https://doi.org/10.2352/EI.2023.35.13.CVAA-210

Recovering lost artworks by deep neural networks: Motivations, methodology, and proof-of-concept simulations

Jesper Eriksson,[a] George H. Cann,[a] Anthony Bourached,[a] and David G. Stork[b]

[a] Oxia Palus, London, UK
[b]Independent consultant, Portola Valley, CA 94028 USA

ABSTRACT

We discuss the problem of computationally generating images resembling those of lost cultural patrimony, specifically two-dimensional artworks such as paintings and drawings. We view the problem as one of computing an estimate of the image in the lost work that best conforms to surviving information in a variety of forms: works by the source artist, including preparatory works such as cartoons for the target work; copies of the target by other artists; other works by these artists that reveal aspects of their style; historical knowledge of art methods and materials; stylistic conventions of the relevant era; textual descriptions of the lost work and images associated with the target's title. Some of the general information linking images and text can be learned from large corpora of natural photographs and accompanying text scraped from the web. We present some preliminary proof-of-concept simulations for recovering lost artworks with a special focus on *textual* information about target artworks. We outline our future directions, such as methods for assessing the contributions of different forms of information in the overall task of recovering lost artworks.

Keywords: deep neural network, computational art analysis, artificial intelligence, computer-assisted connoisseurship, art recovery, lost art

1. INTRODUCTION

The problem of lost cultural patrimony is profound, in regards to both quantity and quality. The lost artworks *that we know of* would fill the walls of every major public art museum, worldwide.[1] Masterpieces of the highest order, such as Diego Velázquez's *Expulsion of the Moriscos*, among many others, have been lost or destroyed. Althought the physical object of an artwork might be lost forever, an approximation of the image depicted in that work might be recoverable from surviving information in a variety of forms. That is the hypothesis that underlies our research program.

We show our preliminary proof-of-concept simulations based on deep neural networks presented with textual descriptions of specific lost works—descriptions of the sort that exist in historical documentation. We find this approach indeed yields images that bear the likely hallmarks of lost works. Future efforts and improvements, vetted and tested on surviving "surrogate lost works," will surely lead to improved results.

In Sects. 1.1–1.3 we present briefly some of the causes of lost cultural patrimony, the scholarly and cultural reasons to computationally recover image approximations of lost artworks (however imperfect), and some of the prior computational research upon which our efforts are based. In Sect. 2 we outline a number of desiderata and criteria that should guide the choice of artworks for digital recovery. Then, in Sect. 3, we discuss our current methodological approach which focuses on the role of textual descriptions. In Sect. 4 we show representative "recovered" images of works known to have been lost from the Western canon. (We call such works "faux"—not "fake"—as our goal is not to deceive viewers as to the authorship of these images.) We conclude in Sect. 5 with a brief summary and next research steps, based on our current results.

Send correspondence to enquiries@oxia-palus.com and artanalyst@gmail.com.

1.1 Leading causes of the loss of art

The following are the leading causes for the loss of cultural heritage of paintings and drawings, with some examples:

Fire On Christmas eve 1743 a fire broke out in the Royal Chambers in the Alcázar Palace in Madrid. Thanks to the speed and courage of ushers and other staff, paintings such as Diego Veláquez's *Las meninas* were cut from their frames and tossed out the window to safety. (Nevertheless, portions of this work along the right and left borders were lost.) However many important works, such as Veláquez's *Expulsion of the Mariscos*, were lost.

War Many tens of thousands of artworks were lost in World War II, for example Gustav Courbet's *The stone breakers*, the earliest masterpiece celebrating common people, and Canaletto's *Piazza Santa Margherita*. In some cases works were deliberately destroyed, such as Gustav Klimt's large *Faculty paintings*, burned in Vienna by the Nazi SS to prevent them from being looted by the advancing Allied Army.[2]

Storm and flood The 1966 flooding of the Arno in Florence led to the loss or severe damage of over 14,000 artworks, including frescoes, easel paintings, and prints. Many more rare books were lost or damaged beyond legibility.

Earthquake The major Italian earthquake of 1997 destroyed the Basilica of Saint Francis of Assisi in its namesake town, sometimes identified as the birthplace of Italian painting. Numerous frescoes were lost, including pivotal works by Cimabue.

