

Signata

Annales des sémiotiques / Annals of Semiotics

14 | 2023

Pour une nouvelle herméneutique des formes symboliques

Nouvelles technologies et formes symboliques comme clés poïétiques et herméneutiques

The (theoretical) elephant in the room

Overlooked assumptions in computer vision analysis of art images

Présuppositions négligées dans l'analyse des images d'art par la vision par ordinateur

CAMILLA BALBI AND ANNA CALISE

<https://doi.org/10.4000/signata.4757>

Abstracts

English Français

Contemporary computer vision software represents an incredible opportunity for both art history researchers and museum practitioners: it is a tool through which images can be described, organized, studied and shared. In this process—the one in which a computer vision software operates over a database of art history images—there are however a variety of dynamics at play. They have to do with theoretical assumptions, historical categories, technological constraints and ideological stances: a set of premises which calls for a closer methodological survey of the process. We propose an account which uses art theory and visual culture studies to scrutinize the different steps and activities which constitute the computer vision analysis: after all, the study of images has historically been a prerogative of art historians. Our intuition is that art images databases somehow provide a “protected environment” in which to observe how old problems, inherent to the discipline, interact with new problems created by the way we consume and design software. The three levels at which we will try to detect biased stances answer three different questions. Which images are we talking about? Which research questions are we asking? Which linguistic and political logics are at play? In order to do so, we will begin the discussion by debunking the myth of a simple parallelism between these new forms of conceptualizing the real and traditional ones, challenging Manovich’s (1999) use of Panofsky’s symbolic form (1927) as a hermeneutic of the database. We will show instead how the art-database logic somehow sticks to the traditional art historical narrative, while at the same time producing new kinds of biases. Then, we will focus on how this technology actually works, and which kind of art historical thought lays behind the algorithm. Our guess is that the praxis of this software is closer to the connoisseurship than to the art historical research. Thirdly, we will analyze the labeling process through which computer vision software creates descriptive metadata of the images in question, using Mitchell’s critical iconology (1994) account to problematize the strong ideological and political stance behind the image-text relationship. Throughout the discourse, and especially in the final paragraph, we will address the transparency and evaluation standards which need to be defined in order to allow a



strict methodological approach to guard and guide the process, at times lacking both in the cultural sector and in the wider visual field. What will emerge is an account of computer vision software and processes which appear to be far from ‘neutral’ or ‘objective’ in their extremely layered functioning, built in the midst of diverse stakeholders’ interests and procedural false steps. Granted that these technologies are however contributing to build the visual culture of our time, we detect a series of overlooked assumptions along the way through the lenses of art theory, hoping to contribute to the design of a clearer view.

La vision par ordinateur contemporaine représente une opportunité incroyable pour les chercheurs en histoire de l’art et les professionnels des musées : c’est un outil grâce auquel les images peuvent être décrites, organisées, étudiées et partagées. Dans ce processus — où un logiciel de vision opère sur une base de données d’images d’histoire de l’art — il y a cependant une variété de dynamiques en jeu. Elles sont en fait associées à des hypothèses théoriques, à des catégories historiques, à des contraintes technologiques et à des positions idéologiques : un ensemble de prémisses qui appelle une étude méthodologique plus approfondie du processus. Nous proposons un compte-rendu qui utilise la théorie de l’art et les études de la culture visuelle pour examiner les différentes étapes et les activités qui constituent l’analyse de la vision par ordinateur : après tout, l’étude des images a historiquement été une prérogative des historiens de l’art. Notre hypothèse est que les bases de données d’images d’art offrent en quelque sorte un « environnement protégé » dans lequel on peut observer comment les anciens problèmes, inhérents à la discipline, interagissent avec les nouveaux problèmes suscités par la façon dont nous consommons et concevons les logiciels. Les trois niveaux auxquels nous essaierons de détecter les positions biaisées répondent à trois questions différentes. De quelles images parlons-nous lorsque nous parlons de bases de données ? Quelles questions de recherche posons-nous aux algorithmes ? Quelles sont les logiques linguistiques et politiques en jeu dans le processus d’étiquetage ? Pour y arriver, nous commencerons la discussion en déboulonnant le mythe d’un simple parallélisme entre ces nouvelles formes de conceptualisation du réel et les formes traditionnelles, remettant en question l’utilisation par Manovich (1999) de la forme symbolique de Panofsky (1927) comme herméneutique de la base de données. Nous montrerons plutôt comment la logique de la base de données sur l’art s’en tient en quelque sorte au récit traditionnel de l’histoire de l’art, tout en produisant de nouveaux types de préjugés. Ensuite, nous nous concentrerons sur la façon dont cette technologie fonctionne réellement, et sur le type de pensée historique de l’art qui se cache derrière l’algorithme. Nous croyons que la praxis de ce logiciel est plus proche de la connaissance que de la recherche historique de l’art. Troisièmement, nous analyserons le processus d’étiquetage par lequel le logiciel de vision crée des métadonnées descriptives des images en question, en utilisant le compte rendu de l’iconologie critique de Mitchell (1994) pour problématiser la forte position idéologique et politique qui se cache derrière la relation image-texte. Tout au long du discours, et surtout dans le dernier paragraphe, nous aborderons les normes de transparence et d’évaluation qui doivent être définies afin de permettre à une approche méthodologique stricte de garder et de guider le processus, qui fait parfois défaut à la fois dans le secteur culturel et dans le domaine visuel au sens large. Il en ressort un compte-rendu des logiciels et des processus de vision par ordinateur qui semblent loin d’être « neutres » ou « objectifs » dans leur fonctionnement extrêmement stratifié, construit au milieu des intérêts de diverses parties prenantes et de faux pas procéduraux. En admettant que ces technologies contribuent cependant à construire la culture visuelle de notre époque, nous détectons une série d’hypothèses négligées en cours de route à travers le prisme de la théorie de l’art, dans l’espoir de contribuer à l’élaboration d’une vision plus claire.

