Yiu, Sheung

One face, millions of faces: Computer vision as hyperobject

Published in:
PHILOSOPHY OF PHOTOGRAPHY

DOI:
10.1386/pop_00048_1

Published: 24/01/2023

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Yiu, S. (2023). One face, millions of faces: Computer vision as hyperobject. *PHILOSOPHY OF PHOTOGRAPHY*, *12*(1-2), 71-91. https://doi.org/10.1386/pop_00048_1

One Face, Millions of Faces: Computer Vision as Hyperobject

Sheung Yiu, Department of Media, Aalto University

Abstract

Borrowing Timothy Morton's notion of hyperobject, this essay explores questions of network and scale in GAN-generated images. In this context, the term network refers to the omnipresence of algorithmic images today and their significant impact on our lives. Such images are massively distributed in time and space beyond any sensible human-scale. Scale, in this context, denotes the relations between different operational layers of algorithmic images, such as the pictorial layer in contrast to the data layer. An algorithmic image is simultaneously a visual image, a symbol, a data point and part of a mass visual milieu. Its meaning is thus polymorhic and can, arguably, never be exhausted. The article explores these terms through analysis of the website *Thispersondoesnotexist.com*.

Thispersondoesnotexist.com is a website containing millions of faces. Created by software engineer Phillip Wang, the website uses a subset of machine learning frameworks called StyleGAN (generative adversarial network) to generate a new image of a human faces with every refresh of the browser. However, each of these photorealistic faces is in fact machine-generated. The person depicted does not exist. The digital image on the screen is an array of pixels, their RGB values calculated by a neural network trained on thousands of real-life examples to generate a pattern that human beings will recognize as faces. The real-life training examples used are pulled together from various available online sources, from photo-sharing platforms from Flickr to the archive of the Metropolitan Museum. The pixel pattern is the result of GAN, a framework consisting of two competing neural networks, a Generator, and a Discriminator, in a reiterative training process that proceeds until the Generator produces what is considered a satisfactory result. All of these are invisible on the web interface. Photography scholars and critical theorists have tended to address new imagining technologies such as this in terms of their invisible elements. For instance, Matteo Pasquinelli has scrutinised the mathematical operations behind AI while highlighting the massive amount of invisible labour entailed in compiling training data. (Pasquinelli, 2020) Wendy Chun has exposed the invisible gendered labour in software development from the perspective of software studies, and arguing that software is as obfuscatory as it is revelatory in visualising certain programming operations while hiding others. (Chun, 2005) Kate Crawford and Trevor Paglen have explored hidden layers of technology in their examination of AI training datasets, revealing the human biases rendered invisible in the black box of facial recognition and computer vision. (Crawford and Paglen, 2019) In his *A Study of Invisible Image*, Paglen also described the images instrumental to the functioning of computer vision as 'invisible images' —images made by machines for other machines, imperceptible to and unintended for the human eye. Specifically for Paglen, this invisibility exposes the inadequacy of traditional photography theory in addressing the rapid technological development of image-making.[1] Examining technology through the framework of the surface and hidden layers, these accounts open up a discourse of image-making and technology beyond the representational, exposing complexities that might otherwise be overlooked.

Critical inquiry into invisible images is also usefully thought of in terms of scale. Scale has recently been described as a fundamental quality of images in the age of computer vision and network culture. (Fisher 2012, p.322) Issues of scale do not only entail questions of visual rescaling, but also entirely new scalar operations. An image functions at a human scale as a visual image, working on the phenomenological and semantic levels. On a pixel scale, an image is an array of values and a repository of data. On a network scale, images connected through metadata and hashtags create the context in which they define and create

[1] "The traditional discourses that we have to think about photographs seem useless today,"

meanings mutually. Computer vision traverses different scales through statistical operations and crowd-sourced labour to invent new operations such as image generation. In these contexts, scale also has implications for knowledge production. Outlining the genealogy of scale in his transdisciplinary framework for media theory, Zachary Horton explains that the physical notion of scale incorporates a notion of field of view. (Horton, 2020, 44) Disciplines divide the world into scales to study a phenomenon. The knowledge produced by each discipline is, therefore, specific to the scale of observation. A quantum physicist observes the universe on the nanoscale and devises explanations in terms of the interactions of subatomic particles; An anthropologist may study the world on the human scale, thus understanding causality through interpersonal relationships. As images expand beyond the moment of capture, are processed en-masse in greater-than-human scale and created through data infrastructure and social networks distributed across time and space, they exceed the disciplinary apparatus that may have previously framed them on the phenomenological scale of the human.

