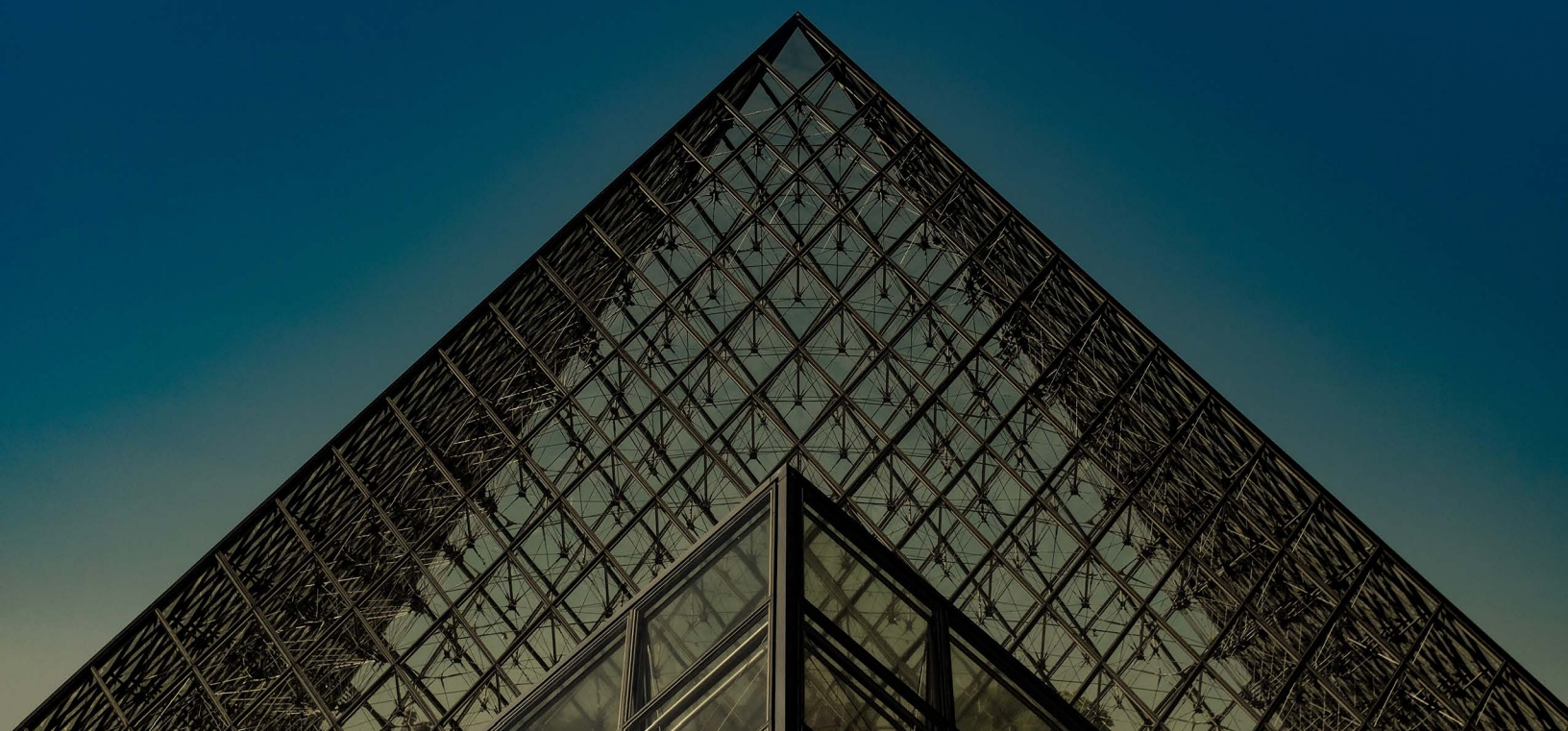




the museum review

VOLUME 5 NUMBER 1 2020

WWW.THEMUSEUMREVIEW.ORG
ISSN 2574-0296



AI sees what? The good, the bad, and the ugly of machine vision for museum collections

BRENDAN CIECKO

Cuseum

The Museum Review

www.TheMuseumReview.org

Rogers Publishing Corporation NFP
5558 S. Kimbark Ave., Suite 2, Chicago, IL 60637
www.RogersPublishing.org

Cover photo: Louvre Museum, Unicorn Cloud.
Image by: Zoltan Tasi.

©2020 The Museum Review

The Museum Review is a peer reviewed Open Access Gold journal, permitting free online access to all articles, multi-media material, and scholarly research. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



AI sees what? The good, the bad, and the ugly of machine vision for museum collections

BRENDAN CIECKO
Cuseum

Keywords Artificial intelligence; machine vision; museum collections

Abstract Recently, as artificial intelligence (AI) has become more widespread and accessible, museums have begun to make use of this technology. One tool in particular, machine vision, has made a considerable splash in museums in recent years. Machine vision is the ability for computers to understand what they are seeing. Although the application of machine vision to museums is still in its early stages, the results show promise. In this article, we will explore the strengths and successes of this new technology, as well as the areas of concern and ethical dilemmas it produces as museums look toward machine vision as a move to effortlessly aid in generating metadata and descriptive text for their collections. Over several months, Cuseum collected data on how machine vision perceives collection images. This study represents a sustained effort to analyze the performance and accuracy of various machine vision tools (including Google Cloud Vision, Microsoft Cognitive Services, AWS Rekognition, etc.) at describing images in museum collection databases. Cuseum's study represents technical analysis, data collection, and interpretation in order to spark a discussion around machine vision in museums and to encourage the community to engage with ongoing ethical considerations related to this technology. While machine vision may unlock new potentials for the cultural sector, when it comes to analyzing culturally-sensitive artifacts, it is essential to scrutinize the ways that machine vision can perpetuate biases, conflate non-Western cultures, and generate confusion.

About the Author Brendan Ciecko is the founder and CEO of Cuseum, a platform that helps museums and cultural organizations engage their visitors, members, and patrons. Ciecko has been building technology since age 11, and has been recognized by Inc. Magazine as one of America's top entrepreneurs under 30. Ciecko has been featured in The New York Times, WIRED, Fast Company, Entrepreneur, TechCrunch, VentureBeat, Esquire, and PC Magazine for his work in design, technology, and business.