Index terms

Mots-clés : intelligence artificielle, archives digitales, image, art, idéologie

Keywords: artificial intelligence, digital archives, image, art, ideology

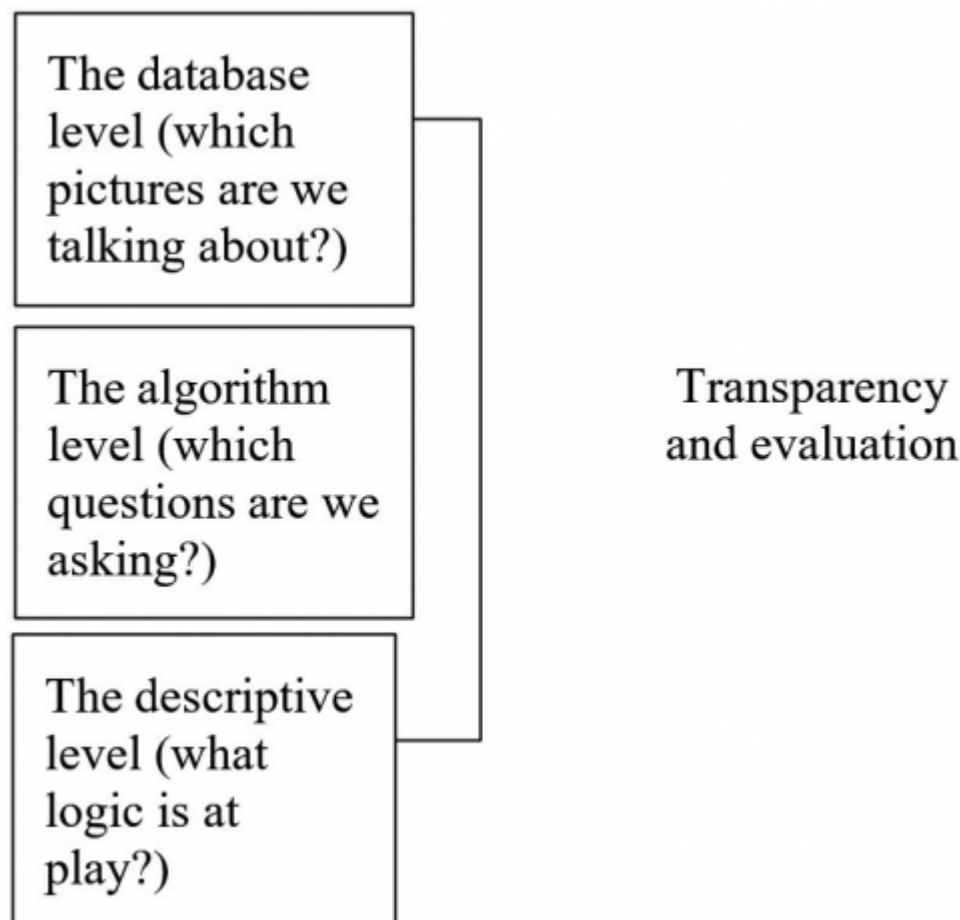
Full text

- 1 Contemporary computer vision software represents an incredible opportunity for both art history researchers and museum practitioners: it is a tool through which images can be described, organized, studied and shared. In this process—the one in which a computer vision software operates over a database of art history images—there are however a variety of dynamics at play. They have to do with theoretical assumptions, historical categories, technological constraints and ideological stances: the objectivity and neutrality that the discourse around technology at times seems to elicit are far from unquestionable. Moreover, the interaction between computer vision and art images occurs within a wider and far more complicated environment: that in which artificial intelligence and digitized images concur to the creation of the visual culture of our time. In this greater and extremely layered battlefield a higher number of

stakeholders and interests are at play, procedural bias is detectable and consequences of visual standard and consumption culture are vast and intertwined.

- 2 The starting point of this research is the conviction that art theory and visual culture can offer a helpful strategy to scrutinize steps and activities that are part of the computer vision analysis process: after all, the study of images has historically been a prerogative of art historians. Art images databases somehow provide us with a “protected environment” in which to observe how old problems, inherent to the discipline, interact with new problems, created by the way we consume and design software. Beginning from this assumption, the paper will follow different moments that constitute the process at hand and try to clarify yet at the same time problematize each of them, through the lenses of art criticism and theory. We will begin with debunking the myth of a simple parallelism between these new forms of conceptualizing the real and traditional ones, challenging Manovich’s use of the concept of symbolic form as a hermeneutic of the database. We will show instead how the art-database logic somehow sticks to the traditional art historical narrative, while at the same time producing new kinds of biases. Then, we will focus on how this technology actually works, and which kind of art historical thought lays behind the algorithm. Our guess is that the praxis of this software is closer to the connoisseurship than to the art historical research. Thirdly, we will analyze the labeling process through which computer vision software creates descriptive metadata of the images in question, using Mitchell’s critical iconology account to problematize the strong ideological and political stance behind the image-text relationship. Lastly, we will address the transparency and evaluation standards which need to be defined in order to allow a strict methodological approach to guard and guide the process, at times lacking both in the cultural sector and in the wider visual field.¹ The different levels of this analysis can be found in the following graph.

Figure 1



Biased levels we need to rethink in the computer vision analysis of art images.

Digital Panofsky, GA&C and the comets of our century

3 In 1924 Hamburg, Erwin Panofsky introduced for the first time a concept intended to abruptly change the course of twentieth-century art theory: the art historical application of Ernst Cassirer's² (2020) symbolic forms (Panofsky 1927).

4 The ambition behind the essay was to denaturalize the use of perspective, deemed a faithful transposition of natural vision since the Renaissance, showing how through perspective "spiritual meaning is attached to a concrete, material sign, and intrinsically given to this sign" (*Ibid.*, trad. 1991, p. 41).

5 A pioneering and wide theoretical framework, which, like all the classics, has known (and allowed) idiosyncratic readings over time, becoming an interpretative paradigm to which different disciplines have resorted. Two theoretical consequences of the essay, the de-naturalization of vision and perception as a result of different ways of approaching reality, and the role of techniques in these changes, however far from the author's original intentions³, have directed the reflections on the text in recent decades. In an age of rapid technological upheaval such as the one we live in, theory seems to have resorted, rightly or wrongly, to the concept of symbolic form to bring any form of technological or media change closer to the Renaissance revolution.

6 The most significant attempt at using the concept of symbolic form as a hermeneutic key to understand new technologies was made by Lev Manovich in the seminal essay *Database as a Symbolic Form* (1999). Although the technologies Manovich refers to, such as the CD-ROM, now belong to the prehistory of digital media, his use of Panofsky's theories is still worth of critical attention and requires to be historicized.