Furthermore, new media articulate new scales. (Horton, 2020, p.28) The advent of computer vision produces a new set of operations on images at both larger, and smaller than human scales. GAN-generated faces exemplify the multifarious nature of images. Depending on the scalar frame, an image is both visual and non-visual, experiential and programmable, an object on its own or a part of an extensive network. Locking eyes with the machine-generated faces in which the human is further obfuscated from its constitutive aspects (Zylinska, 2017, p.20), one is reminded that a picture is a superficial visual layer sustained by a deeper network, be it social, technical, or linguistic. Digital images embedded in a planet-wide computational infrastructure are distributed across networks and beyond the human scale. Computer vision articulates a new scale that introduces new modes of image-making, new forms of knowledge production as well as new constraints. Larger-than-human scales do not always guarantee more clarity. Each scale is an abstract space and time in which certain sets of relations take place while others are obscured.[2] But the larger than human, here, does suggest new posthuman or non-anthropocentric approaches to image studies, that focus on these questions of network and scale using the tools presented by recent speculative ontologies.[3]

The Network

An algorithmic image, such as the GAN-generated faces on *thispersondoesnotexist.-com*, is larger than its surface. The visual image is a human-machine interface of a larger network of things invisible to the observer (Chun, 2005, 26-51). Already in the 2000s, theorists such as Daniel Rubinstein and Katrina Sluis, in their discussions of metadata, stock images and algorithmic photography (Rubinstein, Sluis, 2008, 9-28), foreshadowed a 'paradigm shift' in photography from the single image to the network. This shift is characterised by understanding images in the digital informational economy as larger than their pictorial surfaces and always only understood in relation to the sea of other images. In this vein, comparing stock photography to the singular advertising image, media scholar Paul Frosh wrote: '[a stock image] erases indexical singularity, the uniqueness of the instance, in favour of uniformity and recurrence – the systematic iconic repetition of staged image types'. (Frosh, 2002, p. 171-96)

---

[2] Treating a digital image of a face as an array of numbers allows computer scientists to apply statistical analysis to apprehend the relations between data points for pattern recognition. However, observing an image at the scale of the pixel conceals the cultural history and media standards underpinning the creation of the same image.

[3] Especially those that highlight a network view in a larger-than-human scale, such as Actor-Network Theory's affirmation of agency in non-human actants, New Materialism's emphasis on emergent material configurations, Agential Realism's account of unfolding reality and resolving cuts, Object-Oriented Ontology's proposition of a flat ontology which puts human among non-human agents in the network of the real, Informational Biology's reformulation of individual living organisms into self-organizing systems, and the concept of emergence.

Attempting to get to grips with the emerging phenomena of digital photography Fred Ritchin coined the term ‘hyperphotography’ to distinguish new “linked, dynamic, node-like photography” from preceding forms of analogue photography. (Ritchin, 2010, 73) In his vision, a digital photograph is a map of connection. Each pixel serves as a hypertext, a channel to new information. The same is true for computer-generated imagery, except that the connection is far more profound and extensive than Ritchin envisioned. An algorithmic image is embedded in networks, whether technical ones such as social media and image datasets, societal ones such as the techno-military industrial complex, or ontological ones.

In *the Anatomy of an AI system (2018)*, Kate Crawford and Vladan Joler conducted an anatomical study of the Amazon AI assistant Echo. They mapped out the entangled network of human labour, machines, and internet infrastructure sustaining Amazon’s smart home device, from the mineral mining that produces its parts to the data preparation that powers the AI system (Crawford, Joler, 2018). Similarly, the human faces of *thispersondoesnotexist.com* entail a combination of extensive neural networks and computational infrastructures. The nonexistent faces are generated by GAN, a subset of deep learning framework especially good at image generation tasks that was pioneered by Ian Goodfellow.(Goodfellow et al, 2014) In his words, GAN is ‘unsupervised machine learning, implemented by a system of two neural networks competing against each other in a zero-sum game framework.’ (Nvidia, 2020) GAN comprises two networks: Discriminator and Generator simultaneously trained using one dataset. Generator produces a range of images based on the data distribution and parameterized modifications. Discriminator distinguishes the Generator’s output from the image in the original dataset. The two networks compete to optimise their tasks: Generators get better at fooling Discriminator. Discriminators improve on detecting real data from fake images. The self-correction process repeats until the network achieves its aim, in this case, generating a convincing human face. Wang's website utilises StyleGAN2, an improved framework developed by Nvidia researchers Tero Karras and others (Karras et al, 2020). Without going too much into the technical details, this team redesigned the network so that it employs new mathematical operations that minimise what they called 'normalisation artefacts' - blob-shaped artefacts are allegedly caused by a statistical operation that normalises the image data, which in turn affects Generator's output. StyleGAN2 is trained on the Flickr-Faces-HQ dataset, commonly-known as FFHQ, a dataset consisting of 70,000 high-quality PNG images at 1024×1024 resolution crawled from Flickr (NVlabs/ffhq-dataset, 2020).

