doi: 10.25038/am.v0i29.534

Jaana Okulov
The Department of Art and Media, Aalto University, Finland

Artificial Aesthetics and Aesthetic Machine Attention

Abstract: The aesthetics of artificial intelligence is often viewed in relation to the qualities of their generated expressions. However, aesthetics could have a broader role in developing machine perception. One of the main areas of expertise in aesthetics is the understanding of feature-based information, which involves how the aesthetics of sensory features can cause affective changes in the perceiver, and the other way around – how affective states can give rise to certain kinds of aesthetic features. This two-way link between aesthetic features and affects is not yet well-established in the interdisciplinary discussion; however, according to perceptual psychology, it fundamentally constructs the human experience.
Machine attention is an emerging technique in machine learning that is most often used in tasks like object detection, visual question answering, and language translation. Modern use of technology most often focuses on creating object-based attention through linguistic categories, although the models could also be utilized for nonverbal attention. This paper proposes the following perceptual conditions for aesthetic machine attention: 1) acknowledging that something appears (aesthetic detection); 2) suspension of judgment (aesthetic recognition); and 3) making the incident explicit with expression (aesthetic identification and amplification). These aspects are developed through an interdisciplinary reflection of literature from the fields of aesthetics, perceptual psychology, and machine learning. The paper does not aim to give a general account of aesthetic perception but to expand the interdisciplinary theory of aesthetics and specify the role of aesthetics among other disciplines at the heart of the technological development of the human future.

Keywords: attention; aesthetics; machine attention; feature-based knowledge; interdisciplinary theories.

1. Introduction: Artificial Aesthetics

The fields of AI art and AI aesthetics are not as new as the current acceleration of AI technologies implies. In fact, the historical link between AI and artistic thinking can be considered to have been founded in the work of mathematician Ada Lovelace, who, in her notes about Charles Babbage's analytical engine, published the first algorithm in 1842.[1] She envisioned a new discipline of thinking, *Poetical Science*, and believed a computer could, with its symbolic logic, solve problems of any complexity,

[1] Luigi Menabrea, "Notions sur la Machine Analytique de M. Charles Babbage," *Bibliothèque Universelle de Genève* 41 (1842): 352–76, trans. by Augusta Ada Lovelace, *Scientific Memoirs* 3 (1843): 666–731.

*Author contact information: jaanajohannaokulov@gmail.com

such as musical relations of harmony.[2] However, AI aesthetics, as conceived for and by the modern, international community, can be considered a product of the past 10 years. In 2015, Alex Mordvintsev released images created using a method developed by Zeiler and Fergus to visualize the learning curve in each of the hidden layers of a convolutional neural network.[3] Mordvintsev's *Cats* shows an output of an optimized layer of a neural network most strongly responding to the word 'cat'.[4] DeepDream, the program created by Mordvintsev, became available to the wider community later that year and is now mainly used as a vintage filter for images and videos that is reminiscent of the past aesthetics of AI art (I say this with gentle irony, as perhaps no other artistic styles or techniques can become vintage and outdated so rapidly as those associated with AI art).

After the release of DeepDream, the highest tide in AI art that has so far been observed cascaded from a multitude of different generative adversarial networks (GANs), which have been released to the public in recent years. These have democratized art, as they can generate new content from training material based on competitive networks, with the result that anyone, without mastering a technical skill, is able to express themselves aesthetically. For example, using a GAN in a platform called Artbreeder allows the user to cross-breed an image with different linguistic labels, tingeing it with a bit of, for example, *flamingoness* or *chaos* using sliders and, through the use of verbal intervention, obtain a sense of agency in the process.

Today, at the end of April 2022, the most current AI art can be found on Instagram with hashtags like #latentdiffusion or #discodiffusion. These refer to Disco Diffusion, a "frankensteinian amalgamation of notebooks, models, and techniques"[5] that is based on a notebook by programmer Katherine Crawson and has been further developed by programmers Daniel Russell, Dango233, nsheppard, Vark, Chigozie Nri, Somnai, and Adam Letts. Disco Diffusion uses a diffusion model that, according to Dhariwal and Nichol, can beat a GAN in image synthesis.[6] Another model, or more specifically a combination of two models, VQGAN+clip, is also widely used by the AI artist community.[7] Its aesthetic, which closely resembles that of Disco Diffusion, results from a combination of tweaked parameters, initiating images given to the model (image prompts), the verbal descriptive acrobatics of the user (text prompts), and

[2] Ada Lovelace's insights about computing are revolutionary even today, as she understood the potential of computers to use symbols of any kind and therefore created a link between mathematical logic and art. Robin Hammerman and Andrew Russell, *Ada's Legacy: Cultures of Computing from the Victorian to the Digital Age* (London: Morgan & Claypool, 2015).

[3] Matthew Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks" in *The European Conference on Computer Vision*, (2013), 818–33.

[4] "Alexander Mordvintsev" (web page), AIartist, https://aiartists.org/alexander-mordvintsev, acc. on April 19, 2022.

[5] "disco-diffusion" (Github repository), alembics, https://github.com/alembics/disco-diffusion, acc. on April 22, 2022.

[6] Prafulla Dhariwal and Alexander Nichol, "Diffusion Models Beat GANs on Image Synthesis," *Advances in Neural Information Processing Systems* 34 (2021): 8780–94.

[7] For example, the Nightcafe studio is widely used to generate images with the VQGAN+clip model. "VQGAN+-CLIP Text to Art Generator," Nightcafe, https://creator.nightcafe.studio/text-to-image-art, acc. on May 9, 2022.

their evolution as guided by 14 million human-annotated images of an image library (ImageNet is the default). The outputs are still images and short videoclips that can be aesthetically engineered to resemble any artistic style known by the model (annotated in the dataset). Especially in the video clips, the recognizable aesthetic characteristics enabled by zoom, keyframes, rotation, and pan features are perceived as strange warping motions that can be felt in the body – the distortion of the contextual information strongly influences the senses – and the viewing experience also involves some nightmarish cognitive pain, as just when the viewer recognizes an object, it disappears or reshapes into something else. One might briefly glimpse threads of ever-evolving hallucinations of *ultra-high-definitioned* building-like structures, *Geigered* animal-like shapes, and *Unreal Engined* landscapes, all of which are characteristic of art produced by an AI from human textual input.

