



# Automatic Computation of Meaning in Authored Images Such as Artworks: A Grand Challenge for AI

DAVID G. STORK

---

We discuss preliminary successes and major outstanding challenges in extracting messages, stories, morals, and especially *meaning* in crafted or “authored” images, such as artworks. Traditional semantic image understanding seeks to summarize an image (such as through a caption), or to answer basic questions expressed in free-text natural language, but can neither infer plausible *reasons* the creator made the artwork nor compute a high-level message or meaning it conveys. Such meaning is often abstract, dependent upon genre, period, and art movement, and exploits visual conventions such as special objects and signs (“signifiers”), composition, and possibly non-realistic style. Several of these properties have no counterpart in the natural photographs that are studied in most semantic image analysis, nor in the application-specific image analysis applied to robotics, autonomous driving, medical diagnosis, or remote sensing. For such reasons, the extraction of meaning from works in a variety of styles and subjects will be both a tool for art scholars and a grand challenge to artificial intelligence research.

CCS Concepts: • Computing methodologies → Artificial intelligence; Computer vision; Computer vision problems;

Additional Key Words and Phrases: Computational visual semiology, deep neural network, art analysis, computer-assisted connoisseurship

**ACM Reference format:**

David G. Stork. 2022. Automatic Computation of Meaning in Authored Images Such as Artworks: A Grand Challenge for AI. *ACM J. Comput. Cult. Herit.* 15, 4, Article 65 (December 2022), 11 pages.

<https://doi.org/10.1145/3471619>

---

## 1 THE PROBLEM OF COMPUTING MEANING IN AUTHORED IMAGES

Consider the painting in Figure 1, Harmen van Steenwyck’s *Still life: An allegory of the vanities of human life*, and how current algorithmic approaches to semantic image analysis, suitably improved and refined, might be used to analyze it and similar two-dimensional artworks. The component pattern recognition and semantic segmentation algorithms would likely identify and localize the books, shell, sword, wine jug, and (given contextual information about 17th-century Netherlands) the musical instruments, thin stream of smoke, and shaft of light [31]. These methods can infer simple spatial relations, such as the shell is *on* the table, the books are *touching* the skull, and related predicates [7, 27]. Analyses of this sort might also support basic classification of the overall scene (open scene or indoors or street, etc.) [8]. The current state of question-answering algorithms, as applied to natural photographs, strongly suggests that these algorithms will be increasingly accurate describing objects and simple relations in works such as *Still life* [15]. These are challenging problems in artificial intelligence and the slow but steady progress toward their full solutions has relied on numerous advances in image processing, computer

65

---

Author’s address: David G. Stork, 111 Crescent Avenue, Portola Valley, CA 94028 USA; email: artanalyst@gmail.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

1556-4673/2022/12-ART65 \$15.00

<https://doi.org/10.1145/3471619>



Fig. 1. Harmen van Steenwyck's *Still life: An allegory of the vanities of human life* (39.2 × 50.7 cm), oil on oak panel (1640), National Gallery London. This painting bears a clear and powerful meaning—a meaning that current techniques in AI are apparently insufficient to extract.

vision, machine learning (especially deep neural networks), and scene analysis. Eventual solutions would satisfy current conceptions of a visual Turing test [13].

Research of this sort nevertheless sheds little if any light upon *why* this image was created, *why* it has been prized for over a third of a millennium and rightly deserves presentation in one of the world's most important art museums... in short what makes such an image *special*. Beyond the evident technical skill of the artist and the work's manifest beauty, the most important property of this work is that it—like most artworks—carries a *meaning*. The artist (at his patron's behest) worked diligently to craft a painting that expressed ideas, lessons, and morals, which in this case supported the teachings of the Protestant Church, ascendant during the Dutch Golden Age. In brief, that message here, referred to as *vanitas*, can be summarized as follows [2, 20, 23]:

Do not concern yourself with the diversions of this world, given that you will surely die—at a time few ever know. Instead lead a sober, humble life in preparation for the eternal life to come.

Every educated—and even most illiterate—viewers of 17th-century Netherlands well understood this meaning as it was conveyed by this and similar *vanitas* paintings. This message was preached from pulpits, written in religious texts and philosophical tracts such as by the contemporary philosopher Erasmus. It figured prominently

in the Low Countries' rebellion from Catholic Spain. Artists working in this genre strove to imbue *meaning* in the “soulless” objects of everyday life. Patrons eagerly sought such artworks for their homes to remind them of this message and how to live a good life, and to demonstrate their moral aspirations to visitors.