In addition to the networked image organised into datasets by a social network of crowd-sourced labour, GAN relies on another type of network: neural networks. The two competing neural networks consist of layers of perceptrons, each trained to identify specific features of the images. Trained using deep learning, the perceptrons adjusted the weights of connection to each other using back propagation. Each perceptron affects the degree to which other perceptrons are activated according to a decision-tree-like logic, which determines the final result. The 'learning process' refers to the neural network adjusting the weight of its connection to produce the desired result given a certain input (ArXiv, 2014).[4] By repeatedly providing input (training data) and judging its output (in this case, the image of a human face), whether, by a human agent or another neural network, the neural network improves its accuracy in per-

[4] There are different approaches to machine learning, briefly summarized in Ian J. Goodfellow et al., "Generative Adversarial Networks."

forming specific tasks. This reciprocal relationship between human and machine and machine and machine continues until the neural network creates results comparable to human visual intelligence.

Training an AI also requires image datasets. Because a neural network learns through repetition, the more images a dataset has, the more useful it is. ImageNet, a computer vision dataset created by Fei-Fei Li, has more than 14 millions photographs organised in more than 21,000 categories according to a linguistic hierarchy (Fei-Fei, 2020). The creation of image datasets is made possible by the network. The Internet infrastructure, cloud servers, and the endless flow of content that we willingly upload online provide a pipeline for large-scale collections of images. The Internet also makes crowdsourcing logistically much more manageable. With platforms such as Amazon Mechanical Turks, Li can mobilise a vast amount of available cheap human labour to hand-label images from the datasets that 'collectively produce a visual map of the world as identifiable objects (ibid).' The network is integral to the image dataset.

But the disciplinary frame needs further expanding to match the enormous scale of this AI system. The computational tasks of computer vision are spread across a planet-wide network infrastructure. The training algorithms may be coded on a laptop, but the training process is executed remotely in a supercomputer at a data center, pulling in image datasets from another server. As Nvidia CEO Jensen Huang remarked in 2020: 'In the next decade, Data-centre-scale computing will be the norm, and data centres will be the fundamental computing unit.'(Huang, 2020) If it was not obvious before, the operational mechanism of computer vision now spans the globe.

Networked digital images exceed the scalar frame of human-centred approaches to photography. GAN-generated faces exemplify the way image-making has shifted from a singular act to a distributed event in a networked and complex technical apparatus. Rather than the singular indexicality attributed to analogue photography, the technological structure of algorithmic images reifies the interconnection of photographic images. An algorithm needs a large group of uniform images of faces to understand human faces. In computer vision, the singular perspective of the artist-author is replaced by a heterogeneous mix of human and computer agents. Computer vision operationalises digital image production to optimise the weights given layers of perceptrons in the neural network. The chosen group of images (i.e., the dataset), in turn, decides what the singular image means. The meanings attributed to images is in constant flux through the injection of meta-data, tagging and hand-labelling, and floats in and out of relevance with every retweet and reblog. Each actor forms a node in the artificial visual cortex. Like in a neural network, the connection with the most weight dictates the predominant meaning but never completely settles on a single one. Subsidiary meanings can take over at any time, given a change in the stimulus and the network.

Even though contemporary images ‘happen’ on a larger scale, this does not necessarily translate to more clarity. Each operative scale is a trade-off that renders certain relations visible while obscuring others. Platform seeing depends on statistical tools and parameters to mediate the scale of the network, making a larger-than-human entity accessible and actionable. However, the fixation on the network over the singular often reduces images to mere data and ignores their historical weight and cultural depth. Computer vision is frequently criticised for downplaying inherent biases, such as racial bias, in the statistical processes. The issue of diversity and racial biases have been addressed within and outside of the field of computer vision (Crawford and Paglen, 2019; Buolamwini and Gebru, 2018) – and are acknowledged in the FFHQ dataset github page – but the problem persists. The FFHQ dataset contains 70,000 real faces, from which StyleGAN2 generates an infinite number of faces. It is an oversimplification to say that the computer-generated faces are the average of those 70,000 – the exact process is convoluted and mathematically rigorous. However, the notion of *average* hints at the mathematical artefice of machine learning frameworks. The 70,000 faces are reduced to pixel patterns and feature maps, subjected to a series of statistical operations: normalisation, regularisation, and modulation, abstracted to mean, variance, and standard deviation, and formatted for computations. The network enables computer vision to externalise human vision by *averaging* the collective act of making photographs. The image sustained by a network represents an "average idea" of a human face.

