Maria Giulia Dondero*

Semiotics of artificial intelligence: enunciative praxis in image analysis and generation

https://doi.org/10.1515/sem-2024-0195
Received November 11, 2024; accepted December 20, 2024; published online February 19, 2025

Abstract: This paper explores the relation between images and databases in a twofold way. The first part examines image databases as sources for *image computational analysis*, while the second part studies image databases as sources for *image generation* (notably through the generative artificial intelligence model Midjourney). Image analysis and image generation will be analyzed through the concept of enunciative praxis. Traditionally, enunciative praxis concerns cultural transformations over the long term (the relation between sedimentation and innovation); in our case it will be used notably to study the generation of new images through old, traditional ones.

Keywords: computer vision; genealogy; enunciative praxis; generative models; midjourney

1 Introduction

This paper focuses on the relation between images and databases, considering databases as the prime instrument (in association with algorithmic models) for understanding visual language today. The essay comprises two parts: first, it studies image databases as sources for *image computational analysis* (deep learning), and secondly as sources for *image generation* (diffusion models). These two topics, image analysis and image generation, will be addressed through the concept of enunciative praxis. This concept is useful for examining algorithms and databases as tools for a new approach to visual language and its functioning.

The author is indebted to the colleagues and friends who contributed to enhance this text thanks to their precious insights: Enzo D'Armenio, Adrien Deliège, Pierluigi Basso, and Andrea Valle.

***Corresponding author: Maria Giulia Dondero**, F.R.S.-FNRS (National Fund for Scientific Research), Brussels, Belgium; and University of Liège, Liège, Belgium, E-mail: mariagiulia.dondero@uliege.be. https://orcid.org/0000-0003-2320-8130

Traditionally, researchers have used enunciative praxis to study cultural transformations (the relation between sedimentation and innovation).[1] Before addressing enunciative praxis, it is useful to present a brief archeology of the concept of enunciation.

2 On databases and enunciative praxis in semiotic theory

According to a very general definition, the theory of enunciation allows us to conceive of a mediation between Saussurean *langue*, i.e., a system of virtualities, and *parole*, i.e., the actualizations of *langue* in discourse – which are supposed to broaden the spectrum of the virtualities of language. Since Benveniste, the theory of enunciation has greatly evolved and has enabled to take into account the fact that Saussurean *langue* is not a system of pure grammatical virtualities to be further performed but a system of historically attested discursive habits, built on the practices of the speakers – what Fontanille (2003) calls "virtualization mode of existence" in the process of enunciative praxis. In Fontanille's work, enunciative praxis is a matter of conceiving of a dynamic between the *sedimentation* of existing schemas of signification (virtualization) and the *creativity* inherent to any ongoing semiotic process (realization). In my opinion, this dynamic of the fundamental modes of existence in discourse practices has the merit of valuing the complexity of our linguistic operations caught between creativity and sedimentation, because it multiplies the steps and the nuances in this process (Figure 1).

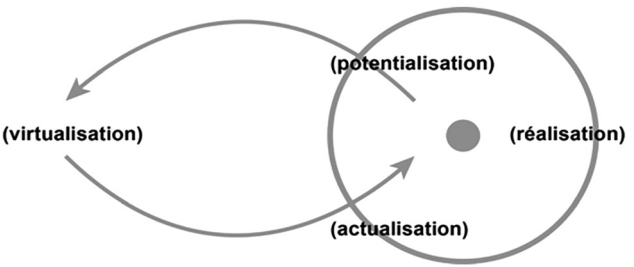

Figure 1: Schema of the modes of existence in enunciative praxis (Fontanille 2003: 199).

1 For a comparison between Fontanille's (2003) general conception of enunciative praxis and a similar concept in linguistic anthropology, especially used in historical analysis of linguistic habits, see Asif's work (2003, 2005) on English pronunciations. On enunciative praxis and C. Goodwin's work in interactional linguistics, see Dondero (2024).

The enunciative praxis is not the sum of all discourses performed, but the locus of a discursive schematization that makes it possible to account for the thickness of our linguistic performances, caught between projection (protention) and backgrounds in memory (retention). This schema highlights the fact that each discourse has a *discursive depth*, based on what Goodwin (1994, 2018) calls the substrate, i.e., a *reservoir* that is *partly removed from the field of practice in course of realization, but that can be partly re-solicited by a movement of appropriation* and *actualization* (Dondero 2016, 2024). This movement of actualization is to be understood as a selection in relation to everything that has previously been accepted by the community of speakers (virtualized) and that is available and re-utilizable by other speakers.

In a nutshell: The process of *virtualization* covers every discourse that has been concretely sedimented and that is solicitable. *Actualization* concerns the process of passing from the reservoir to action through the recalling of acquired competences and skills, while *realization* is the action of putting something into discourse (*mise en discours*). Finally, *potentialization* is the reverse process which follows realization, that is, a process of putting significations *on standby* so that they may subsequently be *virtualized* (or not: they may disappear, only being a single-use phrase or a single-use strategy in a work of art).

3 Databases and algorithms between image analysis and generation

As I wrote in other texts (Dondero 2019a, 2019b; D'Armenio et al. 2024), databases and algorithms today allow us to have a hold on virtualization: before the availability of big data, virtualization was a very abstract concept and, moreover, we could not test nor operationalize it. This means that it is now possible to rethink and readdress some of the questions having long interested linguistics and semiotics, thanks to databases that may be considered as the concretization and totalization *in a same place* of all major visual cultural productions (and of their formal characteristics). When I write "all cultural productions," I mean all canonized artistic images produced in the Western world. And all these images have been "translated" into lists of numbers, meaning that they are all now readily *comparable and manipulable*. I put forth the hypothesis that Fontanillian virtualization is represented through what, in computer vision, is called the "latent space," that is "the abstract, multidimensional space in which deep-learning algorithms turn digital objects into latent representations so that they can be processed and *used to analyze or to generate* new digital

objects (e.g., new images and new texts). Latent representations are made of vectors; that is, long lists of numbers that define the coordinates of the digital objects encoded and embedded in latent space and their relations of distance and proximity within it, just as the three coordinates x, y, and z define the position of a physical object in three-dimensional space and its relations to other physical objects."[2]

Therefore, we can state that today, the realm of the virtualized (previously Saussurean *Langue*) has been concretized in databases and is available for our manipulation (as analysts and as generators of new images from old ones). As I will attempt to show, we can now better answer the old question having obsessed several linguists and semiologists such as Emile Benveniste,[3] Roland Barthes,[4] and René Lindekens (1976): Does a visual *langue* exist? In other words: Does a global *reservoir* of visual marks exist, which we can consider as an ensemble of distinguishable and differentiated marks able to construct a strict identification on the plane of expression and a meaning on the plane of content when they are assembled as happens in natural language? Let us delve into the core of the problem.

Big data have transformed the practices of many disciplines specialized in the study of images, such as art history, philosophy, aesthetics, and semiotics. As already mentioned, I will consider two manners of developing the current availability of visual big data and their constitution into large collections (for example, large collections of paintings). In fact, not only do databases play a crucial role in *studying the evolution of styles* in the history of painting and for retracing new genealogies of

2 On this topic, see Somaini (2023). Not only is this latent space the one to prepare the action of generation, but it is also the one to allow analysis.

3 According to Benveniste, systems which do not have units and rules to govern their relations would be dependent upon natural language for any description or for any "interpretance." As Benveniste stated in "Semiology of Language" (1981: 16), verbal language is the system for interpreting other signs – and society more generally. For Benveniste, the handicap of non-verbal signs is indeed *the lack of distinct constitutive units* and rules which manage the linguistic system, such as the *rules of selectivity* (on the paradigmatic axis) and *recurrence* (on the syntagmatic axis). The problem, for theorists of the relations between the verbal and the visual such as Benveniste, is the presumed *liberty of non-verbal signs*. Gestures, sounds, and images would all lack a *repertoire*, a system of distinct signs, and syntactic rules to govern their syntagmatic dimension – there would be no grammatical rules that would guarantee the intelligibility of gestural or pictorial statements. Non-verbal signs would be unable to ensure the predictability of their occurrences or their transmissibility through a universally accepted system of notation. This problem also concerns plastic arts: "Therefore, the meaning of art may never be reduced to a convention accepted by two partners. New terms must always be found, since they are unlimited in number and unpredictable in nature; thus they must be redevised for each work and, in short, prove unsuitable as an institution. On the other hand, the meaning of language is meaning itself, establishing the possibility of all exchange and of all communication, and thus of all culture.

4 A discussion on visual *langue* according to the problems raised by Benveniste and Barthes can be found in Dondero (2020).

forms (*analysis*), but they also are at the foundation of some artificial intelligence models serving in the *automatic generation of images*, such as Midjourney and DALL·E (*production*). The following pages are thus firstly devoted to the study of the analysis of large collections of images, by considering crucial past and current research projects in digital art history, and, secondly, to the generation of images through Midjourney, in particular taking into account experiments on pictorial stereotypes.

The first part focuses on the analysis (and visualization) of large collections of images, as they have been conducted in recent years. Before presenting my own project on this matter, I will briefly recall the specificities of three ongoing or recently finished research projects that inspired me to make some observations and critiques, specifically the *Media Visualization* project by Lev Manovich devoted to visual formal similarities, the *Replica* project launched by Benoit Seguin at the EPFL's Digital Humanities Lab in Switzerland on migration of motifs, and the project held by Leonard Impett on the modelization of gestures in paintings (*Totentanz*).

The second part concerns the generation of new images from existing images stored in databases via algorithm models such as Midjourney and on the process leading to new image production, a production based on the sedimentation of stereotypical forms in visual culture.

