Tagging the Mead Art Museum Collection: Using Google Cloud Vision API

Reihaneh Iranmanesh ‘25, Adam Rogers ‘24, Advisors: Prof. Mihaela Malita, Miloslava Hruba (Mead Art Museum)

Department of Computer Science, SURF 2022, Amherst College

Introduction

What do you first notice when you look at this painting to the right?
The woman and the children around the table? The dog?
The hat hanging on the wall, the wood ceiling, or maybe the pendulum clock?

Humans tend to interpret and describe artworks according to their personal experience and interests. While humans judge artistic creation subjectively, computers process images objectively, without considering the context, makers, or their motivations.

In order to enhance Mead’s metadata and increase the searchability of its database, we used computer vision and machine learning to generate tags for the collection. For example, users interested in “dogs” and their depiction in different artworks could easily search and see all the images tagged with this term.

Google Cloud Vision API

The Google Cloud Vision API is a computer vision tool trained with millions of realistic images. It applies machine learning to identify visual trends and classify images into thousands of objects and labels with certain confidence scores.

Methods

1. Building common terminology between MIMSY[1] and Vision

The terms used for tagging Mead’s collection differ from those Vision uses to classify images. Many Vision terms are semantically the same and can be put under a general tag. Vision and MIMSY[2] also have many gender-specific tag terms (e.g., women and actresses). Since gender expressions differ among cultures and time periods, we removed all gender-specific Vision tags. The result was a Python dictionary of 1,237 (out of 19,985 total) distinct Google tags mapped to 630 (out of 999 total) MIMSY tags.

MIMSY Term	Vision Term
birds	bird, blackbird, bluebird, hummingbird, …

MIMSY Term	Vision Term
People	Person (all other gender-specific tags were removed)

2. *Tagging Mead’s collection*

We **tagged Mead’s collection of 21,996 web images twice.** Two distinct API requests, *label detection*, and *object localization*, were sent simultaneously to Vision for each image. The former request is for assigning labels and the latter is for object recognition and localization. For each API request for a particular image, by default, ten tag terms with the highest confidence scores are returned.

- First, we tagged the collection with all MIMSY tags that map to Vision tags returned in the API response, regardless of their confidence scores **(no filtering)**.
- Subsequently, we filtered labels and objects depending on their confidence scores **(filtering)**. Using a fixed constant threshold for filtering the Vision tags seemed ineffective, so we considered the highest confidence score (HCS) returned in each label and object API response and included any matching Vision term in the range of [HCS - 0.15, HCS] for each API response.

3. *Script Multiprocessing*

Tagging the whole database would have taken nearly nine hours, so we divided the images into six batches to multi-process all six Python scripts concurrently. As a result, **the whole collection was tagged in approximately half the time!**

1. Acronym for Museum Integrated Management System.
2. MIMSY tags are pre-defined and fixed, approved by the Five Colleges and Historic Deerfield Museum Consortium.

Results

1. A similar tagging between Mead and Vision

Karl Hetz (German, 1828-1899)
Peasant Family at Supper, 1875, oil on canvas
Gift of Miner Tuttle, Class of 1913
AC 1940.1

Vision (no filtering)	Vision (with filtering)	Mead
dogs; art; hats; painting; rooms; people; furniture	dogs; art; hats; painting; people	domestic space; eating; families; food; interiors; women; children

2. A correct example of tagging by Vision

Vision is much more accurate when presented with photorealistic artworks.

Vision (no filtering)	Vision (with filtering)	Mead
water; boats; sky; clouds painting; lakes; animals; landscapes; people	water; boats; sky; painting; art; people	NULL

Bryson Burroughs (American, 1869-1934)
Departure at Dawn, 1907, oil on canvas
Gift of Caleb R. Woodhouse
AC 1991.63

3. An incomplete example of tagging by Vision

In this painting, Vision could not detect the exact animal species.
It identifies the object as an animal, but it does not identify the animals as dogs, horses and elephants:

Unknown, Indian, Moghul
A Hunting Scene, 17th century
Opaque watercolor on paper
Gift of Alban G. Widgery
AC 1967.48

Vision (no filtering)	Vision (with filtering)	Mead
botany; painting; art; plants; trees; animals	botany; painting; art; plants; trees; animals	animals; dogs; hunting elephants; horses; trees

4. An incorrect example of tagging by Vision

Error in describing the exact animal species. In the following painting, the owl is identified as a cat.