*This article was published on April 27, 2020 at www.TheMuseumReview.org
A version of this article was published in MW 2020, January 15, 2020.*

Artificial intelligence (AI) is already reshaping all aspects of society, business, and culture. From offering personalized Netflix recommendations to auto-completing sentences in Gmail, AI underlies routine aspects of our lives in ways the public does not realize.

AI has transformed the commercial sector in myriad ways. While chatbots and predictive engines may be familiar, AI goes far beyond this. From offering contextual marketing messaging, transferring and cross-referencing data, deciding personal injury claims for insurance firms, to enabling financial fraud detection, innovative applications of artificial intelligence technology are popping up routinely.

Recently, as AI has become more widespread and accessible, museums have begun to make use of this technology. One tool in particular, machine vision, has made a considerable splash in museums in recent years. Machine vision is a computer's ability to understand what it is

seeing. Although the application of machine vision to museums is in its early stages, the results show promise. In this paper, we will explore the strengths and successes of this new technology, as well as the areas of concern and ethical dilemmas it produces as museums look toward machine vision as a move to aid in the generation of metadata and descriptive text for their collections.

To advance our understanding of machine vision's potential impacts, over the course of several months, Cuseum collected data about how machine vision perceives collection images. This study represents a sustained effort to analyze the performance and accuracy of various machine vision tools (including Google Cloud Vision, Microsoft Cognitive Services, and AWS Rekognition) at describing images in museum collection databases. In addition to thoroughly assessing the AI-generated outputs, Cuseum shared the results with several museum curators and museum digital technology specialists, collecting expert commentary from the museum professionals on the fruits of this research.

The Cuseum study represents over 100 hours of time invested in technical analysis, data collection, and interpretation to help advance the conversation in the museum, art, and cultural heritage field.

In conjunction with strides in digitizing collections, moving toward open access, and linking open data, as well as the growing application of emerging digital tools across the museum sector, machine vision has the potential to accelerate the value created from these important foundational initiatives.

The goal of Cuseum's exploration of this technology is to spark a discussion around machine vision in museums, and to encourage the community to engage with ongoing ethical considerations related to this technology. While machine vision may unlock new potentials for the cultural sector, it is essential to scrutinize the ways that machine vision can perpetuate biases, conflate non-Western cultures, and generate confusion. when analyzing culturally-sensitive artifacts

What is machine vision?

Machine vision is quickly becoming one of the most important applications of artificial intelligence (Cognex, 2019). In the most simple terms, machine vision can be understood as "the eyes of a machine." According to Forbes, this technology has a variety of applications in business including for "quality control purposes," and helping businesses in many ways today for "identification, inspection, guidance and more" (Marr, 2019). Machine vision is the underlying technology behind facial recognition, such as that of Facebook face tagging and Apple's novel methods of unlocking iPhones, Google's Lens for visual search, and even autonomous vehicles. Like many emerging technologies, the average consumer is likely to interact with machine vision daily and might not even know it.

Though it may appear that machine vision has only recently emerged, this technology has been in development since the 1960s (Papert, 1966). Now, almost sixty years later, we are still developing this technology and unlocking interesting new use cases. Every major technology company has leveraged machine vision to advance their own products and

services, and have made their platforms commercially available to enhance the appeal and power of their cloud-computing solutions.

Machine vision in museums

Recently, as machine vision and AI have become more widespread and accessible, museums have also begun to make use of this technology. Several museums, including The Metropolitan Museum of Art (New York), the Barnes Foundation (Philadelphia, PA), and Harvard Art Museums (Cambridge, MA), have employed machine vision to analyze, categorize, and interpret their collection images. Although the application of machine vision to museums is in its early stages, the results show promise.

From basic subject detection to complex semantic segmentation aided by deep learning, optical character recognition, and color composition, there are different ways in which machine vision can be used. As accuracy improves and more sophisticated models of machine vision are developed, it will almost certainly change the way museum collection images can be explored, dissected, and disseminated.

Museums often have thousands of objects logged in their collection databases, with limited information about the objects. For collections to become easier to analyze and search, it is essential to collect or generate metadata on these objects.

What is metadata?

In simple terms, metadata is data that describes data. In television and film, a piece of data might be the name of a movie. Increasingly sophisticated metadata can be collected and attached to a film based on the content, theme, and emotive response it may generate. For example, a movie can be tagged with “satire,” “war,” and “comedy.” This means that when a user logs onto Netflix, for example, to search for a movie, options can be suggested to fit the viewer’s interests based on algorithmic analytics paired with the wealth of metadata associated with the viewer’s past viewing behavior.

This is equally important in the context of museums. Metadata can reflect three different features about objects: content, context, and structure (Baca, 2008). Creating robust datasets that describe museum collections is essential because without them, even open-access collections are limited in value. Only a robust tagging system that describes various features and contexts of artworks and artifacts enables them to become searchable and discoverable within databases. In other words, metadata can amplify the value of existing data sets, making them usable by researchers, curators, artists, historians, and the public.

There are an increasing number of approaches available to generate and expand metadata within museum collections. Over a decade ago, a project called “steve.museum” began as a cross-institutional experiment in the world of “social tagging,” also known as folksonomy, for museum objects. Over the course of two years, over 2,000 users generated over 36,000 tags across 1,782 works of art (Leason, 2009). While that initiative ended, the ambition around enriching metadata through external sources remains. Even now, museums like the Philadelphia Museum of Art include user-generated social tags on their object pages. Now, advanced machine vision is emerging as a promising tool to *automatically* generate

discoverable descriptive text around museum collections with very little limitation on speed or scale.

Generating metadata and descriptive text

Machine vision has become advanced enough to detect the subject matter and objects depicted in any type of 2-dimensional or 3-dimensional object, including paintings, photographs, and sculptures. This can help expand and enrich existing meta tags, as well as fill gaps where meta tags are lacking or completely void. For instance, many objects in museums collections are logged simply as “Untitled” or “Plate” or “Print” even if the objects contain many significant identifying details. These objects are virtually invisible on databases as they lack any level of sufficient terms to aid in their discoverability.

Such an object might be of unknown origin, yet contain important images, symbols, carvings, or details. To make such objects discoverable for research purposes, it is essential that they are tagged with information that can offer greater insight and specificity into their visual contents.

