Rodrigo Setti

Generative Dreams from Deep Belief Networks
Paper

Topic: Generative Robots

Authors:
Rodrigo Setti, M.S.
Sr. Software Engineer
USA

Main References:

[1] Geoffrey Hinton, et al. “The" wake-sleep" algorithm for unsupervised neural networks”, Science, 1995.

[2] http://googleresearch.blogspot.com/2015/06/inceptionism-going-deeper-into-neural.html

[3] Geoffrey Hilton, et at. “A fast learning algorithm for deep belief nets”, MIT Press, 2006.

[4] Andrew Ng, et al. “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations”, Proceedings of the 26th Annual International Conference on Machine Learning, 2009.

Abstract:

According to the “Memory Consolidation Theory of Dreaming”, dream exists as a way to process and consolidate information that we have acquired during our waking lives. In that perspective, it’s right to say that minds that produce dreams are also minds capable of learning. If we include in the category of capable of learning, not only living creatures, but also artificial systems, we can explore the consequences of those systems being able to dream as well, and, most interestingly, what do they dream about.

If machines can dream, they can also be creative, and even produce art. Deep Belief Networks are artificial systems Inspired in the brain, and capable of learning representations of data with multiple levels of abstractions. These methods have dramatically improved the state-of-the-art in speech recognition, visual object detection, and many other domains such as web search and genomics. These artificial minds are composed of multiple processing layers, much like how visual cortex of humans are structured. One of the remarkable properties of the learning algorithm called “wake-sleep” [1] used to train these systems is that it has to have a “dreaming” period. This dreaming period is necessary for effective learning, and it’s when the neural network generates signals from within, without any external input. An interesting analogy with the psychological theory.

We introduce to the reader how these artificial neural networks are structured, and how they are able to learn images hierarchically within they many layers. Then, following the steps revealed by Google Scientists' “Inceptionist” blog post [2], we explore how we can probe into their dreams, after being trained with million of photographs, and show that it hallucinates fantastic realms, of bizarre and psychedelic variations of reality (figures). The aesthetics and creativity extent of such machine dreams are discussed in the paradigm of generative art. Interpreting each dream as one of the endless expressions realized by the particular artificial mind's anatomy (neural network topology) and experience (trained images), recognizable as its particular vision of the world as has been shown to it.

Example of generative dreams: people and cars (left); animal forms (right).

Contact:

rodrigosetti@gmail.com

Keywords:

Artificial intelligence, neural networks, deep learning, dreams, generative, image recognition, machine learning, deep belief networks

Generative Dreams from Deep Belief Networks

Rodrigo Setti, BSc, MSc.

United States

Email: rodrigosetti@gmail.com

Abstract

According to the “Memory Consolidation Theory of Dreaming”, dream exists as a way to process and consolidate information that we have acquired during our waking lives. In that perspective, it is right to say that minds that produce dreams are also minds capable of learning. If we include in the category of capable of learning, not only living creatures, but also artificial systems, we can explore the consequences of those systems being able to dream as well, and, most interestingly, what do they dream about.

If machines can dream, they can also be creative, and even produce art. Deep Belief Networks are artificial systems inspired in the brain, and capable of learning representations of data with multiple levels of abstractions. These methods have dramatically improved the state-of-the-art in speech recognition, visual object detection, and many other domains such as web search and genomics. These artificial minds are composed of multiple processing layers, much like how visual cortex of humans are structured. One of the remarkable properties of the learning algorithm called “wake-sleep” (Hinton, Dayan, Frey, & Neal, 1995) used to train these systems is that it has to have a “dreaming” period. This dreaming period is necessary for effective learning, and it is when the neural network generates signals from within, without any external input. An interesting analogy with the psychological theory.

We introduce to the reader how these artificial neural networks are structured, and how they are able to learn images hierarchically within their many layers. Then, following the steps revealed by Google Scientists' “Inceptionist” blog post (Mordvintsev, Olah, & Tyka, 2015), we explore how we can probe into their dreams, after being trained with million of photographs, and show that it hallucinates fantastic realms, of bizarre and psychedelic variations of reality (figures). The aesthetics and creativity extent of such machine dreams are discussed in the paradigm of generative art. Interpreting each dream as one of the endless expressions realized by the particular artificial mind's anatomy (neural network topology) and experience (trained images), recognizable as its particular vision of the world as has been shown to it.

