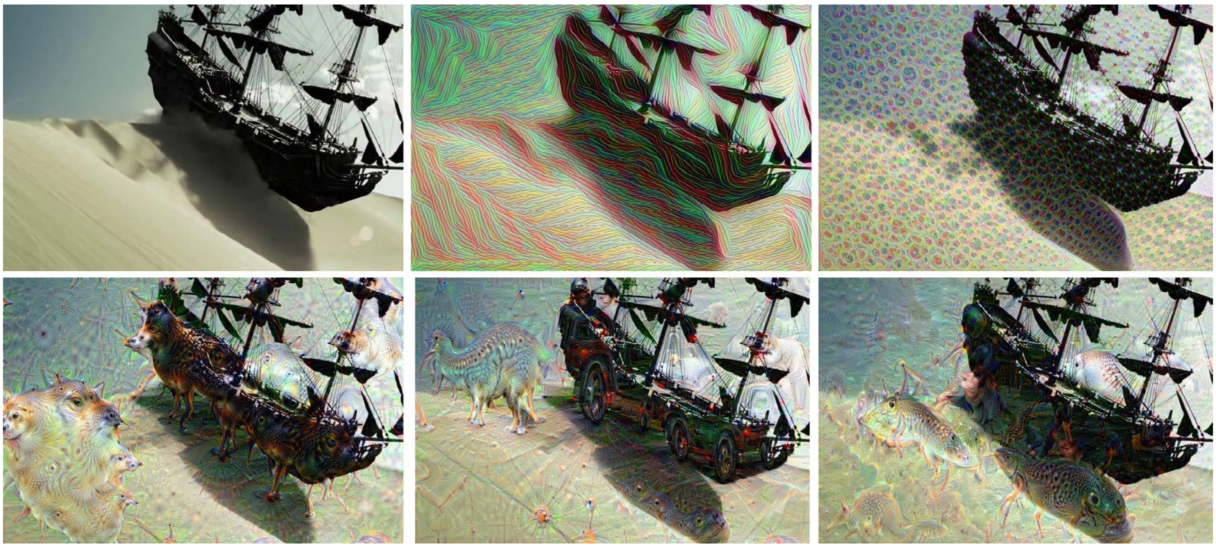

Figure 1: Pictures produced by deep dream. Left-top corner: Original image. From left to right and from up to down the target-layers were changed from lower-level to higher-level layers while keeping all other parameters constant. Enhanced features rise in complexity accordingly. It should be noted how especially in the last two exemplars the enhanced features are integrated into the scene. Also of interest is the fact that the enhanced animals appear to be hybrids. This is presumably the result of maximizing several units in one layer, which are specialised in recognizing different animals. Given this interpretation, the hybrids can be considered concept-blends. Best viewed on screen using zooming.

Sutskever, and Hinton 2012), which, incidentally, is also the case in the first mammalian visual cortex. Zeiler and Fergus have also shown, that features from higher levels exhibit many interpretable properties like compositionality and invariance to spatial operations like mirroring. Effectively it seems that features from lower layers represent properties of *image appearance*, while higher layers represent more and more abstract notions of the *image content* (Mahendran and Vedaldi 2015). If higher-level features are used to reconstruct an image they "invert back to a composition of parts similar but not identical to the ones found in the original image" (Mahendran and Vedaldi 2015). This all indicates that layers in ConvNets are indeed hierarchical, and that higher-level layers are specialised: they identify complex, interpretable objects like e.g. houses or faces. Thus, they exhibit the first two properties of a model of hallucination. What is lacking is a way to generate interpretable, visual representations by increasing the activity of the specialised layers.