How does this work convey this message and bear this meaning? What elements—visible and invisible—work together to these ends? What can current algorithmic image analytic methods reveal about such works? Most importantly, what problems must AI solve if automatic methods are to “understand” an artwork to the level illustrated in the summary above?

## 2 SEMIOLOGY, SIGNIFIERS, AND SIGNIFIEDS

There are two broad classes of tasks and goals in extraction of meaning from artworks. One class is the computational inference of a meaning as directed from the artist to the viewer, as described above. This class includes, for instance, deducing that the pocket watch symbolizes the passage of time and the transience of human life. Another class of tasks concerns the analyses by scholars and historians of art about what a work may mean in a broader cultural context. Thus, van Steenwijk's *Still life* is an illustration of the Protestant rebellion from Catholic Spain. While these two classes of problems may in some cases intersect, our focus here is the first class of problems: How does an artist (“author”) convey a message or meaning *to the viewer* through visual methods?

The range of messages and meanings that have been—and someday may be—expressed through visual arts is vast, indeed almost overwhelming, spanning domains from the personal and inter-personal, to societal, historical, esthetic, religious, and even philosophical matters. The variety of techniques employed by artists to express meanings in these manifold domains is also itself vast and extends from the art medium, overall dimensions, siting, and of course the objects, figures, placements, poses, gestures, and so forth. Every computational effort must at minimum take account of such a context, which will generally limit the scope and domain of the technical results. We illustrate just a few of these issues in fairly narrow classes of visual art—specifically Christian religious art and Dutch vanitas paintings—and focus on one method by which meaning is generated, specifically through the use of reference between signifiers and signifieds.

The most direct and widely used technique for conveying meaning in artworks is to include signs or symbols that refer to another object, idea, or concept. Semiotologists refer to such signs as *signifiers* and the concepts to which they refer as *signifieds* [28]. There are three categories of signifiers: *indexes*, *icons*, and *symbols*; these have the following properties and functions:

**Index:** an indirect or abstract pointer to its associated signified, where the linking relation can in principle be anything. A footprint in the sand is an index to the person who walked the beach.

**Icon:** an image that resembles its signified, but is far more abstract and schematic. Examples include the simplified outline of a man or a woman on many restrooms doors, or numerous icons in graphic user interfaces, such as a tiny printer, paper mail envelope, or file folder.

**Symbol:** an abstract and arbitrary social convention. For example \$ and £ refer to money, + refers to addition, and so on.

Semiotologists and art scholars claim that meaning does not inhere in an artwork such as a painting taken alone, but is constructed by the interaction from the viewer, the artwork, and contextual information and knowledge. For this reason there may be several convincing “readings” or interpretations of a work. Nevertheless, some readings are more compelling and explain more of the visual and associated evidence than others, including the artist's own stated goals, intentions, and explicit message [9, 45].

Consider again the painting in Figure 1 and how the reading, summarized above, might arise in part through its use of objects as signs, as well as its composition and style.

**Objects (as signs).** Nearly every object depicted has symbolic meaning in service of the message of the painting. The skull is clearly a symbol of mortality and the universality of death, and its

location at the center of the work and the fact that its empty eye sockets seem to be looking toward the viewer highlight its importance. The lovely shell at the left would have been rare and expensive, and illustrates earthly pleasures, much as would jewelry. The large open pocket watch on the table symbolizes the passage of time, and its placement close to the skull illustrates that death is always near. The wooden flute on the table, the bell of the sackbut (early trombone), and the belly of a lute (early strummed stringed instrument) elevated at the rear of the composition together represent various expressions of culture, which may lead us from the central concerns of death and the life to come. The Japanese samurai sword, a product of the Netherlands' recent opening of trade routes with Asia, would have been quite novel, rare, and expensive; it illustrates the worldly interests in both travel and luxury. Moreover, the two books touching the skull signify worldly learning. The fine pink silk scarf (also from the newly opened trade with Asia) and the wine jug reference sensual pleasures. The oil lamp at the center represents the passage of time; the fact that its wick has recently burned out—as evidenced by the thin trail of smoke (a semiotic *index*)—shows that death and extinguishment are inevitable. In some interpretations of this painting, the smoke refers to an invisible omniscient being (*signified*) who snuffed out the lamp, conveying a clear theological message [22]. There are numerous other prominent signifiers used in Dutch vanitas paintings, for instance, lemons (which appear sweet yet taste sour) [30], plane and convex mirrors (tools for those preoccupied with appearances, and vanity itself) [18], and even tobacco (a pleasure of life that literally goes up in smoke) [17].