The two parts are linked together by the theoretical and methodological tool that is enunciation. As already stated, a database such as Wikiart contains a large number of artistic paintings, photos, and drawings previously enunciated and performed and then retained in the history of Western art. Moreover, confronted with a database, we face not the substance of the visual, but a substance made of numbers, namely a substance that presents itself as digital representations manipulable by algorithms. Now, what happens once all the images have been represented into strings of numbers within a database? This database becomes a sort of closed system (*langue*/virtualization) that makes the images (*parole*) commensurable with one another. Consequently, a collection of texts is made co-present and made commensurable through chains of numbers, and manipulable through operations on these chains of numbers.

The commensurability of images contained in databases makes them manipulable by algorithms. But how to describe this commensurability between images? The formal and plastic characteristics of each image (colors, forms, topology, etc.) can be quantified and measured against the characteristics of other images to find a system of structural differences and similarities within the collection. In general terms, each database image can be compared with all the images in the collection on the basis of a chosen number of parameters that have been calculated, such as light intensity, color, contours, and the like. In feature extraction methods, these parameters have to be chosen by the researchers according to a series of analytical objectives on the

selected collection. Conversely, in deep learning realm, the relevant categories are based on the *computation process on a training set* of images (and not on parameters chosen by the scholar). As previously stated, this comparability of images according to various parameters makes every database suitable for use in at least two ways that we will examine in the following pages.

4 The analysis of large collections of images

In this first section devoted to the analysis of large collections of images, I take into account two analytical strategies for image collections: the first is feature extraction, used by Lev Manovich in his Media Visualization Project. This method extracts features from the contents of databases based on rules "manually" predefined by the researcher who dictates the instructions to be followed to perform the task, for example the decision regarding the features to be extracted, such as the gradients of color and luminosity in the paintings by abstract painters of the early twentieth century (Manovich 2020).

The second strategy is deep learning, a learning algorithmic model[5] that provides the machine with datasets through which the machine must spot similarities/dissimilarities. When we use a deep learning algorithm, what we have at hand is no longer the *extension of the eye of the researcher* who decides what the machine should find in the collection of images, as was the case with feature extraction. Instead, what we have is the *extension of the database* used to train the algorithm. By using deep learning, we're letting the algorithm decide what it needs to calculate in order to perform its task in the best way possible. The researcher only gives the model feedback on its result, enabling it to self-correct without telling it exactly what calculations it should have made. Indeed, *the quality of the database will condition the model's ability to learn more or less correctly how to perform its task*. It is clear, for example, that if the algorithm has been trained on a dataset of ordinary images representing everyday objects, it will be very difficult to obtain good results in a

5 Deep learning is a learning algorithm where the general principle is to define a model (usually a neural network) with a set of parameters and to try to find by "gradient descent" the best parameters for performing the task. Progress in the search for these best parameters is guided by a feedback mechanism, often within the supervised learning framework. The fact that learning is "deep" is initially linked to the fact that it is possible to train neural networks with a large number of layers/parameters, and the more layers there are, the "deeper" we go into the network, hence this notion of "depth" (linked to the vast number of parameters we can simultaneously optimize using this technique).

search for artistic pictures.[6] In other words, the dataset on which the algorithm is trained must have sufficient affinities with the database it will later be presented with to enable it to be analyzed in terms of similarities/ dissimilarities.

The tasks that the machine performs "in our place" concern image analysis. Definitely, the machine is asked to do more than we ourselves can do, notably to organize large collections of visual data (thousands of digitized[7] images) according to their similarities/dissimilarities.

The analysis of large collections of data allows to relaunch research projects that could not have been accomplished in the past. I'm referring in particular to the projects of Henri Focillon and above all Aby Warburg. Focillon (1992) theorized the genealogy of forms as a process of bifurcation/ramification/stratification/disappearance of forms in his seminal book *The Life of Forms in Art*, originally published in 1934, without being able to trace a genealogy of forms in its entirety. For his part, in the 1920s Aby Warburg studied images through images in his *Atlas Mnemosyne* (2012 [1924–1929]), using as a means of investigation the visualization of images organized in such a way that images with common characteristics, according to their compositional resemblance, were arranged together within large black panels in order to illuminate the relation "of the nearest neighbor" in compositional terms, and to distance in a graduated manner those that are opposed to or in conflict with its plastic characteristics.[8] This is precisely what was done by researchers such as Manovich (2020) in the early 2000s following this project and ensuring that databases and the algorithms that apply to them can group data according to their similarities and differences following Warburg's "nearest neighbor" rule.

Let us consider the visualization of thousand images conceived of as results of analysis of large collections of images in two recent research projects*:* Media Visualization and Replica. The main difference between the two is that Media Visualization processes images by means of techniques based on the extraction of their low-level features (Figure 2), while the Replica project uses deep learning methods (Figure 3). In the first case, the researcher's biases come into play, whereas in the second, it is the database's biases that are at work. In fact, when using a deep learning algorithm, the researcher's eye is no longer extended, but the database used to train the algorithm is. In the case of the supervised method used by

6 In fact, in general, the scholar starts by compiling the dataset he/she wants to analyze, before extracting part of it, annotating it and using it for training purposes. The rest of the database will be used to check the results. If this annotation of the initial dataset is well done, then the scholar gives him or herself a better chance of obtaining a model that generalizes well to his/her data of interest.
7 I write "digitized," and not "digital," because I'll be focusing on the analysis of large collections of images that belong to the Western artistic heritage, and not on the data produced by digital technology itself, such as all the data we produce every day using social networks, emails, etc.
8 https://warburg.sas.ac.uk/archive/bilderatlas-mnemosyne/final-version.

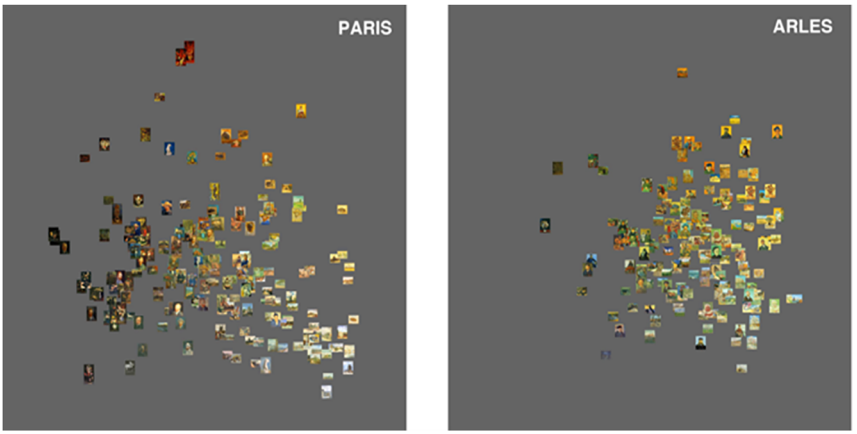

Figure 2: Media Visualization. Manovich and Cultural Analytics Lab. Comparison of the paintings produced by Van Gogh in Paris (left) and Arles (right) with respect to brightness and to saturation. X-axis: brightness average; Y-axis: saturation average (Manovich et al. 2011).

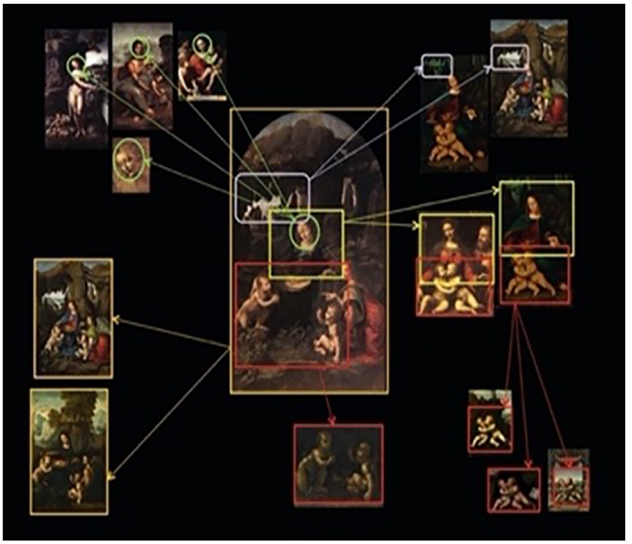

Figure 3: Replica Project. Migration of motifs. See Seguin (2018).

Manovich, the researcher himself defines the characteristics to be calculated by the algorithm, for instance, extracting the brightness and saturation of colors in abstract paintings.

The two projects also share similarities. Firstly, the two projects have in common that they reject metadata such as the date and place of production, the genre, and so on to analyze large collections of images. As stated by Manovich (2015), the main flaw of classification by standard metadata is that the terms used in the classification of images do not reflect the characteristics of visual language, namely the forms of organization *specific to images*.[9] Secondly, both projects produce visualizations of images which function as *meta-images*, that is to say, the visualizations themselves manifest processes of analysis (Dondero 2017, 2019b). Such analysis is to be understood as the result of *mereological operations* of division, grouping, superimposition, etc., mereology studying the relations between a totality (the collection) and its parts (the single images composing the collection) and between the parts themselves.[10]

In the *Replica* project (Figure 3), led by Seguin (2018), the image's analytical visualization, the so-called "morphograph," displays the visual connections which highlight the morphological relationships between a given painting and other similar paintings (*pattern propagations*). More broadly, this project aims at identifying the *repetition* of *recognizable separate visual elements* and/or of iconographical motifs (the kneeling woman, the reclining Venus, and so on). In this sense, the *Replica* Project is more interesting than *Media Visualization* for studying genealogies of visual forms because it tries to trace the path going from an original iconographic motif to the remake/rework of this motif in another painting of a later period.

In the automatic visualizations of images by *Replica* as well, one might recognize the legacy of the panels of images by Aby Warburg. As early efforts to visualize images on the basis of their similarities in terms of forms and of inherent dynamics, Warburg's panels represented a first attempt to understand the *influence of images on images*, through a panel providing a surface helping to apprehend the *graduated relation* between visually close and distant images.