– How to Build an Artificial Mind

They say that when the apprentice surpasses the master, then the later has fulfilled his duty as a teacher, and reached his greatest achievement. In that sense, one can say that mankind greatest achievement is about to come when we are able to build machines smarter than us. What the prominent futurist Ray Kurzweil (Kurzweil, 2005) calls the "singularity", a point around 2045 when machine intelligence will be infinitely more powerful than all human intelligence combined.

An artificial mind exhibiting general intelligence is yet 30 years ahead, but advances in computer science has led to remarkable progress in artificial systems capable of performing tasks likewise or better than humans.

One of these capabilities, which we are going to explore in this paper, is visual recognition (object classification), and its artistic dual: visual creativity. Building a machine capable of identifying hundreds of classes of images is not a small feat, and in fact it took years of the most brilliant scientists, as well as good progress in hardware performance to finally produce some real world applications.

Errore. L'origine riferimento non è stata trovata., **Errore. L'origine riferimento non è stata trovata.**, and **Errore. L'origine riferimento non è stata trovata.**, demonstrate some incredible capabilities of such systems by correctly labelling the images. Those examples were taken from the online "Image Identification Project" from Wolfram Research, available at www.imageidentify.com.

Figure 11 "coupe"

Figure 12 "double bass"

Figure 13 "grey wolf"

One does not approach this kind of problem, as would otherwise with other software engineering tasks. It is impractical for someone to explicitly program the rules of image interpretation that will allow the machine to differentiate between an image of a "dolphin" and a "tree". Thousands of shapes, colours, forms, shades, poses, all intertwined and interrelated in extremely complex ways defines the boundaries of the difference between a "dog" and a "cat".

Instead, scientists approach this problem by building an artificial brain capable of learning. Then, they feed this knowledge-hungry empty brain with thousands of training examples, *i.e.* associations of images with their corresponding labels, *e.g.* examples of "dogs", and examples of "dolphins". If the brain learns well, it will successfully tell the correct answer the next time it sees an image of a "dog", even if it has never seen that particular dog before. That is pretty much the way you can tell that something is a tree even if is the first time ever you're ever seen that particular tree. The reason you can do it is because you have been exposed to countless different variations of trees before, so you know what kinds of patterns are unique signatures of trees.

The mathematical model scientists are using for image classification are the so-called "neural networks". Heavily inspired in their biological counterparts, they are beautiful abstractions that are helping us not only to build fantastic applications, but also to understand better ourselves.

– The Artificial Neuron

When building an artificial mind, if we want to be inspired by biology, that is, the human brain and how from billions of neurons and trillions of synapses emerges intelligence, then is reasonable to focus on the functional aspects of the neurons and the brain – i.e. the workings in terms of information processing, rather than modelling specifics to the biological substrate.

With that perspective, one can say Warren McCulloch and Walter Pitts introduced the pioneer work in artificial neural networks in 1943 (McCulloch & Pitts, 1943). They proposed a mathematical model for the neuron. In their model, the neuron is an information-processing device that takes signals from other neurons connected to it through synapses and produces an output signal of activation, which can be -1 or 1, inactive or active respectively.

The synapses themselves encode the amount of "inverse-resistance" of signal as an amount called weight, which is valued between 0 and 1, full-resistance (no signal) and zero resistance respectively. The reason the weights encode the inverse-resistance is because it makes calculations simple.

Finally, the body of the neuron, takes all input signals, multiplies by their respective weights, sums, and decides whether or not it passes a certain threshold (specified by the neuron). If it does, then it fires a positive 1 signal, otherwise, it fires -1. Figure 14 depicts graphically the McCulloch-Pitts model. Equation (1) formalizes the artificial neuron processing.

