Antonio Somaini

# Film, media, and visual culture studies, and the challenge of machine learning

2021-12-13

https://doi.org/10.25969/mediarep/17289

Veröffentlichungsversion / published version
Zeitschriftenartikel / journal article

**Empfohlene Zitierung / Suggested Citation:**
Somaini, Antonio: Film, media, and visual culture studies, and the challenge of machine learning. In: *NECSUS_European Journal of Media Studies*. #Futures, Jg. 10 (2021-12-13), Nr. 2, S. 49–57. DOI: https://doi.org/10.25969/mediarep/17289.

**Erstmalig hier erschienen / Initial publication here:**
https://necsus-ejms.org/film-media-and-visual-culture-studies-and-the-challenge-of-machine-learning/

EUROPEAN JOURNAL OF MEDIA STUDIES
www.necsus-ejms.org

# Film, media, and visual culture studies, and the challenge of machine learning

Antonio Somaini

NECSUS 10 (2), Autumn 2021: 49–57
URL: https://necsus-ejms.org/film-media-and-visual-culture-studies-and-the-challenge-of-machine-learning/

The history of visual cultures is periodically marked by the appearance of new images and new technologies of vision: images that introduce new forms of representation, and technologies that introduce new ways of seeing, extending, and reorganising the field of the visible. In some cases, such changes produce only marginal transformations, while in others the transformations are vast, tectonic shifts. This is what happened during the 1990s and early 2000s, when digital visual technologies gradually replaced analog ones, and a faster transmission of data across the internet opened the way for an increased circulation of digital images. And this is what is happening again today, as artificial intelligence — in particular, that area of AI known as machine learning — is profoundly transforming the ways that images are produced, modified, circulated, and seen. Three phenomena in particular deserve our closest attention, and constitute a new challenge for the field of film, media, and visual culture studies: the new technologies of machine vision based on artificial neural networks; the presence on the internet of trillions of images that are machine-readable, in the sense that they can be processed and analysed by technologies of machine vision; and the genuinely new types of images that may be produced through processes of machine learning.

Considered from the perspective of a history of images and visual media, the appearance of these three phenomena raises a large series of aesthetic, epistemological, historical, and political questions. Their impact on contemporary film, media, and visual culture is so deep that we must ask ourselves what we mean by the notions of 'vision' and 'image' in the age of machine learning. The very status of moving images, as well as their various forms of production, editing, and reception, are being affected. The traditional

boundaries between fixed and moving images are put into question, as is the distinction between images that are the result of optical recording and those that are entirely computer-generated. Key concepts in film and media theory, such as realism, need to be reevaluated when dealing with technologies that entirely reconfigure the relationship between images and profilmic reality.

## The impact of machine learning technologies

First tested in the late 1950s with image recognition machines such as the Perceptron (invented by the Cornell Aeronautical Laboratory in 1957), and then developed during the 1960s and 1970s as a way of imitating the human visual system in order to endow robots with intelligent behavior, machine vision technologies entered a new phase with the development of machine learning processes, and the possibility of using immense image databases, accessible online, as both training sets and fields of application. The training sets are organised according to precise taxonomies — such as ImageNet, in which 14 million images are arranged according to 21,000 categories derived from the WordNet hierarchy (a large lexical database of English nouns, verbs, adjectives, and adverbs)[1] which allow a rapid increase in the precision of all operations of machine vision.

Among such operations we find pixel counting; segmenting, sorting, and thresholding; feature, edge, and depth detection; pattern recognition and discrimination; object detection, tracking, and measurement; motion capture; color analysis; optical character recognition (this last operation allowing for the reading of words and texts within images, extending the act of machine seeing to a form of reading). For several years now, these operations – which expand the field of image operations that began to be investigated by Harun Farocki in the early 2000s[2] – have been applied to the immense field of machine-readable images. A field whose dimensions can be imagined only if we understand that any networked digital image — whether produced through some kind of optical recording, or entirely computer-generated, or a mix of the two, as is often the case — may be analysed by machine vision technologies based on processes of machine learning, such as Generative Adversarial Networks (GAN).[3]

In recent years, smartphone producers have equipped their devices with cameras and image processing technologies that turn every photo we take

into a machine-readable image, and internet giants such as Google and Facebook, as well as a host of state agencies and private companies, have developed machine vision systems. Taken together, these systems are turning the contemporary digital iconosphere into a vast field for data mining and data aggregation. Faces, bodies, gestures, expressions, emotions, objects, movements, and places may be identified, labeled, stored, organised, retrieved, and processed as data that can be quickly accessed and activated for a wide variety of purposes: from surveillance to policing, from marketing to advertising, from the monitoring of industrial processes to military operations, from driverless vehicles to drones and robots, from the inspection of the inside of the human body (medical imaging) all the way to the study, through satellite images, of the Earth's surface and climate change.