Now turning towards *Media Visualization* (Figure 2), we see that the images under study that share the same features are located in the same areas of the diagram into which they are organized; the ones that do not are placed in other areas. The position of each image within a group of images offers a precise characterization of its plastic properties and shows their relation (gradual distance or/and oppositions) to the qualities of other groups of images. This is an important step for developing the study of corpora. Even if Media Visualization analyzes exclusively what we call in semiotics the *plane of expression* of images (Floch 1985; Greimas 1989),

9 On visual metalanguage, and against the imperialism of natural language as translator of the meaning of images, see Fabbri (1998) and Dondero (2020) who propose images as translators and interpreters of other images.

10 See the semiotic work on mereology by Bordron (2013).

that is, what is *perceptible in an image*,[11] my observation is that this kind of analysis made by operations of division and grouping may be used for the *comparative* study of the *plane of content* of the different groups of images. In a sense, the automatic analysis of the plane of expression can be useful for the discovery of similarities or dissimilarities on the plane of content of images. To reach the objective of analyzing the plane of content as well, it is necessary to observe the different *groups of images distributed* on the surface of the visualization and to question the content characterizing these groups. If the images that belong to a group displayed in the lower-left part of the visualization are considered – concerning their plane of expression – as completely different from the ones situated in the upper-right part of the visualization, what about their plane of content? Do the images belonging to the same group on the plane of expression, belong as well to the same genre (portrait) in contrary to the opposite group (landscape, still life)? Finally, the visualization can become a *map* displaying not places but groups of images defined by their formal features to be used to test their content (genre, for instance). In this sense, tentative corpora can be constituted.

In a sense, we could state that this kind of map visualizes the operations constituting the analysis (division and grouping) previously performed on the image collection. In Figure 2, some images occupy the determinate positions that were available, that is, that were virtualized in the database (for instance, that of maximal saturation in color and lowest intensity in brightness), but some other virtual positions are unoccupied by images (i.e., the blank, grey space). What do these unoccupied positions mean? They mean that some pictorial virtualities were available in the realm of painting at the time but had not been selected and actualized by Van Gogh. Another visualization by Manovich compares the style spaces of Piet Mondrian and Mark Rothko (Figure 4), two painters who evolved from figuration to abstraction during the same period. This figure shows that some (extreme) virtualities had been selected by Mondrian – strong brightness, very week saturation (I refer to the *Composition* of 1913 that is visible at the right bottom corner of his style space) – which, on the contrary, Rothko never conquered (he however came close to doing so during his figurative mythological period, but not as an abstract painter). Conversely, the virtuality of strong color saturation with very weak brightness, as shown in the style space of Rothko, was very rarely realized by Mondrian during his long career.

The two limitations that can be ascribed to the Media Visualization project are firstly that the visual similarities identified in large corpora have little historical relevance, in the sense that they do not produce a genealogy of forms but only a

11 For this reason, this kind of analysis is often violently criticized by art historians and scholars in humanities. See for instance Bishop (2018).

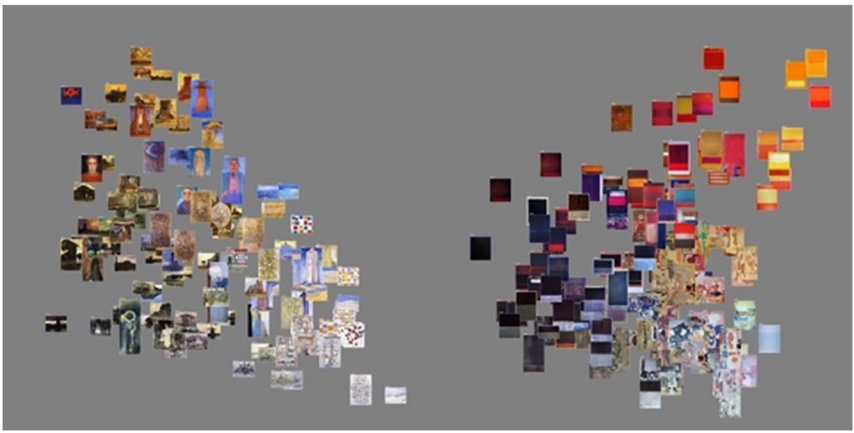

Figure 4: Lev Manovich 2011. Style space: Mondrian and Rothko in contrast.

system of similarities, and secondly, that this method only takes account of *averages* of features (that is, averages in terms of brightness, saturation, and hue), *the problem being that averages do not take account of the actual configuration of the singular image composition*.

Manovich's method identifies low-level features, which Algirdas Julien Greimas's semiotics (Greimas 1989; Floch 1985) calls "plastic properties," identifying three kinds which it has used exclusively for analyzing a singular image or small corpora: *topological categories*, some "rectilinear" (such as upper/lower or left/right) and others "curvilinear" (peripheral/central or enclosing/enclosed), *eidetic categories* (pointed/rounded or contoured/flat), and *chromatic ones* (degrees of saturation of colors).

Unlike Media Visualization, in Greimasian plastic semiotics, the three categories mentioned (topological, eidetic and chromatic) are put in action in image analysis using the structuralist principle of oppositions (upper vs. lower, left vs. right, enclosing form vs. enclosed form, and so on) or using the principle – borrowed from the tensive semiotics of Fontanille and Zilberberg (1998) – of intensity graduation in a categorial scale (from upper to lower light intensity, from a right-situated saturated color to a left-situated unsaturated color, from a light concentration to an intensive shadow, for instance). Through the oppositions and the graduated scale located on the plane of expression of images, the analyst is able to study every image neither as an average of features nor as an accumulation of isolatable objects but *as a singular object made of tensions between opposing characteristics and opposing directions*.

4.1 Analyzing bodily gestures

The third relevant ongoing project in computational image analysis concerns the analysis of the *representation of bodily gestures* in images. On this topic, Impett and Moretti (2017) formalizes the gestures of the bodies found in Warburg's *Atlas Mnemosyne* panels (Figure 5). The attempt is remarkable because it tries to measure and compare the movements of bodies represented in various images in the *Atlas* in order to arrive at the formal description of *Pathosformeln*, the forms of pathos in gestures.

The objective is not to build a genealogy of these gestures but to formalize every isolated gesture to measure the intensity of feelings of each body present in the collections of images selected. My critical observation is that this method reduces the body to a skeleton and so to a movement that is totally abstracted from the rest of the painting and the other bodies represented (humans or objects). In fact, the body conceived of as a geometric figure made of line segments is completely detached from the environment which surrounds it. This also means that this formalization does not take into account that the body is a volume which plays an interconnected role in the whole composition conceived of as a global structure made of an equilibrium of forces.

Impett (2020) tries to complexify the model of the body by inserting the parameters of directionality and the rhythm of moving, as you can see in the following schema (Figure 6).

Through this schema, we see that not only the skeleton has been complexified but also that Impett tries to calculate the rhythm of the movement and the displacement of the body.

This study has been an inspiring foundation for me and for my current research project entitled *Towards a Genealogy of Visual Forms* which has the objective of tracing an innovative genealogy of visual forms inside a more larger image collection

Figure 5: Impett and Moretti (2017).

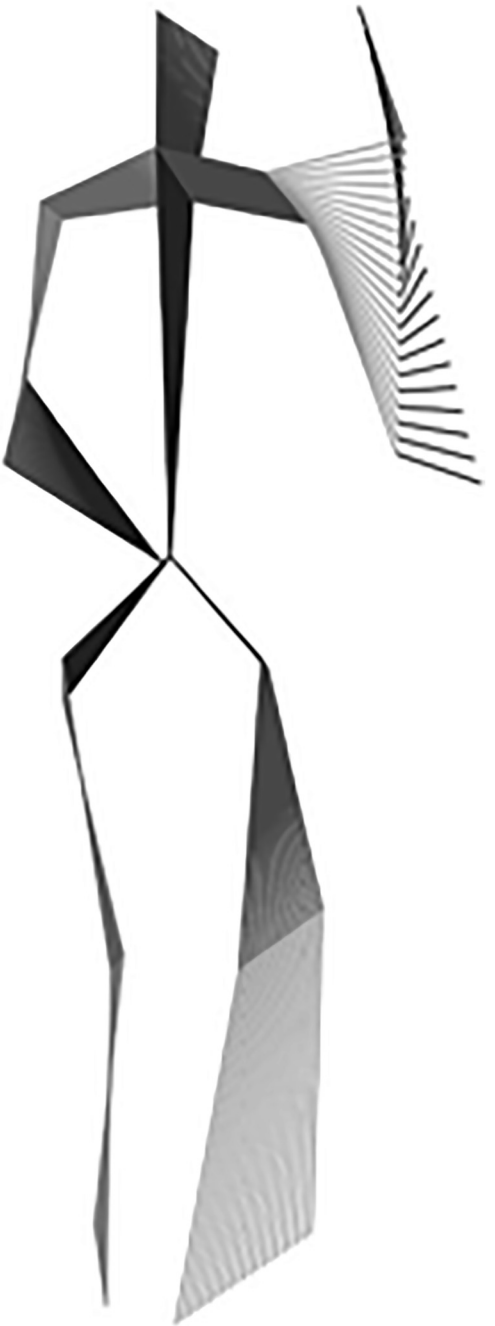

Figure 6: Impett (2020). Principal component analysis on the poses of the Atlas' Panel 46, capturing the strongest morphological feature of the panel concerning the Nymph.

and through other analytical and visualization methods.[12] The project focuses on the relation between contemporary approaches in Image Processing and the genealogy of visual forms in art history as conceived of by Aby Warburg and by Henri Focillon. The main idea of this project is to analyze the genealogy of visual form, from the Renaissance and Baroque painting to the contemporary fashion photography focusing on the gestures and poses represented. It involves a particular conception of forms and forces in images, and strives to make *continuous bodily gestures* analyzable in large collections of paintings and photographs.[13] It aims to trace and cluster poses, gestures, and other kinds of movement and *dynamics of forces* in still images.

