Convolutional Neural Networks as a Computational Model for the Underlying Processes of Aesthetics Perception

Joachim Denzler, Erik Rodner, Marcel Simon

Computer Vision Group,
Friedrich Schiller University Jena, Germany
{firstname.lastname}@uni-jena.de
http://www.inf-cv.uni-jena.de

Abstract. Understanding the underlying processes of aesthetic perception is one of the ultimate goals in empirical aesthetics. While deep learning and convolutional neural networks (CNN) already arrived in the area of aesthetic rating of art and photographs, only little attempts have been made to apply CNNs as the underlying model for aesthetic perception. The information processing architecture of CNNs shows a strong match with the visual processing pipeline in the human visual system. Thus, it seems reasonable to exploit such models to gain better insight into the universal processes that drives aesthetic perception. This work shows first results supporting this claim by analyzing already known common statistical properties of visual art, like sparsity and self-similarity, with the help of CNNs. We report about observed differences in the responses of individual layers between art and non-art images, both in forward and backward (simulation) processing, that might open new directions of research in empirical aesthetics.

Keywords: Aesthetic perception, empirical aesthetics, convolutional neural networks

1 Introduction

Today, researchers from a variety of disciplines, for example, psychology, neuroscience, sociology, museology, art history, philosophy, and recently mathematicians and computer scientists, are active in the area of understanding, modeling, or identifying processes related to aesthetic and aesthetic perception. The reason for such a still increasing interest in aesthetics arises from several questions:

1. How do artists create artwork? What are the underlying processes during such an artistic creativity?
2. What are the underlying processes in our brain leading to an aesthetic perception of specific images, text, sounds, etc.?
3. Can we compute a universal aesthetic value for art not being biased by cultural or educational background?

4. Are we able to optimize creation of art? Can we even support users of cameras to optimize the artistic value of the images and videos they record?

The first two questions resides more in neuroscience and psychology, with the goal to *propose* and *verify* models for the underlying processes, and to explain certain observations in aesthetic perception. The third question seems to be a machine learning problem. However, training a discriminative classifier from data will always suffer from the possible bias in the training set. The forth question would benefit from an available *generative model* that allows the creation or modification of images with certain aesthetic properties. Besides commercial interest, artificial creation of visual art with specific aesthetic values, will be helpful for psychological studies as well to verify answers for the first question.

Our observation is that a computational framework to *model* aesthetic perception is still missing, although the joint efforts in the intersection of experimental aesthetics, computer vision, and machine learning lead to numerous interesting findings. While researchers from computer vision and machine learning are satisfied with accurate prediction of aesthetic value or beauty of images, researchers from empirical aesthetics hunt for findings and interesting properties that allow differentiation of art from non-art. However, the connecting element, a *computational model* is still missing that would be of significant help towards answering at least three of the four questions from above.

In this study, we want to show perspectives towards model building in empirical aesthetics. The main motivation of our work arises from recent progress of and insight in deep learning methods and convolutional neural networks (CNN). CNNs are multi-layer neural networks with convolutional, pooling, and fully connected layers. Currently, these methods define state-of-the-art in many different computer vision and machine learning tasks, like object detection and classification. More details can be found in Section 3. Due to parallels of the processing architecture of CNN and that of the visual cortex [1], such models might be an ideal basis for further investigation of properties of visual art, as well as using CNNs models to verify hypotheses of empirical aesthetics. In addition to the arrival of CNNs, also recently various rated datasets of visual art have become publicly available and enabled further research in the field. Examples are the AVA dataset [2], the JenAesthetics dataset [3], and - although without rating - the Google Art Project [4].

2 Progress in Computational and Empirical Aesthetics

One area of research is computational aesthetics. According to encyclopedia britannica [5], computational aesthetics, is

"a subfield of artificial intelligence (AI) concerned with the computational assessment of beauty in domains of human creative expression such as music, visual art, poetry, and chess problems. Typically, mathematical formulas that represent aesthetic features or principles are used in conjunction with specialized algorithms and statistical techniques to provide numerical aesthetic assessments. Those assessments, ideally, can be shown to correlate well with domain-

competent or expert human assessment. That can be useful, for example, when willing human assessors are difficult to find or prohibitively expensive or when there are too many objects to be evaluated. Such technology can be more reliable and consistent than human assessment, which is often subjective and prone to personal biases. Computational aesthetics may also improve understanding of human aesthetic perception".

Obviously, understanding of human aesthetic perception is not in the main focus. Successful results from this area of research are measuring aesthetic quality of photography [6–9] and paintings [10, 11], quality enhancement of photos [12, 13], analysis of photographic composition [14–16], classification of style [17–19], and composition [20, 21], and painter [22]. Most recently, related work has been published for videos as well [23–26].

Some works present results for automatic creation of art [27, 28], for measuring emotional effects of artwork on humans [29–31], and for improving and quantifying quality of art restoration [32]. Commercial use of those results for building intelligent cameras can be found in [33, 34]. Systems that provide a (web-based) rating tool of photographs are [35, 36].

The second main area of research related to aesthetics is empirical aesthetics. The aim of empirical aesthetics is to develop and apply methods to explain the psychological, neuronal, and socio-cultural basis of aesthetic perceptions and judgments. Compared to computational aesthetics, the aim is more understanding of the processes in aesthetic perception, i.e. why do people perceive music and visual art as varying in their beauty, based on factors such as culture, society, historical period, and individual taste. Some researchers are also interested in general principles of aesthetic perception independent from the so called cultural or educational filter [37].