Exactly that is accomplished by *deep dream*, a third approach to understand ConvNet feature representations, that focuses on visualising what was learned by individual layers (Mordvintsev, Olah, and Tyka 2015): An input image is forward-propagated through a fully-trained network. Starting from the layer to be analysed, back-propagation is performed in a way as to maximize the Euclidean Norm of activations in the target layer. However, unlike in usual training, the parameters of the network remain unchanged and a gradient ascent step is instead applied to the input image. Basically, the input-image is trained to maximize target-layer activation. To achieve better visibility of the changes in the image, Mordvintsev, Olah, and Tyka iteratively repeat this step several times while regularly increasing the scale of the input image. This results in a visual enhancement, as well as adaptation, of features that were already present in the input, and the produced pictures have been described as "trippy" and "visually pleasing" (Koch 2015). What type of features are affected depends on the choice of target layer (see fig. 1). When lower-level layers are targeted, primitive features like oriented gratings are enhanced. Mid-level layers enhance simple objects like eyes and geometric forms, while high-level layers enhance complex objects like buildings or animal-blends in a pareidolia-like fashion.

Applying deep dream to a ConvNet results in a hierarchical system for visual processing, where increasing the activity of a layer results in input-data augmentations that are related to the knowledge encoded in the respective layer. It thus displays all three properties of a functional model of hallucination, and we argue that the images generated by deep dream can be considered computational hallucinations. Indeed Koch reports that a remarkable resemblance between the produced images and hallucinations induced by LSD has been widely noted on the internet. Details on how to affect the visuals created by the model, and how to simulate several phenomena that are related to hallucination, will be discussed in the next section.

Discussion

The following results were all achieved using Berkley Vision and Learning Center's open-source reimplementation[2]

[2]https://github.com/BVLC/caffe/tree/master/models/bvlc_googlenet

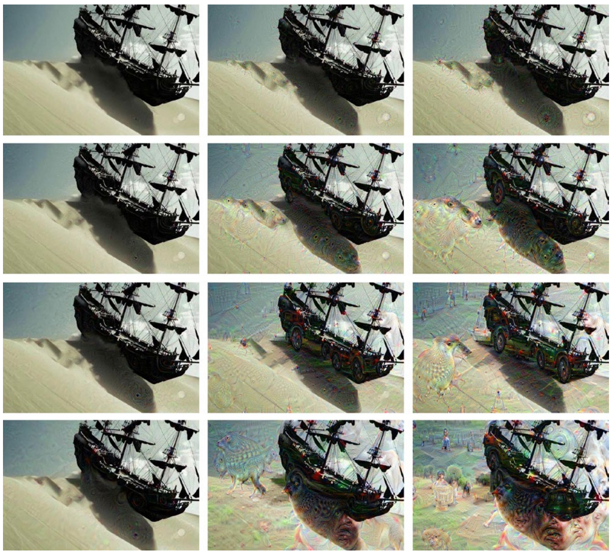

Figure 2: Effect of different scale and iteration settings. Rows (from up to down): 1, 3, 5 and 7 scales. Columns (from left to right): 1, 10 and 19 iterations. Each image was produced by an individual run with the given parameters and the target-layer *inception (4c)*. Best viewed on screen using zooming.

of GoogLeNet (Szegedy et al. 2015), one of the winning models of the ILSVRC 2014 classification challenge (Russakovsky et al. 2014). The architecture is a ConvNet with 22 layers. It employs convolutional layers of varying filter-sizes alternating with pooling-layers for spatial down-sampling. The last layer is a fully-connected soft-max classifier, that follows a 40% dropout layer (Hinton et al. 2012) to prevent overfitting. Furthermore, rectified linear activation (Krizhevsky, Sutskever, and Hinton 2012) is used. Like the original network, the reimplementation was trained on the 1.2 million labeled training-images of the ILSVRC 2014 dataset. For details, especially on the combined convolutional layers called inception and introduced by GoogLeNet, please refer to the original publication.