What do I mean by "the dynamics of forces within a still image"? The forces in an image can partly be identified with directionality: the direction of a glance, of a raised hand, of a pointed finger, but also with the directions given by components of the image which are not figurative but formal/plastic: the change of the luminosity within a painting works as a kind of arrow, the change in saturation is able to produce a force of elevation or of fall.[14] The geometry of a gesture also counts: a gesture composing a circular figure on the plane of expression reflects stability and calmness on the plane of content; on the contrary, a gesture composing an irregular triangle will reflect disrupting directions, a conflict on the plane of content.

As already mentioned, this project is an attempt at pursuing Warburg's work on the genealogy of forms, particularly the forms of pathos, or *Pathosformeln*, which we can find in the poses and gestures of the figures depicted in paintings throughout the ages. However, this project also joins one of the major current projects of visual semiotics, namely the study of temporality and rhythm in the still image. This involves analyzing how temporal and aspectual deixes are signified by means of a still materiality such as that of a painting. While enunciation concerning the communicational attitude of human subjects represented in pictures[15] has significantly been developed in visual semiotics, with several works on the face view and the side view (Beyaert-Geslin 2017), in addition to spatial enunciation, thanks to

12 This four-year project is financed by F.R.S.-FNRS (2022–2025).

13 This conception finds its roots not only in the French School of Semiotics, but also in the philosopher Gilles Deleuze's work on pictorial diagrams in Francis Bacon's painting (Deleuze 2003; Dondero 2023), and in the mathematician René Thom's idea of conflicting forces in painting (Thom 1983). They all argue that the image is not a matter of the representation of recognizable, isolated objects, but rather a matter of dynamic relations and oppositions between forms (and forces supporting them) present within the surface of an image.

14 An example of plastic oppositions can be seen in Tintoretto's *Tarquin and Lucretia* (1578–1580). This painting shows the oppositions between forces of elevation (Lucretia's body is suspended and pulled upwards by Tarquinio) and forces of fall: the pearls of the necklace are falling down, as are the small sculpture in the right side of the painting and the pillow as well. On this issue, see Dondero and Deliège (2025).

15 On data visualization and enunciation, also see Drucker (2019).

several works on perspective (Fontanille 1989; Marin 1993), temporal and aspectual enunciation has not been sufficiently investigated. By temporal enunciation, I mean that the action represented may be oriented towards the future or towards the past or inhabit the present of the act of observation. But time is an encompassing category which also includes aspectual enunciation. The latter concerns the inchoative or terminative moments chosen by the producer within an action represented in an image, or its durativity. In other words, aspectuality concerns the moment of the action that has been focused on by the painter or by the photographer, as well as the rhythm of the action's unfolding, which may be accelerated or calm.[16]

4.2 Towards a genealogy of visual forms

In the framework of the project *Towards a Genealogy of Visual Forms*, Adrien Deliège and I are working to formalize the human body, considering it not as an isolated object but as part of the complexity of a collective body within the painting, regarded as a whole shaped by a global dynamic of forces.

For that purpose,[17] we collect all religious paintings (11,980 available when conducting this research) of WikiArt. Then, on these images, we run MMPose,[18] an algorithm which estimates bodily poses using a skeleton-like schema of 17 keypoints, and we filter out the images not containing at least one pose having all its keypoints confidently detected. This leaves us with 5,269 images, containing 8,599 individual poses. Each individual pose is redrawn in a separate image, with normalized[19] keypoints coordinates to allow for meaningful comparisons between images. This process is illustrated hereafter (Figure 7).

We first analyze all these individual poses before moving to the collective ones. We define the distance between two poses as the sum of the distances between their corresponding keypoints. We then use this distance metric in the UMAP dimensionality reduction module of the PixPlot software to produce a visualization of the individual poses organization (Figure 8), that is, similar poses are located close to each other, far away from dissimilar poses.

Within this collection, it is possible to follow the variety of poses according to the organization established by the algorithms. As examples, each green circle in the general collection diagram (top left corner of each subfigure of Figure 9) refers to a group of paintings the machine recognizes as belonging to a specific pose via MMPOse, which is zoomed in to show via Pixplot the validity of the approach.

16 On this topic, see Colas-Blaise (2019), Dondero (submitted), Groupe Mu (1998), and Petitot (2004). On experimental computational analysis of temporality in still images, see Deliège et al. (2024).

17 From this point onwards until Figure 13, the text is authored by Maria Giulia Dondero and Adrien Deliège.

18 https://github.com/open-mmlab/mmpose.

19 More technical details are given in our blog post here: https://ceserh.hypotheses.org/3929.

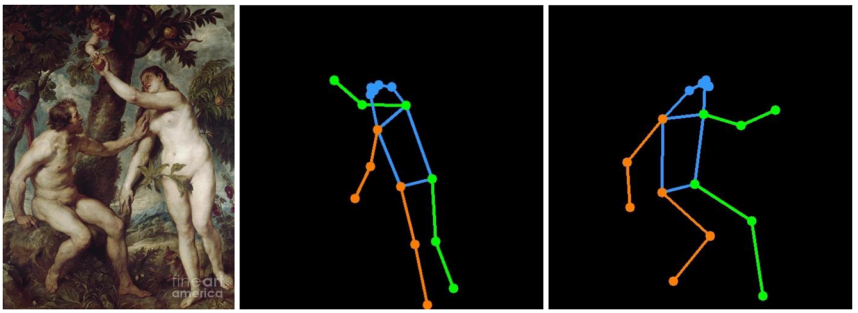

Figure 7: Adrien Deliège and Maria Giulia Dondero 2024. Example of individual poses extracted from an original painting, Peter Paul Rubens, *Adam and Eve in the earthly paradise*, ca. 1628.

Figure 8: Adrien Deliège and Maria Giulia Dondero 2023. Visualization of 8,599 individual poses from 5,269 religious paintings contained in the Wikiart database and organized by pose similarity. An interactive web application that allows navigating through this "image cloud" (with zooming in and out functionalities) is available at https://adriendeliege.z6.web.core.windows.net/outputs/WikiArt_religious_painting_solo_poses/index.html.

It can be observed that the center of the visualization tends to regroup poses that are relatively neutral, depicting a person standing, facing the viewer. As we move away from the center, the poses continuously vary, reaching completely different poses in the corners of the meta-image, such as characters lying down, sitting, falling,

Figure 9: Adrien Deliège and Maria Giulia Dondero 2023. Examples of specific poses shared by certain paintings, circled in green (top left corner) in the general visualization, and zoom in these delimited regions to show the similarity and variations of similar poses belonging to a group of images.

etc. A cluster of representations of Jesus on his cross is also clearly visible, as this pose is common among religious paintings. We can also spot a cluster of images where the character is seen from the back, which is completely, and rightfully, dissociated from the rest of images (on the far left of the visualization). Let us also note that a body lying

down with the head on the left is a completely different pose (according to the metric used) than if the head is on the right of the image. The opposite poses are represented in opposite parts of the visualization, namely top and bottom in this case. Finally, as for every large-scale automated analysis, there are some unfiltered errors that sneaked through this visualization; in this case as a small cluster of poses with legs cut at the knees, corresponding to characters that are not completely shown on the images (on the middle top surrounded by a void neighborhood in the visualization).

The structure of the analytical visualization shows and explores the whole collection, which we call the *reference corpus* (it is all-encompassing); we can select within it a series of (groups of) images that we call the *working corpus*. This group of images can be used to describe the relationship between the plane of expression of images and their plane of content, following a semi-symbolic[20] analysis that establishes that an opposition on the plane of expression is correlated to an opposition on the plane of content of images. Two simple examples could be the following: *arms up vs. arms down = prayer vs. rest*, or *standing body vs. reclined body = arrogant pose vs. pious pose*.

Furthermore, beyond the visualization of the organization of the poses contained in the reference corpus, we can derive a query-and-rank process from our computations. We can submit a query image which may or may not belong to the reference corpus, extract the pose of a character in it, and rank the images of the reference corpus by their pose similarity. However, it should be noted that several possibilities exist to create such rankings. Indeed, with the normalized (location independent and scale-independent) poses, the comparison is made as if the characters where completely detached from their context within the image. An alternative is to keep the scale dependency such that the space occupied by a character in a painting is also taken into account in the comparison, which results in grouping prominent characters together (sometimes at the detriment of the pose similarity itself). To balance the pose similarity with the scale of the pose, we can also produce a third ranking, obtained as an average of the two previous rankings. Along with minor filtering processes,[21] this last option gives good query-and-rank results for a single query pose (exemplified in Figure 10 where the reference corpus is composed of mythological paintings). Finally, a ranking based on the position of the pose within the image can also be envisaged, and combined with the scale (in)dependency of the pose. In our case, we observed lower-quality rankings when using this modality, but still, it might make sense for some specific applications.

20 Semi-symbolism is an analytical tool that has been conceived of by Greimas (1989) and it concerns the relation between categorial oppositions, including the homologation between oppositions on the plane of expression (what is visible/perceivable in an image) and oppositions on the plane of content (its theme or its meaning). For example: "above: below = celestial: terrestrial". See also *supra* about plan of expression and plane of content relation.