Nothing in the Metropolitan Museum’s metadata or description of a “Plate” makes it searchable via the term “horse,” which is the plate’s focal point (figure 1). Yet, Microsoft Computer Vision tags this image with the term “horse,” and does so with 98% confidence. This means that the plate can become searchable and discoverable based on its visual elements, rather than just a title offering little specific information. An image of an object entitled “Plate” in the Metropolitan Museum’s collection contains no other metadata or description (figure 2). Microsoft Machine Vision returns object detection that there are two people depicted in the work, with a high level of confidence that they are women. Google Vision gets even more specific, tagging this image with the term “geisha.”



Figure 1. “Plate” ca. 1820-50, American. The Metropolitan Museum of Art.



Figure 2. “Print” by Utagawa Kunisada. Edo Period, Japan. The Metropolitan Museum of Art.

This type of precise tagging is key to making objects that are functionally invisible some level of discoverable. These examples of objects simply identified as “plate” or “print” are representative of millions of objects in international museum collections. Digitizing collections and making them “open access” is the first step in making collections more accessible. To unleash the untapped potential of digital collections, and to augment and transform human knowledge of cultural relics, it is necessary to make data *searchable*. Machine vision is increasingly making this possible.

Over the past few years, museums including the Museum of Modern Art (MoMA) (New York), San Francisco Museum of Modern Art (SFMOMA), the Barnes Foundation, Harvard Art Museums, Auckland Art Gallery (Australia), and National Museum in Warsaw (Poland) have made headlines for taking advantage of machine vision to enrich and supplement their metadata. The early application of this technology in these museums has already shown enormous promise.

Research results: how well does machine vision perform?

It has become clear that machine vision has a number of clear and beneficial use cases with regard to museum collections. The question now arises: how well does machine vision do? Can it offer accurate tags? Is the metadata generated useful and correct?

In recent years, the accuracy of machine vision has improved significantly. According to research by Electronic Frontier Foundation, which measures the progress of artificial intelligence, the error rate has fallen from ~30% (2010) to ~4% (2016), making the error rate on-par with humans (figure 3).

In 2016, members of the Cuseum team began to explore the capabilities of machine vision in museums, and published initial findings and predicted use cases. Years later, Cuseum is now expanding upon that primary investigation.

Over the course of several months, Cuseum collected data on how machine vision perceives collection images. This study represents a sustained effort to analyze the performance and accuracy of various machine vision tools (including Google Cloud Vision, Microsoft Computer Vision, AWS Rekognition, etc.) at describing images in collection databases at the Metropolitan Museum of Art (New York), Minneapolis Institute of Art, Philadelphia Museum of Art, and the Art Gallery of Ontario (Toronto).

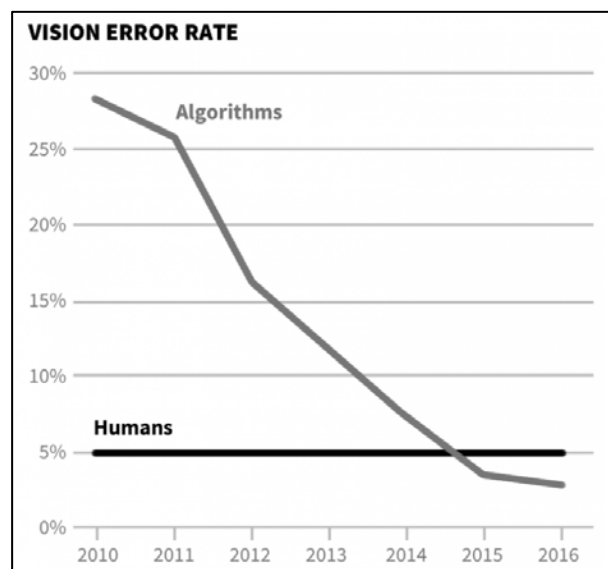


Figure 3. Vision Error Rate graph, Electronic Frontier Foundation.

By running a set of digitized collection images from each of these institutions through six major computer vision tools, the study assessed the accuracy, potential, and limitations of a range of machine vision platforms.

Which machine vision services were evaluated?

Numerous commercially available machine vision services exist. Many of these services offer free trials and are accessible – users do not require advanced computer science experience or access to sophisticated hardware in order to use these. It is possible to tap into the power of these services via online interfaces, APIs, and other easy-to-use methods.

These six machine vision solutions were selected for testing based on Cuseum’s past familiarity and experience:

- Google Cloud Vision
- Microsoft Cognitive Service
- IBM Watson Visual Recognition
- AWS Rekognition
- Clarifai
- CloudSight

Recent third-party industry research and evaluation by Forrester Research reinforces our overall selections, with the exception of CloudSight (figure 4).



Figure 4. Computer vision platforms, Q4 2019. The Forrester New Wave.

All machine vision services were given the same image files, and were tested in an “as-is,” “untrained” capacity; the services were not fed the images or their accurate adjacent data sets in advance.

Successful results

One of the greatest merits demonstrated by this study was the ability of machine vision to accurately identify places and people depicted in works of art. Microsoft Computer Vision successfully identified Canaletto's "Piazza San Marco" painting as "a group of people walking in front of Piazza San Marco" (figure 5). Other machine vision services offered similarly accurate, yet less specific tags, like "building," "architecture," "tower," "urban," "plaza," and "city." Microsoft Computer Vision was also able to recognize Augustus John's "The Marchesa Casati" painting as "Luisa Casati looking at the camera" (figure 6). Other machine vision services tagged this image as "woman in white dress painting" and "retro style," and most understood this image to be a painting of a person.



Figure 5. "Piazza San Marco" by Canaletto. 1720s. The Metropolitan Museum of Art.



Figure 6. "The Marchesa Casati" by Augustus Edwin John. 1919. Art Gallery of Ontario.

Machine vision proved an excellent tool for identifying key pieces of information like style and time period. One prime example is the Doryphoros, a sculpture at the Minneapolis Institute of Art (figure 7). Microsoft Computer Vision returned the description "a sculpture of a man." Services like Clarifai and Google Vision were able to identify this as a "classic" object. The sculpture was overall accurately examined by machine vision with the majority of machine vision services labeled the object with tags such as "art," "standing," "male," "human body," "sculpture," "person," "statue," "marble," "nude," and other accurate terms.