In order to fully understand the impact of AI and machine learning on contemporary visual culture, we need to add those images produced by processes that either transform pre-existing images in ways that were impossible until quite recently, or create entirely new images, never before seen.

Examples of transformation include: producing 3D models of objects from 2D images; altering photographs of human faces in order to show how their appearance might change with age (as with FaceApp), or be merged with another face (Faceswap); animating the old photograph of a deceased person in a highly realistic way (Deep Nostalgia, developed by MyHeritage);[4] taking any given video and upscaling it by increasing its frame rate and definition. An emblematic example of this last application, which in the long run may alter significantly our experience of visual documents of the past, are the videos realized by Denis Shiryaev in which, through a machine learning process, a Lumière film such as *Arrival of a Train at La Ciotat* (1896) is transformed from the original 16 frames per second to 60 frames per second, from the original 1.33:1 format to a contemporary 16:9 format, and from the original, grainy 35mm analog film to a 4K digital resolution.[5]

Other examples of transformation are far more radical, as happens with so-called deepfakes: videos that use neural networks such as autoencoders or GAN to manipulate the images and sounds of pre-existing videos, producing new videos that have a high potential to deceive, thereby further destabilising our trust in recorded images. Among the many examples that can now be found across the internet, videos in which faces of celebrities are placed onto the bodies of porn actors, or speeches by public figures such as Barack Obama and Queen Elizabeth,[6] the content of which has been completely

altered in such a way that the movements of their mouths, thanks to a program called Face2Face, perfectly match the invented words uttered by someone else. In the case of image creation, we are dealing with entire images or sections thereof: examples include modeling patterns of crowd motion in films and videogames; producing photorealistic images of objects and environments for advertising; and inventing highly realistic faces of people who do not actually exist.[7]

To these widespread applications of machine learning we may add the hybrid, unprecedented imagery produced by the popular computer vision program Deep Dream Generator, created in 2015 by the Google engineer and artist Alexander Mordvintsev.[8] This is a program that uses neural networks in order to enhance patterns in any given image, creating a form of algorithmic pareidolia (the impression of seeing a figure where there is none) generated by a process which repeatedly detects and enhances patterns and shapes that the machine vision system has been trained to see. The result of such a recursive process are images that recall a psychedelic iconography that spans cinema, photography, the visual arts, and even art brut: images presented as a dream belonging to the machine itself.

## A new set of questions for theory

The widespread diffusion of machine vision technologies, machine-readable images, and the new images produced by processes of machine learning raises a series of theoretical questions. Some of these are related to the broad field of a theory of media and visual culture, while others are specifically related to film theory. What is vision when the human psycho-physiological process of seeing is reduced, in the case of machine vision technologies, to entirely automated operations of pattern recognition and labeling, and when the various applications of such operations (face and emotion recognition, object and motion detection) may be deployed across an extremely vast visual field (all the still and moving images accessible online) that no human eye could ever attain? By using the term 'vision' within the concept of machine vision, are we mistakenly using an anthropocentric term that should be discarded in favor of a different set of technical terms, specifically related to the field of computer science and data analysis?

Artist-researchers such as Francis Hunger and scholars such as Andreas Broeckmann (with his notion of 'optical calculus' as 'an unthinking, mindless

mechanism, a calculation based on optically derived input data, abstracted into calculable values, which can become part of computational procedures and operations'), Adrian MacKenzie, and Anna Munster (who speak of a 'platform seeing' operating within 'image ensembles' through an 'invisual perception'), Fabian Offert and Peter Bell (according to whom the 'perceptual topology' of machines is irreconcilable with human perception), have all argued for the necessity of moving beyond anthropocentric frameworks and terms, highlighting the fact that machine vision poses a real challenge for the humanities.[9]

Can we still use the term 'image' for a digital file, encoded in some image format,[10] that is machine-readable even when it is not visible by human eyes, or that becomes visible on a screen as a pattern of pixels only for a small fraction of time, spending the rest of its indefinite lifespan circulating across invisible digital networks? Can concepts such as that of 'iconic difference',[11] which highlights the fundamental perceptual difference between an image and its surroundings (its 'off frame'), still be applied to machine-readable images? And how to assess the various attempts — through concepts such as 'iconic turn' and 'pictorial turn' — to underline the necessity of developing concepts for image and visual culture theory that are not derived from language-based disciplines such as semiotics, when the new technologies of machine vision are entirely based on a strict interrelation between words and images? And what is the status of the entirely new images produced by processes of machine learning? These images are not produced through traditional forms of lens-based analog or digital optical recording, nor through traditional computer-generated imagery (CGI) systems, but rather through processes belonging to the wide realm of artificial intelligence (AI). What do such images represent, what kind of agency do they have, what is their temporal status, and how do they mediate our visual relation to the past, the present, and the future?