21 Detailed in our blog post Deliège et al. (2024): https://ceserh.hypotheses.org/3929.

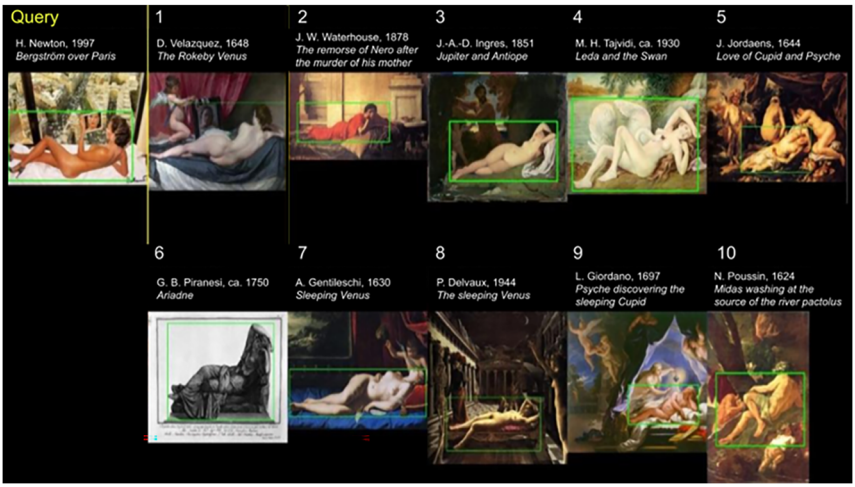

Figure 10: Top left: query image. Right: filtered nearest mythological paintings in term of average combination of scale-dependent and scale-independent rankings.

In order to consider the image as a whole and not only as an ensemble of detached-from-the-background poses, we can extend our study to groups of poses, but this setting comes with additional challenges. A first difficulty is to determine which poses are compared. For instance, if an image contains 4 characters, and another image contains 6 characters, which subgroups of characters should be compared? Generally speaking, it makes sense to compare only subgroups with the same number of characters. In this example, beyond comparing individual poses, we can compare each subgroup of 2, 3, 4 characters of the first image with each subgroup of 2, 3, 4 characters of the second image. While this is theoretically feasible, a combinatorial explosion prevents us from comparing all subgroups of all the paintings. Indeed, if N characters are present, then the total number of subgroups we can form is of the order of 2^N, which grows exponentially with N. Therefore, computational constraints limit us to manageable values of N, for the purpose of illustration, we use at most groups of $N = 3$ poses.

As already outlined in the case of individual poses, several modalities can be compared, related to positions and scales of the poses. So, let us assume that we have two groups of characters from different paintings to compare. How do we compare them? We identified 7 criteria that allow each a unique way of comparing the poses. The first 3 criteria focus on the configuration of the group of poses: the poses are reduced to their average point, these average points give the configuration of the group of poses. We analyze these configurations by looking at (1) its average relative

position in the image, (2) its shape (e.g. triangle pointing upwards) independently of its scale, that is, for example, an upwards equilateral triangle of side 2 and another one of side 5 are both upwards equilateral triangles and are thus equivalent for the analysis, (3) its shape dependently of its scale, such that the two upwards equilateral triangles are not considered equivalent anymore. The next 4 criteria focus on the poses themselves and not on the configuration of the whole group of poses. Again, we can either keep (4)(5) or remove (6)(7) the dependency to the scale of the poses, by letting them as is or normalizing them, just as discussed for the configuration of the group of poses and for individual poses. Besides, when comparing two groups of equally-numbered poses, we must decide which pose of the first group is compared with which pose of the second group. The matching between the poses of the two groups can be done either by matching the poses by their appearance (which pose of the second group is the most similar, as in the individual poses analysis, to which pose in the first group?), or by their localization (which one of the second group is the closest to which one of the first group in term of relative position in the image?). We name the first matching "best pose matching" (4)(6) and the second one "location-based matching" (5)(7). Hence, our 4 pose-based criteria for comparing groups of poses are defined by the 2×2 combinations of scale dependence/independence and best pose matching/location-based matching. In summary, the 7 criteria available are:

1. Configuration comparison: localisation of the group of poses within the image
2. Configuration comparison: shape of the group of poses, scale-dependent
3. Configuration comparison: shape of the group of poses, scale-independent
4. Poses comparison: scale-dependent, best pose matching
5. Poses comparison: scale-dependent, location-based matching
6. Poses comparison: scale-independent, best pose matching
7. Poses comparison: scale-independent, location-based matching

These choices are embodied in Figure 11.

Let us note that this naturally extends the case of individual poses comparisons, where criteria 2 and 3 then become irrelevant since there is no such thing as "shape of the group of poses," and the matching does not matter anymore since there is only one pose to compare with another one, thus criteria 4 and 5 are equivalent, as well as 6 and 7. Thus, one needs to compute only criteria 1, 4, 6 in the case of individual poses.

As in the case of individual poses, each of the 7 criteria can lead to a PixPlot visualization and leads to a ranking of the reference corpus with respect to a query image, as shown in Figure 12. These 7 rankings can be aggregated at will by weighting the ranks as deemed appropriate to produce a final ranking. While many choices can be made, it seems to us that, in the case of individual poses, a combination of equal

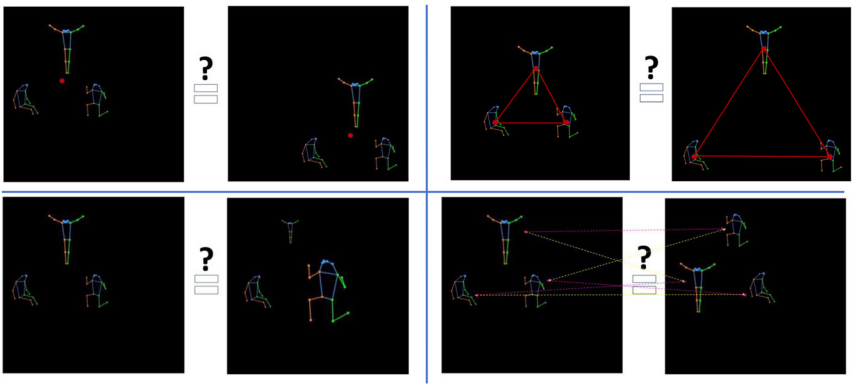

Figure 11: Criteria for comparing groups of poses. Top left: (1) Configuration comparison: should we take into account the localization (red dot) of the group of poses within the image? Top right: (2) and (3) Configuration comparison: should we take into account (2) or not (3) the scale of the shape formed by the group of poses (in this case the red triangle)? Bottom left: (4), (5), (6), (7) Poses comparison: should we take into account (4, 5) or not (6, 7) the scale of the individual poses? Bottom right: (4), (5), (6), (7) Poses comparison: should we compare poses that are the most similar (best pose matching, yellow arrows, 4, 6) or that are located at the most similar place in the group configuration (location-based matching, pink arrows, 5, 7)?

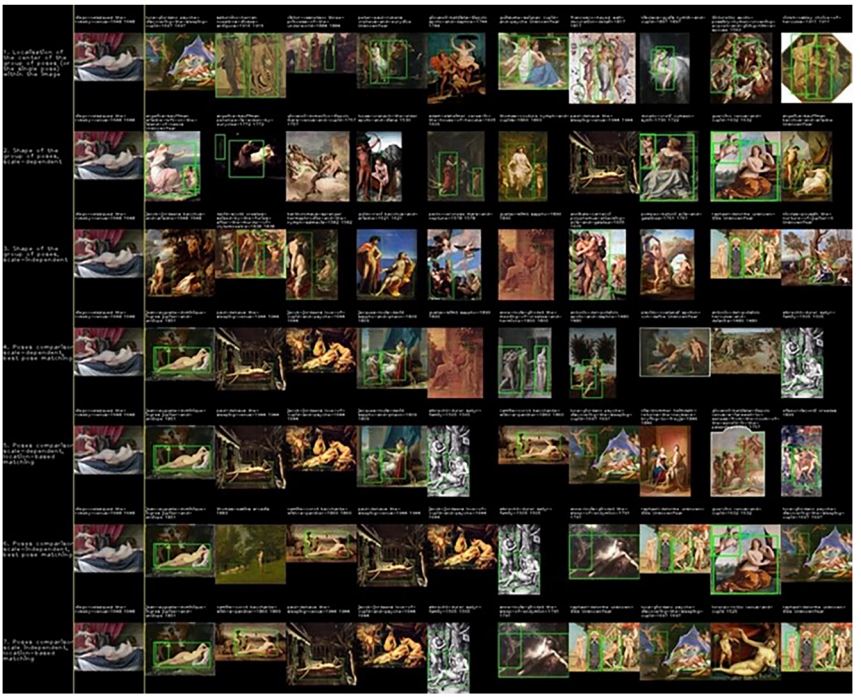

Figure 12: Adrien Deliège and Maria Giulia Dondero 2023. View of parameters organizing pose similarities. The query image is Velazquez' *The Rokeby Venus*. Two characters are detected: the Venus, and the Angel. Thus, groups of 2 poses are compared, each line corresponding to one of our 7 criteria.

weights of criteria 4 and 6 (pose comparisons, with scale-dependence and independence, equivalent to 5 and 7) is appropriate, as already shown previously. This prevents putting too much emphasis on poses that look very different but have the same scale as the query image (which acts thus in its favor in the ranking while being not relevant) as well as putting too much emphasis on poses that look very similar to the query pose but are way too different in term of scale/prominence in the image (which may thus convey a different interpretative meaning). Besides, criterion 1 (the position of the pose in the image) does not seem that useful to us, as we believe desirable to consider as similar poses that are the same, with the same scale, but at different positions in the image. In the case of multiple poses, the same motivations regarding the scale of the poses lead us to consider criteria 5 and 7, with a location-based matching preferred over a pose-based matching. Indeed, we believe that, if the same poses are permuted within the pose configuration, this yields a different image, which is not captured by the pose-based matching. We also consider criteria 2 and 3 in the mix, to incorporate the shape aspect of the configuration in the final comparison, with both of them equally weighted for similar reasons as those motivating the choice of 5 and 7. We neglect criterion 1 again for the same reason as previously. This gives the image shown in Figure 13.