While computational aesthetics can be interpreted as an application-driven procedure, like a discriminative classifier, empirical aesthetics aims more at observation and verification of hypothesis, i.e. parallels can be drawn to generative classifiers. A lot of findings have been reported about, for example, common statistical properties of visual art and natural scenes [38, 39], certain unique properties of art work, like anisotropy in the Fourier domain [40] or self-similarity in the spatial domain [41], verified for different artwork, like faces/portraits [42], text/artistic writing [43], print advertisement/architecture [44], as well as cartoons/comics/mangas. However, the main shortcoming of most of the work from empirical aesthetics is the observation-driven approach without succeeding model building. Although we have observed and identified that there is a difference between images of visual art and arbitrary images, we do not know how this differences leads to aesthetic perception in humans. This finding would be a preliminary to finally understand the process of art creation. In other words, it is necessary to come up with an initial mathematical and computational model of aesthetic perception that can be verified and tested. As in many other disciplines such a model can be used to iteratively gain more insight into the involved processes, to verify hypotheses, to adapt and improve the model itself, and to even synthesize artwork as feedback for human raters in psychological studies. To the

best of our knowledge, there exists no (mathematical or computer) model for aesthetic perception capable to match with findings from empirical aesthetics, i.e. more effort must be put into developing such a model that can be used to explain known, unique properties of visual art and to relate those properties to processing principles in our visual cortex/brain.

In this study, we investigate the potentials of CNNs with respect to model building for aesthetic perception by asking the following questions:

- Is there any difference in the representation of images of artwork in CNNs compared to standard images (Section 6)?
- Which representation level (layer) of a CNN shows the most prominent differences (Section 6)?
- Is there any difference in the findings, if the CNN is trained with images from natural scenes compared to one that is trained on ImageNet? Can we confirm the hypothesis that the human visual system is adapted to process natural scenes more efficiently and that this adaptation builds the basis for aesthetic visual perception (Section 6)?
- Can we confirm some of the hypotheses from prior work in terms of sparse coding/processing for art, natural scenes, and general images (Section 7)?
- Can CNNs serve as a computational model to generate or modify images with certain universal aesthetic properties (Section 8)?

3 Convolutional Neural Networks

Convolutional neural networks are parameterized models for a transformation of an image to a given output vector. They are extensively used for classification [45], where the output vector consists of classification probabilities for each class, and detection problems [46, 47], where the output vector may additionally contain the position of certain objects [47]. The model itself is comprised of a concatenation of simple operations, the so called layers.

The number of parameters of such a model in typically used architectures is often in the order of millions. Thus, training from a large-scale dataset is required. The interesting aspect that motivated our research is that there is strong evidence that such models can indeed learn very typical structural elements of natural images and objects. This was shown for example in the works of [48] for a CNN learned from ImageNet, which is the common "birthplace" of CNNs in vision currently. Furthermore, there has been further empirical evidence that these models are better suited for modeling the visual processing of human vision. Agrawal et al. [49], for example, investigates CNNs for neuroscience studies, where human brain activity is predicted directly from features extracted from the visual stimulus. In addition, Ramakrishnan et al. [50] compares CNNs with other layered vision models in an fMRI study.

Let's walk through a sample architecture, which is sketched in Fig. 1. The input for our network is in our case a single image. The first layer is a convolutional layer, which convolves the image with multiple learned filter masks. Afterwards, the outputs at neighboring locations are optionally combined by

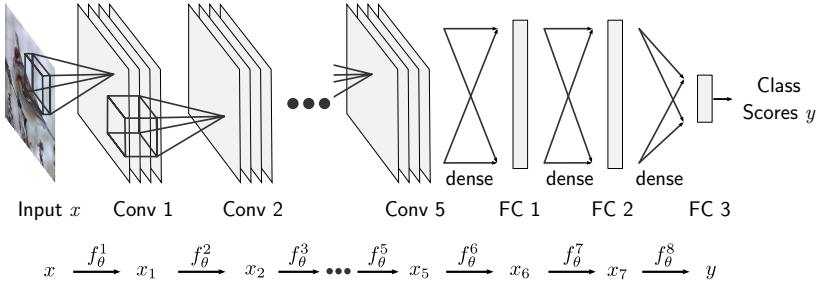

Fig. 1. Example of a convolutional neural network architecture.

applying a maximum operation in a spatial window applied to the result of each convolution, which is known as max-pooling layer. This is followed by an element-wise non-linear activation function, such as the rectified linear unit used in [45]. The last layers are fully connected layers which multiply the input with a matrix of learned parameters followed again by a non-linear activation function. The output of the network are scores for each of the learned categories. We do not provide a detailed explanation of the layers, since there is a wide range of papers and tutorials available already [45]. In summary, we can think about a CNN as one huge model that maps an image through different layers to a semantically meaningful output. The model is parameterized, which includes the weights in the fully connected layers as well as the weights of the convolution masks. All parameters of the CNN are learned by minimizing the error of the network output for an training example compared to the given ground-truth label.

Interestingly, many parts of the model can be directly related to models used in neuroscience [51] for modeling the V1-layer for example.

4 Dataset

We use all images of the JenAesthetics dataset [3], which is a well established dataset in the area of computational aesthetics. The dataset contains images of 1625 different oil paintings by 410 artist from 11 different art periods/styles.

The content or subject matter of a painting plays a crucial role in how an observer will perceive and assess a painting. 16 different keywords identify the most of the common subject matters. The subject matters are: abstract, nearly abstract, landscapes, scenes with person(s), still life, flowers or vegetation, animals, seascape, port or coast, sky, portrait (one person), portrait (many person), nudes, urban scene, building, interior scene, and other subject matters. 425 paintings have 3 and 1047 paintings have 2 subject matter keywords.