Vandalism In Florence in 1497 Fra Girolamo Savanorola urged mobs to destroy artworks that failed to live up to his notions of decorum and propriety (for instance ones that depicted mirrors, which he felt encouraged vanity). Many works were deliberately burned in this way, leading to the episode's moniker, "Bonfire of the vanities."

Iconoclasm Mao Zedong is surely the single most destructive agent in the history of art, where during the Chinese Cultural Revolution of 1949 he ordered the destruction of millions of traditional artworks and had them replaced by hack Communist propaganda "art."

Theft Over millennia, a great deal of artwork has been stolen from private collectors, commercial art dealers, and public and private museums. The largest material theft in history of any objects, in financial terms, was the 1990 theft of works from Boston's Isabella Stewart Gardner Museum, which included Johannes Vermeer's *The concert*, Rembrandt's *Christ in the Storm on the Sea of Galilee*, and other priceless works.

Personal reasons Some artists and patrons destroy works, most frequently for personal rather than financial reasons. Winston Churchill's wife, Baroness Clementine Spencer-Churchill, destroyed the portrait of her husband by Graham Sutherland because she and Winston thought it insufficiently flattering. Jasper Johns, at age 24, destroyed all his works in order to forge a new Pop Art style free from his earlier work. Toward the end of her life, Georgia O'Keeffe purchased back a number of her early works and then destroyed them, all in order to leave an artistic legacy that better reflected her later views and values.

1.2 Reasons to digitally recover lost artworks

In Sect. 3 we shall discuss approaches to the computational estimation or "recovery" of lost art but we must consider first the reasons to infer the appearances of lost works. Such reasons include:

Art scholarship Computed images of recovered lost works may give us a deeper understanding of the target artist's oeuvre and the critical and artistic responses to the lost work, and may generally provide a richer, more complete history of art.

Develop AI techniques The problem of recovering images of lost works will lead to new classes of algorithms not addressed in traditional AI research, for instance methods for integrating information in disparate types (image, text, historical documents, material studies, and so on), all amidst variations in style, meanings, and so forth.

Alert the public to specific lost works The computed works, when displayed to scholars and the general public, may lead to the recovery of the missing *physical* artwork (thought lost), as the public recognizes the computed image as one they might have seen in person. Eugene Delacroix's *Women of Algiers in their apartment*, thought lost for 170 years, ultimately hung on walls of a Parisian woman unaware that the painting was by the illustrious master. Upon seeing an exhibition of Delacroix's works, she considered her painting anew and brought it to the attention of art scholars, who ultimately authenticated it as an autograph Delacroix.

Alert the public to the extent of lost art The overall research program will help alert scholars, students, and the general public to the extent of lost cultural patrimony. Such awareness, in turn, may affect scholarship, research funding, inspire additional efforts to recover art, and ultimately enrich the study of art more broadly.

1.3 Prior computational work related to recovering lost artworks

Computational methods for recovering lost art will rely on a wide range of techniques from computer vision, machine learning, natural language processing, and artificial intelligence and of course traditional art scholarship, as outlined in Sect. 3. Prior work most closely associated with the task falls into three general categories:

- generating digital images of new works in the style of a period or candidate author
- recovering missing aspects of a (possibly surviving) work, such as missing passages in a work or overall coloration in an achromatic version of a lost, polychrome work.
- using textual descriptions about a particular artwork or genre to generate an image.

An example of generating a single work in the style of a particular author is *The next Rembrandt project*.[3] This research effort collected a large number of images of autograph portraits by Rembrandt and used a deep neural network to merge, assemble, and unify image snippets to form a faux portrait in the style of that artist. The resulting image possessed features, such as a large dark hat, that were not mere duplicates from the collected dataset. Note particularly that this method did not exploit textual descriptions of Rembrandt's or indeed any artist's oeuvre.