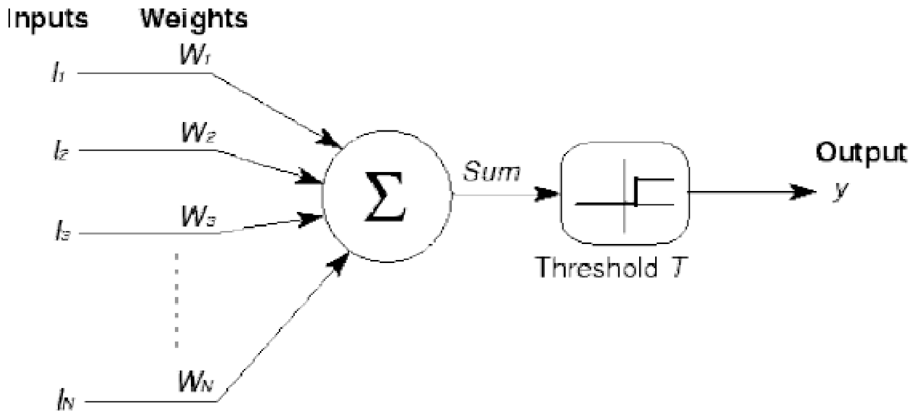

Figure 14: model of McCulloch-Pitts neuron

$$y = \begin{cases} \mathbf{1}, & \sum_{j=1}^{N} I_j W_j \geq T \\ -\mathbf{1}, & \textit{otherwise} \end{cases} \qquad (1)$$

– The Memorable Machine

One cannot learn if one cannot remember. Our neural network must be able to be exposed to training examples and then remember patterns to recall them the next time it sees a real example.

Traditionally, the memory in the brain works much differently than memory in your computer, the key distinction is on how the information is retrieved. In your computer, data is **addressable**, and one can look up the data if one knows the address location of the data. In your brain, on the other hand, memory is **associative**, and therefore accessed by parts of the original content, *i.e.* in order for you to remember something, you need to start by something else that is associated with the former. Think about how you can remember an entire song by just initiating with the first word, or a location by experiencing the same smell. The brain stores information in a way that makes it easy to access by using a subset of that original information.

Let's introduce some interesting mathematics that shall be useful to have a better sense of what we are trying to achieve in the next sections, starting with the Hopfield network. Popularized by John Hopfield in 1982 **(Hopfield, 1982)**, a Hopfield Network is a mathematical structure that can be implemented in a computer, which exhibits the property of information storage and retrieval using associative memory.

Each neuron in a Hopfield network is a McCulloch-Pitts neuron; also, on top of processing information, they also have state, in the form of a property of "excitement", specifying if that particular one is "active" or "inactive". This property is modeled as a number that can have a value of -1, for inactive, and 1, for active.

Finally, each neuron in the Hopfield network is connected to all other neurons thru the synapses both for input and output. With the additional constrain of being symmetric, that is a synapse from neuron A to B always have the same weight as the synapse from B to A.

The network updates itself throughout time, by following the McCulloch-Pitts signal processing, and setting their activation state with the resulting output.

The collective (binary) information of all neurons activations is what the network is "thinking about" (the pattern). One interesting property of the network is that it converges to stable configurations of low energy. If you start “near” a low energy pattern – *i.e.* with a subset of neurons activations in synchrony with the pattern – after some cycles of update, the network will eventually converge to a low-energy attractor pattern and stabilizes there. This is, in essence, associative recall. Equation (2) formalizes exactly what we mean by measure of energy.

$$E = -\frac{1}{2}\sum_{ij} W_{ij}s_i s_j + \sum_i W_i s_i \qquad (2)$$

Where s_i is the state of the *i*-th neuron. The equation states that pairs of neurons active together contribute to a lower energy the bigger the synaptic weight between them. Likewise, pairs of neurons that are out of sync, contribute to a higher energy (because the states are negative), the bigger the synaptic weight between them. **Figure 15** is a (one dimensional) visualization that energy equation as a function of the neurons state, highlighting the current state being attracted to a local minima of low energy during update – once reached the local minima, the network stabilizes.

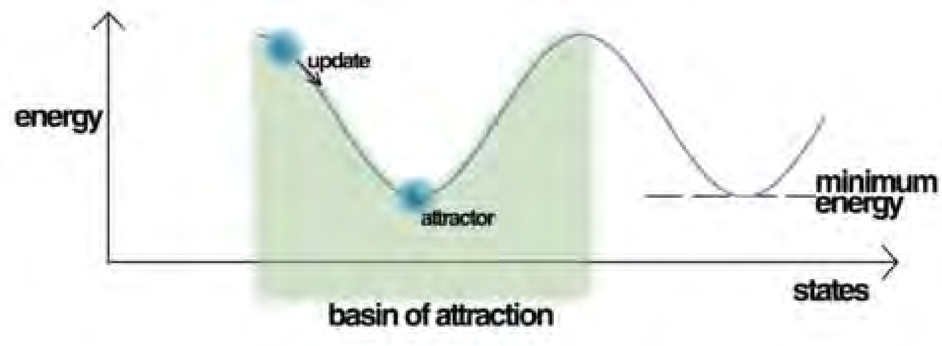

Figure 15 Energy landscape of a Hopfield network. Local minima states will attract the current state during the update (Wikipedia).