To this series of questions related to the vast field of a theory of images and visual culture, we may add questions that are more specific to film theory. How to assess the impact of the new images produced by processes of machine learning on filmmaking and film viewing? The different forms of computational filmmaking that are becoming more and more widespread — technologies that allow cameras to add, in real time, stock images, filters, digital artifacts, and effects to what they are recording — redraw the fine line that separates optical recording from computer generated imagery, and re-

define traditional forms of editing. What happens to the contingent, unpredictable movements that cameras once captured in the profilmic world — the movements of crowds, the fluttering of leaves, the rippling of waves — when such movements are simulated through AI?[12] In what ways is the material and temporal status of historical audiovisual documents altered by machine learning processes that allow us to upscale them? What kind of realism can be detected in images that have a high degree of resemblance and trustworthiness, while being at the same time entirely deceptive? And will the deepfakes that today modify speeches by Barack Obama or Queen Elizabeth allow film directors one day to 'resuscitate' dead actors and have them perform in new films, as is happening in the musical realm with the voices of dead singers?

Other important questions for film studies are raised by machine vision technologies. Are these technologies going to change the way in which we study cinema history (as they are currently changing art history) by allowing researchers to tackle vast corpuses of films? Will we be able to scan through archives, through traditions and genres, searching for faces, expressions, emotions, objects, spaces, environments, atmospheres, frame compositions, color schemes, light configurations, camera movements, editing styles? How are such technologies transforming our relation to film archives? The EYE Filmmuseum in Amsterdam, for example, recently developed a project aimed at 'bring[ing] film heritage to the algorithmic age'. The project is called Jan Bot, capable of generating several found footage videos every day (inspired by trending topics in the news), editing images stemming from the films preserved in this archive.[13]

## A media-archaeological approach, and the role of contemporary artists-researchers

How to tackle these new phenomena, and the challenge that they constitute for our field? Two approaches seem to be particularly promising. We can adopt a media-archaeological approach, which may help us in reconstructing the multiple, interwoven genealogies in which all these developments are inscribed. How, for instance, do deepfakes sit within the tradition of optical media aimed at producing different forms of illusion and deception, from trompe-l'œil paintings to 3D simulations and various forms of digital anima-

tion? And how does machine vision, as a new form of automated seeing, relate to the ideas, hopes, and fears that appear throughout the history of film theory concerning the experience of seeing through the non-human eye of a machine and the possibility of producing new forms of automated perception? The idea that some kind of mechanical vision may either extend the field of human vision beyond the limits of the organic eye, or displace and decenter the human viewpoint by introducing a different, non-human perspective, has triggered reactions dating back to the early years of photography and cinema, which have then been rehearsed and reformulated throughout the second half of the 19th and the entire 20th centuries, up until today.

We also need to pay close attention to the work of contemporary artist-researchers such as Trevor Paglen, Hito Steyerl, Grégory Chatonsky, and many others.[14] Their recent works – fixed and moving images, video installations, texts – present us with an initial series of explorations of the ways in which processes of machine learning are gradually transforming the domain of contemporary images — their relation to profilmic reality, their temporal status, and the ways in which they can be edited through montage. This field is evolving quickly: artists and filmmakers are beginning to test the possibilities that AI introduces, raising questions that theories of film, media, and visual culture studies will need to continue addressing.