One of the hypothesized limitations of machine vision was its limited capacity to flag and tag more abstract works of art, which is a more challenging task than identifying photographs, portraits, or landscapes. All services understood that Emily Carr's "Upward Trend" was a work of art, a painting, etc. (figure 8). Yet CloudSight was able to identify this as "green and blue abstract painting," while Google Cloud Vision tagged this as "painting," "acrylic paint," "art," "water," "watercolor paint," "visual arts," "wave," "modern art," "landscape," and "wind wave."

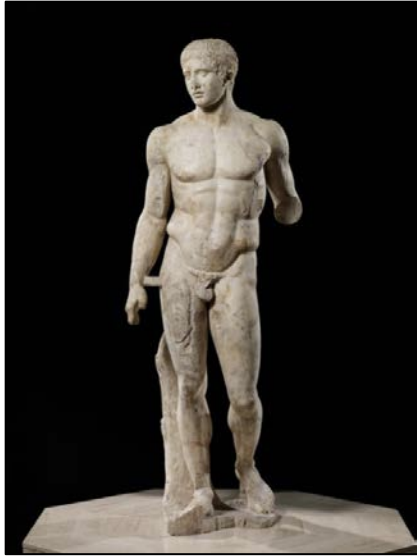


Figure 7. "The Doryphoros" by Unknown Roman. 120-150 BCE. Minneapolis Institute of Art.



Figure 8. "Upward Trend" by Emily Carr. 1937. Art Gallery of Ontario.

Machine vision proved remarkably successful at identifying religious, particularly Christian, iconography. Various tools tagged the Metropolitan Museum's "Madonna and Child Enthroned with Saints" painting accurately, using phrases that included "saint," "religion," "Mary," "church," "painting," "God," "kneeling," "cross," "chapel," "veil," "cathedral," "throne," "aura," and "apostle" (figure 9).

In general, machine vision tools produce more accurate tags for photographs, as opposed to paintings or sculptures, for the simple reason that most algorithms and programs are primarily developed using photographs. Microsoft Computer Vision gave a particularly apt description of this photograph by Edward Burtynsky of the Art Gallery of Ontario, labeling it: "A large brick building with grass in front of a house with Art Gallery of Ontario in the background" (fig. 10).



Figure 9. "Madonna and Child Enthroned with Saints" by Raphael. ca. 1504. The Metropolitan Museum of Art.



Figure 10. "Art Gallery of Ontario: Toronto View from Grange Park" by Edward Burtynsky. 2008. Art Gallery of Ontario.

Poor accuracy

While machine vision tools proved adept at recognizing and generating accurate metadata on certain kinds of images, in other cases, these tools produced misleading tags. Many machine vision tools mistook El Greco's "Christ Driving the Money Changers from the Temple" as a screensaver (figure 11). It is possible that the machine vision services are picking up on all the colors and many images typically labeled as "screensaver" in training datasets often based on monetizable use cases and products, which is why this work may be mistaken for a brightly-colored LCD. Further, this painting contains many complexities and figures, rather than one object of focus. In general, machine vision struggles more to describe such works accurately. In these cases, the services will cast a wide net, which can generate conflicting tags. In casting a wide net, some of the tags still turn out to be accurate. For example, one machine vision service tagged the El Greco as "Renaissance" and "Baroque." This work, authored in 1568, indeed straddles the Renaissance and Baroque periods in European art history.



Figure 11. "Christ Driving the Money Changers from the Temple" by El Greco. ca. 1570. Minneapolis Institute of Art.

Abstract art is one area where machine vision proves to be less accurate. "Some/One" by Do Ho Suh is an abstract sculpture inspired by the artist's time in the Korean military (figure 12). It resembles a jacket or uniform. Machine vision did manage to generate some accurate tags, including "design," "fashion," "decoration," "chrome," "sculpture," "metal," "silver," and "art." However, machine vision programs also returned completely inaccurate tags. Microsoft Computer Vision offered labels such as "elephant," "desk," "mouse," "cat," "computer," "keyboard," and "apple." IBM Watson Visual Recognition generated similarly inaccurate tags, including, "pedestal table," "candlestick," "propeller," "mechanical device," "ax," "tool," "cutlery." In cases like these, machine vision struggles to find anything non-abstract other than the primary *material* of this object. Many tools were able to understand the object as metallic in composition; however, this resulted in a series of inaccurate associations with



Figure 12. "Some/One" by Do Ho Suh. 2005. Minneapolis Institute of Art

common objects of a similar material. Amazon Rekognition flagged this image as “sink faucet,” Google Vision as “bar stool”, and CloudSight as “gray and black leather handbag.”

Minneapolis Institute of Art’s “Jade Mountain” sculpture, while not abstract, is more monochrome with details difficult to decipher by a computer (figure 13). Of the six different machine vision services used, four returned terms related to “food,” “cake,” and/or “ice cream.” It is likely that services are matching the picture with ice cream due to similarities in shape and color, while completely missing the small details of trees and houses due to their limited contrast. Only Google Vision returned remotely satisfactory terms, including “stone carving,” “sculpture,” “carving,” “rock,” “figurine.”

A brazier, a vessel used for heating or cooking, is uncommon within the image datasets used to train various machine vision services. These image datasets likely have very few braziers but quite a few tables, due to their common occurrence in everyday life and in e-commerce datasets. Additionally, their shapes are similar enough that all of the machine vision tools mistook the “Brazier of Sultan al-Malik al-Muzaffar Shams al-Din Yusif ibn ‘Umar” from the Metropolitan Museum of Art for a table (figure 14). While humans may quickly understand context and scale, computers do not yet have this ability. Even though this brazier is not a common object we see, the human eye is likely to discern that it is a diminutive object (13” width x 12” depth x 16” height), and smaller than a typical table. Only Clarifai returned useful terms such as “metalwork,” “art,” “antique,” “gold,” “ornate,” “luxury,” “gilt,” “ancient,” “bronze,” and “wealth.” All of the other machine vision services returned incorrect tags and terms related to furniture, like “brown wooden wall mount rack,” “furniture,” “settee,” and “a gold clock sitting on top of a table.”



Figure 13. “Jade Mountain Illustrating the Gathering of Scholars at the Lanting Pavilion” by Unknown Artist. 1790, China. Minneapolis Institute of Art.