Moreover, recognising contemporary images as distributed events does not only entail the invocation of the network, it raises questions as to whether the zooming out from the singular image is ever sufficient to render the entire network visible. Addressing a common misconception about scale, Horton invokes an argument of Alexander Galloway and Eugene Thacker: "networks are a matter of scaling, but a scaling for which both the 'nothing' of the network and the 'universe' of the network are impossible to depict. One is never simply inside or outside a network; one is never simply 'at the level of a network." (Horton, 2020, p.31) The vastness of networks challenges anthropocentric positions in photography theory. Images and humans are enmeshed in platforms, reciprocally involved in becoming and unbecoming. Networks complexify image production. They add unpredictability to the meaning of an image. An image becomes ontologically inexhaustible and seemingly gains an equal footing to the ‘agency’ of humans. In light of this, the concepts of ‘flat ontology’ and ‘object’ developed in speculative realism may help us to navigate this posthuman condition of contemporary images in algorithmic culture.

The Object

Discussion of the network behind GAN affirmed that images are greater than their pictorial surfaces. The question ‘What is an algorithmic image?’ demands a scalar frame that is not of the order of representation. The study of images in algorithmic culture would not be complete without an non-visual framework that takes into account the *network* and *scale o*f images and which I will call a *hyperimage*.

A network and scalar theory of image takes the algorithmic image to operate on different levels. Graham Harman’s conception of the object, I argue, allows one to address this complexity. His ontology maintains that objects are inexhaustible entities, with which humans can only interact ‘vicariously’ (Harman, 2018, 153). He further characterises objects as *quadruple* (Harman, 2018, 157) as a fourfold of: real object, real qualities, sensual object and sensual qualities. A real object exists in its own right and is inaccessible to other objects, but we can vicariously interact with it through its connection with real and sensual qualities. This framework cultivates the awareness of *the invisible*. In this vein, *thispersondoesnotexist.com* is an object and the image on the screen is one of its many sensual objects. We can access the sensual object through our direct experience of its sensual qualities such as the colour of the pixels and the light intensity of the screen. However, sensual objects also exhibit real qualities, such as the *idea* of a human face, while other real qualities, such as its algorithmic nature and the high dimensional data, remain inaccessible to the senses. Finally, real objects such as

GAN, the technological network which contains all iterations of fake faces, are never directly accessible to humans.

Harman's notion of object as 'anything that cannot be reduced either upwards or downwards' is useful for our analysis. In his account, an event can be an object and an object can contain other objects (The group, *The Beatles* is an object made up its members who are themselves also objects). Harman argues that thinkers tended to either reduce an object to its constituent components (what he calls 'undermining'), or to its effects and relations to others ('overmining'). He proposes that an object is 'more than its pieces and less than its effect' (Harman, 2018, 53). This critical lens can be applied to the case of *thispersondoesnotexist.com*. Focusing on one of its visual images would be to reduce the image to pictorial representation; to undermine it. Dismissing the visuality of the image entirely would be to overmine *thispersondoesnotexist.com*.

Object oriented ontology (OOO) holds that objects are inexhaustibly deep. They are, "dark and stormy events locked in a network with other such events." (Harman, 2007, 24) The claim that no object, even humans, can have complete knowledge of another object, inaugurates OOO as *flat ontology* (Harman, 2007, 54). In the context of the analysis on *thispersondoesnotexist.com*, flat ontology and the 'obscurity' of objects it rests upon, coincide with the notions of network and scale discussed above. The parallel reveals that images have potential beyond their indexical singularity. Photographs of faces, captured by human actors, are archived on the Internet. The images are then organised into datasets and made machine-readable by extensive manual labelling. Through deep neural networks, these images are transcoded into mathematical models, which are then developed into Deepfake and facial recognition softwares. The softwares, to which the images act as an interface, form a control mechanism based on visibility. Each image in the dataset defines what is visible to the machine and renders the outliers invisible (Buolamwini, Gebru, not dated, 15). In response, humans make adjustments either by feeding more images into the algorithms in an effort to enhance it (Yang et al, 2019), augmenting their images and bodies (Belluz, *Vox*, 2018) to conform to it or to leverage on its weakness to bypass computer vision entirely (Harvey, 2013 and Goodfellow et al, 2015). The action of human agents in turn changes the images that enter the network. Each of these steps generates its own feedback loop. Each object exerts an asymmetrical yet reciprocal influence on every other to the extent that human vision and computer vision come to shape each other.