We can wrap this pose analysis by reflecting on one of our initial questions, about assessing motion in paintings. More precisely, how much motion is conveyed within poses of the characters? We could argue that some poses definitely embody

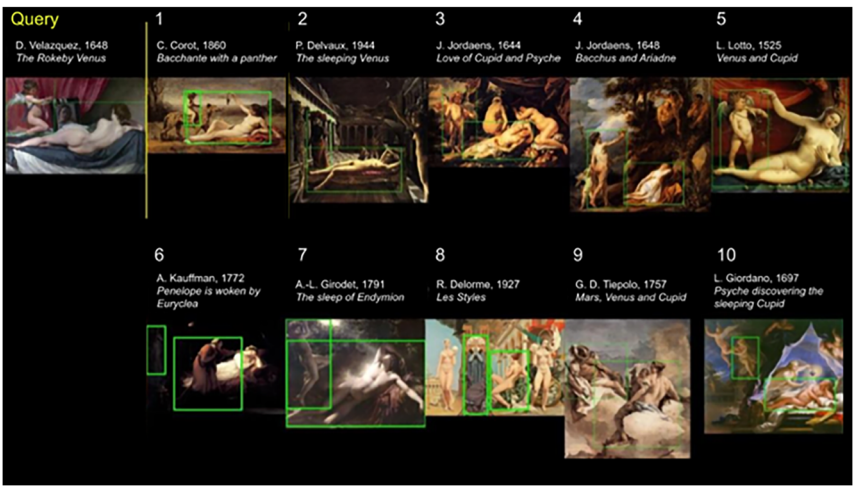

Figure 13: Ranking among the reference corpus "Mythological paintings," by comparing subgroups of 2 poses against the pair Venus-Angel of Velazquez.

some dynamicity, especially when said poses differ heavily from the most common "calm" poses such as standing, sitting, lying down. Using the PixPlot visualization and the distance metric defined previously, we could select representative "calm" poses and define how different from these poses another pose should be to be called "agitated," or "in motion." However, some seemingly calm movements can in fact convey a sense of motion by their relation with other movements, which makes it difficult to make a clear cut between calm and agitated motions. Finally, the poses computed, even when considered "agitated," do not bring much information about the directionality of the motion, the character that perpetrates it, and the type of motion. Therefore, despite being an informative element about motion in paintings, finer analyses or different techniques are needed if one wants to fully characterize motion in images.[22]

5 Generating images from databases

As mentioned in the introduction of this text, the compatibility of images with strings of numbers and their manipulability through algorithms enable the automatic generation of images. In fact, databases and artificial intelligence methods are used not only to analyze collections, but also to produce new images from all the images stored and annotated according to style, author and genre in available databases.

The processes of image generation have been in fact enabled by three major elements: (1) the creation of large databases of images accompanied by textual descriptions (which result from the analyses of these images), (2) the emergence of efficient feature extraction methods trained by deep learning algorithms, (3) the development of powerful and specialized hardware for model training and inference.

In this section, I will describe the operations realized by AI generative models. Image generators such as Midjourney or DALL•E can produce images from descriptions via natural language, and the opposite is also possible: to obtain a description of a given image in verbal language. All these operations start with a more fundamental kind of translation: that of images into numbers and verbal texts into numbers, resulting in lists of numbers known as "embeddings" (Figure 14).

Image generation models (diffusion models) use a "large language model" component, or at least a model that "understands" natural language (e.g., CLIP), in

22 For this purpose, in another text on a different corpus (Deliège et al. 2024), one composed of futurist paintings (Giacomo Balla, Umberto Boccioni, Gino Severini) and of paintings belonging to the "Return to Order" art movement (Felice Casorati, Achille Funi, Mario Sironi), we use edges computing to measure the dynamicity in still images.

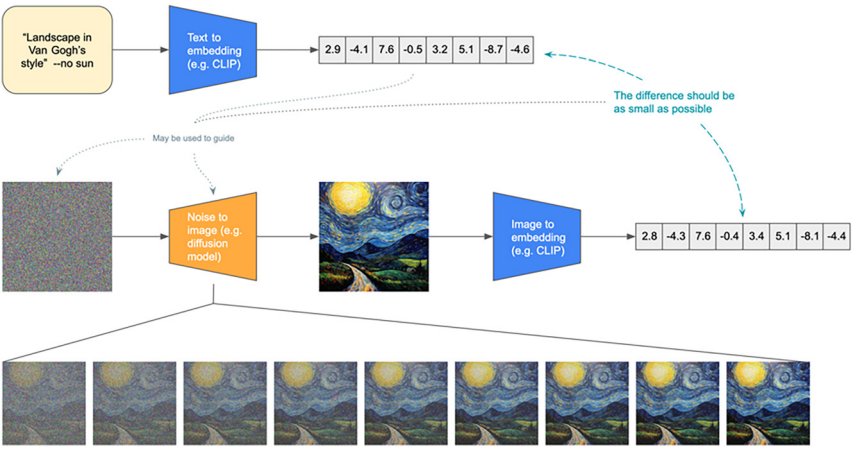

Figure 14: Adrien Deliège 2023. Translation between the embeddings of verbal texts and of images.

order to transform prompts into embeddings (lists of numbers) that can be used by the machine. These learning models, which enable the translation between verbal and visual languages (coordination between two embeddings), are determined by the organization of the database contents.

As Somaini (2023) explains, the diffusion model operates in two phases: "forward diffusion" and "reverse diffusion." The first phase, forward diffusion, turns the images of the dataset used for training into indistinguishable "noise images" (i.e., random pixels distributed within the grid-like image space). It does so by recursively adding a bit of "noise" to the image (i.e., the image becomes more and more noisy as noise continues to be added). During the second phase, the phase of reverse diffusion, starting from the "noise images," the model learns how to recover the initial images of the training set. It does this through a "noise predictor" that learns to predict how much noise was added to the initial images. This allows the model to subtract, for each given "noise image," the layers of noise that were added until it finds the initial image. This double process trains the machine to overcome some obstacles (the noise) in order to learn how to reconstruct the composition of an image through prompts that guide the generation process.

What is important to me is that the new images are generated through operations performed on all the images already produced, stored, and annotated according to style, author, and genre, within available databases such as Wikiart, Artsy, Google Art and Culture, etc. These digitized stocks include all the most famous[23] styles and

23 In order to test the machinic capabilities, I started with very famous painters and artists to see how each one is more or less plastic than the other, and the conclusions to be drawn are obvious: it is

authors in the history of art and contemporary photography. So, we can test what stereotypes of famous painters the database has learned, and test combinations of styles that reveal, at least in part, how algorithms work not only on the translation between verbal and visual languages, but also on the modification/stabilization of past styles through the machine's work. As stated by Wilf (2013: 186) in a paper inspired by Peircian semiotics written ten years before the diffusion of ChatGPT and Midjourney, entitled "From Media Technologies That Reproduce Seconds to Media Technologies That Reproduce Thirds: A Peircean Perspective on Stylistic Fidelity and Style-Reproducing Computerized Algorithms," although these generative systems.

> much like a CD or an MP3 file, are technologies of reproduction, they do not reproduce specific texts, or Seconds, but styles, or Thirds. *Their object of reproduction is the principle of generativity* that is responsible for producing the specific texts that are the object of reproduction of the kind of media technologies that have traditionally stood at the center of linguistic and semiotic anthropological research. These systems' reproduction of style consists both in their ability to *abstract a style from a corpus* of Seconds and to generate new and different Seconds or texts in this style, indefinitely so. (Wilf 2013: 186, my emphasis)

It is without a doubt only through the forms already known and settled in our cultural perception that it is possible to understand the work of the machine – not only the degree of its much-questioned "creativity," but also the way in which it transforms the forms we know by making them into *averages*.[24] But there are two things that this process of transformation of styles into averages of multiple singular images does not prevent: the first is that various averages can produce new forms, as demonstrated by many competitions won by AI-generated images (which also allows us to value the part of the aleatoric that accompanies all image generations); the second is that, although the machine can extract and mimic various styles, the hand of the machine always remains visible. Thus, we can study the *enunciative opacity*

---

almost impossible to extract Klee's style but very easy to extract Van Gogh's (art gadgetization in museum bookshops and in fashion have already shown this functioning). Of course, testing only great painters may be questionable, as Schneider (2022) points out regarding automatic traditions and automatic text generation. In fact, only texts from a certain section of society and only in certain languages are used and thus, their influence increases, while the presence of minority languages and expressions becomes even smaller: "This does not necessarily happen because of a desire of actors in the digital language industry to enforce social and sociolinguistic hierarchies but is coproduced by the affordances and materialities of digital language technology, as machine learning tools require a predefined data set to be trained and amplify what is dominant" (2022: 381). In this preliminary part of my analysis on the opacity of the hand of the machine, using famous artists was necessary.

24 We could state that the machine operates by averages accompanied by the aleatoric, and that on the contrary, a painter will use a diagrammatic device to produce something new out of past forms. On diagrammatic production, see Deleuze (2003) on Bacon's paintings and Dondero (2023) for a comparison between the diagrammatic thinking in the works of Peirce, Deleuze, and Goodman.

(Marin 1993)[25] of the hand of the machine due to the fact that we know the reference styles on the basis of which the machine works, and we can observe the deformation of these known styles due to the specificity of the model, that has its own style. In this sense, the enunciative opacity of each generative model is recognizable through comparisons with other generative models, for instance Stable Diffusion or DALL•E.[26] The visual results of a prompt, especially if the latter is repeated many times, manifest a stylistic tendency of the model that is executing them. Partly, this style is determined by the region of the database that is solicited by the prompt, that I call "regional style." Indeed, this local style is determined by the annotations of images enclosed in the database that a prompt actualizes. But, transversal to the different regions of databases, it is possible to recognize a recurrent style which is specific to each model and that I call the "global style" of a model. It identifies the style of the model and is transversal to every kind of prompt and database zone solicited.