**Composition.** Of course, a painting is a static object, but in the West viewers tend to “read” a work from left to right (matching the flow of printed text), and thus the shaft of light is interpreted as coming from above, flowing downward to the right. For this reason this shaft signifies heavenly light. This shaft, directed toward the skull, is clearly a reference to the religious notion of an omnipotent being in heaven guiding worldly affairs, including man’s ultimate destiny, as signified by the skull.

**Style and format.** Several formal properties of the work reinforce the work’s message. First, the painting is roughly life size, making the objects, including the skull, read as present before the viewer. (If the work were a miniature or a large mural, then this aspect of the meaning would be diminished or lost.) Moreover the style is extremely realistic, with subtleties of lighting, reflection, and texture, all expertly rendered, thereby reinforcing the notion that the message is present before you, as real as the common objects that surround us all, including those in the patron’s room where the painting hung. The color scheme is dark, sober, even somber, engendering quiet contemplation and serious reflection by the viewer, again, essential to the meaning of the painting.

*Still life* was very carefully planned and executed by the artist, sometimes called an “author” in such contexts. Its rich, almost overwhelming, message rarely if ever appears in the casual photographs or application-specific videos that dominate research in automatic image analysis and understanding. Even state-of-the-art image interpretation, question-answering, and other methods of automatic semantic analysis provide almost no extraction of *why* an image such as was created, and especially what its author *means*, beyond perhaps simply documenting a scene [21]. The role of *style* in the extraction of meaning is almost entirely ignored in current AI methods, given that photographs are stylistically far more uniform than paintings from all eras and regions across the world.

The vast majority of research in computer vision, deep learning, and so forth seeks to interpret the physical world, be it a natural photograph of children at play, or a real-time video from the front of an autonomous vehicle of the roadway, or an x-ray of a malignant lung tumor. This “surface-based” analysis—as challenging as it may be—provides but initial steps toward addressing the deeper problem of inferring an artist’s likely intentions or *meaning* conveyed in an artwork [19].

One of the most important features of two-dimensional artworks that frequently contributes to their message and meaning is *style*, and because this aspect is not adequately addressed by current AI techniques, it bears some discussion. For instance, the rapid, free brushstrokes of Impressionism are part of the message of that art movement, which celebrated the transient, even ephemeral observation of the contemporary and quotidian world. Despite their stylistic differences, many of the paintings by Pierre Renoir, Claude Monet, Paul Cézanne, and others exploited such rapid and relatively “free” brushstrokes, a dramatic departure from the refined and generally invisible brushstrokes of the prevailing Academic movement, led by William Adolphe Bouguereau, Jean-Léon Gérôme, Jean-Auguste-Dominique Ingres, and others.

The realistic style and symbolic color schemes exploited in Neoclassicism contributed to the permanent, even eternal, messages and morals that were conveyed in such art. Thus, Jacques-Louis David’s *Oath of the Horatii*, which celebrated honor, duty, service (even to one’s death) to the political state (Rome), was painted in a style of strong lighting, crisp contours, and bold symbolic colors. The vigorous, immediate and even wild marks in Abstract Expressionism were perhaps the principal technique by which the meanings of this art—the primacy of emotions and existential confrontations in an uncertain world—and were expressed most clearly by Jackson Pollock, Joan Mitchell, and Arshile Gorky. (As the art critic Meyer Shapiro said: even if a painting did not depict any objects it could still convey a *meaning*.) The Pop artist Roy Lichtenstein’s *Brushstroke* depicts a single large, bold “dripping” yellow brushstroke of Abstract Expressionist art but rendered in the cool, flat, industrial style of comic books (complete with large blue BenDay dots prominent in newsprint)—a savage and ironic parody of the earlier art movement.

Such semantic function of style generally holds—by definition—regardless of the explicit objects and spatial relations that may or may not be depicted in an artwork. (The Lichtenstein *Brushstroke* is an example where the meaning of the painting is enhanced due to the relation between the style and the content.) Most relevant here is that such semantic function of style has almost no relevance to the properties of natural photographs and thus has received minimal attention from the AI community. Note too that the extensive successes in learning and transferring style (e.g., transferring Impressionist style to a photograph) have no direct bearing upon inferring meaning in either the source painting nor (of course) the target photograph [24].

A closely related property of much art that distinguishes it from natural photographs is the depictions of non-existent objects, such as dragons, monsters, halos, winged angels, centaurs, unicorns, and so forth. Such imaginary objects do not appear in the billions of natural photographs typically used for training deep networks. Likewise, some artworks depict objects with altered material properties under conditions that violate the laws of the natural world. For instance, paintings depict flying infants (*putti*), floating gods and angels, pocket watches dripping like molasses, holy infants glowing a golden light, and so forth—all important components in the message of an artwork.