The Hyperobject

Scaling up the disciplinary lens on images reveals the otherwise invisible network of algorithmic images, the distributed nature of which means it can be thought of as a hyperobject. Timothy Morton defined the term as follows:

> In The Ecological Thought I coined the term hyperobjects to refer to things that are massively distributed in time and space relative to humans. A hyperobject could be a black hole. A hyperobject could be the Lago Agrio oil field in Ecuador, or the Florida Everglades. A hyperobject could be the biosphere, or the Solar System. A hyperobject could be the sum total of all the nuclear materials on Earth; or just the plutonium, or the uranium. A hyperobject could be the very long-lasting product of direct human manufacture, such as Styrofoam or plastic bags, or the sum of all the whirring ma-

chinery of capitalism. Hyperobjects, then, are “hyper” in relation to some other entity, whether they are directly manufactured by humans or not. (Morton, 2013, p.1)

In this light, rather than being a singular image, an algorithmic image appears as a rhizomatic entity. The image is a symptom of a large ecology of agents, an interface of a deep object. As an object it entails a global network of agents distributed across time and space much larger than the human scale. Algorithmic images exemplify an operationality that shifts from singular indexicality to *distributed vision* a term denoting the fact that algorithmic images are generated not from a single computer, but an intercontinental computational infrastructure. Data collected from the Internet at one end of the world is processed by Mechanical Turks at the other. The instruction executed from one laptop evoked an AI model developed from code written in another laptop, but the computation may happen elsewhere in a datacentre thousands of kilometres away. The computational infrastructure itself is also a hyperobject with its own modes of impacting the world. For instance, the components of computation require rare metals obtained from commercial mining and the cooling of data centres consumes electricity and emits greenhouse gases. (Hogan, 2018, 631) Their impact on the environment, in turn, changes what the world looks like and what images will be produced from it.

Distributed seeing refers to the phenomenon that every image has the potential to influence any other image. While this might also ring true in certain ways for traditional pictorial media, the concept is mathematically formulated and technologically embedded in algorithmic images. Since every available digital image can potentially be organised into a training dataset, each image, regardless of the time and place it is taken, impacts the network and by extension, influences future modes of image production. A candid and trivial photograph of a dog taken in someone's backyard and uploaded online is potentially the ingredient of a computer vision system that identifies other dogs in other photographs. A headshot of a Japanese woman taken in the 1990s is part of the dataset that trained machines to identify emotions (Lyons et al, 1998, 200-205). More companies are using these systems to screen application CVs and filter out 'hostile' faces (Harwell, *Washington Post*, 2019/10/22). A digital photograph is an array of RGB values, from whose pattern features such as edge and pattern can be extracted. In deep learning, an image is treated as a set of high-dimensional[5] vectors and compressed into a latent space where data points are situated according to their similarities. In more technical terms, basic deep neural networks are trained using an algorithmic technique of brute force approximation called gradient descent (Pasquinelli, Joler, 2020, 11), optimising its ability to meaningfully distinguish one pattern from another in a latent space, be it different digits, objects or faces. A portrait of a celebrity at a movie premiere is hiding somewhere behind the multi-dimensional neural network that produces the faces of a person that does not exist. Morton further explains how the high-dimensionality of hyperobject escapes human comprehension:

[5] High dimension in this sense refers to multiple data dimensions rather than spatial dimensions. For example, a colour in RGB scale is three-dimensional as it is represented by three numbers for each red, blue and green channel respectively.