These images will serve as a representative sample of the category "art". Images not related to art paintings ("non-art") show various semantic concepts like plants, vegetation, buildings etc. Specifically, we used 175 photographs of building facades, 528 photographs of entire buildings, mostly without the ground

floors to avoid the inclusion of cars and people, 225 photographs of urban scenes. We also included an additional dataset [52] with 289 photographs of large-vista natural scenes, 289 photographs of vegetation taken from a distance of about 5-50m, and 316 close-up photographs of one type of plant. A detailed description of the used data can be found in [53, 42, 52].

5 Analyzed Models and Experimental Setup

Learning convolutional neural networks is done in a supervised fashion with pairs of images and the corresponding category labels [45]. Relating this to the processing in the brain, supervised learning of networks can be seen as teaching with different visual stimuli towards the goal of categorization in a specific task.

To study visual processing for different types of stimuli in the teaching phase, we train CNNs with a common architecture from three different datasets. We make use of the AlexNet [45] architecture. The first model is trained on roughly 1.5 million images and 1000 common object categories of the ImageNet Large Scale Visual Recognition Challenge 2012 [54] dataset, and is denoted by `imagenet_CNN` [45]. Second is the `places_CNN` [55], which is trained on over 7 million images divided into 205 scene categories including indoor and outdoor scenes, comprised by natural and man-made scenes. Third and last is a CNN trained on 125.000 images showing 128 categories of natural scenes. These images were taken from ImageNet and the categories were manually selected. We refer to this network as `natural_CNN` and it achieves an accuracy of almost 70% on the 1280 held-out test images of the natural scene categories. The reason we added this network to our analysis is that it allows us to study a model completely learned with natural non-human-made visual stimuli. In addition, we also experiment with the deeper architecture VGG19 proposed by [55].

In the layer names used in the following, the prefix `conv` refers to the output of convolutional layers and `fc` correspond to the output of a fully connected layer.

6 Separation of Art vs. Non-Art at Different Layers

In the beginning, we asked whether there is any difference in the representation of images (art vs. non-art) over the individual layers, i.e. at which level of the abstraction of an input image do we observe the largest difference. We also want to test whether there are differences in processing art images, if we initially train the CNNs on different datasets. With this experiment we want to verify whether the adaptation of the visual system of humans during evolution towards natural scenes plays a role for the underlying processes of aesthetic perception.

Measuring the differences is a non-trivial task since the output of a layer is high-dimensional. Therefore, we decided to use a classification approach, where we estimate the differences between both categories (artwork and all other images) over the individual layers by classification performance. The idea is, if the feature representations of the two categories are similar, the classifier would not

be able to separate the two classes. In particular, we learn a linear support vector machine classifier using the layer outputs for each image of the two categories. The classifier is learned on 25% of the data and the classification performance is measured using the remaining 75%. As a performance measure we use the area under the ROC curve, which is well suited for binary classification problems, since it's value is invariant with respect to the distribution of the categories in the test dataset. To increase the robustness of our estimates, we also sample 5 different splits of the data into training and testing subset. The SVM hyper-parameter is tuned using cross-validation for each run of the experiments. We restrict our analysis to linear layers, i.e. fully connected and convolutional layers.

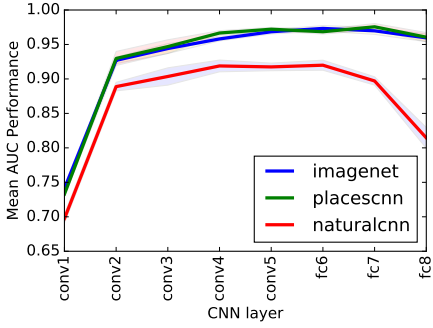

Fig. 2. Which layers of a CNN show the highest differences between artwork and all other images? We evaluate the separation ability of (1) `imagenet_CNN` (2) `natural_CNN` and (3) `places_CNN`.

The results of our analysis are given in Fig. 2 for all three of our networks. Regarding the maximum absolute performance, the `imagenet_CNN` showed the best performance. For `imagenet_CNN` and `places_CNN`, the differences between artwork and non-artwork increase up to the `conv4` layer and stay constant for later layers, which is not the case for `natural_CNN`. Interestingly, `natural_CNN` shows the worst performance, i.e. the statistics of images from natural scenes are not as well suited for separation of art and non-art images later on. This observation seems to be contradictory to the hypotheses that the adaption of the visual system toward natural scenes during evolution plays some role in explaining aesthetic perception. However, a more technical explanation is more likely. Since the art images under investigation show basically objects and scenes that are also present in ImageNet and Places data set, the representational power of those CNNs are superior to the one trained solely on natural scenes.

7 Are Artworks Characterized by Sparse Representations?

Next, we asked whether hypotheses from [37] and findings from [38] can be verified for representations in CNNs as well. One hypotheses is that a universal model of aesthetic perception is based on sparse, i.e. efficient coding, of sensory input [37, Chapter 4]. If activities in the visual cortex can be coded with sparse representations, they allow for efficient processing with minimal energy. Comparing statistics of natural scenes and visual art showed that these two categories of images share a common property related to sparsity in the representation [38, 39]. Hence, we analyze next the sparsity of the output representations in different layers of a CNN. Sparse CNN representations of visual stimuli correspond to only a few activated neurons with non-zero output for which we first need to define a mathematical measure.

Sparsity measure As a sparsity measure for a representation, we use the ℓ_1/ℓ_2 value given in [56], which we additionally normalize as follows to compare values of this measure also for vectors of different dimensionality. Let $\boldsymbol{x} \in \mathbb{R}^D$ be a vector of size D, our sparsity is therefore defined as follows:

$$\text{sparsity}(\boldsymbol{x}) = \frac{1}{\sqrt{D}} \frac{\|\boldsymbol{x}\|_1}{\|\boldsymbol{x}_2\|_2} = \frac{\sum_{k=1}^{D} |x_k|}{\sqrt{D \cdot \sum_{k=1}^{D} x_k^2}} \leq 1 \ . \tag{1}$$

This sparsity measure is small for sparse vectors, e.g. $\text{sparsity}([1, 0, \ldots, 0]) = \frac{1}{\sqrt{D}}$, and high for non-sparse vectors, e.g. $\text{sparsity}([1, \ldots, 1]) = 1$. In contrast to the standard ℓ_0 sparsity measure, where simply non-zero components are counted, our sparsity measure has the advantage of being smooth and taking into account approximate sparseness, e.g. with vectors having values close to zero relative to the overall magnitude.