Third, text-to-image model, DALL·E 2, which is accessible through a waiting list and used by a small number of AI artists, such as Memo Akten, Sofia Crespo, and Holly Herndon, was also released in April. The experiments with DALL·E 2 that the artists have published are unexpected in the sense of the extensions the model brings to text-to-image generation: There is more internal coherence in the shape, content, and lighting conditions of the still images, a higher resolution or even infinite scale,[8] and the user can make edits to some specific areas of the images with text.[9] These three models represent the state of the art in AI art today. Due to the reliance of these models on textual input—although GAN models have also been widely utilized also for image generation directly from learning material—they have interestingly turned AI art from a field focused on visual expertise toward one much more integrated with the linguistic realm.

The future of text-to-image models will have a drastic impact on image creation; it is possible that any image one can think of and describe with language will someday be generated with hardly any perceptual flaws, similar to the evolution that has happened with face generation.[10] In 2014, artificial faces were somewhat recognizable as faces but lacked the realistic touch; in 2022, they are almost impossible to differentiate from images of real faces. In some cases, only their irregularly shaped pupils[11] or the covert fingerprints the models leave behind[12] reveal them to be fake.

[8] "Infinite Images and the Latent Camera," Holly Herndon and Mathew Dryhurst, last modified on May 6, 2022, https://mirror.xyz/herndondryhurst.eth/eZG6mucl9fqU897XvJs0vUUMnm5OITpSWN8S-6KWamY, acc on May 6, 2022.

[9] "DALL·E 2," OpenAI, last modified April 6, 2022, https://openai.com/dall-e-2/, acc. on May 9, 2022. Similar model is also developed by independent research lab Midjourney (https://www.midjourney.com/home/).

[10] The authors show on page 13747 of the publication how the aesthetics of face generation has developed. Keyang Cheng et al., "An Analysis of Generative Adversarial Networks and Variants for Image Synthesis on MNIST Dataset," *Multimedia Tools and Applications* 79, 19 (2020): 13725–52.

[11] Hui Guo et al., "Eyes Tell All: Irregular Pupil Shapes Reveal GAN-Generated Faces," in *ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing* (2022), 2904–8.

[12] Tianyun Yang et al., "Learning to Disentangle GAN Fingerprint for Fake Image Attribution," arXiv Preprint, submitted 2021, *arXiv:2106.08749*; Ning Yu, "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), 7556–66.

I wonder if the ability to produce photorealistic content from one's imagination will pivot AI aesthetics from representational forms of expression towards deeper, artificial hallucinations, the same way that painting cut lose from its burden to represent the world accurately when photography and accessible equipment were developed in the 1830s. Photography could document the world, so painting, and interestingly, later also photography, could concentrate on capturing impressions. When artificial art can simulate existing reality, will the novelty of expression and thinking start to have more value? Will there be an explosion of techniques to modify the affective and aesthetic qualities – the impressions – of AI aesthetics outside of the textual realm?

1.1 Aesthetic agency

One of the most essential questions in AI aesthetics revolves around the role played by aesthetic agency in creation, which makes it possible to unravel the collaborative aspects of aesthetics that lie behind the final artwork. Derived from programmer and AI artist Helena Sarin's thinking on artistic process with an AI,[13] aesthetic agency can be considered to arise through three different trajectories: *algorithmic complexity*, *data ownership*, and the *idea*. The first two relate strongly to AI aesthetics and are discussed below; the last influences the resulting aesthetics but in versatile and unexpected ways. As the *idea* is more connected to the creative use of an algorithm or dataset than to a general discussion of AI aesthetics, it will not be addressed further.

Algorithmic complexity means that the artist codes the algorithm from scratch, uses pre-trained models as they are, adjusts those models' parameters, or combines multiple (pre-trained) models. The current trend is the use of online platforms for creation.[14] The user sometimes only writes a text prompt, and an image is generated without any need to understand, or even the possibility to see, the background code. The main technical demand and tool for aesthetic influence in platforms that generate from text is knowledge-based: the user must have a historical understanding of the existing artistic and aesthetic styles and the linguistic abilities to describe the desired results. The aesthetic process with a pre-trained model can be considered a kind of search task in a closed system – the image is sculpted from the grey mass of existing aesthetic alternatives. Tweaking the parameters of an algorithm influences various aesthetic qualities of the outcome; for example, adjusting training steps can enhance the image quality and improve how closely it represents the object the user is looking for. Although the influences are aesthetic, the main goal is often to get the best possible outcome in quality, resolution, and representation.

[13] "Playing a Game of GANstruction; Eyeo 2019 – Helena Sarin," June 5, 2019, video, https://vimeo.com/354276365, acc. on May 16, 2022.

[14] Github (https://github.com) is a development platform and a cloud service that enables the user to maintain, develop, and share their code. Google Colab (https://colab.research.google.com) enables the user to execute and share python code in a browser. AI demos for almost any model can be found online; for example, Pollinations.ai (https://pollinations.ai) hosts trending AI models for art creation on their website, which has a simple interface.

The highest aesthetic autonomy is gained by creating and training models. By modifying how the model processes information, aesthetics can be addressed at a more fundamental level. In general, information processing is not recognized as an aesthetic task, although aiming to create a model that produces photorealistic results, for example, is certainly an aesthetic choice. Understanding aesthetics as an attentional mode relating to information broadens the possible roles that aesthetic theory might play in the design of perceptual models for machines. This aspect is discussed in more detail below.