7 Moving from messianic assumptions typical of the postmodernist season, Manovich sees in new technologies a radical upheaval of our way of thinking about reality, in which:

Following art historian Ervin Panofsky's analysis of linear perspective as a 'symbolic form' of the modern age, we may even call the database a new symbolic form of the computer age (or, as philosopher Jean-Francois Lyotard called it in his famous 1979 book *The Postmodern Condition*, 'computerized society'), a new way to structure our experience of ourselves and of the world. Indeed, if, after the death of God (Nietzsche), the end of great Narratives of Enlightenment (Lyotard) and the arrival of the web (Tim Berners-Lee), the world appears to us as an endless and unstructured collection of images, texts and other data records, it is only appropriate that we will be moved to model it as a database. (*Ibid.*, p. 81.)

8 The main outcome of this argument, as Manovich (2000, p. 183) puts it elsewhere is that:

In new media, the database supports a range of cultural forms which range from direct translation (i.e., a database stays a database) to a form whose logic is the opposite of the logic of the material form itself—a narrative. More precisely, a database can support narrative, but there is nothing in the logic of the medium itself which would foster its generation.

9 By saying that, Manovich's discourse limits itself to an analysis of database technology, without dwelling on its empirical usage practices, and presents us with only one side of the coin. In fact, the idea of a database as "unmarked term" *Ibid.*, at least as far as databases of artistic images are concerned, seems to present itself as essentially utopian: there are actually *many* elements in the logic of the medium (or, *behind* it) that foster the generation of narratives.

10 The question should now be: what narratives are these?

11 Following Johanna Drucker (2013, pp. 5,7), we might answer: the old ones. Somehow, narratives even older than the theories that shaped our field since the Eighties:

In the 1980s, traditional art history was upended. Semiotics, structuralism, post-structuralism, psychoanalysis, Marxism, cultural and critical studies, and feminist thinking sharply divided art historians. At conferences, scholars would stand and bear witness to their love of “objects,” generating applause for their defense of traditional approaches. Many deplored such retrograde thinking, nixing the false binarism between theory and objects it introduced. The effect of theory was profound [...] But no particular changes of thought or critical stance come with this convenience, even if the days of slide-hoarding possessiveness are mercifully past, and the range of images and resources available are also more varied as well as more numerous. Challenges remain before the full corpus (however defined) of art historical images will be online—if ever. But even if this conversion into digital access and delivery has wrought substantive changes in the world of visual resources management, it has not had a ripple effect on the intellectual foundations of art history.

- 12 At the base of this hiatus there are also structural problems, such as the fact that, as Shelbert (2017, pp. 4,7) observes, these research in practice are conducted more in the museum and engineering sphere than in the theoretical and art-historical sphere. The result, as will be discussed in the following pages, is that “digital art history” is indeed Panofskyan, but not in the sense in which Manovich intends it. On the one hand, the type of categories through which “classical” art history is understood and periodized are the ones that are taught to machines, replicating a series of standardized observations already familiar to the iconological method, and blind to the challenges of contemporary theories.⁴ On the other hand, Panofsky himself (1927, trad. 1991, p. 34), in his essay on perspective, seems to recognize how any technology of ordering and visualizing reality inevitably accounts for blind spots:

What is most interesting is that Kepler fully recognized that he had originally overlooked or even denied these illusory curves [of comet trajectory being curved, ed.] only because he had been schooled in linear perspective. He had been led by the rules of painterly perspective to believe that straight is always seen as straight, without stopping to consider that the eye in fact projects not onto a *plana tabella* but onto the inner surface of a sphere.

- 13 Similarly, the images that come to compose the largest and most widespread digital datasets seem to reproduce the same biases, not only categorical but also content-related, of Western art history. Digital art history, to date, leaves non-western art in a blind spot.
- 14 These limitations are evident in platforms such as Google Arts and Culture (GA&C), probably the leading online platform of high-resolution images and videos of artworks and cultural artifacts from partner cultural organizations throughout the world.
- 15 Even if the Google team declares—in what we may consider a paradigmatic definition of a, naive or disingenuous, “digital universalism” (Chan 2014)—to have among its purposes a world-wide impact, in which “together, our mission is to preserve and bring the world’s art and culture online so it’s accessible to anyone, anywhere”⁵ in reality, these goals are far from being achieved.
- 16 A recent study (Kizhner *et al.* 2021) has indeed shown that the distribution of digitized art objects in GA&C represents for the 93.4% only 5 countries (UK, USA, Netherlands, Italy, South Korea), leaving to the other 190 countries from the United Nations list only the 6.6% of the exhibited content (123 of these don’t even publish samples of their institutional collections via GA&C). One of the paradoxes of this state of things is that the artifacts of some countries are represented only through their historical colonial presence in Western country databases.
- 17 Of course, this is not a problem inherent within the logic of the database, whose “neutral” nature as a tool is not disputed here, and it does not concern only the perpetuation of western canons in the art historical discourse. Rather, it is a matter of social and structural issues, involving “data transparency, open access policies, and having previously digitized content to facilitate data transfer” (*Ibid.*, p. 24).
- 18 Yet, even if we are not really talking about mere critical or art historical choices, we are talking about the premises of tomorrow’s art history. When an aggregator of cultural

content contains an overflow of artworks, images and histories belonging to a situated idea of culture, it is this model of culture that will be accessed, studied and disseminated, shaping a new (and very much old) digital canon. A canon possibly more “armored” than the old one, since assimilated and reinforced by deep learning and AI.

New skin for the old ceremony: the digital connoisseur

19 In the past few years, the number of digitized fine-art collections has been growing briskly. With the accessibility of such a massive dataset of digitized artworks, the need to develop new systems to archive and retrieve this pool of data has also risen. This is a key aspect for a “digitized” art history that, as Drucker claims (2013, p. 7), aims to become a “digital art history”.⁶ A similar task requires art historians to step out of their comfort zone, without limiting themselves to consider technology as a “black box”⁷, and strive to understand critically how these digital systems work.