There have been several efforts to recover missing portions or missing features of lost (or partially lost) artworks. Rembrandt's *The Night Watch* was trimmed along all four edges in 1715 to fit its new home in Amsterdam's Town Hall; the trimmed strips have been lost. Gerrit Lundens, a contemporary of Rembrandt, was commissioned to make a small copy of the original (untrimmed) *The Night Watch*; this copy survives. The missing strips of Rembrandt's work were computationally estimated using the corresponding passages from Lundens' copy refined by the computational mapping of Rembrandt's autograph style.[4]

Another example of the digital recovery of information in the lost artwork involves the recovery of original colors. Gustav Klimt executed *The faculty paintings*, large polychrome works to decorate the ceilings in the University of Vienna faculties of Medicine, Jurisprudence, Religion, and Philosophy. During World War II, the Nazi SS destroyed these works, lest they fall into the hands of the advancing Allied Army. Before they did so, however, they took black and white photographs of the works; these photographs survive. A full-color image of the lost painting was computed based on the style of surviving works by Klimt as well as textual descriptions of the color in his contemporaneous oeuvre, mapped onto the design in a black and white digital image of the photograph.

Several paintings by some artists—most notably the early Pablo Picasso and Vincent van Gogh—were executed on re-used canvas supports. Infrared reflectography and x-radiography reveal that some of these works bear entire designs or "ghost paintings" hidden behind the artwork that is currently visible. These imaging modalities do not reveal the *colors* of the hidden designs, however. One can use computational style transfer to map likely colors to such underdrawings and ghost paintings, in essence recovering an approximation of the appearance of the hidden work.[5–7]

Recent artificial intelligence research has led to systems that can generate images from textual input, such as "A woman riding a horse next to a tree." These systems can also render images in a variety of styles, for example Impressionism, as we discuss in Sect. 3. Such textual information is what we exploit in our work reported here. To our knowledge, our work is the first to use art historical and scholarly text to generate images of specific lost artworks.

We call our computed works "faux," as they are meant to be representative of the works of specified authors, yet are not passed as genuine—there is no effort to deceive the viewer (the defining characteristic of a fake work).

2. DESIDERATA FOR SELECTING LOST WORKS TO DIGITALLY RECOVER

There are several desiderata and criteria that should guide the choice of candidate lost artworks for digital recovery. These desiderata help ensure that any effort is likely to produce images that approximate the lost works that will be of art-historical interest to scholars and the general public alike.

The artwork existed We have secure evidence, from documentary records such as sales and display records, copy artworks, and so on, that the candidate lost artwork indeed existed.

Preparatory sketches survive Preparatory sketches, studies, cartoons, and similar works for the lost work survive.

Copies survive Copies and other derivatives of the target work—by its author and by followers—survive.

Textual descriptions of the work survive There are surviving textual descriptions of the lost work, for instance describing the number of actors, their relation, the colors, style, relation to other works, and so on.

Many works by the same author or followers There are many works in the style of the lost work by the same author and his or her followers; such works are useful for learning style, such as color palette, mark style, composition strategies, and so on.

Visually informative title We know a title that, in context, informs or constrains the appearance of the lost artwork. While "Untitled" provides no reliable information about a work, "Supper at Emmaus" does, as this refers to the Biblical story and implies the scene depicts the interior of an inn, the actors include Christ and two or three apostles, an innkeeper, and other objects associated with such a scene at mealtime.

Technically feasible Contextual information suggests that the image in missing work is "simple enough" to be generated by current computational methods.

Sheds light upon art questions The recovered work will shed light upon problems in the history and interpretation of art.

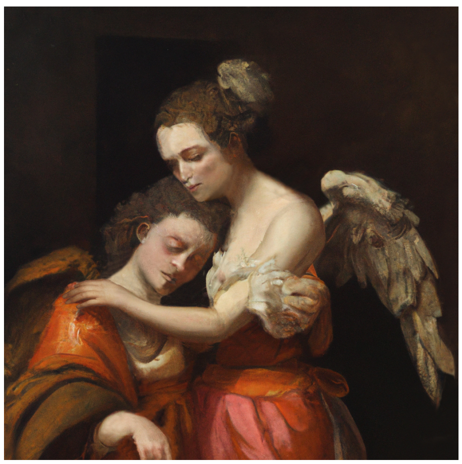
Of course it is highly unlikely that any lost work satisfies all these desiderata, and thus any choice of target work will involve an implicit weighting of the criteria, particularly in the context of the art-historical questions at hand. One natural candidate for demonstration and validation is ?Leonardo's *Salvator mundi*, as this work fulfills many of the criteria or desiderata listed above.[8] It could be considered "lost" and serve as ground truth image for validating and improving the algorithmic and traditional scholarly methods.