The second approach to aesthetic agency in AI is *data ownership.*[15] AI artists can use pre-trained models, in which case they probably do not know the original images in the dataset, they can choose an open data library as a whole or from which to curate, they can scrape images from any source, or the artists can create their own dataset on which to train their model. As many models demand an excessive amount of data for training and significant processing capacity from the computer, the use of pre-trained models is currently the most popular option.

Curation of data should itself be considered an aesthetic act, as it determines the aesthetic and ethical latent space in which the model can travel and from which it can generate results. For many artists, discussion of the role of data connects their artistic practice to ethical questions; for example, artist Jake Elwes challenges the use of cis and heteronormative representations of humans in datasets with their artwork *Zizi – Queering the Dataset* from 2019, which used only images of drag queens as training data.[16] Similarly, artist and researcher Mimi Ọnụọha discusses data that has been excluded from society in her artwork *Library of Missing Datasets*.[17] Both Elwes' and Ọnụọha's artistic acts necessitate the creation of their own datasets, whereas artists such as Memo Akten and Sofia Crespo mainly use large existing datasets, such as satellite images from NASA[18] or a collection of images of endangered species.[19] Their aesthetic autonomy arises mainly from their creative ideas rather than from the creation of bespoke datasets.

Artistic freedom granted using tailored datasets and pre-trained models has also led to fundamental debates on the ownership of the artwork.[20] From an aesthetics perspective, all elements influencing the resulting aesthetic can be considered to

[15] "Playing a Game of GANstruction; Eyeo 2019 – Helena Sarin."

[16] Jake Elwes, *Zizi – Queering the Dataset*, 2019, 30-second extract of single channel, https://www.jakeelwes.com/ project-zizi-2019.html, acc. on April 19, 2022.

[17] Mimi Ọnụọha, *Library of Missing Datasets*, 2016, mixed-media installation, https://mimionuoha.com/the-library-of-missing-datasets, acc. on April 19, 2022.

[18] Memo Akten, *Learning to See*, 2017, video series, https://www.memo.tv/works/learning-to-see/, acc. on April 19, 2022.

[19] Sofia Crespo, *Critically Extant*, 2022, a collection of video works, https://criticallyextant.com/, acc. on April 19, 2022.

[20] See, for example, the case of Robbie Barrat and Obvious collective: in 2018, a painting created with Barrat's code and the same training material as the original model was sold in an art auction for a high price by Obvious collective, and the artwork itself was signed by the algorithm. Robbie Barrat (@videodrome), Twitter, October 25, 2018, https://twitter.com/videodrome/status/1055360024548012033.

define the ownership; the person who developed the dataset, the programmer responsible for the code, and the curator who decided on the outcome would thereby gain aesthetic ownership. As AI aesthetics most often arises through communal knowledge and effort, it should always be discussed as a process, not only as a final object.

Therefore, I argue that aesthetics should be framed more broadly in the discussion of AI to make it an applicable and interdisciplinary concept. To provide a more static definition for aesthetics, I propose three conditions for the evolution of aesthetic machine attention: 1) acknowledging that something appears (aesthetic detection); 2) suspending judgments (aesthetic recognition); and 3) making the incident explicit with expression (aesthetic identification and amplification). In the next chapter, I frame the issue through the interdisciplinary theory of attention from machine learning, philosophy, and psychology.

2. Aesthetic Machine Attention

This chapter approaches the question of aesthetic machine attention by considering attention as a mode of thinking,[21] and it concentrates on how different modes of attention can affect perceptual content differently, an idea first presented by philosopher Bence Nanay.[22] To bring these philosophical stances closer to machine learning and psychology, attention can be further specified as a mode of knowledge production; with different modes of attention, sensory stimuli are processed differently, and information can emerge in different forms. Aesthetic attention would, according to this definition, necessitate a certain kind of mode for knowledge production, a hypothesis that is further clarified by theory from psychology.

Researcher Marisa Carrasco categorizes theories of attention through two metaphors, *filtering* and *amplifying*, emphasizing that information is not only filtered out from the sensory signals but that sensory signals can also be enhanced in the early stages of sensory processing[23] and that attention can concretely alter perceived appearance.[24] These alterations, although they often go unnoticed, drastically influence

[21] Philosopher John Locke describes attention as a selective mode of perception; that is, one that is mediated by ideas. According to Locke, we attend to our ideas about things. Locke argues that we are not able to recognize complex ideas, but complexity gets constructed from a combination of simpler ideas. This represents an internalist view. My view is a combination of internalist and externalist positions. I believe that features can directly activate our exogenous attention (externalist view); however, these features are aesthetically modulated through our affective and homeostatic states and preconceptions even before our conscious awareness takes place (a variation of an internalist view). As a result, the subjective features can give rise to attention in an involuntary manner. Therefore, the aesthetic mode of attention would mean attending to the subjective and altered nature of the experience, either selectively or automatically. John Locke, *An Essay Concerning Human Understanding* (London: Penguin Books, 1997), 214; See also Matthew Stuart, "Locke on Attention," *British Journal for the History of Philosophy* 25, 3 (2017): 487–505.

[22] Bence Nanay, "Attention and Perceptual Content," *Analysis* 70, 2 (2009): 263–70.

[23] Marisa Carrasco, "Visual Attention: The Past 25 Years," *Vision Research* 51, 13 (2011): 1484–525. Simone Schnall, "Embodiment in Affective Space: Social Influences on Spatial Perception," in *Spatial Dimensions of Social Thought*, ed. by A. Maas and T. Schubert (Berlin: De Gruyter Mouton, 2011), 129–52.

[24] Marisa Carrasco et al., "Attention Alters Appearance," *Nature Neuroscience* 7 (2004): 308–13.

interpretations and impressions of a scene.[25] For example, if a person believes there exists a social division between "us" and "them," they overestimate their perceptual distance to "them"[26]. I believe this altered perceptual distance can be considered a starting point for a determination of the aesthetics of that situation.