More generally speaking about the concept of style, Roland Meyer states that:

> Historical as well as contemporary, collective as well as individual forms of representation can seemingly be detached at will from their time and place of origin and the work of their authors. It is not least for this reason that O'Reilly and others speak of plagiarism. More importantly, the logic of the prompt radically expands and de-hierarchizes the notion of style … *Thus 'style' ceases to be a historical category and becomes a pattern of visual information to be extracted and monetized.* (Meyer 2023: 106–107, our emphasis)

I will return on this problem of style in the following pages but for now let us examine the process that leads us to generate images. When an instruction is given to the Midjourney platform (prompt), four visual translations are by default obtained through the production of four original[27] images (which can be understood as different *optimizations* of the instruction given), each being differentiated according to luminosity, colors, the positioning of the objects, and so on. The experimenter can choose the version that suits him or her best and decide to continue searching for the image believed to be attainable by giving further instructions: he or she can modify the prompt serving as input or the image through which the quest will continue. Production can thus be described in terms of *decision operations* and transformation requests realized through verbal and visual instructions supported by the system of embeddings that correlate them. Indeed, visual instructions, not contained in the prompts, are also possible: Midjourney introduced a tool for refining a part of the image already produced, by allowing the experimenter to select a part of the image

25 Enunciative opacity, according to Marin, can be defined as the way in which the representation not only represents something of the outer world, but also presents itself to the observer as a representation.

26 For a comparison between Midjourney and DALL·E functioning, see D'Armenio et al. (2024).

27 For a general discussion about originality, fake and machinic creativity, see Leone (2023).

through the "Vary Region" command. This function allows to circle/highlight the part to be modified and to enter a prompt corresponding to what the user wants to see appear for this part of the image. For instance, the user can select an empty part of the image and request the addition of a visual object in order to improve the overall composition. If the user aims at localized modifications in an image, this kind of instruction is much more efficient than propositing a new modifying a prompt. It is also a tool which serves to minimize the statistical bias or the aleatory process at the basis of every computational modification within Midjourney.

In addition, the experimenters, if they are programmers, can decide to fine-tune an existing neural network through annotations, building finer correspondences between the lists of numbers that identify natural language descriptions and the lists of numbers that identify images. Another way of limiting the machine's automatisms and minimize the bias or noise produced by overly generic databases, is not only to indicate the styles of one or two artists, or to fine-tune the network, but also to explicitly indicate to the machine the technique to be used, such as "chalk drawing," "oil painting," "fresco," and so on.

Let us consider some examples of this functioning. If we ask Midjourney to generate stereotyped images of Van Gogh via the prompt: "A landscape in Van Gogh's style," we realize that it is difficult to get rid of particular objects, including the sun and the moon (Figure 15).

This is because, based on the correspondence between image embeddings and image descriptions that have been coded, the sun is probably considered a predominant feature in Van Gogh's work. A first, perhaps naive attempt to make the sun disappear is to add to the prompt the words "without sun, without moon" (Figure 16). We can see that the images produced keep the sun (or the moon – it is hard to tell), as Midjourney isn't designed to really "think" about the meaning of the prompt, nor to distinguish between the positive and negative meanings of our requests. As stated in the Midjourney documentation, a word that appears in the prompt is in fact more likely to end up represented in the image.

Midjourney seems for the moment (experiment made with Midjourney 4) incapable of reasoning in a meta-semiotic way, namely to operate a negation of an element expressed in the prompt. In order to eliminate an element, the user must use the special command "–" (no sun, no moon; Figure 17).

If we repeat the prompt "Landscape in Van Gogh style," interestingly, the results are very always different, but what counts for me is that each set of images produced are from a scientific point of view (the one I am interested in, in contrast to the aesthetic one) more interesting as *samplings of regions of the dataset* than simply as isolated images (Figure 18).

In other words: the images produced by the generative models count more as extractions of features typical of a region of the dataset – this is why I called this style

Figure 15: Maria Giulia Dondero, Enzo D'Armenio, and Adrien Deliège 2023. Midjourney. Prompt: "A Landscape in Van Gogh's style."

resulting from a region of a data base "regional style." A region of a dataset encompasses examples of patterns produced by the work of algorithms *exploring* (Meyer 2023) certain domains of the dataset decided by annotations and operationalized by embeddings – definitive correspondences between certain words and certain shapes do not exist in this dimension.

Midjourney experiments are especially important for testing not only the stereotypes of various famous painters but also to reflect on the idea of composition that the machine develops. This is possible when the user mixes the styles of different painters: several interesting situations concerning compositionality then emerge. Lev Manovich's experiments are crucial in this respect. The next figure (Figure 19) shows the result of mixing Bosch and Malevich. Bosch's figures change according to

Figure 16: Maria Giulia Dondero, Enzo D'Armenio, and Adrien Deliège 2023. Midjourney. Prompt: "landscape in Van Gogh's style *without sun, without moon*."

the positions they occupy within the landscape, whose coordinates are given by Malevich-inspired geometries.

In the case of another experiment by Manovich, which mixes Brueghel and Kandinsky (Figure 20), it can be argued that the machine uses abstract artists such as Malevich and Kandinsky as landscape artists offering the overall topology of the image, which hosts the figures of painters such as Bosch and Brueghel who, in the machine's perspective, are painters of small characters (yet traditionally considered landscape artists themselves!).

Together with Adrien Deliège and Enzo D'Armenio, we experimented various mixings of different painting styles. The results are either irritating or amusing, as in

Figure 17: Maria Giulia Dondero, Enzo D'Armenio, and Adrien Deliège 2023. Midjourney. Prompt: "landscape in Van Gogh Style" *–no sun, no moon*.

the case of the mix between Leonardo's *Mona Lisa* and the prompt: "Mona Lisa in manneristic Pontormo style" (Figure 21).

It is evident from these results that Pontormo's database contains collective scenes (or paintings annotated as "collective scenes") and not portraits; in fact, the Mona Lisa is surrounded by various characters (even more numerous when zooming out).

Next, we tried to "coalesce" the style of Leonardo and Rothko because the two painters, while separated by a few centuries, were recognized as specialists in atmospheric perspective, a technique that builds up the depth of the landscape through the gradual addition of a mist effect. Some of the results are irritating, as in the case of the image where a rectangle of Rothko's color is banally superimposed over the Mona

Figure 18: Maria Giulia 2024. Midjourney. Prompt: "Landscape in Van Gogh Style."

Figure 19: Lev Manovich and Emanuele Arielli 2021–2024. Midjourney. Prompt: "Painting by Malevich and Bosch."

Lisa, but the results are more interesting when Rothko's strata of color, sometimes verging on transparency, are superimposed onto the atmospheric perspective of Leonardo's landscape. Note that in all four images, the addition of blurring and

Figure 20: Lev Manovich, fall 2022. Midjourney. Prompt: “Painted by Brueghel and Kandinsky.”

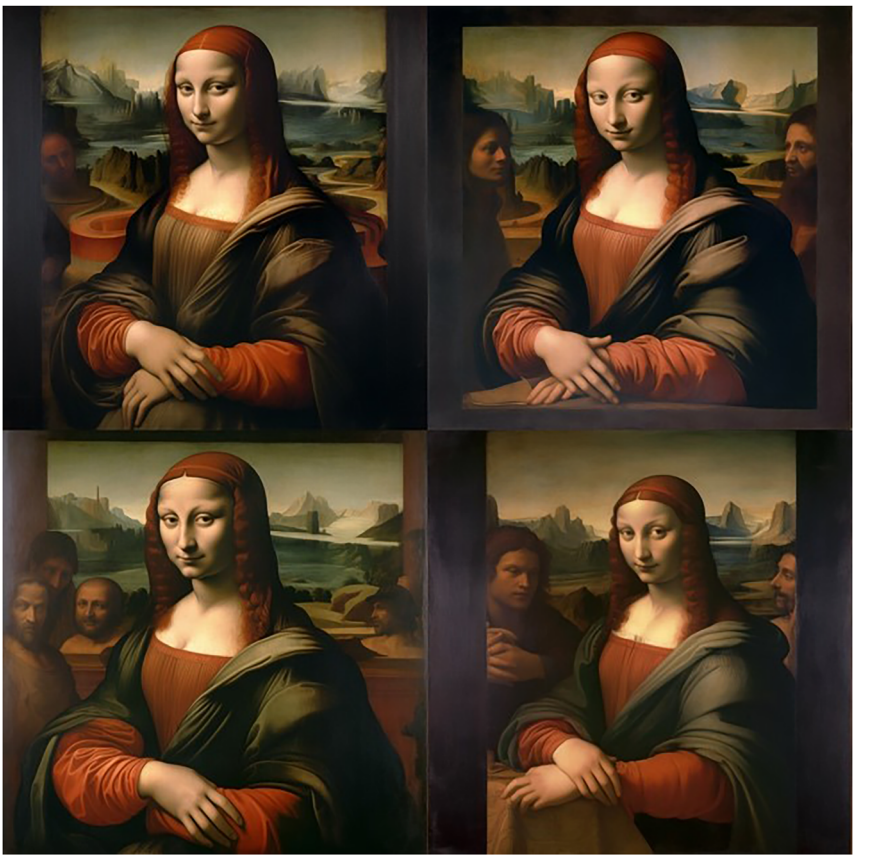

Figure 21: Maria Giulia Dondero, Enzo D’Armenio, and Adrien Deliège 2023. Midjourney. Prompt: “Mona Lisa in manneristic Pontormo style.”

transparency onto the image's outlines turns Leonardo's landscape from vague to sharp, causing it to resemble the hyper-realist American paintings of the 1970s. Is Midjourney programmed to always balance the vague and the sharp, the blurred and the detailed?