20 An interesting aspect of the algorithms that organize image databases is that they work, as is to be expected, in terms of tracing similarities⁸: be it stylistic (Lecoutre *et al.* 2017; Karayev *et al.* 2014; Arora *et al.* 2012) or iconographic (Madhu *et al.* 2019; Crowley *et al.* 2014). This procedure, though connatural to “classical” art history, as Claire Bishop (2018, p. 125) notes, is problematic in itself, as it ends up perpetuating “uncritical assumptions about the intrinsic value of statistics”. Secondly, as these algorithms are produced in a market-oriented, neoliberal, and corporate environment, the objectives of these procedures have often very limited theoretical ambitions, as in the end “there is a need to develop automated recommendation systems that can retrieve ‘similar’ paintings that the user might like to buy” (Saleh *et al.* 2017, p. 71).

21 In this paragraph I will consider as a case in point the machine developed by Babak Saleh and Ahmed Elgammal (*Ibid.*), whose goal is to:

make aesthetic-related semantic-level judgments, such as predicting a painting’s style, genre, and artist, as well as providing similarity measures optimized based on the knowledge available in the domain of art historical interpretation. (*Ibid.*, p. 71.)

22 The type of analysis proposed by the authors, which moves from low-level elements to “high”, semantic-level, elements, is singularly close to the first procedures of the classic iconological analysis (except for the moment of cultural intuition that constituted, significantly, the culmination of the latter). On the one hand, this appears to be an interesting finding for the art theorist, as it shows how the vitality of iconology (and its ambition to work on vast *corpora* of works) almost naturally meets the needs of digital art history.

23 On the other hand, as we will see, it is necessary to ask how these premises are followed up and exploited, *which* history of art we are actually teaching to our algorithms, and if/how it dialogues with the most recent assumptions of art theory.

24 When an art historian is introduced to “visual features” categorized in “low” and “high” levels, where “low-level features are visual descriptors that there is no explicit meaning for each dimension of them, while high-level visual features are designed to capture some notions (usually objects)” (*Ibid.*, p. 78), what comes to mind is indeed the traditional combination of a formalistic approach to an iconographical one in the description of the work of art. In practice, this is of course an imperfect conversion.

25 It is in fact worth noting that the technologies employed to such goals are far removed from any art historical practice and rooted in applications with a shadier pragmatic background such as space and object recognition.⁹ They mimic the historical-artistic practice, without having structural reasons to reproduce it: for the machine, concepts like “style”, “genre”, or “subject” (or even the idea of “levels of analysis”) are entirely surreptitious, technicians provide them with pre-packaged aesthetic concepts, and train

them to recognize them. While at the beginning of this paragraph we noticed how these technologies appear to mimic iconology, a closer look seems to reveal a reductive employment, in its goals, of the iconological approach, aligning the work of these machines more to that of the connoisseur than the art historian. All summed up, we are currently teaching our algorithms to become XIX Century connoisseurs, where, as Ebtiz (1988, p. 206) puts it:

The value of connoisseurship to the history of art is judged by the practical result of the connoisseur's activity, namely the attribution of works of art to a particular artist, school or workshop, and to a particular time and place.

26 It is no coincidence that, historically, the connoisseur, and not the art historian, was the figure of reference for the art market, and still today, as Sachs (2020, p. 1692) explains:

As users navigate their own trajectories through webs of relations in the image data, they learn to place their tastes within a context of symbolic value, as sketched by the art experts. The organizational objective is to channel potential consumers into the emerging online market for art by helping them develop the confidence needed to embrace more risk, and of course to connect them to that market.

27 Computer vision seems to be fulfilling these goals with thrilling results, as it proves itself extremely useful in tasks like comparing details, finding the same motives in different techniques and the variations in a particular iconography, visualizing the characteristic style of an artist or a workshop and so on (Bell *et al.*, 2010). In this context, similarity is shaped as a market-value.

28 While some experiences (e.g. Manovich, 2013) have shown how these capabilities can also be used profitably by art historical research, if we try to ask ourselves—with Cuno (2012)—“what Panofsky or Aby Warburg could have done with our technology”¹⁰ the question—due to the lack of methodological or theoretical efforts for shaping new methods and new questions around them—remains unattended.

Icons, texts, and the battleground of digital labeling

29 As introduced in the previous paragraph, there are various ways in which artificial neural networks and deep learning technologies are employed in the visual arts, and more specifically on artistic image datasets. As Santos *et al.* (2021) summarize in their overview of the scientific literature on this topic, almost 50% of studies from 2012 to today (42 out of 90)¹¹, are engaged in either object detection—identifying people, animals, artifacts—or style and visual similarities detection—grouping images based on shared visual characteristics (*Ibid.*, p. 123). This operation—having computer vision software identify visual elements inside images—serves an extremely practical objective: producing descriptive metadata¹² which can help cultural institutions analyze and organize their content. As Craig (2021, pp. 198-199) states, referring to the whole sector: “most libraries, archives and museums (LAMs) face the constant and overwhelming struggle of ever-increasing backlogs of data to process [...] as acquisitions continue while immediate access to them is often blocked by lack of funds, time and labor”. The need to employ technology to facilitate processes and researches is shared by both big institutions—among the ones that have recently been taking advantage of AI for enriching and supplementing their metadata there are the Museum of Modern Art (MoMA), the San Francisco Museum of Modern Art (SFMOMA), the Harvard Art Museum and many others¹³—and smaller ones, although these last ones often, as Vlachidis *et al.* (2022, p. 106)¹⁴ note, struggle to undertake technological endeavors due to the lack of funds and resources.

30 While AI seems very helpful towards solving this problem, “care must be taken before accepting algorithmic description as a simple and unquestionable solution to undescribed materials, as that could leave room for significant errors to go undetected” (Craig 2021, p. 202). Even more so if the mission of LAMs is taken into account: institutions built on principles of equity and accessibility, democracy and cultural citizenship; values which strongly clash with biased and opaque processes.