Vision (no filtering)	Vision (with filtering)	Mead
branches; cats; art; painting; animals	branches; cats; art; painting	animals; birds; flowers; trees

Hirose Bihō (Japanese, born 1873-n.r.)
Untitled (Owl and Cherry Branch), ca. 1910
Woodblock print
Gift of William Green
AC 2005.176

Statistical Data

As you can see, the number of detected flowers and dogs by Vision is much less than the same tags done by humans. Vision is more prone to use general terms like “plants” or “animals.” Interestingly, out of 61 images tagged with “dogs” by Mead, Google could only detect 8. However, it identified that 34 (~56%) of those images include an animal, but not specifically a dog.

An example of three tags and the number of their usage, both for Vision and Mead

Google Vision API (out of 21,770 images)		**Mead, done by humans** (out of 6,247 images)		**% of Mead tags detected by Vision**	
Tag Term	*Count*	*Tag Term*	*Count*	*Ratio*	*%*
animals	2660	animals	565	241/565	43
flowers	264	flowers	324	47/324	15
dogs	53	dogs	61	8/61	13

Findings

In the Mead database, 29% of 21,996 total images are tagged, an initiative of the museum over the last three years largely engaging student interns. The **remaining 71% of records remain searchable by often limited cataloguing information**. The assigned tags are subject to human interpretation and error.

The Vision API is mostly trained on realistic images from recent decades. It struggles when presented with various artistic mediums and styles across different time periods. As a result, **Vision returns less descriptive tags that do not convey specific types of objects**.

Mead	Vision
• Tagged 29% of the database (6,247 objects) • 802 distinct tags • Assigned 32,761 total tags • 5 tags per image • 5 minutes to tag one image • About 2000 hours to tag collection • More descriptive • Detects meaning of art • Subject to human error	• Tagged 99%* of the database (21,770 objects) • 308 distinct tags • Assigned 91,237 total tags • 4 tags per image • 1.5 seconds to tag one image • About 5 hours to tag collection • More general • Classifies objects • Consistent strategy of tagging

Not surprisingly, **the Google Vision API is not adequately trained to accurately label art**. Vision and similar computer vision tools (Amazon Rekognition, IBM Watson, Microsoft Azure) are trained to accurately detect objects in realistic images. When presented with creative rendering of objects, these programs struggle to assign accurate yet detailed labels. **While programmatically tagging artworks can be completed in a fraction of the time, computer vision tools cannot yet fully and correctly absorb the diversity of human creation.**

* The other 1% are either corrupted URLs (100) or Vision could not detect any labels or objects that had an acceptable confidence score (according to the range defined above in our program).

Further Research

Existing computer vision tools have been trained on stock image photography and do not work well with art objects such as prints, paintings, sculptures, textiles, ceramics, decorative arts, etc. For achieving better accuracy, museums should build their own machine learning models. **The biggest challenge is a lack of training data.** Thus, museums should use synthetic data (artificially annotated data that is generated by computer algorithms or simulations), GANs (generative adversarial networks) for data augmentation, and style transfer to augment their limited machine learning datasets with thousands of additional examples.

Acknowledgements

Thank you to the Mead Art Museum for access to their database, to MIMSY staff for their guidance, to SURF for funding our research, to Google Cloud and the DALL-E team for allowing us free access to their tools, to the Metropolitan Museum of Art for their openness and insight, and to the creators of the following Python libraries: pickle, pandas, TensorFlow, google-cloud-vision, requests, multiprocessing.
For viewing the source code on GitHub, please scan the following QR code.