Sparsity values for art and non-art images and different CNNs Fig. 3 shows the distribution of sparsity values for pairs of layers and different networks. The figures reflect our results obtained in Section 6: the discrimination ability increases for `imagenet_CNN` and `places_CNN` in later layers. It is indeed interesting that this is reflected in the sparsity values as well. Art images show more sparse representations at layer fc6 than non-art images. The lower representational power of the `natural_CNN` is confirmed in this analysis as well. Images from art and non-art show no significant difference in terms of sparsity over the individual layers. However, the representation in the intermediate layers (see conv1 vs. conv5) is systematically more sparse for `natural_CNN`, ranging from values 0.09 to 0.14 compared to values from 0.14 to 0.28 for the other two CNNs. So far, no conclusion seems be to directly possible. However, this experiment shows that there are differences in sparsity between layers and networks when comparing art and non-art images.

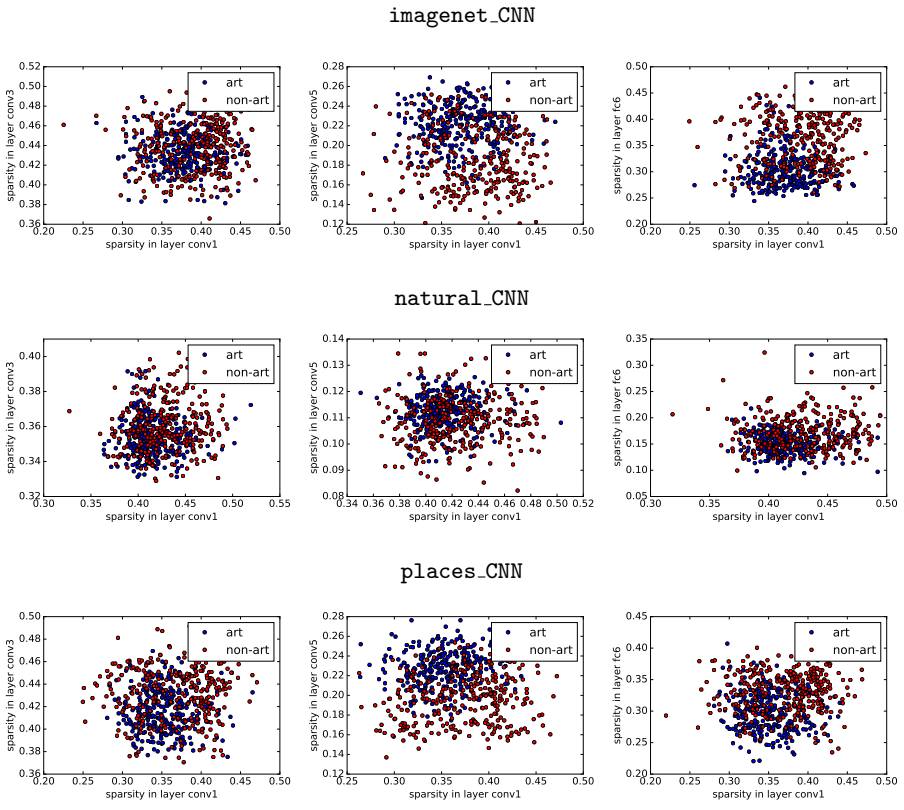

Fig. 3. Distribution of sparsity scores for art and non-art images computed for the outputs of two layers. Columns: conv1 vs. conv3, conv1 vs. conv5, conv1 vs. fc6. Rows correspond to different networks: `imagenet_CNN`, `natural_CNN`, and `places_CNN`. Smaller values correspond to higher sparsity. Best viewed in color.

8 CNN as Generative Model: Transferring the Statistics of Artworks

In the following, we analyze the change of intrinsic statistics of images, when we apply methods that allow for optimizing common images towards being "art-like". This includes the texture transfer method of [57] as well as the method of [58], which we modified to maximizing the probability of the image for belonging to the "art category". We can indeed show that transferring images towards art-like also transfers intrinsic statistical properties, like self-similarity or sparsity, towards art-like.

8.1 Texture transfer

The work of [57] presented an approach for transferring the style from a painting to a different image. In this section, we will use this idea to visualize and

understand the type of style information encoded in each layer of the CNN. It will turn out that each layer captures a fundamentally different aspect of the style, which can also be connected to the observations concerning sparsity of the previous sections.

The style transfer approach of [57] takes two images as input. One image provides the content and the second one the style, as shown in Figure 4. Starting from a white noise image, we now try to find a new image, which matches the content of the first and the style of the second image. This is done by changing the image step-by-step such that the neural activations at selected layers match the content and the style image, respectively. For the style image, the entries of the Gram matrix $G^l = \sum_{i,j}(x^l_{i,j}) \cdot (x^l_{i,j})^T$ of the style layer should match instead of the activations itself. Here, $x^l_{i,j} = (x^l_{i,j,k})_k$ denote the feature descriptor at position (i, j) of layer l. As an example, for the first output image in Figure 4, the white noise image is optimized such that the activations of layer conv4_2 match the content image and the Gram matrix of the activations of layer conv1_1 match the style image.