31 The issue with ‘algorithmic description’, as Craig puts it, is an issue about labeling: the textual descriptions, hence the metadata, that the computer vision software provides of the images in question. This problem, far from being exclusive to artistic research, is one pertaining to the automated interpretation of images in general. In their provocative piece *Excavating AI, the politics of images in machine learning training sets*¹⁵ Kate Crawford and Trevor Paglen identify the problematic nature of this process in an epistemic assumption: that which entails the correspondence between images and concepts, appearances and essence. The computer vision process operates on the underlying premise that there are a series of consistent concepts which unite instances of a kind, each expressing visually in a clear and determined way. This assumption is built in the anatomy of the training sets on which each software operates; it is scripted in the different levels of the categorical taxonomy which allows the system to work. Structurally, these training sets (the image databases discussed in the first section of this paper) are organized into different hierarchical levels: there is a first level divided into a group of classes (eg. fruit, people, furniture), followed by a second level in which each class is divided in more narrow categories [fruit (apples, pears, oranges), people (astronauts, muslims, kleptomaniacs), furniture (tables, chairs, beds)], and then by a third ultimate level which is constituted by the images themselves, to which these labels are attached.¹⁶ This process, however, leads to the definition of a “‘visual essence’ of concepts, which is determined through statistical methods which look for formal patterns across labeled images” (Crawford, 2019). As the examples in brackets anticipate, there is an important difference between defining the visual correspondence of an apple and that of a kleptomaniac, or a muslim. These linguistic choices, and their use within this technical process, have very delicate and dangerous political implications, they carry a wide social impact.¹⁷ By making machines which operate through a visual logic communicate their findings in a linguistic one a problematic process seems to be set in motion.

32 Far from being an issue distant from artistic research and theory, the question of the relationship between images and their descriptions has been a central and vastly debated one across the years. In *Picture theory: essay on verbal and visual representation*, W.J.T. Mitchell (1994, p. 9) addresses the iconological debate claiming that “the very notion of a theory of pictures suggests an attempt to master the field of visual representation with a verbal discourse”. He proceeds by criticizing Panofsky’s iconological method claiming that “Panofsky’s is an iconology in which the “icon” is thoroughly absorbed by the “logos” understood as a rhetorical, literary, or even (less convincingly) a scientific discourse” (*Ibid.*, p. 28). This approach, which repressed the image in the name of language, has been historically followed by the postmodernist one, which moved from the opposite stance and repressed language in the name of the image: an epoch which Mitchell defines as: “the absorption of all language into images and “simulacra”, a semiotic hall of mirrors” (*Ibid.*). This tension between icon and logos, Mitchell argues, is however “embedded in the very grammar of ‘iconology’” (*Ibid.*) which is “a fracture concept, a suturing of image and text” (*Ibid.*), thus irreconcilable in principle.

33 Mitchell’s path towards a solution is what he defines as ‘critical iconology’: an account which begins from admitting the ideological premises which lay behind any iconological attempt:

The point [...] is not simply to make iconology “ideologically aware” or self-critical, but to make the ideological critique iconologically aware. Ideological critique cannot simply enter the discussion of the image, or the text-image difference, as a super method. It intervenes and is itself subjected to intervention by its object.

That is why I've called this notion of iconology critical and dialectical. It does not rest in a master-code, an ultimate horizon of History, Language, Mind, Nature, Being or any other abstract principle, but asks us to return to the scene of the crime, the scene of greeting between the speaking and the seeing subject, [...] the ideologist and the iconologist. (*Ibid.*, p. 30.)

- 34 Ultimately, what seems evident is that the relationship between an image and its description is a complex one, which cannot be reduced to a linear identification between image and text. How does this then relate to the descriptive practice which a computer vision software operates on images, be them artistic ones or not? There seems to be a case for extreme caution in accepting both the intrinsic value of this process and the textual metadata that comes from it without carefully scrutinizing every step of the way. By endorsing that images cannot be simply translated in text on the one hand, and accounting for the situated and material aspects of the descriptive process on the other, the grounds for a more cautious and aware endeavor can be set.

Transparency and evaluation: a new narrative for the blind spots of (computer) vision

- 35 It seems at this point evident that when machine learning algorithms are operating over an image database there are a variety of delicate aspects to take into account, each of them accounting for a possible biased beginning, biased process, biased results. As Bode (2020, p. 96) argues in her *Why you can't model away bias*:

These gaps and biases can be random or systematic, and they can arise from multiple sources, including historical conditions, cultural and institutional practices, economic factors, and/or technological processes. Investigating them requires attention to the histories of transmission and “infrastructures of knowledge-making” through which [...] data are constituted: a process in which meaning is inevitably transformed, if not lost entirely.

- 36 Even though the author is here discussing databases of literary items her claim holds equally applicable towards artistic image sets, highlighting the inevitable concurring factors which dissolve the hope for an unbiased journey. In the end the illusory understanding of a *pure* and *neutral* essence or use of technology seems to be—hopefully—discarded, yet the research and creative potential that these tools still hold is also evident. A middle route must be drawn, one which uses theory and critical thinking to problematize the diverse steps of the process.

- 37 In these regards, as Smits *et al.* (2021) suggest, a first step could be to endorse the utmost transparency when working with a computer vision software. According to Craig (2021, pp. 204-205), this could entail the following steps: informing the public that the descriptive metadata of a collection has been generated through computer vision, in order to foster clarity and awareness; allowing users to control the parameters of the computer vision algorithm, clarifying the guidelines through which the software operates and the extent of its engagement; applying a tracking feature to object description which would allow users to understand how an object metadata may change over time; and emphasizing the degree of certainty of each description and attribution the machine has made.

- 38 The issue of transparency becomes even more fundamental as often institutions and researchers cannot control every step of the process, they are engaged in. It may happen that they are forced to operate with a software that has already been trained by other companies, as training one's own software is in itself a costly and technical operation. This means that a museum might be using a specific software to create metadata for its own collection, without having clear information on how this software was built: on which data set did it train, engaging with which low level and high-level features, using

which taxonomy of categories and so on: all the problematic levels highlighted in the previous paragraphs concur to this scenario without institutions being necessarily aware of them. An example is the case of the the Minneapolis Institute of Art which used the Clarifai¹⁸ software to create metadata for its collection. As Ciecko presents in his research (2020), the software analyzed two portraits by photographer Chuck Close depicting a man (the photographer himself) and a woman with the same identical expression. Both portraits, furthermore, shared the same artistic take and stylistic features, being part of a series shot in analogous conditions. The computer vision software, however, described the two images with very different words: “cute”, “pretty” and “sexy” the female figure, “funny” and “crazy” the male one. As these are not accurate art historical categories and refer to a clearly gendered vision of female and male bodies, it seems that the instrument used did not receive the right training for the job.