> *Think of a daffodil. A daffodil flower is a three-dimensional map of an algorithm executed by DNA and RNA in the daffodil's genome. The tips of the crinkly stem show the latest state of the algorithm unfolding in three-dimensional phase space. The base of the flower is a plot of the beginnings of the flower algorithm. Your face is a map of everything that happened to it... We only see snapshots of what is actually a very complex plot of a super complex set of algorithms executing themselves in a high-dimensional phase space...What you once thought was real turns out to be a sensual representation, a thin slice of an image...A process just is a real object, but one that occupies higher dimension than objects to which we are accustomed (Morton, 2013, 70).*

Embedded in every image is a network of other images. A flat ontology of the image acknowledges that every image can form part of an automaton that generates other images. Putting images and humans on the same ontological footing has several implications. For one, both entities affect the becoming and unbecoming of each other. For another, images stare back at humans as they are being looked at. An image is not just the product of a singular perspective, but also the culmination of a collective seeing carried out by the whole of human civilization across time and space. A singular account of an image does not suffice to take in account the multidimensionality and network nature of images. Behind one face are millions of faces.

The posthuman paradigm needs to understand the image as a *hyperimage*. As Morton has pointed out, the tricky part of a hyperobject is its bigness, which is something humans are not adept at contemplating. But the bigness of a hyperobject does not imply infinity. Morton distinguishes bigness and infinity, arguing that infinity is a much more easily graspable concept than bigness for human minds.

> These gigantic timescales are truly humiliating in the sense that they force us to realize how close to Earth we are. Infinity is far easier to cope with. Infinity brings to mind our cognitive powers, ...There is a real sense in which it is far easier to conceive of "forever" than very large finitude. Forever makes you feel important. One hundred thousand years makes you wonder whether you can imagine one hundred thousand anything (Morton, 2013, 60).

It is this bigness that forces a paradigm shift. A single random event is unpredictable but a large enough number of random events forms a pattern. Bigness *per se* is a quality. Hyperconnectedness brings forth a new, previously sensually inaccessible aspect of images. Objects behave differently at different scales and require new analytical lenses that operate at scales, not necessarily mutually exclusive. The microscopic and macroscopic are useful for navigating tensions between the individual and the whole in a hyperobject. Inverting the common saying *the whole is greater than the sum of its parts*, Morton suggests that *the whole is also less than the sum of its parts.* He used the relationship between weather and climate change as an example to illustrate objects' resistance to *duomining*:

> Weather is a symptom of climate. Obviously, everything that happened happens because of global warming…Global warming is the biggest scale thing within which all the other stuff is happening…but the thing is weather does so much more than being just a symptom of global warming. It is this lovely soft sensation on my skin. It is this bath for these toads outside my house (Morton, *Youtube*, 2020).

To expand on this example, the mass of photographs we upload online - around 1.8 billion per day and embedded with all sorts of metadata (Eveleth, *Atlantic*, 2015) - forms a complex network of data. From this mass of photographs, a pattern emerges. Using specialised mathematical models and computational structure, computer vision exploits the emergent phenomena to achieve facial recognition, object detection and GAN-generated faces. *thispersondoesnotexist.com* is a system in which the computer-generated faces alongside training photo datasets of ordinary people are symptoms. It is therefore bigger than the sum of its parts. But the photographs are also deeper than the data and metadata which computer vision exploits. For example, they have a phenomenological potential. Each photograph is also a medium of interpersonal memory, of human perception and of psychology. Thus, one can also say that computer vision is *smaller than the sum of its parts* for objects are inexhaustably deep and only interact with each other vicariously.

Hyperimage approaches the posthuman condition of contemporary images through the notion of network and scale. Hyperimage prioritises the network, the system, the model, the relationships between images and between images and the apparatus over the singular, the representational and the purely aesthetic. Hyperimage recognises the oscillation between the individual and the network underpinning the new image paradigm. Hyperimage cannot be *duomined* (Harman, 2018, p. 54) – it cannot be undermined to its pictorial surface or its constituting pixels; it cannot be overmined to its societal effects and operational technicalities. Hyperimage operates on least two scales. On the one hand, it acknowledges the networked image. Each image is a symptom of the invisible network of agents that sustains it. Each image injected into the system (both technological and cultural) constitutes a hyperobject, which in its bigness affects the whole of computer vision and human vision. On the other hand, each image is more complex than just a symptom of the network. Images exhibit their qualities in relation to other objects, they operate interobjectively (Morton, 2013, 1). Images have an experiential dimension. They contain cultural memory and shared history. In the human-image interactions, the real qualities and the sensual qualities affect us on an interobjective level.

Every image exists on a scale larger than humans and ‘massively distributed in time and space relative to humans.’ (Morton, 2013, p.1) Every image itself is a picture *and* a part of a larger network of images. Each machine-generated face of *thispersondoesnotexist.com* is simultaneously a representation of the idea of a human face and an event of distributed seeing. In other words, it is a hyperobject.