In the end, it appears that the machine replicates the style of each painter. In the case of Van Gogh, for example, Midjourney uses the painter's typical textures and mimics a sensori-motricity that is quite similar to the rhythm of his touch. At the same time, the machine itself possesses a sort of standard style, namely *a certain opacity of the hand* (a sort of average hand) that is Midjourney's own, and which seems to be akin to the American pictorial expressionism of the 1970s.

It is only by producing a multitude of images by varying styles or production techniques, and by iterating our requests using slight variations in the prompts or by using the automatically generated prompts of GPT-4 that it will be possible to have an overview on the annotations linked to styles and to understand the virtual mathematical space behind these productions. Starting from this multitude of generated images, it will be possible to make hypotheses about the database Midjourney has been trained on, and thus, about its (undisclosed) model.

6 Final remarks

In these final lines, I'd like to return to the aforementioned concept of enunciative praxis that makes it possible to understand the relation between singular images and image databases, both in the case of analytical visualizations, as already seen, and in the case of current automated image and text generation.

The theory of enunciative praxis is useful for understanding the cultural process of production of new forms and their later stabilization/sedimentation/disappearance. This theory can be operationalized and made methodologically beneficial through its modes of existence (actualization, realization, potentialization, and virtualization). As already stated, the database is constituted by all the available images produced, digitized, and recognized in Western culture and it coincides with the moment of virtualization as it encompasses all such images that are available to be studied or to be mixed and reproduced. Actualization concerns the possibility of production, that is, the skill to produce a new image (in the case of Midjourney, the functioning of a database in relation to and to the prompt algorithms, and the good correspondence between embeddings). The realization coincides with the images generated by Midjourney that stem from the mixing of past productions. In other words, the prompt that engages the translation between embeddings can be seen as an actualization, which is realized in the images generated. As far as potentialization is concerned, the graphic statements generated through our prompts will not

immediately (and perhaps never) be repeated and accepted into the database, which, in the case of Midjourney, is stabilized and encompasses images that have a history, contrary to Stable Diffusion that has an open-source and an ever-changing database.[28] In the case of Midjourney, we have to become recognized artists for our images to be able to take part in its database and to participate in the transformation of what is now sedimented, thence becoming virtual possibilities in the generation of new images.

Research funding: This work was supported by Fonds De La Recherche Scientifique - FNRS (T.0065.22 - R.FNRS.5453 entitled â€œPour une gÃ©n).

References

Asif, Agha. 2003. Social life of cultural value. *Language & Communication* 23(3–4). 231–273.
Asif, Agha. 2005. Voicing, footing, enregisterment. *Journal of Linguistic Anthropology* 15(1). 38–59.
Benveniste, Émile. 1981. The semiology of language. *Semiotica* 37. 5–23.
Beyaert-Geslin, Anne. 2017. *Sémiotique du portrait. De Dibutade au selfie*. Louvain-la-Neuve: De Boeck Supérieur.
Bishop, Claire. 2018. Against digital art history. *International Journal for Digital Art History* 3.
Bordron, Jean-François. 2013. *Image et vérité : Essais sur les dimensions iconiques de la connaissance*. Liège: Presses Universitaires de Liège.
Colas-Blaise, Marion. 2019. Comment penser la narrativité dans l'image fixe ? La "composition cinétique" chez Paul Klee. *Pratiques*. 181–182. http://journals.openedition.org/pratiques/6097 (accessed 10 January 2025).
D'Armenio, Enzo, Adrien Deliège & Maria Giulia Dondero. 2024. Semiotics of machinic co-enunciation: About generative models (Midjourney and DALL·E). *Signata* 15. https://doi.org/10.4000/127x4.
Deleuze, Gilles. 2003. *Francis Bacon: The logic of sensation*. London: Continuum.
Deliège, Adrien, Maria Giulia Dondero & Enzo D'Armenio. 2024. On the dynamism of paintings through the distribution of edge directions. *Journal of Imaging* 10(11). 276.
Dondero, Maria Giulia. submitted. *Enunciación temporal en imágenes fijas*. Mexico: Tópicos del seminario.
Dondero, Maria Giulia. 2016. L'approche sémiotique de Charles Goodwin: langage visuel, énonciation et diagramme. *Tracés* 16. https://doi.org/10.4000/traces.6548.
Dondero, Maria Giulia. 2017. The semiotics of design in media visualization: Mereology and observation strategies. *Information Design Journal* 23(2). 208–218.
Dondero, Maria Giulia. 2019a. Visual semiotics and automatic analysis of images from the Cultural Analytics Lab: How can quantitative and qualitative analysis be combined? *Semiotica* 230(1/4). 121–142.

28 In his paper, Somaini (2023) explains that some models, such as Stable Diffusion, that have been trained through large datasets such as LAION-5B, could also be fed with the databases of all the images produced by the users.

Dondero, Maria Giulia. 2019b. De l'énonciation et du métavisuel dans le cadre des Big Data. In Estay Stange Verónica, Pauline Hachette & Raphaël Horrein (eds.), *Sens à l'horizon ! Hommage à Denis Bertrand*, 285–298. Limoges: Lambert Lucas.

Dondero, Maria Giulia. 2020. *The language of images: The forms and the forces*. Cham: Springer.

Dondero, Maria Giulia. 2023. The experimental space of the diagram according to Peirce, Deleuze, and Goodman. *Semiotic Review* 9. https://semioticreview.com/ojs/index.php/sr/article/view/79 (accessed 14 July 2024).

Dondero, Maria Giulia. 2024. The face: Between background, enunciative temporality and status. *Reti, linguaggi, saperi* 1. 49–70.

Dondero, Maria Giulia & Adrien Deliège. 2025. The semiotic and computational analysis of represented poses in painting and photography. In Pietro Conte, Anna Caterina Dalmasso, Maria Giulia Dondero & Andrea Pinotti (eds.), *Algomedia: The image at the time of artificial intelligence*. Cham: Springer.

Drucker, Johanna. 2019. *Visualization and interpretation: Humanistic approaches to display*. Cambridge, MA: MIT Press.

Fabbri, Paolo. 1998. *La svolta semiotica*. Roma & Bari: Laterza.

Floch, Jean-Marie. 1985. *Petites mythologies de l'œil et de l'esprit. Pour une sémiotique plastique*. Paris & Amsterdam: Hadès-Benjamins.

Focillon, Henri. 1992. *The life of forms in art*. New York: Zone.

Fontanille, Jacques. 1989. *Les espaces subjectifs. Introduction à la sémiotique de l'observateur*. Paris: Hachette.

Fontanille, Jacques. 2003. *Semiotics of discourse*. Berlin: Peter Lang.

Fontanille, Jacques & Claude Zilberberg. 1998. *Tension et signification*. Liège: Mardaga.

Goodwin, Charles. 1994. Professional vision. *American Anthropologist* 96(3). 606–633.

Goodwin, Charles. 2018. *Co-operative action*. Cambridge: Cambridge University Press.

Greimas, Algirdas J. 1989. Figurative semiotics and the semiotics of the plastic arts. *New Literary History* 20(3). 627–649.

Groupe, Mu. 1998. L'effet de temporalité dans les images fixes. *Texte* 21–22. 41–69.

Impett, Leonardo. 2020. Analyzing gesture in digital art history. In Kathryn Brown (ed.), *Routledge companion to digital humanities and art history*, 386–407. New York: Routledge.

Impett, Leonardo & Franco Moretti. 2017. Totentanz. Operationalizing Aby Warburg's Pathosformeln. *Literary Lab Pamphlet* 16. https://litlab.stanford.edu/LiteraryLabPamphlet16.pdf (accessed 10 January 2025).

Leone, Massimo. 2023. The spiral of digital falsehood in deepfakes. *International Journal for the Semiotics of Law* 36. 385–405.

Lindekens, René. 1976. *Essai de sémiotique visuelle. Le photographique, le filmique, le graphique*. Paris: Klincksieck.

Manovich, Lev. 2011. *Style Space: How to Compare image sets and follow their evolution*. https://manovich.net/index.php/projects/style-space (accessed 22 January 2025).

Manovich, Lev. 2015. Data science and digital art history. *International Journal for Digital Art History* 1(1). 3–35.

Manovich, Lev. 2020. Computer vision, human senses, and language of art. *AI & Society* 36. 1145–1152.

Manovich, Lev & Emanuele Arielli. 2021–2024. *Artificial aesthetics: Generative AI, art and visual media*. Manovich Web Site. https://manovich.net/index.php/projects/artificial-aesthetics (accessed 22 January 2025).

Manovich, Lev, Jeremy Douglass & Tara Zepel. 2011. How to compare one million images? In David Berry (ed.), *Understanding digital humanities*, 249–278. London: Palgrave Macmillan.

Marin, Louis. 1993. *De la représentation*. Paris: Seuil.

Meyer, Roland. 2023. The new value of the archive: AI image generation and the visual economy of "style." *IMAGE* 19(1). 100–111.
Petitot, Jean. 2004. *Morphologie et esthétique*. Paris: Maisonneuve et Larose.
Schneider, Britta. 2022. Multilingualism and AI: The regimentation of language in the age of digital capitalism. *Signs and Society* 10(3). 362–387.
Seguin, Benoit. 2018. *Making large art historical photo archives searchable*. Lausanne: École Polytechnique Fédérale de Lausanne dissertation.
Somaini, Antonio. 2023. Algorithmic images: Artificial Intelligence and visual culture. *Grey Room* 93. 74–115.
Thom, René. 1983. Local et global dans l'œuvre d'art. *Le Débat* 2(24). 73–89.
Warburg, Aby. 2012 [1924–1929]. *L'atlas mnémosyne*. Paris: Éditions Atelier de l'écarquillé.
Wilf, Eitan Y. 2013. From media technologies that reproduce seconds to media technologies that reproduce thirds: A Peircean perspective on stylistic fidelity and style-reproducing computerized algorithms. *Signs and Society* 1(2). 185–211.