Figure 14. “Brazier of Sultan al-Malik al-Muzaffar Shams al-Din Yusif ibn ‘Umar” by Unknown Artist. 13th c., Egypt. The Metropolitan Museum of Art.

Problematic results

While machine vision showed great promise for many works of art, research illustrates its limitations and biases, particularly gender and cultural biases. With significant efforts being made towards equity, diversity, and inclusion, the chance of an insensitive or potentially offensive tag presents new risks for museums.

Gender bias

Take, for example, two portraits by Chuck Close (figures 15 and 16). “Big Self Portrait” depicts the artist with a cigarette in his mouth, and a second portrait depicts a young woman with a cigarette in her mouth. Both individuals have nearly identical expressions. While the man was tagged by Clarifai using descriptors like “funny” and “crazy,” the woman was tagged by the same tool as “pretty,” “cute,” and “sexy.” Chuck Close’s painting at the Minneapolis Institute of Art, entitled “Frank,” is similar (figure 17). Clarifai tagged the work, as “writer” and “scientist,” drawing attention to the fact that men in glasses are often associated with intellectualism. When examining a similar work by Close, “Susan,” depicting a woman in glasses, Clarifai flagged the work as “model,” “smile,” “pretty,” “beautiful,” “cute,” and “actor” (figure 18).



Figure 15. “Big Self-Portrait” by Chuck Close. 1967-68. Walker Art Center.



Figure 16. “Untitled” portrait of a woman, Chuck Close.



Figure 17. “Frank” by Chuck Close. 1969. Minneapolis Institute of Art.



Figure 18. “Susan” by Chuck Close. 1971. Museum of Contemporary Art Chicago.

Western cultural bias

While machine vision tools were able to frequently identify Christian iconography, they were inept in identifying non-Western art, particularly Asian and African art. Take, for example, the Yoruba terracotta shrine head at the Minneapolis Institute of Art (figure 19). This Nigerian work was mistaken as “Buddha” across various machine vision services. The Japanese samurai suit of armor at the MIA was also mistaken as “Buddha” by machine vision services (figure 20). This suggests a pattern of conflating non-Western cultures.



Figure 19. “Shrine Head” by Unknown Yoruba Artist. 12th-14th c., Nigeria. Minneapolis Institute of Art.



Figure 20. “Red-and-blue-laced Suit of Armor from the Kii Tokugawa Family” by Unknown Artist. Mid-17th c, Japan. Minneapolis Institute of Art.

Delicate topics

Should museums set boundaries around what types of images are analyzed by machine vision? With complex and challenging topics including colonization, slavery, genocide, and other forms of oppression, it may be advised that the use of machine vision be avoided altogether.

During the Yale-Smithsonian Partnership’s “Machine Vision for Cultural Heritage & Natural Science Collections” symposium in 2019, Peter Leonard, Director of Yale’s Digital Humanities Lab, discussed scenarios where machine vision could go wrong. Leonard ran an image from the Museum of Applied Arts and Sciences (Sydney, Australia) through a machine vision service that returned terms including “Fashion Accessory” and “Jewelry,” when in fact the object was iron ankle manacles from Australia’s convict history (figure 21). Leonard noted, “you can only imagine the valence of this in an American context with African American history” (Leonard, 2019). In the near future it is likely that museums dedicated to non-Western art and culture or those focused on objects of a sensitive nature or relating to historically marginalized

communities will steer clear of machine-generated metadata and descriptive text. However, every type of museum should consider these potential risks and plan accordingly.



Figure 21. “Convict iron leg manacles” by Unknown. ca. 1772- 1886. Australia. Museum of Applied Arts and Sciences.

Potential for success through proper training and hybrid models

Despite success in certain areas, no single machine vision service performs with complete or near complete accuracy at this time. If a museum wants to leverage machine vision and increase overall precision, could the answer be as simple as training with large sets of high-quality, human-verified data from cultural institutions? Or could one aggregate the output and results across a plurality of machine vision tools, and only accept terms of a specific frequency, the threshold of accuracy, and/or human verification via Mechanical Turks, volunteers, or the broader public?

Hybrid approaches will likely emerge that allow museums to easily test a machine vision model with their collection images and adjacent data, and leverage human verification and the results of multiple machine vision services.

How have museums begun to use machine vision?

Given the enormous potential of various machine vision services to assist in generating metadata for museum collections, institutions have begun to harness these tools in effective and creative ways. Examples include the following.

Harvard Art Museums

Harvard Art Museums exhibits one of the best displays of machine vision to generate metadata in museums. The museums use multiple machine vision tools to start tagging the 250,000 works in the collections, with the hope of eventually using “AI-generated descriptions as keywords or search terms for people searching for art on Harvard’s databases” (Yao, 2018).

Museum of Modern Art

The Museum of Modern Art (New York) partnered with Google Arts and Culture “to comb through over 30,000 exhibition photos” using machine vision. “A vast network of new links” were created between MoMA’s exhibition history and the online collection (Google).

Cleveland Museum of Art

Cleveland Museum of Art’s Art Explorer is powered by Microsoft’s Cognitive Search, which uses AI algorithms to enrich the metadata for the artworks.

The Barnes Foundation

The Barnes Foundation has a program that interprets and pairs digital works of art together to recognize art style, objects, and other basic elements (Jones, 2018). It is a notable step in art-historical analysis.

Auckland Art Gallery

One recent example of the merits of AI-generated tags is seen at the Auckland Art Gallery (Australia). This organization has utilized more than 100,000 human-sourced and machine-generated tags to categorize works of art as part of a larger chatbot initiative (Auckland Art Gallery, 2018).

An array of museums including the Art Institute of Chicago, the Cooper Hewitt, M+ (Hong Kong), as well as Google Art and Culture, and Artsy, have offered new pathways into their collections by leveraging machine-extracted color metadata like palette, partitions, and histogram data.

Perspectives from the museum community

With any new technology, a variety of perspectives and opinions follow. Artificial intelligence is a growing topic of interest amongst museum technologists, the art world, and curators alike. In SFMOMA’s renowned project, Send Me SFMOMA, the museum explored the possibility of leveraging machine-generated tags, but found the terms to be too formal, literal, and uninspired. In an interview with Vox, a museum representative remarked that “the intuition and the humanness of the way that our staff has been tagging” is what is interesting, “versus the linearity of the computer vision approach, [which] just makes you miss out on all of the sublime” (Grady, 2017).