Training a Hopfield network involves lowering the energy of states that the network should "remember", and this is precisely solved by a learning rule introduced by Donald Hebb in 1949 **(Hebb, 1949)** (a.k.a. Hebbian learning rule). Basically, during the training phase, the network is presented with a pattern to remember, this is done by setting the activation levels of all neurons to be like the presented pattern (*e.g.* an image), and then we follow just two simple rules of learning:

- Decrease the synaptic weight between neurons that out of sync.
- Increase the synaptic weight between neurons that are active together.

This rule is often summarized as "Neurons that fire together, wire together. Neurons that fire out of sync, fail to link". These modifications will transform the network to have these training patterns as low-energy attractors.

The capacity of the Hopfield network is proportional to the number of neurons, and is considered a corner stone to more sophisticated algorithms used to build modern neural networks.

– Visual Cortex

One of the major milestones in image recognition was the refinement, by Yann LeCun, *et. al.*, of the Convolutional neural network (LeCun, Bottou, Bengio, & Haffner, 1998). The invention of this technique was principal to the major developments of today’s applications in image recognition.

Intuitively, the conception of the Convolutional neural network comes from the realization that many visual patterns found in images are repeated in different positions of the image, for example, when identifying a tree in a figure, leaves, branches and roots are found at different places (and at different scales across the visual space).

It is a waste of computing to force a system to learn the same patterns at different places of the visual field, in fact, without Convolutional networks, is impractical to learn patterns from relatively larger images effectively. The way those networks works is by learning what is called “convolutional kernels”, which are smaller versions of neural networks, specialized at learning one specific parameter. A typical system will have several of these kernels, each concentrated in learning one specific type of pattern. Since those patterns can be “everywhere”, the kernels are replicated across the image and fed with the input from each replicated region.

Figure 16 depicts an example where, starting from an original image (left), three different convolution kernels are applied. The kernels are numeric matrixes (on the top), which are applied (multiplied by the image pixels) across the entire image. Convolutional neural networks use all convolution output in order to learn the relation of different aspects of one image. For example, detecting horizontal lines in one kernel, and vertical lines in another kernel in order to make higher-level considerations of full object boundaries.

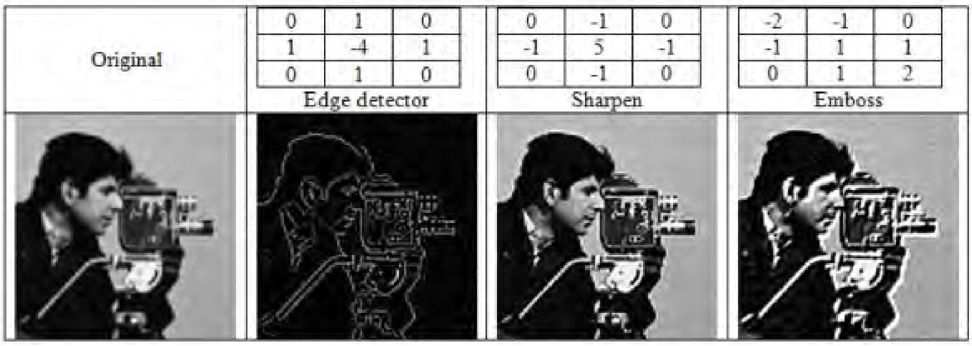

Figure 16 example of three convolution kernels applied to an image

– Deep Learning Revolution

Learning how to “see” is one of the most remarkably tasks our brain does. There is a big neural path dedicated to processing images – the visual cortex, and research has shown that its structure is hierarchical.

It has been shown that different layers of the hierarchy of the cortex in mammals are responsible for leaning one specific feature of images, and the higher the layer, the higher the conceptual space of the features. Lower levels, right after the retina, are responsible for understanding edges and lines, after that, subsequent layers processes more sophisticated patterns like perspective, shadow, *etc.* Moving further, upper layers identify full objects like an eye or a nose. Finally, top layers are the ones putting all those pieces together to identify a person or a full scene, composed of numerous objects.