The amplifier aspect of attention is an essential part of perceptual processing in general;[27] not only a property of artistic and aesthetic experiences, attention-based amplification carries radical potentiality for aesthetics as a discipline. It can resolve what aesthetics means from a perceptual point and specify the role of aesthetics among other disciplines that study perception. For aesthetic attention to arise from the amplification effect, attention must be directed toward the subjective nature of percepts; the emerging aesthetics of the situation, according to studies, seems to originate from affective and homeostatic states.[28]

In machine learning, different attentional modes are seldomly addressed, but the chosen data determine what kind of information emerges. For example, in models using textually annotated images,[29] the mode of attention could be considered ob-

[25] See Chaz Firestone and Brian Scholl, "Cognition Does Not Affect Perception: Evaluating the Evidence for 'Top-Down' Effects," *Behavioral and Brain Sciences* 39 (2016) and its broad commentary for a review.

[26] Simone Schnall, "Embodiment in Affective Space: Social Influences on Spatial Perception," in *Spatial Dimensions of Social Thought*, ed. A. Maas and T. Schubert (Berlin: De Gruyter Mouton, 2011), 129–52.

[27] Attention can directly boost impressions, such as apparent contrast, motion, spatial resolution, and size. Katharina Anton-Erxleben et al., "Attention Changes Perceived Size of Moving Visual Patterns," *Journal of Vision* 7, 5 (2007): 1–9; Marisa Carrasco, "Visual Attention: The Past 25 Years," *Vision Research* 51, no. 13 (2011). It alters how something appears; for example, an object can look more saturated, bigger, and faster if suddenly attended. Marisa Carrasco and Antoine Barbot, "Spatial Attention Alters Visual Appearance," *Current Opinion in Psychology* 29 (2019); Marisa Carrasco, "Cross-Modal Attention Enhances Perceived Contrast," *Proceedings of the National Academy of Sciences of the United States of America* 106, 52 (2009); Jared Abrams, "Voluntary Attention Increases Perceived Spatial Frequency," *Attention, Perception & Psychophysics* 72, 6 (2010): 1510–21.

[28] Aesthetic modulations can happen through affective states; for example, depressive disorder weakens contrast sensitivity. Emanuel Bubl et al., "Vision in Depressive Disorder," *The World Journal of Biological Psychiatry* 10 (2009); Marisa Carrasco and Antoine Barbot, "Spatial Attention Alters Visual Appearance," *Current Opinion in Psychology* 29 (2019). Positive mood directs the gaze more towards the periphery, broadening the perceptual receptive field, which leads to an even wider action repertoire. Kai Kaspar and Peter König, "Emotions and Personality Traits as High-Level Factors in Visual Attention: A Review," *Frontiers in Human Neuroscience* 6 (2012). It also likely creates an impression that the scene is more spacious. Interestingly, imaginary brightness influences physical reactions and pupil size similarly to real brightness. Matthias Hartmann and Martin Fischer, "Pupillometry: The Eyes Shed Fresh Light on the Mind," *Current Biology* 24, 7 (2014). In fact, even bright thoughts make pupils larger. Weizhen Xie and Weiwei Zhang, "The El Greco Fallacy and Pupillometry: Pupillary Evidence for Top-Down Effects on Perception," *Behavioral and Brain Sciences* 39 (2016). They can amplify the apparent brightness of a scene, and some studies describe how aesthetic features influence the body: When a person views their own body through a minifying lens, their subjective experience of pain is weakened. Charles Spence, "Multisensory Perception," *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* 2 (2018): 1–56. Additionally, when a person views 'hot' images, for example of a desert, their core body temperature adaptively decreases. Jun´ya Takakura et al., "Nonthermal Sensory Input and Altered Human Thermoregulation: Effects of Visual Information Depicting Hot or Cold Environments," *International Journal of Biometeorology* 59, 10 (2015). These examples show an important link between aesthetics and subjective (bodily) states, which seem to function both from features into the creation of affective states and from different states to the alteration of perceptual features.

[29] See, for example, Xiaodong He, "Deep Attention Mechanism for Multimodal Intelligence: Perception, Reasoning, & Expression," March 12, 2018, https://www.youtube.com/watch?v=YYKpS-Y75LY, acc. on April 19, 2022.

ject-based. According to psychology, object-based attention results from conceptual learning, and two terms, language-based and category-based, are used as synonyms for object-based attention.[30] According to Nanay, aesthetic attention is signified by its ability to simultaneously focus on a single object and be distributed across its properties.[31] I partially disagree with Nanay on this postulation. I consider affective states to be essential motivation for perceptually altered aesthetic reality and argue that aesthetics emerge through or with featural information; therefore, the role of cognitive object formation in information processing becomes less important for the establishment of aesthetic attention. I do not assume that all aesthetic phenomena stay in featural form, but I argue that it should be set as the starting point for aesthetic machine attention.

In psychology, this mode of attending is called feature-based attention. It can be directed to features such as colors, shapes, directions of motions, and particular orientations, and it can spread globally across the receptive field to reach features, regardless of their location.[32] When a feature is given attention, perception becomes sensitized to similar properties and is tuned towards the attended feature,[33] amplifying its aesthetic presence in the scene. Therefore, aesthetic machine attention should not concentrate on a defined object, for example a *tree*, and then conclude that the color of a tree is tinted with warm light; instead, it should notice how warm light paints everything in the scene. As a result, the affective and attended quale itself becomes aesthetically salient and is amplified, whereas object-based knowledge is attenuated in the process.