Taking a contrary point of view, Jeff Stewart, Director of Digital Infrastructure and Emerging Technology at Harvard Art Museums, believes that machine-extracted terms and vocabularies can sometimes be more human than the lofty or excessively intellectual statements written by an academic or a curator. Indeed, a painting in the Harvard collection, “Still Life with Watermelon” by Sarah Miriam Peale, was described as “juicy,” “sweet,” and “delicious,” suggesting machine vision’s ability to provide approachable and sensational dynamics to the equation in ways that curators may not.

Upon receiving a computer-generated description of Miro’s “Dog Barking at the Moon,” a notable work from the Philadelphia Museum of Art’s collection, Michael Taylor, Chief Curator and Deputy Director of Virginia Museum of Fine Arts, shared his positive and amused reaction.

The machine-generated description was “white, blue and brown two legged animated animal near ladder illustration,” which is fairly accurate to the literal depiction, but does not capture the conceptual, abstract nuances, or well-studied interpretation of the work that a curator can provide. Compare that to Taylor’s description of the same work:

At once engaging and perplexing, Joan Miró’s “Dog Barking at the Moon” exemplifies the Spanish artist’s sophisticated blend of pictorial wit and abstraction. Like many of the works that the artist painted in Paris, this work registers Miró’s memories of his native Catalan landscape, which remained the emotional center and source of his imagery for much of his life. The work’s genesis can be found in a preparatory sketch showing the moon rejecting a dog’s plaintive yelps, saying in Catalan, “You know, I don’t give a damn.” Although these words were excluded in the finished painting, their meaning is conveyed through the vacant space between the few pictorial elements that compose this stark, yet whimsical image of frustrated longing and nocturnal isolation. Against the simple background of the brown earth and black night sky, the artist has painted a colorful dog, moon, and a ladder that stretches across the meandering horizon line and recedes into the sky, perhaps suggesting the dream of escape. This remarkable combination of earthiness, mysticism, and humor marks Miró’s successful merging of international artistic preoccupations with an emphatically regional outlook to arrive at his distinctively personal and deeply poetic sensibility.

One could surmise and agree that artificial intelligence and computer vision are currently unable to deliver a comparable description, rich with art historical context and deep interpretation.

Machine vision bias: ethical considerations

While machine vision may unlock new potential for the cultural sector, it is essential to scrutinize the ways that machine vision can perpetuate biases, conflate non-Western cultures, and generate confusion. Two recent projects amplified the topics of AI bias within the art and cultural worlds. Google’s Art Selfie project, an app that matched a user’s face with a similar work of art, was a viral phenomenon in 2018, yet it faced criticism from people of color due to limited results, many of which exemplified racial stereotypes or otherwise produced inappropriate and offensive “lookalikes.”

In 2019, an art project by researcher Kate Crawford and artist Trevor Paglen called “Training Humans” sparked a dialogue about the problematic bias of facial recognition software. The project led ImageNet, one of the leading image databases used to train the machine vision model, to remove more than half a million images.

Advancements in AI technology will help but not necessarily solve these issues. In fact, one prediction by technologist Roman Yampolskiy published in the *Harvard Business Review* warned that “the frequency and seriousness of AI failures will steadily increase as AIs become more capable.” However, there is an upside. According to the *New York Times*, “biased algorithms are easier to fix than biased people” (Cook, 2019). While it can prove difficult to

“reprogram” our hearts and minds, software can be updated when biases are uncovered. This suggests that discrimination and bias in AI can be detected and remedied, helping to overcome some of its biggest challenges.

Conclusion

Cuseum’s broad research and experience suggest that technologies like machine vision are getting better, faster, and more accessible, and that AI’s bias is wholly recognized and is being addressed. In the years to come, we can anticipate an increase in the use of machine vision in museum collections, as well as the heightened accuracy and decreased bias of machine vision tools. This has the potential to make collections discoverable in new ways, unlocking the full value of digitization initiatives by creating a body of metadata that will make collections exponentially more searchable. According to Hans Ulrich Obrist, curator and co-director of The Serpentine Galleries, “we need new experiments in art and technology” (Selvin, 2018). Machine vision may prove one of the key tools that will advance the museum field.

It is equally important to consider the potential of AI and machine vision to generate new insights that are different or that go beyond what a human eye or mind might generate. According to technologist and AI expert Amir Husain in his book *The Sentient Machine*, “Too often, we frame our discussion of AI around its anthropological characteristics: How much does it resemble us?” Further adding “Do we really imagine that human intelligence is the only kind of intelligence worth imitating? Is mimicry really the ultimate goal? Machines have much to teach us about ‘thoughts’ that have nothing to do with human thought” (Husain, 2017).

This introduces the idea that the end goal of harnessing machine vision in museums may not even be about *mirroring* what a human, curator, educator, or scholar could do. Machine vision opens up doors for a new kind of analysis, and introduces a different type of interpretation and understanding of a work that may or may not reflect the cognitive limitations or learning frameworks of the human mind.

In the next decade, the computing power and abilities of machine vision will be great multiples more significant than they are today. Today machine vision is just at the beginning. If prioritized, budding partnerships between museums and technology companies will help build models and algorithms for cultural and artistic use, and help alleviate some of the obstacles holding the sector back today.

In anticipation of coming advances, now is the time to act and steer the future of collections. To ensure the accurate and ethical application of machine vision to museums, standards and policies that will guide how this technology is employed must be set and followed. By entering into a thoughtful dialogue, collections management practice can be reshaped to enable discovery and object analysis on a new level. To maximize the position of museums in this rapidly changing landscape, there is no better time to discuss, challenge, and explore the value of new technology.

Museums should proceed with caution and thoughtful consideration, but should not act out of fear. The more organizations and key stakeholders that are involved in this exploration, the greater the value that will be created and shared across the museum field. This will assist in

making cultural heritage accessible to as many people as possible. This journey remains in the preliminary and experimental stages. The future of museums will be many things – and will require both vision and machines to get there.

List of Figures

Figure 1. “Plate” ca. 1820-50, American. The Metropolitan Museum of Art.

Figure 2. “Print” by Utagawa Kunisada. Edo Period, Japan. The Metropolitan Museum of Art.