39 Critical awareness on the items and data one is working with is key: artist and PhD student Minne Atairu emphasizes it in her research project, *Interrupting Benin Art Historical Archives with AI and AR*.¹⁹ In trying to build a dataset of images of Benin bronzes with which she planned to develop new AI generated content that would help restore a temporal absence in the production of these statues, she found the Google Image search tool deeply flawed in detecting the images of these artworks. The set she retrieved from the search was approximate at best, expliciting an important weakness in one very popular engine. Which is why the artist proceeded to compile an accurate dataset herself as the starting ground for her artistic project. When compared with the previous example concerning Google Images (that which analyzed the dataset produced and promoted by GA&C), the care in scrutinizing the different stages and materials one is producing is clear: while GA&C seems to be carelessly producing biased portraits of reality Atairu is rowing in the opposite direction.

40 The Minneapolis Institute of Art case and the *Interrupting Benin Art Historical Archives with AI and AR* project exemplify how artists and institutions need to be part of the open dialogue which surrounds these technologies and their use, ensuring that they are not complicit in expanding the impact of opaque and potentially dangerous dynamics and results.

41 Transparent processes, furthermore, need to be followed by exhaustive and methodical evaluative criteria. The historical rhetoric surrounding the humanities as ‘soft sciences’ which work with less strict standards and measures, escaping the rigor of numbers and so called ‘hard evidence’ risks to shadow the use of these kinds of technologies in the cultural domain, especially if corroborated by the mined field that the process itself has proved to be. As Vlachidis *et al.* (2021) have shown it is however possible and instrumental to work towards a shared system for the collection, integration and reuse of machine learning produced data in the cultural sector. In their paper *Semantic metadata enrichment and data augmentation of small museum collections following the FAIR principles*, they detail their attempt to introduce FAIR (Findable, Accessible, Interoperable and Reusable) data standard to the sector. The FAIR data principles have the goal of creating a set of guidelines and best practices which would facilitate both computer agents and humans in the discovery and management of scholarly data (Wilkinson *et al.*, 2016). Through the case of the CrossCult project organized within the Archeological Museum of Tripoli, the authors detail how the technological, methodological and cultural challenges encountered in the design and implementation of the programme paid off with the creation of a data set which made it possible to connect the museum with a wider and well-established network of other institutions and their publics.

42 While successful case studies must be celebrated, documenting difficulties in research to the scientific community is also of the utmost importance: it is vital to testify where and how these systems are failing. Claudia Engel *et al.* (2019) in their *Computer Vision and Image Recognition in Archaeology*, report on the use of this technology over a database of 150.000 images collected between 1993 and 2017 by an international team of archeologists of a UNESCO World Heritage site in Turkey: the 9000-year-old

neolithic settlement of Çatalhöyük. As the authors state in their introduction, despite the use of two different computer vision tools, Clarifai and Google Vision, they encountered three fundamental challenges: “Firstly, the metadata recorded with the images are incomplete and inconsistent. Secondly, researchers require access to information captured in the images that is not contained in the metadata. And lastly, the number of images cannot be reasonably processed by hand” (*Ibid.*, p. 1). Despite attempting different strategies the archeologists still found themselves with a problem: a set of images that they couldn’t possibly analyze without the help of technology, yet no technology that could offer consistent and useful data. It is through research like this one, which look for solutions within the steps of the process, that this technology will hopefully build solid progress.

43 Transparent methodology and severe evaluation standards seem, ultimately, to be necessary factors if computer vision analysis is to provide useful data and possible answers in the cultural community. Overall, it seems that an art theory perspective can guide this analysis, informing the discussion on the overlooked assumptions throughout the computer vision process.

Bibliography

ALLOA, Emmanuel (2015), “Could Perspective Ever be a Symbolic Form? Revisiting Panofsky with Cassirer”, *Journal of Aesthetics and Phenomenology*, 2:1, pp. 51-71.

ARORA, Ravneet Singh & ELGAMMAL, Ahmed (2012), “Towards automated classification of fine-art painting style: A comparative study”, in *ICPR*, pp. 3541-3544.

BACA, Murtha (2016), *Introduction to Metadata*, Los Angeles, Getty Research Institute.

BISHOP, Claire (2018), “Against Digital Art History”, *International Journal for Digital Art History*, 3, pp. 123-131.

BODE, Katherine (2020), “Why you can’t model away bias”, in *Modern Language Quarterly*, 81:1, pp. 95-124.

DOI : 10.1215/00267929-7933102

CAMPOLO, Alex & CRAWFORD, Kate (2020), “Enchanted Determinism: Power without Responsibility in Artificial Intelligence”, *Engaging Science, Technology, and Society*, 6, pp. 1-19.

DOI : 10.17351/ests2020.277

CASSIRER, Ernst (2020), *The Philosophy of Symbolic Forms (III Vol)*, London, Routledge.

DOI : 10.4324/9780429284922

CHAN, Anita (2014), *Networking peripheries: Technological futures and the myth of digital universalism*, Cambridge, MIT Press.

DOI : 10.7551/mitpress/9360.001.0001

CRAIG, Jessica (2021), “Computer vision for visual arts collections: looking at algorithmic bias, transparency and labor”, *Art Documentation: Journal of the Art Libraries Society of North America*, .40:2, pp. 198-208.

CROWLEY, Elliot & ZISSERMAN, Andrew (2014), “The State of the Art: Object Retrieval in Paintings using Discriminative Regions”, *Proceedings British Machine Vision Conference*, pp. 1-12.

DRUCKER, Johanna (2013), “Is there a ‘digital’ art history?”, *Visual Resources*, 29:1-2, pp. 5-13.

DOI : 10.1080/01973762.2013.761106

EBTIZ, David (1988), “Connoisseurship as Practice”, *Artibus et Historiae*, 9:1, 1988, pp. 207-212.

DOI : 10.2307/1483344

ENGEL, Claudia & MANGIAFICO, Peter & ISSAVI, Justin & LUKAS Dominik (2019), “Computer Vision and Image Recognition in Archaeology”, *AIDR*, pp. 1-4.

HUB, Berthold (2010), “Perspektive, Symbol und symbolische Form. Zum Verhältnis Cassirer – Panofsky”, *Eстетика: the European Journal of Aesthetics*, XLVII:2, pp. 51-71.

DOI : 10.33134/eeja.69

IVERSEN, Margaret (2005), “The discourse of Perspective in the Twentieth Century: Panofsky, Damisch, Lacan”, *Oxford Art Journal*, 28, pp. 191-202.