2.1 Acknowledging that something appears (aesthetic detection)

To acknowledge that something appears, machines must possess prior knowledge about what *could* appear. Object detection in machine learning includes two steps, localization and recognition. It answers a computer vision problem: *What objects are where?*[34] For aesthetic detection, this question seems less relevant. As discussed already, aesthetic attention should not initially be stated as an object detection task, as aesthetic attention is a feature-based mode of attention in which awareness can spread across the scene. To formulate a computable problem for aesthetic machine attention, perhaps a more relevant question would be *What features are present, and how do they appear?* To explore this cryptic question further, the assigned task

[30] Gary Lupyan and Emily Ward, "Language Can Boost Otherwise Unseen Objects into Visual Awareness," *Proceedings of the National Academy of Sciences* 110, 35 (2013); Timo Stein and Marius Peelen, "Object Detection in Natural Scenes: Independent Effects of Spatial and Category-Based Attention," *Attention, Perception, & Psychophysics* 79, 3 (2017): 738–52.

[31] Bence Nanay, "Aesthetic Attention," *Journal of Consciousness Studies* 22, 5–6 (2015): 96–118.

[32] Marisa Carrasco, "Visual Attention: The Past 25 Years," *Vision Research* 51, 13 (2011): 1484–525.

[33] Ibid.; Veldri Kurniawan, "The Neural Basis of Multisensory Spatial and Feature-Based Attention in Vision and Somatosensation" (PhD thesis, School of Psychology, Cardiff University, 2012).

[34] Zhengxia Zou et al., "Object Detection in 20 Years: A Survey," arXiv Preprint, submitted 2019, *arXiv:1905.05055*.

should be the study of the percept and its versatile features (appearances) to understand how it resonates within the system.

To resonate in a system, in this case, means that it catalyzes reactions in other modalities; for example, visual features could lead to bodily reactions.[35] To quote philosopher Dieter Mersch, aesthetic research "seeks out the unexpected or the strange, and rather than hope for progress in knowledge, an increase of objectivity, and stable models, it induces the oscillation of phenomena and instigates a moment of transformation, a *conversio* or *inversion* in observers"[36]. This is how I believe aesthetic attention resembles aesthetic research: both are investigatory. Also, Mersch highlights oscillation rather than stability, and aesthetic detection can be considered a dynamic phenomenon that evolves and can be attuned to over a certain duration of time.[37] In machine learning, detection is determined by the training data. For object detection, the training data includes those instances that are relevant for the task. In surveillance, for example, faces often function as the main object, which can then be linked with real identities through the training data. For aesthetic attention, the underlying logic for detection could be multimodal links in nonverbal data – features that share the same affective foundation.[38]

2.2 Suspension of judgment (aesthetic recognition)

Aesthetic recognition emerges from an act during which instances are investigated, not judged. To elaborate upon what this means concerning perceptual tasks, I borrow researcher Don Ihde's phenomenological method and his three hermeneutic rules: 1) attend to all the phenomena as and how they show themselves; 2) describe, don't explain; and 3) horizontalize all phenomena.[39] This approach would revolutionize the premises of machine vision, which is currently heavily based on

[35] According to neurological studies, aesthetic perception is linked to activation in the motor cortex, and interpretation of movement in an image participates in its aesthetic comprehension (in the study, the researchers used human and nature content paintings). Cinzia Di Dio et al., "Human, Nature, Dynamism: The Effects of Content and Movement Perception on Brain Activations During the Aesthetic Judgment of Representational Paintings," *Frontiers in Human Neuroscience* 9, 705 (2016).

[36] Dieter Mersch, *Epistemologies of Aesthetics* (Zurich: Diaphanes, 2015).

[37] Expert viewers of art that can be assumed to use aesthetic visual strategies gaze at images globally, with a wider scope in eye movements towards the periphery, whereas novice viewers focus on the most semantically salient aspects with fewer eye fixations. Stine Vogt and Svein Magnussen, "Expertise in Pictorial Perception: Eye-Movement Patterns and Visual Memory in Artists and Laymen," *Perception* 36, 1 (2007): 91–100. This might explain why aesthetic attention requires a temporality element to be included if it is to be computed. In neural attention models and saliency models, the generated attentional heat maps/saliency maps do not show attention gradually shifting between instances in time but give a static representation of gaze-attracting objects. Scanpath modeling records eye movements in time and tracks their paths and directions.

[38] Crossmodal correspondences and multimodality are widely studied topics in perceptual psychology. According to studies, multimodality is an innate state and is hypothesized to be based on an affective logic. See Spence, "Multisensory Perception,"; also Charles Spence, "Crossmodal Correspondences: A Tutorial Review," *Attention, Perception & Psychophysics*, 73, 4 (2011): 971–95, for a review and Kelly Whiteford et al., "Color, Music, and Emotion: Bach to the Blues," *I-Perception* 9, 6 (2018): 2041669518808535 on affectivity.

[39] Don Ihde, *Experimental Phenomenology: Multistabilities* (New York: Suny Press, 2012).

classification.[40] Ihde's method describes how a phenomenon should be approached when it appears—without assigning assumptions to it or subjecting it to any theory, idea, concept, or construction that attempts to go beyond the phenomenon and without arranging the observations into any hierarchical order in relation to each other. In short, beliefs should be suspended to allow the full range of appearances to show themselves.[41] Although Ihde argues against hierarchical order, the hierarchy of information can serve a purpose for machine learning. In hierarchical machine learning models, information is considered to emerge differently at low and high levels of the hierarchy; information in the low-level perception is feature-based and cumulates towards semantic objects at the high end of the continuum.[42] Suspension of judgment concerning hierarchy would mean that attention would be sustained at the level of information where it has not yet taken the form of a learned semantic concept or an object category. Holding attention at the level of affects would mean inspecting the textures, colors, motions, and other aesthetic features that resonate across the modalities. In other words, aesthetic recognition could involve acknowledgement of the affective qualities of feature-based information.