Figure 3. Vision Error Rate graph, Electronic Frontier Foundation.

Figure 4. Computer vision platforms, Q4 2019. The Forrester New Wave.

Figure 5. “Piazza San Marco” by Canaletto. 1720s. The Metropolitan Museum of Art.

Figure 6. “The Marchesa Casati” by Augustus Edwin John. 1919. Art Gallery of Ontario.

Figure 7. “The Doryphoros” by Unknown Roman. 120-150 BCE. Minneapolis Institute of Art.

Figure 8. “Upward Trend” by Emily Carr. 1937. Art Gallery of Ontario.

Figure 9. “Madonna and Child Enthroned with Saints” by Raphael. ca. 1504. The Metropolitan Museum of Art.

Figure 10. “Art Gallery of Ontario: Toronto View from Grange Park” by Edward Burtynsky. 2008. Art Gallery of Ontario.

Figure 11. “Christ Driving the Money Changers from the Temple” by El Greco. ca. 1570. Minneapolis Institute of Art.

Figure 12. “Some/One” by Do Ho Suh. 2005. Minneapolis Institute of Art

Figure 13. “Jade Mountain Illustrating the Gathering of Scholars at the Lanting Pavilion” by Unknown Artist. 1790, China. Minneapolis Institute of Art.

Figure 14. “Brazier of Sultan al-Malik al-Muzaffar Shams al-Din Yusif ibn ‘Umar” by Unknown Artist. 13th c., Egypt. The Metropolitan Museum of Art.

Figure 15. “Big Self-Portrait” by Chuck Close. 1967-68. Walker Art Center.

Figure 16. “Untitled” portrait of a woman, Chuck Close.

Figure 17. “Frank” by Chuck Close. 1969. Minneapolis Institute of Art.

Figure 18. “Susan” by Chuck Close. 1971. Museum of Contemporary Art Chicago.

Figure 19. "Shrine Head" by Unknown Yoruba Artist. 12th-14th c., Nigeria. Minneapolis Institute of Art.

Figure 20. "Red-and-blue-laced Suit of Armor from the Kii Tokugawa Family" by Unknown Artist. Mid-17th c, Japan. Minneapolis Institute of Art.

Figure 21. "Convict iron leg manacles" by Unknown. ca. 1772-1886, Australia. Museum of Applied Arts and Sciences.

References

"Art+Tech Summit at Christie's: The A.I. Revolution." *Art+Tech Summit at Christie's 2019*. Accessed January 2020. <https://www.christies.com/exhibitions/art-tech-summit-ai-revolution>

Auckland Art Gallery (2018). "Auckland Art Gallery's new chatbot demonstrates artificial intelligence to give new access to 17,000 artworks." Accessed January 2020. <https://www.aucklandartgallery.com/page/auckland-art-gallery-s-new-chatbot-demonstrates-artificial-intelligence-to-give-new-access-to-17000-artworks>

Baca, M. (2008). "Introduction to Metadata." *The Getty Research Institute*. Accessed January 2020. <https://www.getty.edu/publications/intrometadata>

Bailey, J. (2019). "Solving Art's Data Problem – Part One, Museums." *Art Nome*. Accessed January 2020. <https://www.artnome.com/news/2019/4/29/solving-arts-data-problem-part-one-museums>

Cates, M. (2019). "The Met, Microsoft, and MIT Explore the Impact of Artificial Intelligence on How Global Audiences Connect with Art." *MIT Open Learning*. Accessed January 2020. <https://openlearning.mit.edu/news/met-microsoft-and-mit-explore-impact-artificial-intelligence-how-global-audiences-connect-art>

CapTech Consulting (2017). "Accuracy of Six Leading Image Recognition Technologies Assessed by New CapTech Study." *CapTech*. Accessed January 2020. <https://captechconsulting.com/news/accuracy-of-six-leading-image-recognition-technologies-assessed-by-new-captech-study>

Ciecko, B.(2017). "Examining The Impact Of Artificial Intelligence In Museums." *Museums and the Web 2017*. Accessed January 2020. <https://mw17.mwconf.org/paper/exploring-artificial-intelligence-in-museums/>

Cognex (2019). "What is Machine Vision." Accessed January 2020. <https://www.cognex.com/what-is/machine-vision/what-is-machine-vision>

Cook, T. (2019). "Biased Algorithms are Easier to Fix than Biased People." *New York Times*. Accessed January 2020. <https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html>

Cooper Hewitt (2013). "All your color are belong to Giv." Accessed January 2020. <https://labs.cooperhewitt.org/2013/giv-do>

Engel, C. & Mangiafico P. & Issavi, J. & Lukas, D. (2019). "Computer vision and image recognition in archaeology." *Proceedings of the Conference on Artificial Intelligence for Data Discovery and Reuse 2019*. Accessed January 2020. <https://dl.acm.org/doi/10.1145/3359115.3359117>

Fenstermaker, W. (2019). "How Artificial Intelligence Can Change the Way We Explore Visual Collections." *Met Museum Blog*. Accessed September 2019. <https://www.metmuseum.org/blogs/now-at-the-met/2019/artificial-intelligence-machine-learning-art-authorship>

Google Arts & Culture. (2018). "MoMA & Machine Learning." Experiments with Google. Accessed January 2020. <https://experiments.withgoogle.com/moma>

Grady, C. (2017). "How the SFMOMA's artbot responds to text message requests with personally curated art." *Vox*. Accessed January 2020. <https://www.vox.com/culture/2017/7/11/15949872/sfmomas-artbot-send-me-text-message>

Harvard Business Review. (2019). "Artificial Intelligence: The Insights You Need from Harvard Business Review." Cambridge: Harvard Business School Press.

Hao, K. (2019). "This is how AI bias really happens—and why it's so hard to fix." *Technology Review*. Accessed January 2020. <https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix>

Husain, A. (2017). *The Sentient Machine*. New York: Scribner.