### 3 EXTRATION OF MEANING IN NATURAL TEXT AND THE VISUAL ARTS

The variety of interpretations and *types* of interpretations of artworks is vast, often subtle and complex, and can involve an extraordinarily diverse set of concerns and likely representations [4, 29]. Clearly inferring meaning in general artworks is AI complete and requires human-level artificial general intelligence [13, 26]. As such, there is no compelling unifying representation or research methodology for addressing the problem in its full generality—other than the vague and daunting suggestion to *solve general intelligence first*. For this reason, we should confront special cases of the general problem that expand the concerns of AI and advance scholarship toward that eventual goal [39].

In many ways, natural language processing in its variety of subtasks has confronted the analogous problem more directly than has image analysis. After all, nearly all the text processed by automatic semantic analysis was written by *authors* with intentions to convey *messages*. Here, for instance, sentiment analysis has been used to extract tone, valence, and related stylistic properties of text, relatively independent of the explicit content or meaning. Nevertheless, such analysis can be relevant in the extraction of the author’s intent or meaning [5].

As just mentioned, natural language processing has, by the very nature of its subject matter, confronted the problem of extracting an author’s intent and meaning far more extensively than has semantic image analysis. This is not the venue for an extensive rehearsal of the methods and accomplishments in natural language processing, but instead a chance to highlight the methods of natural language processing that show the greatest promise in advancing automatic inference of meaning in authored images such as artworks.

Such methods include sentiment and form analysis, which in several ways is analogous to style analysis in artworks. Such language analysis focusses on valence words and structures (nearly) independent of higher-level topic and content of a message. Here the statistical properties of lengths of sentences, proportion of anaphora, valence of adjectives and adverbs, and so on are estimated, and such information forms a component of higher-level semantic interpretations. Linguistic processing can infer rudimentary meanings based on symbolic words and phrases, analogous to the signifiers in certain paintings, presented above [16].

A key subproblem in efforts of extraction of high-level meaning involve knowledge representation, and numerous structures—semantic nets, frames, rules, ontologies—have been developed to this end [43]. Of course in many deep network or belief network approaches for use with natural language text as well as for static images knowledge is represented in highly distributed form through millions of trained connection weights [40].

For the problem of extracting meaning from an artwork that is based on some story or moral expressed in natural language, we might first explore representations that have shown value in knowledge and meaning in both natural language and image domains, perhaps one based on logic and predicates. Such knowledge might be represented in predicate form such as Eat(Eve, apple) or Touch(book, skull) [6]. Other, more distributed and statistical representations trained with massive data sets are likely to provide a foundation as well.

#### 4 ONE-SHOT OR FEW-SHOT LEARNING

Given the recent and dramatic successes in object classification, event classification, semantic segmentation, automatic question-answering, and related problems provided by deep networks [14], we might expect that training such networks on a corpus of art images (and possibly accompanying textual descriptions and analyses) would provide a foundation for extraction of meaning from artworks. However, this approach immediately confronts a problem: the number of available artworks is far smaller than the number of photographs used to train deep networks, which may reach several hundred million or even a billion. Even a very prolific fine artist such as Pablo Picasso executed roughly 13,500 paintings, while Johannes Vermeer executed fewer than 34, and Leonardo da Vinci left us just 16.

The obvious approach, then, is to employ *transfer training* in which, here, a network trained with a large corpus of natural photographs is further trained with art images. Alas the number of art images is rather small and would not lead to acceptably high performance as judged on artworks. Such transfer learning is almost certain to fail, catastrophically, on highly stylized paintings and abstract art.

Consider Willem de Kooning’s *Two women in the country* in Figure 2. State-of-the-art object recognition, semantic segmentation, and summarization algorithms trained on natural photographs would likely fail on such an image. Clearly the statistics of color, shape, and so on in the artwork differs markedly from the large data sets of natural photographs used to train current image segmentation and analysis systems, and that would predict such poor performance. Nevertheless, upon seeing even just *one* painting by this artist, nearly all humans can learn this artist’s style and recognize his other works very reliably, as the reader can confirm by searching for his paintings online. This fact exposes another—well-known—limitation of current learning-based approaches to image understanding: the implementation of one-shot (or few-shot) learning [38, 44]. It appears that art paintings provide a rich domain for the development of methods for one-shot learning.

While robust and accurate such algorithms are being developed, we can explore alternate methods, ones that exploit knowledge that links art images to associated stories and meanings expressed in text. We now describe one such method, presented recently [36].