2. 3. Making the incident explicit with expression (aesthetic identification and amplification)

How, then, could attention emerge from the process described above? In machine learning, the act of attending means to *express* what draws attention. In neural attention models, attention is expressed via heatmaps showing the hottest spots of attention,[43] and in saliency models, the saliency maps show the locations that people find the most important and, in free-viewing situations, toward which they will direct their attention.[44] Interestingly, even just the movement of the eyes can be understood as expressive.[45] In the field of machine attention, the eyes express the attentional locations that draw the viewer's gaze, but they also express the mode of attention that will be established. In models trained with the human gaze, the 'freeness' of the free-viewing protocol most often means using object-based visual strategies, as human attention

[40] Kate Crawford, "NIPS 2017 Keynote Lecture: Trouble with Bias," December 10, 2017, https://www.youtube.com/watch?v=fMym_BKWQzk, acc. on April 19, 2022.

[41] Don Ihde, *Experimental Phenomenology: Multistabilities* (New York: Suny Press, 2012).

[42] Honglak Lee et al., "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations," *Proceedings of the 26th Annual International Conference on Machine Learning,* (2009); Sarthak Mittal et al., "Learning to Combine Top-Down and Bottom-Up Signals in Recurrent Neural Networks with Attention Over Modules," *International Conference on Machine Learning*, 2020, 6972–86.

[43] See, for example, Teng Wang et al., "Image Caption with Endogenous-Exogenous Attention," *Neural Processing Letters* 50, 1 (2019): 431–43, for visualizations of attentional heatmaps.

[44] See Ming Jiang et al., "Salicon: Saliency in Context," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, 1072–80, as an interesting exception to typical saliency models that are trained with eye-movement data. In Salicon, digital pen movements investigating an image are used as training data.

[45] Studies by Yarbus showed back in the 1960s how the performance of eye movements changes according to a given task. The expressions of the eyes can be studied to understand what kind of cognitive premises they exhibit. Alfred Yarbus, *Eye Movements and Vision* (New York: Plenum, 1967).

is commonly focused on semantic objects in images;[46] in attention models that are trained with linguistic annotations, the heatmaps again express object-based understanding.

How about attention models that are trained using, for example, a combination of visual and audio streams for machine attention? Could they be considered aesthetic models? A model from Min et al. uses multimodal information, but the machine vision task is again to detect objects in a scene.[47] Similarly, a model from Zhang et al.[48] using audio and video for emotion recognition disappointingly starts its recognition task from seven verbal categories of emotions and then uses them to classify emotions from the combination of visual and audio signals. Currently, it seems impossible to escape the use of verbal categories in machine learning identification tasks.

For aesthetic machine attention, one must ask which other modalities, other than language, are available for expression. The example presented at the beginning of this chapter was gaze; however, when gathering training data with gaze detection, the situation should be carefully designed. One possible method would be to let experienced art observers train the model to ensure that the established machine attention is feature-based and global.[49] Another would be to set out clear instructions regarding how the participant might notice the affective nature of the stimuli. However, as these methods might not lead to easily generalizable results, one option could be to use aesthetic stimuli that are abstract and avoid semantically recognizable objects, for example, images including low- and mid-level features such as shapes and patterns of different colors. The lack of any attention-drawing semantic objects could naturally lead to an aesthetic mode of attention. The use of other modalities to support the expression of attention, for example singing the attended affective qualities, could help to make the aesthetic salience of an incident explicit. Also, training the model with pen expressions could not only locate attentional hotspots similarly to gaze[50] but could also describe the quality of attention; whether the drawn gestures are gentle or rough and fast or slow in relation to the stimuli should be considered informative regarding the quality of the features.

This is how artistic expression inherently links to aesthetic attention. Artistic practices make it possible to express the subjective nature of attention without subsuming the phenomenon under categorical and conceptual thinking. If the attended features were amplified following the amplifier theory of attention, artistic expression would be given a new role in relation to aesthetic attention. With generative artificial

---

[46] Matthias Kümmerer and Matthias Bethge, "State-of-the-Art in Human Scanpath Prediction," arXiv Preprint, submitted 2021, *arXiv:2102.12239*.

[47] Xiongkuo Min et al., "A Multimodal Saliency Model for Videos with High Audio-Visual Correspondence," *IEEE Transactions on Image Processing* 29 (2020): 3805–19.

[48] Yuanyuan Zhang et al., "Deep Fusion: An Attention Guided Factorized Bilinear Pooling for Audio-Video Emotion Recognition," *2019 International Joint Conference on Neural Networks*, 2019, 1–8.

[49] See Vogt and Magnussen, "Expertise in Pictorial Perception."

[50] See Jiang et al., "Salicon: Saliency in Context," for more on how pen movements can replace gaze detection in the training of a saliency model.

aesthetics, the attended aesthetic features and their subjective qualia are translated again, or for the first time of this described perceptual process, into an explicit form that aesthetically shows how affective states influence perception.

Conclusion

This paper described how, in current discussions of AI aesthetics, aesthetics is often addressed from the point of view of created artworks. The term aesthetic agency was coined to locate the possibilities for aesthetic influence that an artist can have when creating with an AI and to argue that AI aesthetics is most often a collective effort. The paper proposed an alternative approach to aesthetics derived from psychology which considers aesthetics emerging as a result of a certain kind of attentional process. Three stages of aesthetic machine attention were developed through study of the machine learning literature enriched with philosophical and psychological theory: 1) acknowledging that something appears (aesthetic detection); 2) suspension of judgment (aesthetic recognition); and 3) making the incident explicit with expression (aesthetic identification and amplification). These steps offer clarity on aesthetic attention in order to make it computable. First, a computable question should be defined; in an aesthetic case it would be: What features are present and how do they appear? Second, to recognize these features, they should be attended with a mode of attention that avoids making judgements or forcing the phenomena into any categories. This requires holding the attention at a level at which the affective quality of feature-based information is acknowledged – before any further cognitive processing of the incident takes place. For machine attention, this means avoiding labeling with linguistic categories in a detection task and attending to feature-based information instead. Third and as a result, to avoid linguistic categorizations, expressions with other modalities could be used to locate aesthetically salient features for aesthetic machine attention. This process gives rise to aesthetic knowledge that is feature-based, multimodal, and global – a useful addition to current computer vision models that are focused on objects or linguistic categories. Through aesthetic machine attention and AI art, the subjective quality of experience could be made explicit.