Parametrization

From the several knobs and levers afforded us by the deep dream algorithm the most relevant for visual appearance were identified as the *target layer*, the *number of iterations per scale* and the *number of scales*. Most importantly, the target layer influences the complexity of the generated hallucinations. For instance the second image of fig. 1 was generated using the second convolution layer and the enhanced features are colourful, oriented gratings. The last image of fig. 1, on the other hand, was generated using the tenth convolutional layer (*inception 5a*) and the enhanced features are hybrid fish-like creatures. As noted earlier, intermediate layers enhance features of varying, but rising complexity. No regularities could be identified between images generated from layers that share a type (like for instance the max-pool layers) but differ in their respective position in the network.

Figure 3: Picture produced by applying deep dream on a white-noise image. The employed parameters were: 8 scales, 50 iterations per scale and 100 repetitions of the algorithm; the target-layer was *inception (5a)*. Several distinct creatures emerged despite a complete lack of statistical structure in the input. This effect simulates the shift in processing balance from input-data to prior-knowledge that happens during hallucinations. Best viewed in colour.

As for the two other parameters, a larger number of iterations increases the intensity of the enhanced features, while a larger number of scales increases the size, number and detail-grade of the enhanced features (see fig. 2).

Simulating Hallucination-Related Phenomena

Apart from following our definition for a functional model of hallucination, and producing qualitatively plausible outputs, the deep dream system is capable of modeling several hallucination-related phenomena:

One is that hallucinations can be induced by sensory deprivation, be it artificial (Merabet et al. 2004) or due to a medical condition (Burke 2002). This can be simulated by repeatedly applying the deep dream procedure to a white-noise image, while significantly increasing the number of scales and propagation steps (see fig. 3). Although a noise-input provides no structure for the ConvNet to detect, network-internal noise nevertheless results in layer activation. This random activation is propagated to the input image and results in discernible but random effects. The stabilisation to a distinct structure is due to a complete shift in processing balance from input data to prior-knowledge encoded in the network.

Another phenomenon is that CVH usually follow a limited number of typologies (Santhouse, Howard, and Ffytche 2000; Mocellin, Walterfang, and Velakoulis 2006), e.g. disembodied, distorted faces or small figures in costumes. The same holds for images produced by the deep-dream system which display some patterns like eyes or dog-shaped creatures more often then others, which is suggestive of basins of attraction.[3] The type of these basins is dependent on the dataset used for training the ConvNet and favours features

[3] As discussed in https://github.com/google/deepdream

Figure 4: Picture produced by guided dreaming at a high-level layer, an effect that can be taken to simulate the interference of temporally coincident perception- and manipulation-processes operating on the same type of data. Bottom-right corner: guiding image. The employed parameters were: 6 scales, 20 iterations per scale and the target-layer was *inception (5a)*. Best viewed in colour.

Figure 5: Picture produced by combining guided dreaming with input-deprivation using the target-layer *inception (4a)*. Bottom-right corner: guiding image. Best viewed in colour.

that were over-represented.

A last phenomenon is connected to perception-based systems in general, and was reported in the section on loose perception: Flowers and Garbin state that interference effects should occur when processing of sensory input coincides with the active manipulation of mental representations of the same type. This can be simulated using a technique called *guided dreaming*. For that, in a preparatory step, a guiding image is forward-propagated through the ConvNet and the layer-activity is noted. In the deep dreaming procedure the optimization objective is then changed to maximizing the dot product of input-image activation and guiding-image activation at the target layer. In that way only features that were detected in both images are enhanced, which especially at higher-level layers result in a spilling-over effect from the guide to the output image (see fig. 4).

While not directly relevant for the simulation of human phenomena, it shall be noted that our model allows the combination of guided dreaming with input-deprivation. This produces an interesting, artistic effect where shapes and features from the guide are transfered and randomly reassembled in the output image, resulting in a "colorful, free improvisation on the theme of the guide" (see fig. 5).