KARAYEV, Sergey & TRENTACOSTE, Matthew & HAN, Helen & AGARWALA, Aseem & DARRELL, Trevor & HERTZMANN, Aaron & WINNEMOELLER, Holger (2014), “Recognizing Image Style”, *British Machine Vision Conference (BMVC)*, pp. 1-20.

KIZHNER, Inna & TERRAS, Melissa & RYMYANTSEV, Maxim & KHOKHLOVA, Valentina & DEMESHKOVA, Elisaveta & RUDOV, Ivan & AFANASIEVA, Julia (2021), “Digital cultural colonialism: measuring bias in

aggregated digitized content held in Google Arts and Culture”, *Digital Scholarship in the Humanities*, 36:3, pp. 607-640.

LATOUR, Bruno (1999), *Pandora's hope: Essays on the reality of science studies*, Cambridge: Harvard University Press.

LECOUTRE, Adrian & NEGREVERGNE, Benjamin & YGER, Florian (2017), “Recognizing Art Style Automatically in painting with deep learning”, *Proceedings of Machine Learning Research*, 77, pp. 327-342.

MANOVICH, Lev (1999), “Database as Symbolic Form”, *Convergence: The International Journal of Research into New Media Technologies*, 5, pp. 80-99.

MANOVICH, Lev (2000), “Database as a Genre of New Media”, *AI&Soc*, 14, pp. 176-184.

DOI : 10.1007/BF01205448

MANOVICH, Lev (2013), “Museum without Walls, Art History without Names: Visualization Methods for Humanities and Media Studies”, VERNALLIS, HERZOG, RICHARDSON (eds.), *Oxford Handbook of Sound and Image in Digital Media*, Oxford, Oxford University Press.

MADHU, Prathmesh & KOSTI, Ronak & MUEHRENBURG, Lara & BELL, Peter & MAIER, Andreas & CHRISTLEIN, Vincent (2019), “Recognizing Characters in Art History Using Deep Learning”, *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents*, pp. 15-22.

MITCHELL, William John Thomas (1994), *Picture Theory: essay on verbal and visual representation*, Chicago, University of Chicago Press.

NÄSLUND DAHLGREN, Anna & WASIELEWSKI, Amanda (2021), “Cultures of Digitization: A Historiographic Perspective on Digital Art History”, *Visual Resources*, pp. 1-21.

OLIVA, Aude & TORRALBA, Antonio (2001), “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope”, *International Journal of Computer Vision*, 42:3, pp. 145-175.

PANOFSKY, Erwin (1927), “Die Perspektive als ‘symbolische Form’”, *Vorträge der Bibliothek Warburg 1924-1925*, Leipzig, Berlin, 1927, pp. 258-330; English trans., *Perspective as Symbolic Form*, New York, Zone Books.

SACHS, Jefferey (2020), “The algorithm at work? Explanation and repair in the enactment of similarity in art data”, *Information, Communication&Society*, 23:11, pp. 1689-1690.

DOI : 10.1080/1369118X.2019.1612933

SALEH, Babak & ELGAMMAL, Ahmed (2017), “Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature”, *International Journal for Digital Art History*, 2, pp. 71-93.

SANTOS, Iria & CASTRO, Luiz & RODRIGUEZ-FERNANDEZ, Nereida & TORRENTE-PATINO, Alvaro & CARBALLAL, Adrian (2021), “Artificial Neural Networks and Deep Learning in the Visual Arts: a review”, *Neural Computing and Applications*, 33, pp. 121-157.

SCHELBERT, Georg (2017), “Art History in the World of Digital Humanities. Aspects of a Difficult Relationship”, *kunsttexte.de – E-Journal für Kunst- und Bildgeschichte*, 4, pp. 1-10.

SMITS, Thomas, WEVERS, Melvin (2021), “The agency of computer vision models as optical instruments”, *Visual Communication*, 21:2, pp. 329-349.

DOI : 10.1177/1470357221992097

TORRESANI, Lorenzo & SZUMMER, Martin & FITZGIBBON, Andrew (2010), “Efficient Object Category Recognition Using Classemes”, in DANILIDIS & MARAGOS & PARAGIOS (eds.), *Computer Vision – ECCV 2010*, Heidelberg, Springer, pp. 776-790.

VLACHIDIS, Andreas & ANTONIOU, Angeliki & BIKAKIS, Antonis & TERRAS, Melissa, (2022), “Semantic metadata enrichment and data augmentation of small museum collections following the FAIR principles”, in GOLUB & LIU (eds.), *Information and Knowledge organization in digital humanities*, London-New York, Routledge, pp. 106-129.

WILKINSON, Mark & DUMONTIER, Michel & AALBERSBERG, Ijsbrand *et al.* (2016), “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, 3, pp. 1-9.

ZWEIG, Benjamin (2015), “Forgotten Genealogies: Brief Reflections on the History of Digital Art History”, *International Journal for Digital Art History*, 1, pp. 39-48.

Notes

1 Although the paper is the result of a collective research and reflection work made by the two authors, the first two paragraphs were written by Camilla Balbi, and the last two by Anna Calise.

2 On the distance between Panofsky's and Cassirer's concept of *symbolic form* see Hub (2010), and Alloa (2015).

3 In fact, as Margaret Iversen notes, in spite of the fact that the premises of the essay seem to lead, with Riegl, to substantially relativistic positions, the spatial concept of perspective seems, for Panofsky, to constitute a stable point of arrival in the human way of thinking space: “Although Panofsky’s “Perspective as Symbolic Form” purports to be a history of the development of single point perspective construction and the various conceptions of space implied by that history, it is in fact structured around a basic binary opposition between two strikingly different sorts of perspective. Antique and Renaissance (or Modern) perspectives stand at the opposite poles of an evolution and all the intervening moments are presented as hardly more than strategic moves and reversals that enable the history to get from A to B” Margaret Iversen (2005, p. 195).

4 In this regard, it is interesting to consider Dahlgren and Wasielewski study, which compares (combining both text mining and close reading) three datasets of art history journal articles published in the last decade: DAH (*International Journal of Digital Art History*, special issues of *Visual Resources*), *Art History*, and *Art Journal*. “As our text mining shows, there seems to be a gap or a disconnect in terms of scholarly interests between those who develop and use digital tools for art historical studies (DAH) and those who do not (*Art History*, *Art Journal*)”. (Näslund Dahlgren & Wasielewski, 2021).