Fig. 2. Willem de Kooning's *Two women in the country* (117.2 × 103.5 cm), oil, enamel, and charcoal on canvas (1954), Hirshhorn Museum and Sculpture Garden. An artwork such as this presents many challenges to automatic analysis: recognizing the figures (two women) and setting (countryside), identifying the style, and crafting an interpretation or “reading” of the work that addresses what the artist is trying to convey. Current AI and image analysis techniques are poorly suited to such tasks, because here plausible meanings are built upon the full range of aspects of the work, from the medium and style up to the composition and relation to other contemporary artworks.

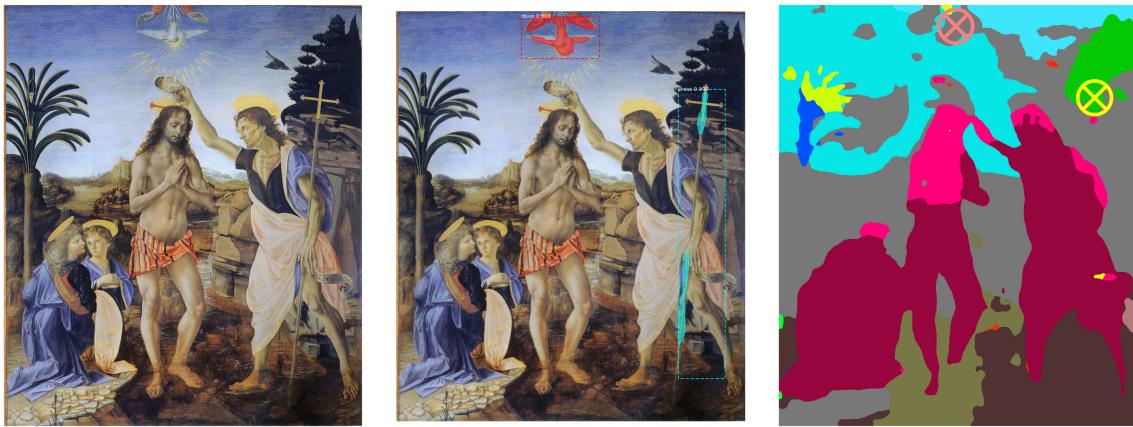


Fig. 3. (L) Andrea Verrocchio’s *Baptism of Christ* (177 × 151 cm), tempera and oil on panel (ca. 1470–1475), (C) bounding boxes for objects recognized as iconographic *attributes* by deep neural network trained on an independent database of such attributes appearing in Renaissance paintings, and (R) semantic segmentation, identifying human figures (maroon and pink), based on natural photographs. The identity of each actors can be inferred from the identity of the closest semiotic *attribute*, and such identification can lead to the recognition of the source story and its meaning. (Figure after Reference [36].)

## 5 AUTOMATIC IDENTIFICATION OF ACTORS IN PAINTINGS

Perhaps techniques from natural language understanding most relevant to inferring messages in artworks concern story analysis, including the inference of the *actors*, their actions, and relations. Recent work, unrelated to art images, has addressed character role assignment based on extraction of character identities from text and existing structure and morphology of folktales [41]. One of the greatest corpora of message-bearing images are religious works from the Western canon. The meaning to be extracted is a story (with associated morals) that lie in Biblical texts. The Church was the primary patron for art for well over a millennium, employing most of the greatest artists; the greatest works were crafted to convey a meaning, message, or moral from the Bible. Thus, religious art provides many examples and varieties of meaning-bearing artworks.

The computational goal is to extract meaning by identify the story being depicted in an artwork, even amidst the great variety in style, composition, and so forth. One component step is to identify the figures, or as art scholars say, the “actors” depicted in an artwork. One can address the problem of identifying actors in artworks much as a traditional classification problem, and use techniques such as deep network directly to this end [25]. We now summarize a different method, recently presented, that is based on the above semiological functions.

The left panel in Figure 3 shows Andrea Verrocchio’s *Baptism of Christ*, which as its title makes clear refers to the Biblical story in which Christ is washed by John the Baptist in the River Jordan. (Similar compositional strategies were pursued by notable artists such as Bartolome Esteban Murillo, Ottavio Vannini, Guido Reni, Nicolas Poussin, Perugino, as well as several others.) The message or moral expressed through this painting is based on the Biblical text:

Saint John baptized Christ under the observation of the Almighty and heavenly angels in order to cleanse him of the sins of mankind that he had borne; moreover, salvation will come only through Christ’s crucifixion.

As stressed above, current semantic image analysis methods fall short of extracting such a message and meaning.