Deep Learning, inspired by the visual cortex architecture, is a fundamental innovation in neural networks, pioneered by Geoffrey Hinton and his team (Hinton & Salakhutdinov, Reducing the dimensionality of data with neural networks, 2006). In this paradigm, the artificial brain is composed of several neural networks stacked on each other – each layer of the stack is responsible for learning a particular pattern in the conceptual hierarchy, interpreting that pattern, and feeding the upper layer with some digested information, very analogous to the biological counterparts.

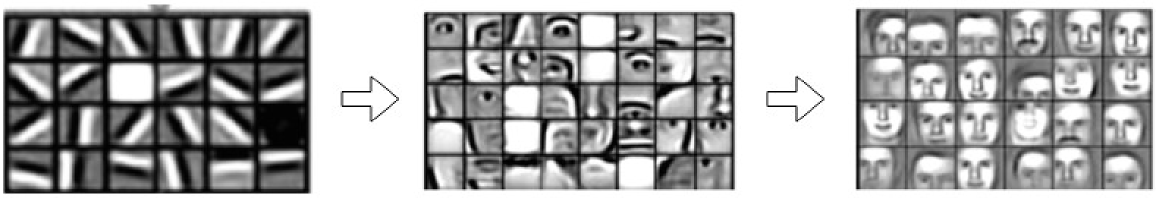

Figure 17 Increasingly complex pattern recognition in the visual cortex hierarchy

That, combined with convolutional neural networks, and an optimization of the previously discussed Hopfield network, called Restricted Boltzmann Machines, introduced by Hilton (Hinton, A practical guide to training restricted Boltzmann machines, 2010), finally allowed computers to perform like or even beat humans in image recognition tasks.

One particular thing to note about the Restricted Boltzmann Machines learning algorithm, is that for effective learning, it has to be submitted through a “dreaming” process, in which patterns are generated by the network without real input – and those patterns are “forgotten” by it, in order to make space for the real stuff, during the “awake” phase. It’s wonderful how we are mimicking biology in all these different facets in order to paint the big picture of artificial intelligence.

– **Inception**

What happens, then, if you take those Deep Learning systems and, instead of feeding them data for classification, look inside in the hope of inspecting what they are “thinking of”?

That is the sort of question made by a group of Google engineers in 2015 (Mordvintsev, Olah, & Tyka, 2015). It is known that after training, each layer progressively extracts higher and higher-level features of the image, until the final layer essentially makes a decision on what the image shows. On the quest to understand what exactly goes on at each layer, the engineers turned the network upside down and asked it to enhance an input image in such a way as to elicit a particular interpretation. Say you want to know what sort of image would result in “banana.” Start with an image full of random noise, and then gradually tweak the image towards what the neural net considers a banana (Figure 18).

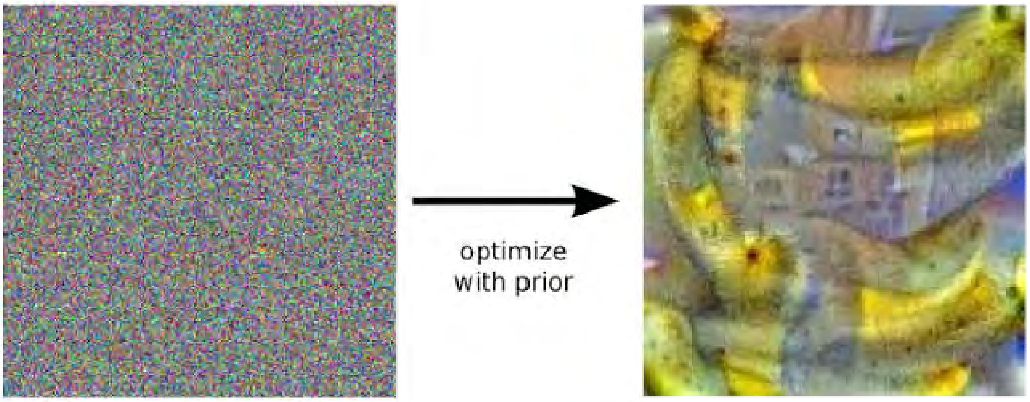

Figure 18 Optimizing a network to "see" bananas in random noise (figure from blog post)

The network used is called “GoogLeNet” (Szegedy, et al., 2014), a 22 layers deep network that was trained on the prominent ImageNet (Russakovsky, et al., 2015) dataset, a collection of hundreds of thousands of images of 200 different classes. From this network, we used the techniques described by the Google engineers to produce wonderful and bizarre images from the “deep dreams” of that network. The remainder of this section will expose a few of these results.