3. GENERAL METHODOLOGY

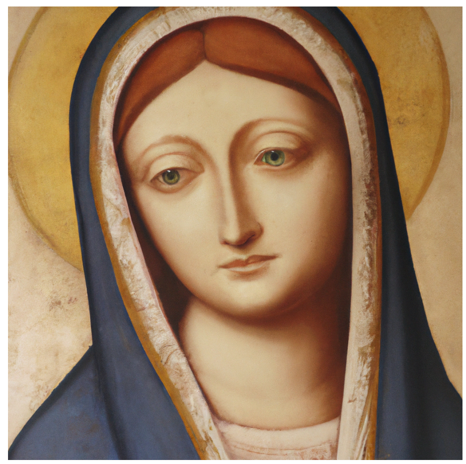
In our work, the core processing is performed by DALL-E 2,[9] which was built upon the highly trained language model GPT-3, the third generation of Generative Pre-trained Transformer systems.[10] We start by researching textual records of the missing work, uncovering any accounts of the motif, colours, techniques, and methods used. We also collect other surviving works the author produced during the same period to give us a benchmark of their output. We then start building appropriate textual prompts for input into DALL-E 2, a step that currently involves a modicum of trial and error.

4. RESULTS AND ANALYSES

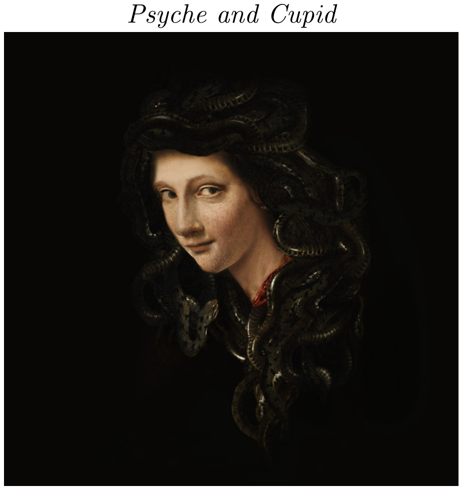
Figure 1 shows five representative images of lost works computed by our methods. The first point to notice is that all are plausible artworks—visually coherent, consistent lighting, figures, and so forth. The most conspicuous "error" is visible in the faux Velázquez's *Expulsion of the Moriscos*, where the faces of the figures are incoherent and in many cases devoid of recognizable features.



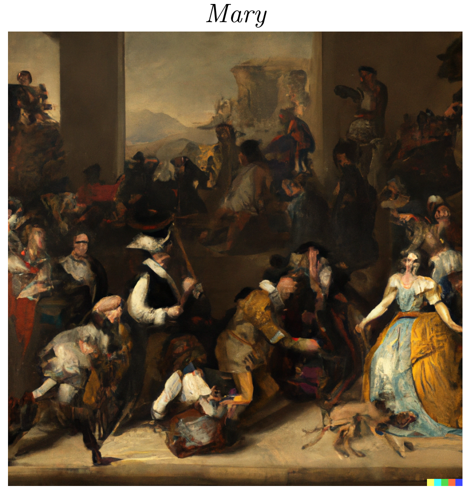
Diego Velázquez
Psyche and Cupid



Giotto
Mary



Leonardo
Medusa



Diego Velázqiuez
Expulsion of the Moriscos

Figure 1. Proof-of-concept computational approximations of four lost artworks.

Sometimes DALL-E 2 provides accurate results from the first couple of prompts, rendering highly plausible imagery in the intended artist's style. However at other times the model produces patently unacceptable results (in style, content, and realistic representation of certain objects, notably faces), which means we have to keep tailoring the prompts to explore DALL-E 2's vector space. For example, DALL-E 2 produced a highly accurate rendition of Giotto's *Virgin Mary* from the basic prompt: "A painting of the virgin Mary by Giotto," with only minor artefacts around the eyes and mouth of the subject. However, the model failed to generate a plausible rendition of Leonardo's *Medusa*, either producing an accurate portrayal of the mythological subject, or adhering to the artist's style, but not both.