Role in Computational Painting

Of course a computational model of hallucination alone is not sufficient for a computational painter. There are two possibilities of incorporating the introduced model in the broader context of the previously outlined model of perception-based creativity. One is to rigidly define all parameters that influence the visual properties of generated images, which seems an obvious approach considering that it corresponds to something we would call perceptual disorder in a human. The other option is to leave these parameters (especially the target-layer for each iteration) variable, and allow them to change depending on the results of the selection step. This creates a feedback-loop and potentially allows for the emergence of a particular style. This approach is less plausible from a psychological perspective since it seems to imply the volitional adaptability of perceptual disorders. From a computational perspective, however, it is a more promising option since it hands over more creative freedom to the system. In order to compare both approaches an implementation of the whole process needs to be realised.

Several options for implementing selection-processes can be explored. This includes training classifiers based on art-theoretic high-level features (Li and Chen 2009) or a recently introduced creativity-score that is a measure for a painting's "abstraction in shape and form" as well as its "texture and pattern" (Elgammal and Saleh 2015). Approaches for implementing executively controlled construction include colour harmonisation (Cohen-Or et al. 2006), composition-enhancement (Bhattacharya, Sukthankar, and Shah 2010) and style-imitation (Gatys, Ecker, and Bethge 2015). It seems promising to use video-input instead of individual scenes, because this provides the selection system with a variety of perspectives to choose from. This not only appears to be a more natural context for a perception-based system but also transfers more artistic responsibility to the software. Such an implementation is currently in progress.

Related Work

Other attempts have been made to create computational models of hallucination (Jardri and Denève 2013). However these models focus on simulating cortical activity at different levels of abstraction. Our approach, on the other hand, does not claim to make predictions about structural properties of hallucination. Instead it operates on a functional level, by modeling the effect of visual hallucination. To the best of our knowledge this work is the first computer model that can simulate the sensation of visual hallucination.

Our overall goal was to discuss the role of visual hallucination in creativity. Several systems exist that are concerned with the computational accounts of the fine arts. One of the earliest is called AARON and has been maintained for over 40 years by artist Harold Cohen (McCorduck 1991). AARON differs from our proposed system in two major ways: First, while it is definitely more than a mere artistic tool, it was not designed to function independently from Cohen, who actively takes part in its creative process. Second, AARON is mainly concerned with figurative and abstract art and does not draw inspiration from real-world scenes.

Our work shares the spirit of the currently most prominent computational painter, Colton's Painting Fool (Colton 2012). Colton's goal is to create "an automated painter which is one day taken seriously as a creative artist in its own right". His work was mostly concentrated on enabling the system to create painterly renditions of photographs simulating different natural media, and on choosing the most appropriate style to do so (Colton, Valstar, and Pantic 2008). This was criticised by artist Faure-Walker as a lack of imagination and creative intent (Colton 2012). The approach proposed here is, on the contrary, concerned with modifying the input in a meaningful way by changing its content and composition based on a systematic misperception. It thus addresses Faure-Walker's criticism by drawing imagination from an unconventional way of perception and making its intention one of sharing this unique type of impression, much in the fashion of modern artistic movements like Post-Impressionism.

Based on this analysis we propose to differentiate between two creative acts: *sketch-composition* (what to draw) and *rendering* (how to draw it). These two differ significantly in the involved problems (e.g. composition, colouring, symbol-language or intention in the first case and material, stroke-type, colour-palette and level of detail in the latter) as well es in the intended outputs (a mental sketch or idea in the first case and an artistic artefact in the latter). A similar distinction is already successfully employed in computational storytelling (Gervás 2009), where creative systems are concerned with creating *fabula* (what is told) or *discourse* (how it is told). Thus our proposal helps to align research in different strands of computational creativity by drawing out the differences between the conception of a work of art, and its implementation. It also helps to disentangle research on computational painting, since such a division is for instance applicable to recent advances on the Painting Fool, because the system judges its rendered artefact by comparing them with previously generated sketches (Colton et al. 2015).

With the distinction between sketch-composition and rendering in mind, a combination of the system proposed here and the Painting Fool becomes plausible. The former can select appropriate scenes and generate a sketch based on loose perception and effortful construction, potentially grounded in art theory. The latter can select an appropriate rendering style and render the sketch accordingly, thus resulting in a more complete model of a human painting process.