## Author

Antonio Somaini is Professor of Film, Media, and Visual Culture Theory at Université Sorbonne Nouvelle in Paris. He is the author of the books *La Glass House de Serguei Eisenstein. Cinématisme et architecture de verre* (Editions B2, 2017), *Cultura visuale. Immagini, sguardi, media, dispositivi* (with Andrea Pinotti, Einaudi, 2016, French translation forthcoming in 2022), and *Ejzenštejn. Il cinema, le arti, il montaggio* (Einaudi, 2011), and the editor of anthologies (in English, French, and Italian) of texts by Walter Benjamin, Sergei Eisenstein, László Moholy-Nagy, and Dziga Vertov. Among his latest edited publications are the collective volumes *Repenser le médium. Art contemporain et cinéma* (with L. Dryansky and F. Casetti, Presses du réel, 2022) and *La haute et la basse définition des images. Photographie, cinéma, art contemporain, culture visuelle* (with F. Casetti,

Mimésis, 2021). In 2020 he was the chief curator of the exhibition *Time Machine. Cinematic Temporalities* (catalogue published by Skira, website www.timemachineexhibition.com)

## References

Boehm, G. 'Ikonische Differenz', *Rheinsprung 11. Zeitschrift für Bildkritik*, 1 (2011), pp. 170-176.

Broeckmann, A. 'Optical Calculus', paper presented at the conference *Images Beyond Control*, FAMU, Prague, 6 November 2020 (recording at: https://www.youtube.com/watch?v=FnAgBbInMfA); A. MacKenzie & A. Munster,

Jancovic, M, Schneider, A. & Volmar, A. (eds.). *Format Matters: Standards, Practices, and Politics in Media Cultures* (Lüneburg: Meson Press, 2019).

Schonig, J. 'Contingent Motion: Rethinking the "Wind in the Trees" in Early Cinema and CGI', in *Discourse*, vol. 40, No. 1 (2018), pp. 30-61.

MacKenzie, A. and Munster, A. 'Platform Seeing: Image Ensembles and Their Invisualities', *Theory, Culture & Society*, Vol. 36, No. 5 (2019), pp. 3-22

Offert, F. and Bell, P. 'Perceptual Bias and Technical Metapictures: Critical Machine Vision as a Humanities Challenge' in *AI & Society* (12 October 2020), https://link.springer.com/article/10.1007/s00146-020-01058-z.

## Notes

[1] http://www.image-net.org

[2] Some major texts by Farocki on operational images: 'Phantom Images', *Public*, no. 29 (2004), and 'War Always Finds A Way' in *HF / RG* [*Harun Farocki / Rodney Graham*] (Paris: Jeu de Paume / Blackjack editions, 2009), p. 107.

[3] Generative adversarial networks are a class of machine learning frameworks that were first designed by Ian Goodfellows and his colleagues in 2014. See Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Asron Courville, Yoshua Bengio, 'Generative Adversarial Nets', in *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Vol. 2, December 2014, pp. 2672-2680.

[4] FaceApp (https://www.faceapp.com/), FaceSwap (https://faceswap.dev/), and Deep Nostalgia (https://www.myheritage.fr/deep-nostalgia).

[5] On the work of Denis Shiryaev, see https://neural.love/. The upscaled version of *Arrival of a Train at La Ciotat* (1896) can be accessed here: https://www.youtube.com/watch?v=3RYNThid23g.

[6] The deepfake video of Barack Obama: https://www.youtube.com/watch?v=cQ54GDm1eL0. The Queen Elizabeth video: https://www.youtube.com/watch?v=IvY-Abd2FfM.

[7] On this last application, see http://doppelGANger.agency.

[8] The program is now open source; see, for example, the website Deep Dream Generator https://deepdreamgenerator.com/. Mordvintsev's website is https://znah.net/.

[9] Broeckmann 2020; MacKenzie & Munster 2019; Offert & Bell 2020

[10] On the theory of formats, see Jancovic et al. 2019.

[11] On the concept of 'iconic difference', see Boehm 2011.

[12] On the status of contingency within computer-generated imagery, see Schonig 2018.

[13] See https://www.jan.bot/livelog.

[14] In recent years, both Trevor Paglen (often in collaboration with Kate Crawford) and Grégory Chatonsky have written important essays on machine vision technologies and images produced through processes of machine learning. See for example: Trevor Paglen, 'Operational Images', *e-flux*, no. 59 (November 2014), https://www.e-flux.com/journal/59/61130/operational-images/; Trevor Paglen, 'Invisible Images (Your Pictures Are Looking At You)', *The New Inquiry*, 8 December 2016, https://thenewinquiry.com/ invisible-images-your-pictures-are-looking-at-you/; Kate Crawford and Trevor Paglen, 'Excavating AI: The Politics of Training Sets for Machine Learning' (19 September 2019), https://excavating.ai; Grégory Chatonsky, 'Après le réalisme : L'espace abstrait de l'intelligence artificielle' (http://chatonsky.net/realisme-ia/ ); 'Hyperproduction: les machines à réalisme' (http://chatonsky.net/hyperproduction-realisme-unige/).