A first step in extracting such a meaning automatically is to recognize the actors. Western Biblical stories—both written and illustrated visually—exploit *attributes* for identifying important actors, such as saints. In this context, the term “attribute” is not related to the common usage in computer vision and pattern recognition [11].

In religious iconography, an attribute is an object closely associated with a given religious actor, almost always deriving from a Biblical or historical story or event. One actor might have several attributes, and any given attribute might be associated with more than one actor, such as a saint. For example, Saint Peter's attribute are keys, Saint Luke's is a bull, Christ's are a crucifixion cross and dove, Saint John the Baptist's is the crucifixion cross, and so on. (Such attributes helped the illiterate in early church history to better understand the stories in paintings and stained glass windows.) The associations between saints and attributes are codified in ecclesiastical books and **The golden legend** [10], and can be easily represented in an associative database.

A deep neural network was trained with examples of several Biblical attributes and when applied to the Verrocchio painting, and this network then identified and localized the dove and crucifixion cross, as shown in the center panel of Figure 3 [36]. The right panel shows the semantic segmentation performed by a separate deep network trained on modern natural photographs, where pink and maroon indicate the label of PERSON and ANIMAL, respectively. Next, the spatial distance between each semiotic attribute and candidate actor was computed. Finally, a lookup table of attributes and saints was used for actor identification. In this way, the central actor is recognized as Christ and the actor to the right as Saint John the Baptist. The average precision and recall of this two-network system were 92% and 81%, respectively, over four actor identities. These recognized actors can then serve as an index or query into a database of Biblical episodes or stories, thereby finding the Baptism story and its associated (text) message or moral.

## 6 CONCLUSIONS AND FUTURE DIRECTIONS

Art images are among the most memorable and culturally relevant images of any sort, and they present numerous challenges to the development of high-level automatic semantic interpretation. The problem of the extraction of meaning from authored visual works requires a wide range of levels of visual analysis (from low-level style up through object recognition and inferring of relations), knowledge of conventions in symbols, context, common sense reasoning, natural language processing, and more. The number of such artworks is small (relative to the databases used in deep learning methods), the objects frequently depicted are imaginary or non-physical, their styles vary widely, and thus they present profound challenges to AI methods.

Of course, artworks such as we address are unique objects and the computational methods must always be applied with an awareness of context. Nevertheless, with very few counter-examples (such as Marcel Duchamp's *Fountain*), there are regularities and commonalities among groups of artworks and these regularities can be exploited. Such regularities and commonalities are what permit scholars to refer to genres, schools, after all. Thus, in the two genres described above, we can say that every depiction of a skull refers to mortality and death and not a single one connotes youth, beauty, and the promise of a long healthy life. Likewise, in Christian art the dove refers to the Holy Spirit and never to the vice of lechery. It is imperative that computational methods exploit such regularities widely embraced by humanist art scholars. In different contexts, such objects may have different meanings, or none at all, so the scholarly use of computational approaches must be sensitive to such contexts.

Computer vision and AI will continue to aid art scholars and have already been decisive in resolving outstanding debates, for instance, over whether some artists in the early Renaissance secretly used optical aids [35, 37]. Close collaboration between art scholars and AI researchers will be increasingly valuable. Fortunately, a growing number of art scholars are open to exploring computer methods in their research, which bodes well for future collaboration between scientists and humanists [12, 33, 34].

Computational semantic analysis of art is in its infancy, and for many years will provide at best modest direct benefit to professional art scholars. By analogy, though, the first computer chess-, go-, and Texas hold 'em-playing systems were pathetically weak compared to average players, and especially to elite players [3, 32, 42]. Nevertheless, now artificial intelligence systems have surpassed every human player in these games and nearly every other board and card game, and continue to improve significantly far faster than the best human experts. Today every international chess grandmaster uses computational aids during training, and some even create their own computational systems [1].

The future value of computational semantic analysis will be improved search, digital curation, and analysis of fine art images, especially through large databases. Thus, ultimately art scholars will be able to ask new classes of questions about databases such as: What is the difference in number, distribution of Biblical stories in Catholic Spain and Protestant North, and how did it develop over the 17th Century? Moreover, computer tools serve as aid for periods and genres. Further, though, such an effort exposes deep processes of cognition, and how some artists exploit them through their works, that simply are not being addressed by mainstream AI. Just as with any new and powerful tools, it is hard to predict what additional problems in art scholarship computational extraction of meaning will serve. As such, these challenging problems posed by the analysis and especially high-level interpretation of art promise to be a fertile ground for continued development of techniques in artificial intelligence—indeed, these problems present a fresh and unique grand challenge to AI.