5 <https://about.artsandculture.google.com/> Last accessed October 10, 2022.

6 We do agree with Benjamin Zweig (2015, p. 41) that “Drucker’s distinction between digitized and digital art history, while imperfect categories (for a critic of which see e.g. Schelbert 2017 and Näslund Dahlgren & Wasielewski, 2021), affords us with a good point of entry from which to understand the history of doing art history digitally”.

7 The term was coined by Bruno Latour (1999).

8 On the difficult relationship between “art historical similarity” and “engineering similarity” see Sachs (2020).

9 This is the case for both low level features, which resort to the *Spatial Envelope* computation model (see Oliva *et al.*, 2001) and high or semantic level features, which use descriptors such as Classemes, which encodes object presence in the image (see Torresani *et al.*, 2010).

10 James Cuno, “How Art History Is Failing at the Internet”, *The Daily Dot*, Nov 19, 2012, (URL: <https://www.dailymail.com/via/art-history-failing-internet/>). Last accessed May 10, 2022.

11 Although scientific literature production does not represent an exhaustive indicator of cultural organizations’ activities—there are various museum programs which employ computer vision softwares and are not however assessed through academic publications—it is still an important indicator of the main interests and research direction of the sector, especially if the strong technical nature of these projects and researches is considered.

12 A useful publication which contextualizes metadata and their use in relation to cultural objects is Murtha Baca, 2016. In the introductory chapter, “Setting the Stage” Anne J. Gillian writes “as all information objects, regardless of the physical or intellectual form they take, have three features—content, context, and structure—all of which can and should be reflected through metadata”. This reference can be useful in exemplifying the wide range of information that museums attempt to gather through computer vision software.

13 Brendan Ciecko, *AI sees what? The good, the bad and the ugly of machine vision for museum collections*, 2020 retrieved at <https://mw20.museweb.net/paper/ai-sees-what-the-good-the-bad-and-the-ugly-of-machine-vision-for-museum-collections/> consulted on the 10th of April, 2022.

14 This article also point out how this difference of scale across museums becomes even more relevant in southern countries. While smaller museums in northern regions (as North Europe) can still to some extent benefit from shared systems and technologies, southern museums of the same continent suffer a greater lack of resources.

15 Kate Crawford, Trevor Paglen (2019), “Excavating AI, the politics of images in machine learning training sets”, *The AI Now Institute*, published at <https://excavating.ai/> , consulted on May 6th, 2022.


16 The number of possible sub groups hierarchically organized is potentially limitless, depending on the complexity of the topic in question. The important aspect, from a logical standpoint, is that each textual label ultimately corresponds to a visual element.

17 The discourse by Crawford and Paglen proceeds by discussing the critical political implications of this dynamic, which have to do with discriminatory practices (with reference to race, gender, intersectional bias and prejudice). For a comprehensive account of the political implication of AI technologies cf. Campolo *et al.*, 2020.

18 Clarifai computer vision software website, consulted on the 8th of May 2022: <https://www.clarifai.com/>.

19 Cf. the artist’s website at <https://igun.minneatairu.com/> or watch a presentation of the project on the website of the conference Art History 2060 by Davidson College where Minne Atairu presented her project, video available from minute 18.6 to minute 36.40 at the following website: <https://www.youtube.com/watch?v=Rd2PXvL9deI&t=1045s>. Links consulted on the 8th of May 2022.

List of illustrations

	Title	Figure 1
Caption	Biased levels we need to rethink in the computer vision analysis of art images.	
URL	http://journals.openedition.org/signata/docannexe/image/4757/img-1.jpg	
File	image/jpeg, 83k	

References

Electronic reference

Camilla Balbi and Anna Calise, "The (theoretical) elephant in the room", *Signata* [Online], 14 | 2023, Online since 06 November 2023, connection on 29 January 2024. URL: <http://journals.openedition.org/signata/4757>; DOI: <https://doi.org/10.4000/signata.4757>

About the authors

Camilla Balbi

Camilla Balbi (1994) is a postdoctoral researcher at the Photography Research Centre of the Institute of Art History, Czech Academy of Sciences. In 2023 she obtained her PhD in Visual and Media Studies at IULM University in Milan. She is also a 2020/2021 Visiting Scholar at New York University's Department of German Studies at. Camilla has always been interested in the intersections between media practices and different artistic languages, and her main research interests include early twentieth-century German-Jewish art theory and curatorial studies. Alongside these interests, which form the subject of her PhD research, she writes and works on political art and eccentric visual cultures, working on the specificities of the Jewish gaze and the female and queer gaze. Her recent publications include *Unburied Iconology, Erwin Panofsky on Photography* (Mimesis, 2023) and *Fake or Fortune? Alexander Droner and the Weimar Reproduction Debate* (Venice Arts, December 2021). In addition to her academic activities, she writes for the Enciclopedia Treccani di Arte Contemporanea and works as an art critic for Flash Art magazine and researcher with the Israel Museum in Jerusalem.
Email: [camilla.balbi1\[at\]studenti.iulm.it](mailto:camilla.balbi1[at]studenti.iulm.it)

Anna Calise

Anna Calise is a PhD student in Visual and Media Studies at IULM University in Milan, and a 2021 Visiting PhD at the University of Amsterdam, at the School for Heritage, Memory and Material Culture. She is researching the digitization of museums and the mediatization of the cultural experience, attempting to define a new portrait of the museum in light of today's technological development. She has a Philosophy degree from King's College London, and two Master degrees in Arts Management (from Federico II in Naples and SDA Bocconi). She has worked on the design of participatory cultural strategies, and coordinated the Matera 2019 Community Projects Program in the year of the European Capital of Culture. Among her recent publications: *Mixed Reality: frontiera dell'educazione museale*, (Pianob. Arti e culture Visive, 2022), and *The digital museum and its power dynamics: the case of The Smithsonian* (Mimesis, 2023). Conferences in 2022: *Art History 2060*, Davidson College, March 2022; *International Conference of Intermedia Studies*, Trinity College Dublin, September 2022; *Transcultural Exchange*, the Colleges of Fenway Boston, November 2022.
Email: [anna.calise\[at\]studenti.iulm.it](mailto:anna.calise[at]studenti.iulm.it)

Copyright



The text only may be used under licence CC BY 4.0. All other elements (illustrations, imported files) are "All rights reserved", unless otherwise stated.