References

Barad, Karen Michelle (2007) *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press.

Belluz, Julia (2018) 'Selfie Face Distortion Is Driving People to Get Nose Jobs.', [Online], *Vox*, 1 March, https://www.vox.com/science-and-health/2018/3/1/17059566/plastic-surgery-selfie-distortion (Accessed: 15 May 2020).

Buolamwini, Joy, and Gebru, Timnit (2018) 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,' [Online], Proceedings of the 1st Conference on Fairness, Accountability and Transparency, http://proceedings.mlr.press/v81/buolamwini18a.html, (Accessed: 15 May 2020).

Chun, Wendy Hui Kyong. (2005) "On Software, or the Persistence of Visual Knowledge." Grey Room 18:26–51. doi: 10.1162/1526381043320741.

Chun, Wendy Hui Kyong (2011) *Programmed Visions: Software and Memory, Software Studies,* Cambridge, Massachussetts: MIT Press

Crawford, Kate and Joler, Vladar (2018) *Anatomy of an AI System*, [Online], http://www.anatomyof.ai, (Accessed: 15 May 2020).

Crawford, Kate and Paglen,Trevor (2019) "Excavating AI: The Politics of Training Sets for Machine Learning" (September 19) https://excavating.ai (Accessed: 15 May 2020).

Dahmani, Louisa, and Véronique D. Bohbot. (2020) "Habitual Use of GPS Negatively Impacts Spatial Memory during Self-Guided Navigation." Scientific Reports 10(1):6310. doi:10.1038/s41598-020-62877-0.

Elkins, James (1997) *The Object Stares Back: On the Nature of Seeing*, [Online], San Diego, New York, London: Harcourt Brace, http://archive.org/details/objectstaresback00elki_0, (Accessed: 15 May 2020).

Eveleth, Rose (2015) 'How Many Photographs of You Are Out There In the World?', [Online], *The Atlantic*, 2 November, https://www.theatlantic.com/technology/archive/2015/11/how-many-photographs-of-you-are out-there-in-the-world/413389/, (Accessed: 15 May 2020).

Fei-Fei, Li (2020) 'Where Did ImageNet Come From?' https://unthinking.photography/articles/where-did-imagenet-come-from, (Accessed January 4, 2020).

Fisher, Andrew. (2012) "Photographic Scale." Philosophy of Photography 3(2):310–29.

doi: 10.1386/pop.3.2.310_1.

Flusser, Vilém (2017) Countervision, In: "'To document something which does not exist.' Vilém Flusser and Joan Fontcuberta: A Collaboration", *Flusser Studies*, vol 13. May 29

Frosh, Paul (2002) 'Rhetorics of the Overlooked: On the Communicative Modes of Stock Advertising Images', *Journal of Consumer Culture* 2, no. 2 (July): p. 171–96, https://doi.org/10.1177/146954050200200202.

Galloway, Alexander R. (2013) 'The Poverty of Philosophy: Realism and Post-Fordism', *Critical Inquiry* 39, no. 2 (January): https://doi.org/10.1086/668529, p. 347–66

Goodfellow, Ian J., Shlens, Jonathon and Szegedy, Christian (2015) 'Explaining and Harnessing Adversarial Examples', *ArXiv:1412.6572 [Cs, Stat]*, March http://arxiv.org/abs/1412.6572 , (Accessed: 15 May 2020).

Goodfellow, Ian J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014) Generative Adversarial Networks. ArXiv:1406.2661 [Cs, Stat]. http://arxiv.org/abs/1406.2661, (Accessed: 15 May 2020).

Hunag, Jensen (2020) 'GTC Digital 2020 Keynote Speaker CEO Jensen Huang',[Online], *NVIDIA*, https://www.nvidia.com/en-us/gtc/keynote/, (Accessed: 20 May 2020).

Harman, Graham (2007) *Heidegger explained: from phenomenon to thing*. Chicago: Open Court.

Harman, Graham (2018) *Object-Oriented Ontology: A New Theory of Everything*. Pelican Book 18. London: Pelican Books.

Harvey, Adam (2013) *CV Dazzle*, https://cvdazzle.com/, (Accessed May 20, 2020).

Harwell, Drew (2019) 'A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job,' [Online], *Washington Post*, 6 November, https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/ (Accessed: 7 February 2021).

Hogan, Mél (2018) 'Big Data Ecologies', *Ephemera* 18, no. 3: 631.