## ACKNOWLEDGMENTS

The author thanks the Getty Research Institute, Los Angeles, for access to its Research Library, where some of the above research was conducted, as well as three anonymous reviewers. Thanks also go to the National Gallery, London, the Hirshhorn Museum and Sculpture Garden, Washington, and the Uffizi Gallery, Florence, for permission to reproduce the paintings in Figures 1, 2, and 3, respectively.

## REFERENCES

- [1] Play Magnus. [n. d.]. Retrieved from [www.playmagnus.com/](http://www.playmagnus.com/).
- [2] Paul Barolsky. 2007. Vanitas painting and the celebration of life. *Source: Notes History Art* 26, 2 (2007), 38–39.
- [3] Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. 2015. Heads-up limit hold'em poker is solved. *Science* 347, 6218 (2015), 145–149.
- [4] Norman Bryson, Michael Ann Holly, and Keith Moxey (Eds.). 1991. *Visual Theory: Paintings and Interpretation*. HarperCollins, New York, NY.
- [5] Erik Cambria, Soujanya Poria, Alexander Gelbukh, and Mike Thelwall. 2017. Sentiment analysis is a big suitcase. *IEEE Intell. Syst.* 32, 6 (2017), 74–80.
- [6] K. R. Chowdhary. 2020. First order predicate logic. In *Fundamental of Artificial Intelligence*. Springer, New Delhi, Chapter 3, 51–88.
- [7] Bo Dai, Yuqi Zhang, and Dahua Lin. 2017. Detecting visual relationships with deep relational networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)*. 3076–3086.
- [8] Nikhil Damodaran, Sowmya Vishvanathan, Govind Divu, and Soman Kp. 2019. Scene classification using transfer learning: Theories and applications. In *Studies in Computational Intelligence*. Springer, Berlin, 363–399.
- [9] Patrick de Rynck. 2004. *How to Read a Painting: Lessons from the Old Masters*. Harry N. Abrams, New York, NY.
- [10] Jacobus de Voragine. 2018. *The Golden Legend* (Modern translation ed.). Patristic Publishing, Omaha, NE.
- [11] Richard O. Duda, Peter E. Hart, and David G. Stork. 2001. *Pattern Classification* (2nd ed.). John Wiley and Sons, New York, NY.
- [12] Ahmed Elgammal, Bingchen Liu, Diana Kim, Mohamed Elhoseiny, and Marian Mazzone. 2018. The shape of art history in the eyes of the machine. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2183–2191.
- [13] Donald Geman, Stuart Geman, Neil Hallonquist, and Laurent Younes. 2015. Visual Turing test for computer vision systems. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12 (2015), 3618–3623.
- [14] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press, Cambridge, MA.
- [15] Daniel Gordon, Aniruddha Kembhavi, Mohammad Rastegari, Joseph Redmon, Dieter Fox, and Ali Farhadi. 2018. IQA: Visual question answering in interactive environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. 4089–4098.
- [16] Ricardo Gudwin and João Queiroz. 2005. Towards an introduction to computational semiotics. In *Proceedings of the International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. 393–398.
- [17] David Harley. 1998. The moral symbolism of tobacco in Dutch genre painting. In *Ashes to Ashes*, Stephen Lock, Lois Reynolds, and E. M. Tansey (Eds.). Brill Publishers, Leiden, Netherlands, 78–88.
- [18] Anthony F. Janson. 1985. The convex mirror as vanitas symbol. *Notes in the History of Art* 4, 2/3 (1985), 51–54.
- [19] Yueying Kao, Ran He, and Kaiqi Huang. 2017. Deep aesthetic quality assessment with semantic information. *IEEE Trans. Image Process.* 26, 3 (2017), 1482–1495.
- [20] Raymond J. Kelly. 2006. *To Be, or Not to Be: Four Hundred Years of Vanitas Paintings*. Flint Institute Press, Flint, MI.
- [21] Sunder Ali Khawaja and Seok-Lyong Lee. 2020. Semantic image networks for human action recognition. *Int. J. Comput. Vision* 128, 2 (2020), 393–419.