References

Abrams, Jared, Antoine Barbot, and Marisa Carrasco. "Voluntary Attention Increases Perceived Spatial Frequency." *Attention, Perception & Psychophysics* 72, 6 (2010): 1510–21. doi:10.3758/APP.72.6.1510.

AIartist. "Alexander Mordvintsev" (web page). https://aiartists.org/alexander-mordvintsev. Accessed on April 19, 2022.

Akten, Memo. *Learning to See*. 2017. Video series. https://www.memo.tv/works/learning-to-see/. Accessed on April 19, 2022.

alembics. "Disco-diffusion" (Github repository). https://github.com/alembics/ disco-diffusion. Accessed on April 19, 2022.

Anton-Erxleben, Katharina, Christian Henrich, and Stefan Treue. "Attention Changes Perceived Size of Moving Visual Patterns." *Journal of Vision* 7, 11 (2007): 5.

Barrat, Robbie (@videodrome). "left: the 'AI generated' portrait Christie's is auctioning off right now right: outputs from a neural network I trained and put online *over a year ago*. Does anyone else care about this? Am I crazy for thinking that they really just used my network and are selling the results?" Twitter, October 25, 2018. https://twitter.com/videodrome/status/1055360024548012033. Accessed on April 19, 2022.

Bubl, Emanuel, Ludger Tebartz Van Elst, Matthias Gondan, Dieter Ebert, and Mark W. Greenlee. "Vision in Depressive Disorder." *The World Journal of Biological Psychiatry* 10, (2009): 377–84.

Carrasco, Marisa. "Cross-Modal Attention Enhances Perceived Contrast." *Proceedings of the National Academy of Sciences of the United States of America* 106, 52 (2009): 22039–40. doi:10.1073/pnas.0913322107

Carrasco, Marisa. "Visual Attention: The Past 25 Years." *Vision Research* 51, 13 (2011): 1484–525. doi:10.1016/j.visres.2011.04.012

Carrasco, Marisa and Antoine Barbot. "Spatial Attention Alters Visual Appearance." *Current Opinion in Psychology* 29 (2019): 56–64.

Carrasco, Marisa, Sam Ling, and Sarah Read. "Attention Alters Appearance." *Nature Neuroscience* 7 (2004): 308–13.

Cheng, Keyang, Rabia Tahir, Lubamba Kasangu Eric, and Maozhen Li. "An Analysis of Generative Adversarial Networks and Variants for Image Synthesis on MNIST Dataset." *Multimedia Tools and Applications* 79, 19 (2020): 13725–52.

Crawford, Kate. "NIPS 2017 Keynote Lecture: Trouble with Bias." December 10, 2017. YouTube video, 49:31. https://www.youtube.com/watch?v=fMym_BKWQzk. Accessed on April 19, 2022.

Crespo, Sofia. *Critically Extant*. 2022. A collection of video works. https://criticallyextant.com/. Accessed on April 19, 2022.

Dhariwal, Prafulla, and Alexander Nichol. "Diffusion Models Beat GANs on Image Synthesis." *Advances in Neural Information Processing Systems* 34 (2021): 8780–94.

Di Dio, Cinzia, Martina Ardizzi, Davide Massaro, Giuseppe Di Cesare, Gabriella Gilli, Antonella Marchetti, and Vittorio Gallese. "Human, Nature, Dynamism: The Effects of Content and Movement Perception on Brain Activations During the Aesthetic Judgment of Representational Paintings." *Frontiers in Human Neuroscience* 9, 705 (2016). doi: 10.3389/fnhum.2015.00705

Elwes, Jake. *Zizi — Queering the Dataset*. 2019. 30-second extract of single channel. https://www.jakeelwes.com/project-zizi-2019.html. Accessed on April 19, 2022.

Firestone, Chaz, and Brian J Scholl. "Cognition Does Not Affect Perception: Evaluating the Evidence for 'Top-Down' Effects." *Behavioral and Brain Sciences* 39 (2016). doi: 10.1017/S0140525X15000965

Guo, Hui, Shu Hu, Xin Wang, Ming-Ching Chang, and Siwei Lyu. "Eyes Tell All: Irregular Pupil Shapes Reveal GAN-Generated Faces." In *Proceedings ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, 2904—8.

Hammerman, Robin, and Andrew L. Russell. *Ada's Legacy: Cultures of Computing from the Victorian to the Digital Age*. London: Morgan & Claypool, 2015.

Hartmann, Matthias, and Martin H. Fischer. "Pupillometry: The Eyes Shed Fresh Light on the Mind." *Current Biology* 24, 7 (2014): R281—R282.

He, Xiaodong. "Deep Attention Mechanism for Multimodal Intelligence: Perception, Reasoning, & Expression." March 12, 2018. YouTube video, 2:12. https://www.youtube.com/watch?v= YYKpS-Y75LY. Accessed on April 19, 2022.

Herndon, Holly, and Mathew Dryhurst. "Infinite Images and the Latent camera" (web page). https://mirror.xyz/herndondryhurst.eth/eZG6mucl9fqU897XvJs0vUUMnm5OITpSWN8S-6KWamY. Accessed on May 9, 2022.

Ihde, Don. *Experimental Phenomenology: Multistabilities* (2nd ed.). Albany, NY: State University of New York Press, 2012.

Jiang, Ming, Shengsheng Huang, Juanyong Duan, and Qi Zhao. "Salicon: Saliency in Context." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, 1072—80.

Kaspar, Kai, and Peter König. "Emotions and Personality Traits as High-Level Factors in Visual Attention: A Review." *Frontiers in Human Neuroscience* 6, 321 (2012). doi: 10.3389/fnhum.2012.0032

Kümmerer, Matthias, and Matthias Bethge. "State-of-the-Art in Human Scanpath Prediction." Preprint, submitted 2021. *arXiv:2102.12239*.