This behavior is not unexpected since there exist much more contextual information about Mary as a subject of Giotto's than there does for Medusa by Leonardo. Recall that Giotto leaves us the full Scrovegni Chapel, while Leonardo leaves us just 19 or so paintings. Medusa is also a much more specific and detailed subject in visual representation than Mary, which places higher demands on the model. Moreover, we have found that DALL-E 2, like many text models, often responds well to certain key phrases rather than lengthy descriptions.

Consider a short prompt crafted to produce an approximation of Diego Velázquez's lost *Cupid and Psyche*:

> "A portrait by Velázquez of a woman called Psyche embracing Cupid in a darkened room"

This prompt led to the generation of scene which matched the colour scheme and general impression of Velázquez's other work during the Spanish Golden Age. The two mythological names "Cupid" and "Psyche" seemed to set the scene for the subject matter. Further, "A woman called Psyche" seemed to clue the model into the female-male relationship between the two subjects. "Embracing" set the detail of that relationship up and also their spatial composition, whilst "darkened room" seemed to provide an accurate scene, which when combined with its training data from other works by Velázquez pulled out a colour scheme similar to what we were looking for.

This stage leads to the next stage, where we manually remove unwanted visual artefacts. With more complex scenes and subjects the model often struggles to generate plausible facial features, hands, and more minute details. If minor, we perform manual brush work to correct these artefacts, or if more major features are concerned, we prompt the model for specific renders of the necessary features, which we then manually stitch together with the original output to create an accurate representation.

5. CONCLUSIONS AND FUTURE DIRECTIONS

We have demonstrated that a highly trained deep network-based system for generating images from textual descriptions can generate images that plausibly resemble missing artworks in their style, design, and content. Of course it is hard to judge these works more rigorously through comparison to lost art given that, by choice, such target works are truly lost. Thus our ongoing work is a systematic study of the above methods on *surviving* artworks, where we can judge the results against such *ground truth*.

We believe computational methods—sometimes called computer-assisted connoisseurship—can address problems of genuine interest in the history and interpretation of fine art, and will progress to the extent that scholars in the humanities collaborate in this program by providing problems, criticism, and expertise to the growing number of scholars from computer science addressing problems in the history and interpretation of fine art.

REFERENCES

1. N. Charney, *The museum of lost art*, Phaidon Press, New York, NY, 2018.
2. M. Ledivelec-Gloeckner, ed., *Gustav Klimt*, Crown Publishers, New York, NY, 1987.
3. B. Korsten and E. Flores, "The next Rembrandt," Retrieved from *www.nextrembrandt.com* , 2016.
4. N. Siegel, "Rembrandt's damaged masterpiece is whole again, with A.I.'s help," 2021. www.nytimes.com/2021/06/23/arts/design/rembrandt-night-watch-artificial-intelligence.html.
5. A. Bourached and G. H. Cann, "Raiders of the lost art," 2019. `arXiv:1909.05677v1`.
6. A. Bourached, G. H. Cann, R. Griffiths, and D. G. Stork, "Recovery of underdrawings and ghost-paintings via style transfer by deep convolutional neural networks: A digital tool for art scholars," in *Computer vision and analysis of art*, D. G. Stork and K. Heumiller, eds., IS&T, 2021.

7. A. Bourached, G. H. Cann, R. Griffiths, and D. G. Stork, "Resolution enhancement in the recovery of underdrawings via style transfer by generative adversarial deep neural networks," in *Computer vision and analysis of art*, D. G. Stork and K. Heumiller, eds., IS&T, 2021.
8. M. Dalivalle, M. Kemp, and R. B. Simon, *Leonardo's Salvator Mundi & the collecting of Leonardo in the Stuart courts*, Oxford University Press, Oxford, UK, 2019.
9. D. Bau, A. Andonian, A. Cui, Y. Park, A. Jahanian, A. Oliva, and A. Torralba, "Paint by word," *CoRR* **abs/2103.10951**, 2021.
10. T. B. Brown and 27 contributors, "Language models are few-shot learners," 2020. https://arxiv.org/pdf/2005.14165.pdf.