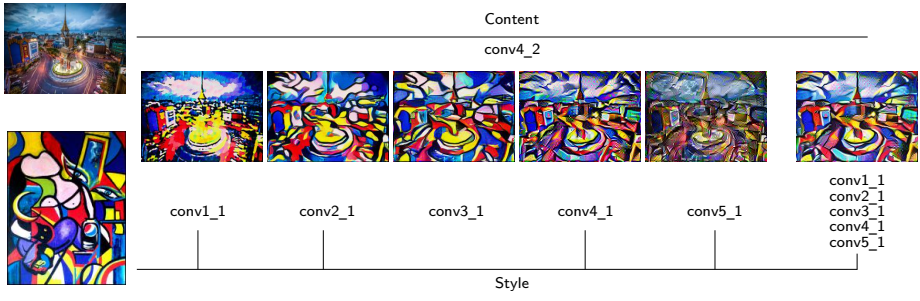

Fig. 4. Texture transfer for the content image shown on the top left and the style given by the image on the bottom left. The content was defined by the activations of the layer conv4_2 for all images. The style was defined by the bilinear activations of different layers as annotated below each image. Best viewed in color.

How does self-similarity change? With the above technique, we analyze the process of transforming an image into a more artistic image by transferring a mean style of artworks to images. We use all images of the JenAesthetics database, compute their mean Gram matrix and use it as the definition of "style" in the above algorithm.

Self-similarity is a well-known measure used in computational aesthetics to characterize an image [41, 42]. The question arises how the values of this measure change while optimizing a regular image towards an artwork. The results are given in Fig. 5, where we refer to the image after the texture transfer as "Art-transfer image". For all the images shown in the Fig. 5, we observe a significant increase in self-similarity after applying the texture transfer technique.

It is worth noting that the self-similarity values after texture transfer are in the range of artwork for not-art images as well. Even art images show an increase of self-similarity (second row, first image). Please also observe the "generation" of a synthetic art image in the last row. In this case, no content image was constraining the generation process. As in the other cases, we started with a white noise image and modified it such that the style matches the mean art feature.

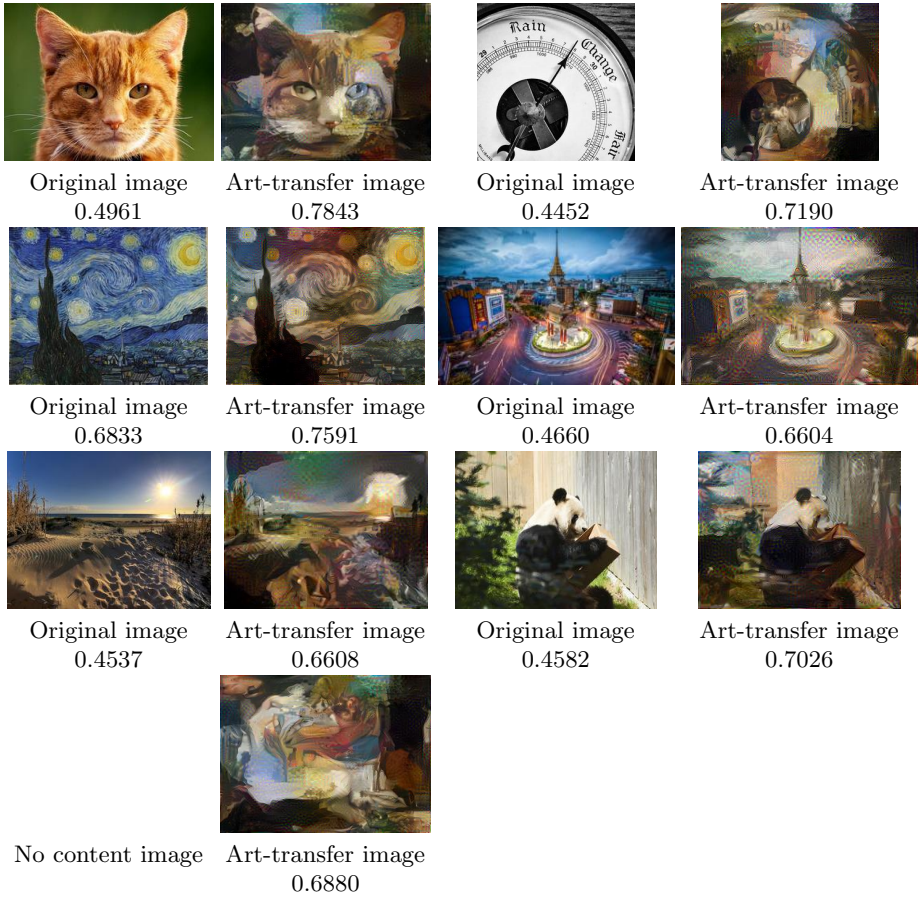

Fig. 5. Self-similarity changes when optimizing regular images towards artworks. We use the texture transfer technique of [57] in this case. The numbers below the images show the self-similarity scores from [42].

A second investigation concerns the change of the self-similarity score over time during the texture transfer. Fig. 6 depicts the progress for three examples. For each example, the plot is shown above the content as well as the final image. The first plot combines the input image with the style of the painting "Trans-

verse Line" by Kandinsky and the second one with the painting "Clin d'oeil à Picasso" by Bochaton. The self-similarity score of the generated image is shown in blue. As shown already in the previous figure, the self-similarity changes when transferring style to a new image. The change, however, is not monotonic but shows in the first ten iterations a dramatic increase from 0.5335 to 0.8056 and 0.8427, respectively, and thus surpasses the self-similarity score of the target style image depicted in green. After the initial overshoot, the score gets closer the one of the painting, but converges at a higher level of 0.7581 and 0.7363, respectively. The third subplot depicts the change over time for an optimization towards the mean art style. Similarly, there is an initial overshoot, followed by a short descent and a strong increase towards the final value of 0.7155.