Kurniawan, Veldri. "The Neural Basis of Multisensory Spatial and Feature-Based Attention in Vision and Somatosensation." PhD diss., School of Psychology, Cardiff University, 2012.

Lee, Honglak, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng. "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations." In *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, 609—616.

Locke, John. *An Essay Concerning Human Understanding*. London: Penguin Books, 1997. First published 1690.

Lupyan, Gary, and Emily J. Ward. "Language Can Boost Otherwise Unseen Objects into Visual Awareness." *Proceedings of the National Academy of Sciences* 110, 35 (2013): 14196—201.

Menabrea, Luigi F. "Notions sur la Machine Analytique de M. Charles Babbage." *Bibliothèque Universelle de Genève* 41 (1842): 352–76. First Translated by Augusta Ada Lovelace, *Scientific Memoirs* 3 (1843): 666–731).

Mersch, Dieter. *Epistemologies of Aesthetics*. Zurich: Diaphanes, 2015.

Min, Xiongkuo, Guangtao Zhai, Jiantao Zhou, Xiao-Ping Zhang, Xiaokang Yang, and Xinping Guan. "A Multimodal Saliency Model for Videos with High Audio-Visual Correspondence." *IEEE Transactions on Image Processing* 29 (2020): 3805–19.

Mittal, Sarthak, Alex Lamb, Anirudh Goyal, Vikram Voleti, Murray Shanahan, Giullaume Lajoie, Michael Mozer, and Yoshua Bengio. "Learning to Combine Top-Down and Bottom-Up Signals in Recurrent Neural Networks with Attention Over Modules." *International Conference on Machine Learning* 2020, 6972–86.

Nanay, Bence. "Attention and Perceptual Content." *Analysis* 70, 2 (2009): 263–70.

Nanay, Bence. "Aesthetic Attention." *Journal of Consciousness Studies* 22, 5–6 (2015): 96–118.

Nightcafe. "VQGAN+CLIP Text to Art Generator" (web page). https://creator. nightcafe.studio/text-to-image-art. Accessed on May 9, 2022.

Ọnụọha, Mimi. *Library of Missing Datasets*. 2016. Mixed-media installation. https://mimionuoha.com/the-library-of-missing-datasets. Accessed on April 19, 2022.

OpenAI. "DALL·E 2" (web page). https://openai.com/dall-e-2/. Accessed on May 9, 2022.

"Playing a Game of GANstruction; Eyeo 2019 – Helena Sarin," June 5, 2019, video. https://vimeo.com/354276365. Accessed on May 16, 2022.

Schnall, Simone. "Embodiment in Affective Space: Social Influences on Spatial Perception." In *Spatial Dimensions of Social Thought*, edited by A. Maas and T. Schubert, 129–52. Berlin: De Gruyter Mouton, 2011.

Spence, Charles. "Crossmodal Correspondences: A Tutorial Review." *Attention, Perception & Psychophysics* 73, 4 (2011): 971–95. doi: 10.3758/s13414-010-0073-7.

Spence, Charles. "Multisensory Perception." *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* 2 (2018): 1–56.

Stein, Timo and Marius V. Peelen. "Object Detection in Natural Scenes: Independent Effects of Spatial and Category-Based Attention." *Attention, Perception, & Psychophysics* 79, 3 (2017): 738–52.

Takakura, Jun'ya, Takayuki Nishimura, Damee Choi, Yuka Egashira, and Shigeki Watanuki. "Nonthermal Sensory Input and Altered Human Thermoregulation: Effects of Visual Information Depicting Hot or Cold Environments." *International Journal of Biometeorology* 59, 10 (2015): 1453–60.

Vogt, Stine and Svein Magnussen. "Expertise in Pictorial Perception: Eye-Movement Patterns and Visual Memory in Artists and Laymen." *Perception* 36, 1 (2007): 91–100.

Wang, Teng, Haifeng Hu, and Chen He. "Image Caption with Endogenous-Exogenous Attention." *Neural Processing Letters* 50, 1 (2019): 431–43.

Whiteford, Kelly L., Karen B. Schloss, Nathaniel E. Helwig, and Stephen E. Palmer. "Color, Music, and Emotion: Bach to the Blues." *I-Perception* 9, 6 (2018): 2041669518808535.

Xie, Weizhen and Weiwei Zhang. "The El Greco Fallacy and Pupillometry: Pupillary Evidence for Top-Down Effects on Perception." *Behavioral and Brain Sciences*, 39 (2016).

Yang, Tianyun, Juan Cao, Qiang Sheng, Lei Li, Jiagi Ji, Xirong Li, and Sheng Tang. "Learning to Disentangle GAN Fingerprint for Fake Image Attribution." Preprint, submitted 2021. /doi: 10.48550/arXiv.2106.08749

Yarbus, Alfred. L. *Eye Movements and Vision*. New York: Plenum, 1967.

Yu, Ning, Larry S. Davis, and Mario Fritz. "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints." In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (7556–66). 2019.

Zeiler, Matthew D., and Rob Fergus. "Visualizing and Understanding Convolutional Networks." In *European Conference on Computer Vision*, 2013, 818–33.

Zhang, Yuanyuan, Zi-Rui Wang, and Jun Du. "Deep Fusion: An Attention Guided Factorized Bilinear Pooling for Audio-Video Emotion Recognition." In *2019 International Joint Conference on Neural Networks, IJCNN*, 2019, 1–8.

Zou, Zhengxia, Zhenwei Shi, Yuhong Guo, and Jieping Ye. "Object Detection in 20 Years: A Survey," 2019. Preprint, submitted 2021. *arXiv:1905.05055*.

Article received: May 10, 2022
Article accepted: July 15, 2022
Original scholarly paper