Conclusion

Starting from the observation that some painters, especially from modern art movements, drew inspiration from natural or artificially induced perceptual disorders we performed an investigation of the role of visual hallucination in creativity. For that the neurological correlates of hallucinations were outlined and criteria were derived that a computational model must meet in order to be considered a functional model of hallucination. Subsequently we argued that deep dream, a technique for ConvNet feature visualisation, meets all the necessary criteria and can be functionally compared to inducing hallucinations by electrically stimulating specialised brain areas. This conclusion was further corroborated by showing how several phenomena connected with hallucination can be simulated using deep dream. On a technical level this might be a straightforward exploration of the deep dream tool. What is relevant here, however, is not *how* deep dream changes images, but rather *what* these changes constitute. The significance of this exploration is on a conceptual, rather then a technical level, by partially validating the proposed model.

Just having hallucinations does not necessarily make an artist. Based on psychological research on the role of perception in creativity we derived a three-step process that illustrates how hallucinations can be used for creative insight. Taking this as a framework we then outlined possible avenues for implementing a misperception-based computational painter. Contrasting this implementations with current work on computational painters allowed us to introduce the distinction between sketch-composition and rendering, two distinct creative acts that are both necessary for a successful painter but involve very different processes.

Thus the contribution of the present work is threefold. First, it demonstrates an algorithm that allows computational painters to draw inspiration from systematically misperceiving input scenes. By that it, second, makes the case for a broader approach to creativity that, instead of renouncing the myth of the mad artist, uses computational methods to simulate abnormal mental patterns to further understand the role that madness might play in creativity. Third, it introduces a theoretic distinction, which helps disentangle different processes involved in implementing computational painters, and aligns research on computational painters and computational storytellers.

References

Aragón, J. L.; Naumis, G. G.; Bai, M.; Torres, M.; and Maini, P. K. 2008. Turbulent Luminance in Impassioned van Gogh Paintings. *Journal of Mathematical Imaging and Vision* 30(3):275–283.

Bhattacharya, S.; Sukthankar, R.; and Shah, M. 2010. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the international conference on Multimedia*, 271–280. ACM.

Blumer, D. 2002. The illness of Vincent van Gogh. *American Journal of Psychiatry*.

Burke, W. 2002. The neural basis of Charles Bonnet hallucinations: a hypothesis. *Journal of Neurology, Neurosurgery & Psychiatry* 73(5):535–541.

Cohen-Or, D.; Sorkine, O.; Gal, R.; Leyvand, T.; and Xu, Y.-Q. 2006. Color harmonization. In *ACM Transactions on Graphics (TOG)*, volume 25, 624–630. ACM.

Colton, S.; Halskov, J.; Ventura, D.; Gouldstone, I.; Cook, M.; Blanca, P.; and others. 2015. The Painting Fool Sees! New Projects with the Automated Painter.

Colton, S.; Valstar, M. F.; and Pantic, M. 2008. Emotionally aware automated portrait painting. In *Proceedings of the 3rd conference on DIMEA*, 304–311. ACM.

Colton, S. 2012. The Painting Fool: Stories from Building an Automated Painter. In McCormack, J., and d'Inverno, M., eds., *Computers and Creativity*. Berlin, Heidelberg: Springer Berlin Heidelberg. 3–38.

Csikszentmihalyi, M. 1997. Flow and the Psychology of Discovery and Invention. *HarperPerennial, New York* 39.

Elgammal, A., and Saleh, B. 2015. Quantifying Creativity in Art Networks.

Felleman, D. J., and Van Essen, D. C. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1(1):1–47.

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

Frey, J. G. 1936. Miro and the Surrealists. *Parnassus* 8(5):13–15.

Gatys, L. A.; Ecker, A. S.; and Bethge, M. 2015. A neural algorithm of artistic style.