Jones, B. (2018). "Computers saw Jesus, graffiti, and selfies in this art, and critics were floored." *Digital Trends*. Accessed January 2020. <https://www.digitaltrends.com/computing/philadelphia-art-gallery-the-barnes-foundation-uses-machine-learning>

Kessler, M. (2019). "The Met x Microsoft x MIT: A Closer Look at the Collaboration." *The Met Blog*. Accessed January 2020. <https://www.metmuseum.org/blogs/now-at-the-met/2019/met-microsoft-mit-reveal-event-video>

Knott, J. (2017). "Using AI to analyze collections." *Museum Association*. Accessed January 2020. <https://www.museumsassociation.org/museum-practice/artificial-intelligence/14022017-using-ai-to-analyse-collections>

Leason, T. and steve.museum, Steve (2009). "The Art Museum Social Tagging Project: A Report on the Tag Contributor Experience." In J. Trant and D. Bearman (eds). *Museums and the Web 2009: Proceedings*. Toronto: Archives & Museum Informatics. Accessed January 2020. <http://www.archimuse.com/mw2009/papers/leason/leason.html>

Leonard, P. (2019). "The Yale-Smithsonian Partnership presents: Machine Vision for Cultural Heritage & Natural Science Collections." Accessed January 2020. <https://www.youtube.com/watch?v=6dmV-ocRJUI>

Marr, B. (2019). "What is Machine Vision And How Is It Used In Business Today?" *Forbes*. Accessed January 2020. <https://www.forbes.com/sites/bernardmarr/2019/10/11/what-is-machine-vision-and-how-is-it-used-in-business-today>

McAfee, A., & Brynjolfsson, E. (2018). *Machine, platform, crowd: harnessing our digital revolution*. New York: W.W. Norton et Company.

Merrit, E. (2017). "Artificial Intelligence The Rise Of The Intelligent Machine." *AAM Center for the Future of Museums Blog*. Accessed January 2020. <https://www.aam-us.org/2017/05/01/artificial-intelligence-the-rise-of-the-intelligent-machine>

Merritt, E. (2018). "Exploring the Explosion of Museum AI." *American Alliance of Museums*. Accessed January 2020. <https://www.aam-us.org/2018/10/02/exploring-the-explosion-of-museum-ai>

Moriarty, A. (2019). "A Crisis of Capacity: How can Museums use Machine Learning, the Gig Economy and the Power of the Crowd to Tackle Our Backlogs." *Museum and the Web 2019*. Accessed January 2020. <https://mw19.mwconf.org/paper/a-crisis-of-capacity-how-can-museums-use-machine-learning-the-gig-economy-and-the-power-of-the-crowd-to-tackle-our-backlogs>

Ngo, T. and Tsang, W. (2017). "Classify Art using TensorFlow." *IBM*. Accessed January 2020. <https://developer.ibm.com/patterns/classify-art-using-tensorflow-model>

Nunez, M. (2018). "The Google Arts and Culture app has a race problem." *Mashable*. Accessed January 2020. <https://mashable.com/2018/01/16/google-arts-culture-app-race-problem-racist>

Papert, S. (1966). "The Summer Vision Project." Accessed January 2020. <http://people.csail.mit.edu/brooks/idocs/AIM-100.pdf>

Rao, N. (2019). "How ImageNet Roulette, an Art Project That Went Viral by Exposing Facial Recognition's Biases, Is Changing People's Minds About AI." *ArtNet*. Accessed January 2020. <https://news.artnet.com/art-world/imagenet-roulette-trevor-paglen-kate-crawford-1658305>

Rao, S. (2019). "Illuminating Colonization Through Augmented Reality." *Museum and the Web 2019*. Accessed January 2020. <https://mw19.mwconf.org/paper/illuminating-colonization-through-augmented-reality>

Robinson, S. (2017). "When art meets big data: Analyzing 200,000 items from The Met collection in BigQuery." Accessed January 2020.

<https://cloud.google.com/blog/products/gcp/when-art-meets-big-data-analyzing-200000-items-from-the-met-collection-in-bigquery>

Ruiz, C. (2019). "Leading online database to remove 600,000 images after art project reveals its racist bias." *The Art Newspaper*. Accessed January 2020. <https://www.theartnewspaper.com/news/trevor-paglen>

Schneider, T. (2019). "The Gray Market: How the Met's Artificial Intelligence Initiative Masks the Technology's Larger Threats." *ArtNet*. Accessed January 2020. <https://news.artnet.com/opinion/metropolitan-museum-artificial-intelligence-1461730>

Selvin, C. (2018). "'Curating Involves a Daily Protest Against Forgetting': Hans Ulrich Obrist Waxes Poetic at Armory Show." *ArtNews*. Accessed January 2020. <https://www.artnews.com/art-news/market/curating-involves-daily-protest-forgetting-hans-ulrich-obrist-waxes-poetic-armory-show-9948>

Shu, C. (2018). "Why inclusion in the Google Arts & Culture selfie feature matters." *Tech Crunch*. Accessed January 2020. <https://techcrunch.com/2018/01/21/why-inclusion-in-the-google-arts-culture-selfie-feature-matters>

Smith, R. (2017). "How Artificial Intelligence Could Revolutionize Archival Museum Research." *Smithsonian Magazine*. Accessed January 2020. <https://www.smithsonianmag.com/smithsonian-institution/how-artificial-intelligence-could-revolutionize-museum-research-180967065>

Summers, K. (2019). "Magical machinery: what AI can do for museums." *American Alliance of Museums*. Consulted January 2020. Accessed January 2020. <https://www.aam-us.org/2019/05/03/magical-machinery-what-ai-can-do-for-museums>

Swant, M. (2018). "How the Cooper Hewitt Museum and R/GA Are Showing the Evolution of Technology and Design." *AdWeek*. Accessed January 2020. <https://www.adweek.com/brand-marketing/how-the-cooper-hewitt-museum-and-r-ga-are-showing-the-evolution-of-technology-and-design>

Trivedi, N. (2019). "The Color of Serendipity: Searching with the Color Wheel." *Art Institute of Chicago*. Accessed January 2020. <https://www.artic.edu/articles/759/the-color-of-serendipity-searching-with-the-color-wheel>

Winsor, R. (2016). "Clarifai vs Google Vision: Two Visual Recognition APIs Compared." *DAM News*. Accessed January 2020. <https://digitalassetmanagementnews.org/emerging-dam-technologies/clarifai-vs-google-vision-two-visual-recognition-apis-compared>

Yao, S. (2018). "A Probe into How Machines See Art." *The Crimson*. Accessed January 2020. <https://www.thecrimson.com/article/2018/10/30/how-machines-see-art>