We belief that these initial experiments with the texture transfer method indicate that CNNs are capable as a generative model in empirical aesthetics and can be exploited to generate images with specific, statistical properties related to an aesthetic value of the image.

8.2 Maximizing art probability

Adapting DeepDream towards optimizing art category probability Instead of indirectly optimizing images towards artworks by transferring the texture as done in the previous section, we can also perform the optimization directly.

First, we fine-tune a convolutional neural network to solve the binary classification task artworks vs. non-artworks. The original DeepDream technique of [58] tries to modify the image such that the L_2-norm of the activations of a certain layer is maximized. We modify this objective, such that the class probability for the artworks category is optimized. The algorithm for the optimization is still a gradient-descent algorithm as in [58] using gradients computed with back-propagation.

Details about the fine-tuning Fine-tuning is done with a pre-trained `imagenet_CNN` model. In particular, we set the learning rate to 0.001 and the batch size to 20 and perform optimization with the common tricks-of-the-trade: (1) momentum with $\mu = 0.9$, (2) weight decay of 0.0005 and (3) dropout with $p = 0.5$ applied after the first and second fully connected layers.

How does the self-similarity score change? Fig. 7 shows quantitative results of our analysis, where we compared the self-similarity scores before and after the optimization towards the artwork and the non-artwork category. As can be seen, the self-similarity scores increase in both cases for a large number of images. This is not the expected result when considering the findings of the texture transfer technique. So far, optimizing towards category probabilities seems not to be a reasonable method for enforcing certain statistical properties of art or non-art images. However, this is not a surprise, considering the amount of information contained in category probabilities compared to the target image style represented by the Gram matrix.

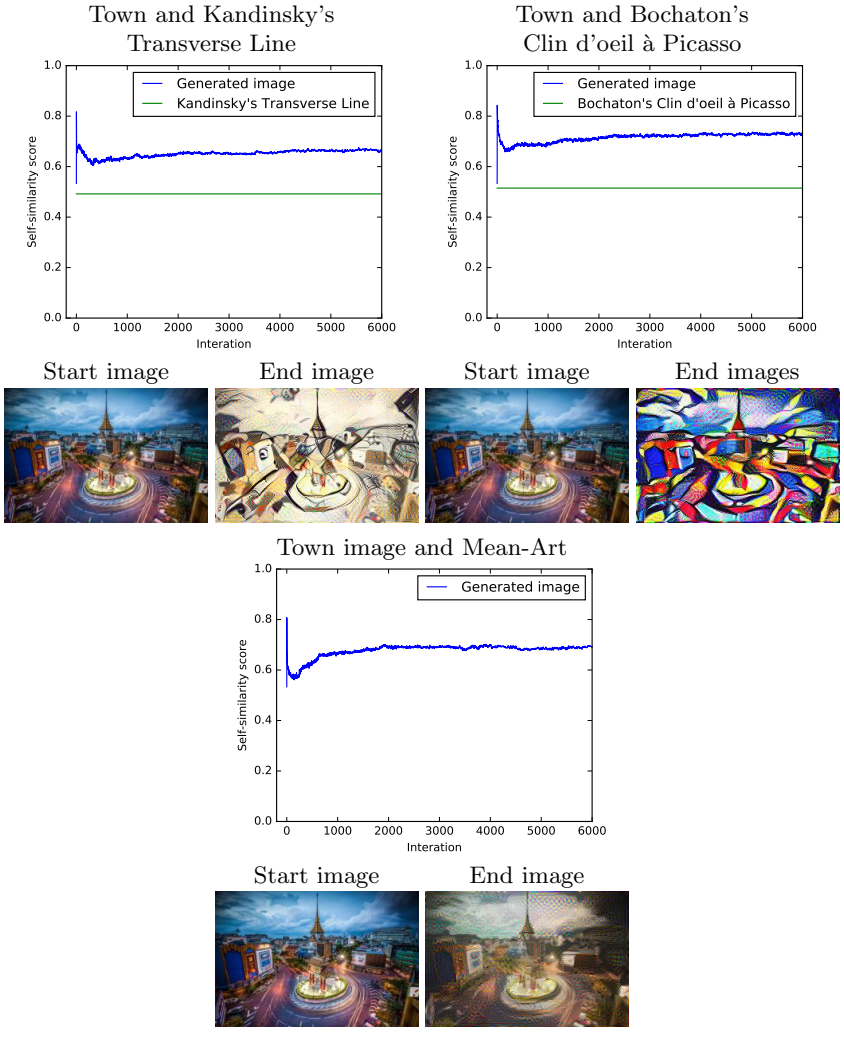

Fig. 6. Change of the self-similarity score over time during the texture transfer given an input image and a painting. We depict the change for two different styles transferred to the input image shown on the left. The plots show the self-similarity score of the image after iteration k in blue and the self-similarity score of the painting in green.

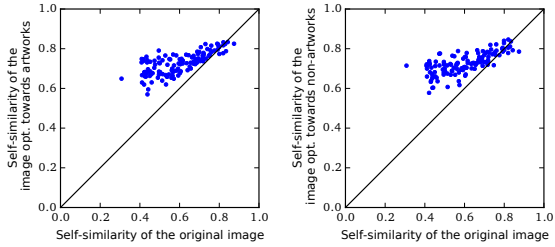

Fig. 7. Self-similarity scores before and after the optimization with respect to the artwork and the non-artwork category.

9 Conclusions

This work started with the observation that a computational model is missing in empirical aesthetic research. Such an initial model would allow the verification of hypotheses, to generate and modify images for psychological studies, to refine hypotheses as well as the model, and to initiate succeeding experiments and investigations along the way to understand the underlying processes in aesthetic perception.