Gervás, P. 2009. Computational Approaches to Storytelling and Creativity.

Hinton, G. E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. R. 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580 [cs]*.

Jardri, R., and Denève, S. 2013. Computational models of hallucinations. In *The neuroscience of hallucinations*. Springer. 289–313.

Keeler, M. H. 1970. Klüver's Mechanisms of Hallucinations as Illustrated by the Paintings of Max Ernst. In M.D, W. K., ed., *Origin and Mechanisms of Hallucinations*. Springer US. 205–208.

Koch, C. 2015. Do Androids Dream? *Scientific American Mind* 26(6):24–27.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324.

Li, C., and Chen, T. 2009. Aesthetic visual quality assessment of paintings. *Selected Topics in Signal Processing, IEEE Journal of* 3(2):236–252.

Luo, W.; Wang, X.; and Tang, X. 2011. Content-based photo quality assessment. In *International Conference on Computer Vision*, 2206–2213. IEEE.

Mahendran, A., and Vedaldi, A. 2015. Understanding deep image representations by inverting them. In *Proceedings of the IEEE CVPR*, 5188–5196. IEEE.

McCorduck, P. 1991. *Aaron's code: meta-art, artificial intelligence, and the work of Harold Cohen*. Macmillan.

Mégevand, P.; Groppe, D. M.; Goldfinger, M. S.; Hwang, S. T.; Kingsley, P. B.; Davidesco, I.; and Mehta, A. D. 2014. Seeing Scenes: Topographic Visual Hallucinations Evoked by Direct Electrical Stimulation of the Parahippocampal Place Area. *The Journal of Neuroscience* 34(16):5399–5405.

Merabet, L. B.; Maguire, D.; Warde, A.; Alterescu, K.; Stickgold, R.; and Pascual-Leone, A. 2004. Visual hallucinations during prolonged blindfolding in sighted subjects. *Journal of Neuro-Ophthalmology* 24(2):109–113.

Mocellin, R.; Walterfang, M.; and Velakoulis, D. 2006. Neuropsychiatry of complex visual hallucinations. *Australian and New Zealand journal of psychiatry* 40(9):742–751.

Mordvintsev, A.; Olah, C.; and Tyka, M. 2015. Inceptionism: Going Deeper into Neural Networks. Google Research Blog.

Phillips, W. 1948. *Partisan Review, February, 1948, Volume XV, Number 2*, volume XV of *Partisan Review*. New York, NY: Added Enterprises.

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. 2014. ImageNet Large Scale Visual Recognition Challenge. *arXiv:1409.0575 [cs]*.

Santhouse, A. M.; Howard, R. J.; and Ffytche, D. H. 2000. Visual hallucinatory syndromes and the anatomy of the visual brain. *Brain* 123(10):2055–2064.

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 CVPR*, 1–9.

Teufel, C.; Subramaniam, N.; Dobler, V.; Perez, J.; Finnemann, J.; Mehta, P. R.; Goodyer, I. M.; and Fletcher, P. C. 2015. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences* 112(43):13401–13406.

van Gogh, V. 1889. Letter 827: To Willemien van Gogh. Vincent van Gogh: The Letters. Van Gogh Museum.

Veale, T. 2012. *Exploding the creativity myth: the computational foundations of linguistic creativity*. London ; New York: Continuum International Pub. Group.

Yao, L.; Suryanarayan, P.; Qiao, M.; Wang, J. Z.; and Li, J. 2012. Oscar: On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision* 96(3):353–383.

Zeiler, M. D., and Fergus, R. 2014. Visualizing and Understanding Convolutional Networks. In *Computer Vision–ECCV 2014*, 818–833. Springer.

Zeki, S.; Watson, J. D.; Lueck, C. J.; Friston, K. J.; Kennard, C.; and Frackowiak, R. S. 1991. A direct demonstration of functional specialization in human visual cortex. *The Journal of Neuroscience* 11(3):641–649.