Horton, Zachary K. (2020) The Cosmic Zoom: Scale, Knowledge, and Mediation. Chicago ; London: The University of Chicago Press.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020) 'Analyzing and Improving the Image Quality of StyleGAN'. ArXiv:1912.04958 [Cs, Eess, Stat]. http://arxiv.org/abs/1912.04958.

Latour, Bruno (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*.

Clarendon Lectures in Management Studies. Oxford ; New York: Oxford University Press.

Lemke, Thomas (2017) ‘Materialism without Matter: The Recurrence of Subjectivism in Object-Oriented Ontology.’, *Distinktion: Journal of Social Theory* 18, no. 2 (May 4) https://doi.org/10.1080/1600910X.2017.1373686, p. 133–52.

Lyons, Michael J., Akamatsu, Shigeru , Kamachi, Miyuki, Gyoba, Jiro
(1998) ‘Coding Facial Expressions with Gabor Wavelets’, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, http://doi.org/10.1109/AFGR.1998.670949, p. 200-05
Open access content available at: https://zenodo.org/record/3430156

MacKenzie, Adrian, and Munster, Anna (2019) ‘Platform Seeing: Image Ensembles and Their Invisualities.’ *Theory, Culture & Society* 36, no. 5 (September), https://doi.org/10.1177/0263276419847508, p 3–22.

Mitchell, William J.(1992) *The Reconfigured Eye: Visual Truth in the Post-Photographic Era*. Cambridge, Massachussetts: MIT Press.

Mitchell, W. J. T. (2005) *What Do Pictures Want? The Lives and Loves of Images*. Chicago: University of Chicago Press.

Morton, Timothy (2013) *Hyperobjects: Philosophy and Ecology after the End of the World*. Posthumanities 27. Minneapolis: University of Minnesota Press.

Nvidia (2020) ‘Deep Learning Glossary,’ [Online], https://www.nvidia.com/en-us/data-center/resources/deep-learning-glossary/
(Accessed: 4 March 2020).

NVlabs/ffhq-dataset. (n.d.). [Online], GitHub. Retrieved May 26, 2021, from https://github.com/NVlabs/ffhq-dataset

Paglen, Trevor, and Kate Crawford. n.d. “Excavating AI.” Excavating AI.
Retrieved March 18, 2020 (https://www.excavating.ai).

Paglen, Trevor (2014) ‘Operational Images’, [Online], *E-Flux*, no. 59 (November).
https://www.e-flux.com/journal/59/61130/operational-images/, (Accessed: 20 May 2020).

Pasquinelli, Matteo (2019) ‘How a Machine Learns and Fails – A Grammar of Error for Artificial Intelligence’, Text. *Spheres* (blog), [Online], 20 November, http://spheres-journal.org/how-a-machine-learns-and-fails-a-grammar-of-error-for-artificial-intelligence/, (Accessed: 20 May 2020).

Pasquinelli, Matteo, and Joler, Vladan (2020) *The Nooscope Manifested Artificial Intelligence as Instrument of Knowledge Extractivism*, [Online], https://nooscope.ai/, (Accessed: 20 May 2020).

Ritchin, Fred. (2010) After Photography. First paperback ed. New York, NY: Norton.

Rose, Todd (2015), *The End of Average: How We Succeed in a World That Values Sameness*. First Edition. New York: HarperOne.

Rubinstein, Daniel, and Sluis, Katrina (2008) 'A LIFE MORE PHOTOGRAPHIC: Mapping the Networked Image', *Photographies* 1, no. 1 (March), https://doi.org/10.1080/17540760701785842, p. 9–28.

Rubinstein, Daniel (2020) 'Fragmentation of the Photographic Image in the Digital Age', ed. Daniel Rubinstein, 1st ed. (New York, NY: Routledge, 2020. | Series: [Routledge history of photography] https://doi.org/10.4324/9781351027946.

The RSA (2018) *Being Ecological | Timothy Morton | RSA Replay - YouTube,* [Online] https://www.youtube.com/watch?v=d_5UWI-SEVE, (Accessed: 20 May, 2020.)

Virilio, Paul (1994) *The Vision Machine*. Indiana University Press.

Yang, Kaiyu, Qinami, Klint, Fei-Fei, Li, Deng, Jia and Russakovsky, Olga, (2019) 'Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy.' *ArXiv:1912.07726 [Cs]*, December 16,. https://doi.org/10.1145/3351095.3375709.

Zylinska, Joanna (2017) *Nonhuman Photography*, Cambridge, Massachusetts: The MIT Press