We started to investigate the potentials of CNNs in this area, extending its previous use as pure classifier. The main goal was to figure out whether already known statistical properties of visual art, like sparsity and self-similarity, are reflected in the representation of images by CNNs as well. In addition, we analyzed two methods to use CNNs for generating new images with specific category properties (DeepDream) or style properties (texture transfer).

Our results indicate that there are statistical differences in the representation over the hierarchies of layers in CNNs. Those differences not only arise from the input image being processed, but also from the underlying training data of the CNN. The main finding is that sparsity of activations in individual layers is one property to be further investigated. This is in accordance with previous findings.

In addition, we applied CNNs as a generative model using techniques from literature. Interestingly, the method of texture transfer is able to modify self-similarity in images, a property that has been previously used to characterize art work. Hence, generating images with aesthetic properties seems to be possible as well.

So far, we only started the investigation. There are several aspects not considered so far, for example, different network architectures, different other statistical properties of aesthetic images, like fractality or anisotropy, and how such properties can be mapped to arbitrary images. These aspects are subject to future work.

References

1. Cadieu, C.F., Hong, H., Yamins, D.L., Pinto, N., Ardila, D., Solomon, E.A., Majaj, N.J., DiCarlo, J.J.: Deep neural networks rival the representation of primate it cortex for core visual object recognition. PLoS computational biology **10**(12) (2014) e1003963
2. Murray, N., Marchesotti, L., Perronnin, F.: Ava: A large-scale database for aesthetic visual analysis. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2408–2415
3. Amirshahi, S.A., Hayn-Leichsenring, G.U., Denzler, J., Redies, C.: Jenaesthetics subjective dataset: Analyzing paintings by subjective scores. In: European Conference on Computer Vision Workshops (ECCV-W), Springer (2014) 3–19
4. Proctor, N.: The google art project: A new generation of museums on the web? Curator: The Museum Journal **54**(2) (2011) 215–221
5. Goetz, P.W., McHenry, R., Hoiberg, D., eds.: Encyclopedia Britannica. Volume 9. Encyclopaedia Britannica Inc. (2010)
6. Ravi, F., Battiato, S.: A novel computational tool for aesthetic scoring of digital photography. In: Conference on Colour in Graphics, Imaging, and Vision, Society for Imaging Science and Technology (2012) 349–354
7. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. In: Computer Vision–ECCV 2006. Springer (2006) 288–301
8. Romero, J., Machado, P., Carballal, A., Osorio, O.: Aesthetic classification and sorting based on image compression. In: Applications of Evolutionary Computation. Springer (2011) 394–403
9. Wu, Y., Bauckhage, C., Thurau, C.: The good, the bad, and the ugly: Predicting aesthetic image labels. In: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE (2010) 1586–1589
10. Wickramasinghe, W., Dharmaratne, A.T., Kodikara, N.: A tool for ranking and enhancing aesthetic quality of paintings. In: Signal Processing, Image Processing and Pattern Recognition. Springer (2011) 254–260
11. Li, C., Chen, T.: Aesthetic visual quality assessment of paintings. Selected Topics in Signal Processing, IEEE Journal of **3**(2) (2009) 236–252
12. Bhattacharya, S., Sukthankar, R., Shah, M.: A framework for photo-quality assessment and enhancement based on visual aesthetics. In: Proceedings of the international conference on Multimedia, ACM (2010) 271–280
13. Zhang, F.L., Wang, M., Hu, S.M.: Aesthetic image enhancement by dependence-aware object re-composition. IEEE Transactions on Multimedia **15**(7) (2013) 1480–1490
14. Escoffery, D.: A framework for learning photographic composition preferences from gameplay data. Technical report, University of California, Santa Cruz (2012) Master Thesis.
15. Jin, Y., Wu, Q., Liu, L.: Aesthetic photo composition by optimal crop-and-warp. Computers & Graphics **36**(8) (2012) 955–965
16. Gallea, R., Ardizzone, E., Pirrone, R.: Automatic aesthetic photo composition. In: Image Analysis and Processing–ICIAP 2013. Springer (2013) 21–30
17. Wallraven, C., Fleming, R., Cunningham, D., Rigau, J., Feixas, M., Sbert, M.: Categorizing art: Comparing humans and computers. Computers & Graphics **33**(4) (2009) 484–495