Figure 19 The strange city of car-men

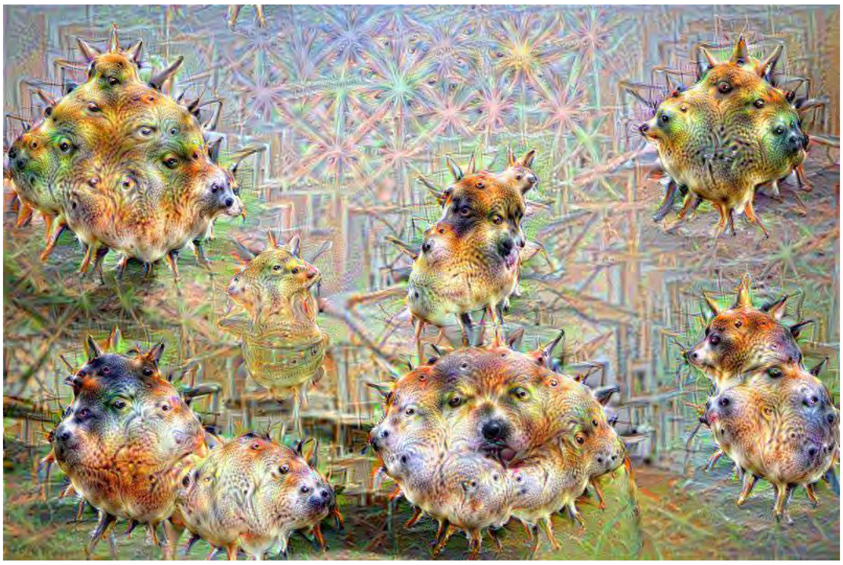

Figure 20 The horrific melted-puppies

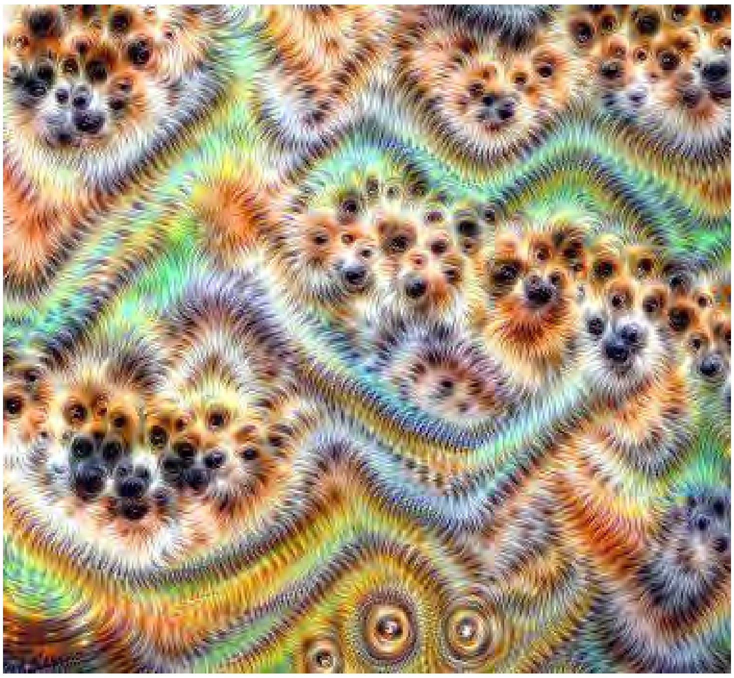

Figure 21 The atrocious puppy-faced carpet

Figure 22 Bizarre containers

Figure 23 Wonderful spider nets

Figure 24 Mister grasshopper frog

Figure 25 The curious two-headed bird

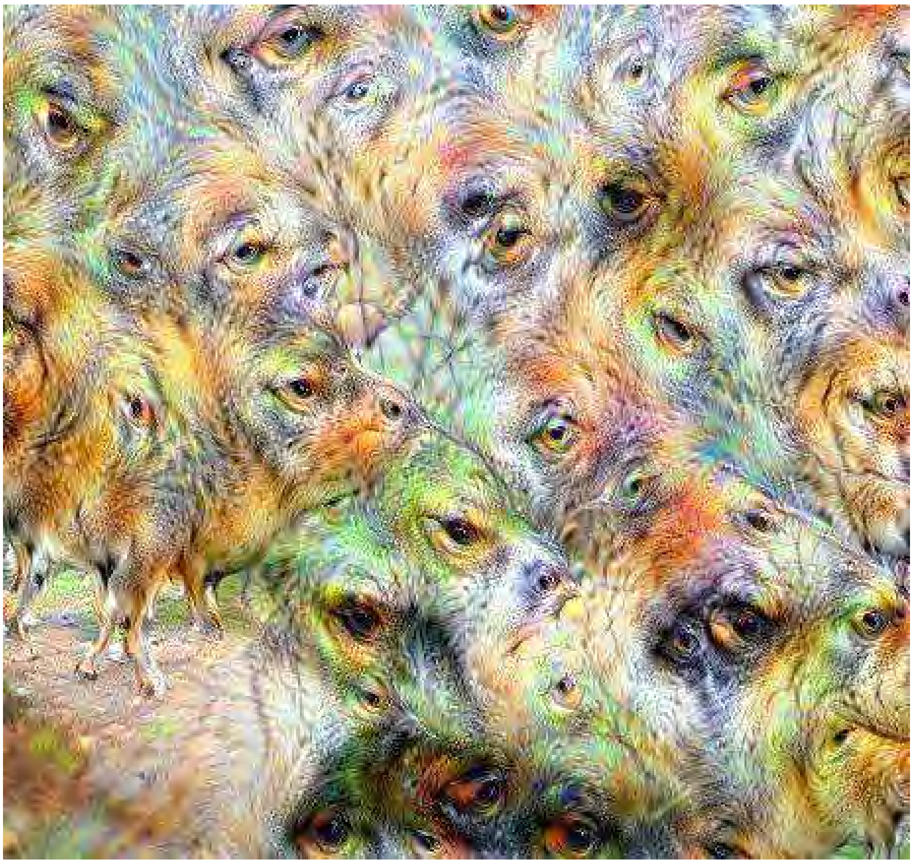

Figure 26 The dreadful wall of judgemental eyes

Figure 27 Several iterations over the original image on the top left (Bathers at Asnieres), over different layers of the network. Starting from lower layers (optimizing edges and curves), down to upper layers (optimizing objects and complex forms)

– Conclusion and Future Work

We would like to explore different Deep Learning topologies, not only GoogLeNet. There are topologies specialized in different domains like hand-writing recognition and face recognition that would definitely bring something new to the generated images.

Another area of exploration is to train the same network (or other networks) with different image datasets (not only ImageNet). For example, one can train one network with works of art of specific category (like impressionism), and hopefully have being provoked this system to dream about impressionist strokes.

On top of different networks and image datasets, there is also a broad space of parameters and techniques when generating the images themselves, layers can be optimized in conjunction, and with different optimization functions – certainly there are interesting discoveries waiting to be made on that front.

Finally, we would like to add that there are other approaches to image classification that don't use neural networks, Support Vector Machines, and K-Means, are among the supervised machine learning algorithms that can be used for that purpose. They also can be used to generate images from trained models, likely with completely different characteristics.

– References

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hinton, G. E. (2010). A practical guide to training restricted Boltzmann machines. *Momentum , 9* (1), 926.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science , 313* (5786), 504-507.

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The" wake-sleep" algorithm for unsupervised neural networks. *Science , 268* (5214), 1158-1161.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation , 18* (7), 1527-1554.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences. 79*, p. 2554-2558. National Academy of Sciences.

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *IEEE. 86*, p. 2278-2324. IEEE.

Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *26th Annual International Conference on Machine Learning* (p. 609-616). ACM.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics , 5* (4), 115-133.

Mordvintsev, A., Olah, C., & Tyka, M. (2015, June 17). *Inceptionism: Going Deeper into Neural Networks*. Tratto il giorno October 25, 2015 da Google Research Blog: http://googleresearch.blogspot.com/2015/06/inceptionism-going-deeper-into-neural.html

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015, April). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* , 1-42.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2014). Going deeper with convolutions. *arXiv preprint arXiv:1409.4842* .