- [22] Kristine Koozin. 1990. *The Vanitas Still Lifes of Harmen Steenwyck (1612–1656): Metaphoric Realism*. Vol. 1. Edwin Mellen Press, Lewiston, NY.
- [23] Walter Liedtke. 2007. *Dutch Paintings in the Metropolitan Museum of Art*. Metropolitan Museum of Art and Yale University Presses, New York, NY.
- [24] Xiao-Chang Liu, Ming-Ming Cheng, Yu-Kun Lai, and Paul L. Rosin. 2017. Depth-aware neural style transfer. In *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*. 1–10.
- [25] Prathmesh Madhu, Ronak Kosti, Lara Mühlberg, Peter Bell, Andreas Maier, and Vincent Christlein. 2019. Recognizing characters in art history using deep learning. In *Proceedings of the Workshop on Structuring and Understanding of Multimedia Heritage Contents*. 15–22.
- [26] John McCarthy. 2007. From here to human-level AI. *Artific. Intell.* 171, 18 (2007), 1174–1182.
- [27] Margaret Mitchell, Jesse Dodge, Amit Goyal, Kota Yamaguchi, Karl Stratos, Xufeng Han, Alyssa Mensch, Alexander Berg, Tamara Berg, and Hal Daumé III. 2012. Midge: Generating image descriptions from computer vision detections. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. 747–756.
- [28] Sandra Moriarty. 2005. Visual semiotics theory. In *Handbook of Visual Communication: Theory, Methods and Media*, Ken Smith, Sandra Moriarty, Gretchen Barbatsis, and Keith Kenney (Eds.). Lawrence Erlbaum, Mahwah, NJ, Chapter 15, 227–241.
- [29] Erwin Panofsky. 1955. *Meaning in the Visual Arts*. Doubleday, New York, NY.
- [30] Mary Piepmeyer. 2018. *The Appeal of Lemons: Appearance and Meaning in Mid Seventeenth-century Dutch Painting*. MS Thesis. University of North Carolina at Chapel Hill, Chapel Hill, NC.
- [31] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander Berg, and Fei-Fei Li. 2015. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision* 115, 3 (2015), 211–252.
- [32] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. 2017. Mastering the game of go without human knowledge. *Nature* 550 (2017), 354–359.
- [33] Emily L. Spratt and Ahmed Elgammal. 2014. The Digital Humanities Unveiled: Perceptions Held by Art Historians and Computer Scientists about Computer Vision Technology. Retrieved from <https://arxiv.org/abs/1411.6714>.
- [34] David M. Stone and David G. Stork. 2011. Computer assisted connoisseurship: The interdisciplinary science of computer vision and image analysis in the study of art. In *Proceedings of the Image Processing for Art Investigations (IP4AI'11)*. Museum of Modern Art, Museum of Modern Art, New York, NY.
- [35] David G. Stork. 2004. Optics and realism in renaissance art. *Sci. Amer.* 291, 6 (2004), 76–84.
- [36] David G. Stork, Anthony Bourached, George H. Cann, and Ryan-Rhis Griffiths. 2021. Computational identification of significant actors in paintings through symbols and attributes. In *Computer Vision and Analysis of Art*, David G. Stork and Kurt Heumiller (Eds.). IS&T, San Francisco, CA.
- [37] David G. Stork, Jacob Collins, Marco Duarte, Yasuo Furuichi, David Kale, Ashutosh Kulkarni, M. Dirk Robinson, Christopher W. Tyler, Sara Schechner, and Nicholas Williams. 2011. Did early renaissance painters trace optically projected images? The conclusion of independent scientists, art historians, and artists. In *Digital Imaging for Cultural Heritage Preservation*, Filippo Stanco, Sebastiano Battiatto, and Giovanni Gallo (Eds.). CRC Press, Boca Raton, FL, Chapter 8, 379–407.
- [38] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H. S. Torr, and Timothy M. Hospedales. 2018. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. 1199–1208.
- [39] Valerio Targon. 2018. Toward semiotic artificial intelligence. *Procedia Comput. Sci.* 145 (2018), 555–563.
- [40] Son N. Tran and Artur S. d'Avila Garcez. 2018. Deep logic networks: Inserting and extracting knowledge from deep belief networks. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 2 (2018), 246–258.
- [41] Josep Valls-Vargas, Santiago Ontañón, and Jichen Zhu. 2015. Toward character role assignment for natural language stories. In *Intelligent Narrative Technologies*. 101–104.
- [42] H. Japp van den Herik. 2018. Computer chess: From idea to DeepMind. *ICGA J.* 40, 3 (2018), 160–176.
- [43] Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter (Eds.). 2008. *Handbook of Knowledge Representation*. Elsevier, Amsterdam, Netherlands.
- [44] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Dean Wierstra. 2016. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, Daniel Lee, Masashi Sugiyama, Ulrich von Luxburg, Isabelle Guyon, and Roman Garnett (Eds.), Vol. 29. Curran Associates, Red Hook, NY, 3630–3638.
- [45] Stefano Zuffi. 2010. *How to Read Italian Renaissance Painting*. Harry N. Abrams, New York, NY.

Received 31 January 2021; revised 17 June 2021; accepted 17 June 2021