18. Condorovici, R.G., Florea, C., Vrânceanu, R., Vertan, C.: Perceptually-inspired artistic genre identification system in digitized painting collections. In: Image Analysis. Springer (2013) 687–696
19. Karayev, S., Hertzmann, A., Winnemoeller, H., Agarwala, A., Darrell, T.: Recognizing image style. arXiv preprint arXiv:1311.3715 (2013)
20. Yao, L.: Automated analysis of composition and style of photographs and paintings. PhD thesis, The Pennsylvania State University (2013)
21. Obrador, P., Schmidt-Hackenberg, L., Oliver, N.: The role of image composition in image aesthetics. In: Image Processing (ICIP), 2010 17th IEEE International Conference on, IEEE (2010) 3185–3188
22. Cetinic, E., Grgic, S.: Automated painter recognition based on image feature extraction. In: ELMAR, 2013 55th International Symposium, IEEE (2013) 19–22
23. Wang, Y., Dai, Q., Feng, R., Jiang, Y.G.: Beauty is here: evaluating aesthetics in videos using multimodal features and free training data. In: Proceedings of the 21st ACM international conference on Multimedia, ACM (2013) 369–372
24. Chung, S., Sammartino, J., Bai, J., Barsky, B.A.: Can motion features inform video aesthetic preferences. University of California at Berkeley Technical Report No. UCB/EECS-2012-172June **29** (2012)
25. Bhattacharya, S., Nojavanasghari, B., Chen, T., Liu, D., Chang, S.F., Shah, M.: Towards a comprehensive computational model foraesthetic assessment of videos. In: Proceedings of the 21st ACM international conference on Multimedia, ACM (2013) 361–364
26. Moorthy, A.K., Obrador, P., Oliver, N.: Towards computational models of the visual aesthetic appeal of consumer videos. In: Computer Vision–ECCV 2010. Springer (2010) 1–14
27. Galanter, P.: Computational aesthetic evaluation: steps towards machine creativity. In: ACM SIGGRAPH 2012 Courses, ACM (2012) 14
28. Zhang, K., Harrell, S., Ji, X.: Computational aesthetics: On the complexity of computer-generated paintings. Leonardo **45**(3) (2012) 243–248
29. Zhang, H., Augilius, E., Honkela, T., Laaksonen, J., Gamper, H., Alene, H.: Analyzing emotional semantics of abstract art using low-level image features. In: Advances in Intelligent Data Analysis X. Springer (2011) 413–423
30. Joshi, D., Datta, R., Fedorovskaya, E., Luong, Q.T., Wang, J.Z., Li, J., Luo, J.: Aesthetics and emotions in images. Signal Processing Magazine, IEEE **28**(5) (2011) 94–115
31. Bertola, F., Patti, V.: Emotional responses to artworks in online collections. Proceedings of PATCH (2013)
32. Oncu, A.I., Deger, F., Hardeberg, J.Y.: Evaluation of digital inpainting quality in the context of artwork restoration. In: Computer Vision–ECCV 2012. Workshops and Demonstrations, Springer (2012) 561–570
33. Lo, K.Y., Liu, K.H., Chen, C.S.: Intelligent photographing interface with on-device aesthetic quality assessment. In: Computer Vision-ACCV 2012 Workshops, Springer (2013) 533–544
34. Mitarai, H., Itamiya, Y., Yoshitaka, A.: Interactive photographic shooting assistance based on composition and saliency. In: Computational Science and Its Applications–ICCSA 2013. Springer (2013) 348–363
35. Yao, L., Suryanarayan, P., Qiao, M., Wang, J.Z., Li, J.: Oscar: On-site composition and aesthetics feedback through exemplars for photographers. International journal of computer vision **96**(3) (2012) 353–383

36. Datta, R., Wang, J.Z.: Acquine: aesthetic quality inference engine-real-time automatic rating of photo aesthetics. In: Proceedings of the international conference on Multimedia information retrieval, ACM (2010) 421–424
37. Redies, C.: A universal model of esthetic perception based on the sensory coding of natural stimuli. Spatial Vision **21**(1) (2007) 97–117
38. Redies, C., Hasenstein, J., Denzler, J.: Fractal-like image statistics in visual art: similarity to natural scenes. Spatial Vision **21**(1-2) (2007) 97 – 117
39. Redies, C., H"anisch, J., Blickhan, M., Denzler, J.: Artists portray human faces with the fourier statistics of complex natural scenes. Network: Computation in Neural Systems **18**(3) (2007) 235–248
40. Koch, M., Denzler, J., Redies, C.: $1/f^2$ characteristics and isotropy in the fourier power spectra of visual art, cartoons, comics, mangas, and different categories of photographs. PLoS ONE **5**(8) (2010) e12268
41. S. A. Amirshahi, M. Koch, J.D., Redies, C.: Phog analysis of self-similarity in aesthetic images. In: IS T/SPIE Electronic Imaging. (2012)
42. Amirshahi, S.A., Redies, C., Denzler, J.: How self-similar are artworks at different levels of spatial resolution? In: Computational Aesthetics. (2013)
43. Melmer, T., Amirshahi, S.A., Koch, M., Denzler, J., Redies, C.: From regular text to artistic writing and artworks: Fourier statistics of images with low and high aesthetic appeal. Frontiers in Human Neuroscience **7**(00106) (2013)
44. J. Braun, S. A. Amirshahi, J.D., Redies, C.: Statistical image properties of print advertisements, visual artworks and images of architecture. Frontiers in Psychology **4** (2013) 808
45. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. (2012) 1097–1105
46. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2014) 580–587
47. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. arXiv preprint arXiv:1506.02640 (2015)
48. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Computer vision–ECCV 2014, Springer (2014) 818–833
49. Agrawal, P., Stansbury, D., Malik, J., Gallant, J.L.: Pixels to voxels: Modeling visual representation in the human brain. arXiv preprint arXiv:1407.5104 (2014)
50. Ramakrishnan, K., Scholte, S., Lamme, V., Smeulders, A., Ghebreab, S.: Convolutional neural networks in the brain: an fmri study. Journal of vision **15**(12) (2015) 371–371
51. Pinto, N., Cox, D.D., DiCarlo, J.J.: Why is real-world visual object recognition hard? PLoS Comput Biol **4**(1) (2008) e27
52. Redies, C., Amirshahi, S.A., Koch, M., Denzler, J.: Phog-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects. In: European Conference on Computer Vision (ECCV) VISART workshop. (2012)
53. Amirshahi, S.A., Denzler, J., Redies, C.: Jenaesthetics—a public dataset of paintings for aesthetic research. Technical report, Computer Vision Group, Friedrich-Schiller-University Jena (2013)
54. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision (2014) 1–42

55. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Advances in Neural Information Processing Systems. (2014) 487–495
56. Hurley, N., Rickard, S.: Comparing measures of sparsity. Information Theory, IEEE Transactions on **55**(10) (2009) 4723–4741
57. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
58. Mordvintsev, A., Tyka, M., Olah, C.: Inceptionism: Going deeper into neural networks, google research blog. Retreived June **17** (2015)