InterPARES Trust AI

Teachable AI for the Archival Professions – Module 4:

AL/ML for Processing Image-based Records in Archives

Kaila Fewster[1] and Richard Arias-Hernandez[2]

This educational module is part of a series of learning materials developed by InterPARES Trust AI[3] researchers and educators to train archival professionals and students to effectively leverage artificial intelligence in their archival work. The current version was completed on August 29th, 2025.

---

[1] InterPARES Trust AI Graduate Academic Assistant, University of British Columbia.

[2] Associate Professor of Teaching, School of Information, University of British Columbia and InterPARES AI Trust Researcher. richard.arias@ubc.ca

[3] This work is an outcome of InterPARES Trust AI, an international research partnership led by Drs. Luciana Duranti and Muhammad Abdul-Mageed, University of British Columbia, and funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). https://interparestrustai.org/

Module 4: AI/ML for Processing Image-based Records in Archives

READ FOR THIS MODULE

Required:

- Bushey, J. (2023). AI-Generated Images as an Emergent Record Format. *2023 IEEE International Conference on Big Data (BigData)*, 2020–2031. https://doi.org/10.1109/BigData59044.2023.10386946

- Lincoln, M., Corrin, J., Davis, E., & Weingart, S. B. (2020). *CAMPI: Computer-Aided Metadata Generation for Photo archives Initiative*. Carnegie Mellon University. https://doi.org/10.1184/R1/12791807.v2

- Yıldız, M., & Rukancı, F. (2025). AI-powered visual classification in archives: A computer vision approach to facial recognition in historical archives. *Archival Science*, *25*(2), 18. https://doi.org/10.1007/s10502-025-09486-w

Recommended:

- Caballar, R. D., & Stryker, C. (2025, February 25). *What Are Vision Language Models (VLMs)?* https://www.ibm.com/think/topics/vision-language-models

- Holt, B., & Hartwick, L. (1994). Retrieving art images by image content: The UC Davis QBIC project. *Aslib Proceedings*, *46*(10), 243–248. https://doi.org/10.1108/eb051371

- Kundu, R. (2023, January 17). *YOLO Algorithm for Object Detection Explained*. https://www.v7labs.com/blog/yolo-object-detection

- Nicoletti, L., & Bass, D. (2023, June 9). Humans Are Biased. Generative AI Is Even Worse. *Bloomberg.Com*. https://www.bloomberg.com/graphics/2023-generative-ai-bias/

- Paolanti, M., Pietrini, R., Della Sciucca, L., Balloni, E., Compagnoni, B. L., Cesarini, A., Fois, L., Feliciati, P., & Frontoni, E. (2022). PergaNet: A Deep Learning Framework for Automatic Appearance-Based Analysis of Ancient Parchment Collections. In P. L. Mazzeo, E. Frontoni, S. Sclaroff, & C. Distante (Eds.), *Image Analysis and Processing. ICIAP 2022 Workshops* (Vol. 13374, pp. 290–301). Springer International Publishing. https://doi.org/10.1007/978-3-031-13324-4_25

OVERVIEW

This module provides an overview of how artificial intelligence (AI) and machine learning (ML) are changing the way image-based records are processed in archives, and it demonstrates how archival professionals can engage critically, but effectively, with these tools. It explores the origins of processing image-based records with AI from early OCR and content-based image retrieval to modern applications of deep learning, convolutional neural networks, and object detection tools like YOLO. Moreover, this module also discusses key algorithms and architectures for processing visual records and highlights their potential for supporting tasks such as description, classification, and diplomatic analysis. Finally, it discusses the concerns around authenticity which have arisen as a result of AI-generated or modified images becoming an emergent record format, as well as critical challenges with AI and ML tools like algorithmic bias and copyright issues.

LEARNING OBJECTIVES

By the end of this lesson, students will be able to:

- Explain the history of AI computer tools in archives with emphasis on processing of image-based records, content-based image retrieval (CBIR), convolutional neural networks (CNNs), YOLO algorithms and computer vision more broadly.
- Use AI tools for the processing of image-based digital records in the archives including digitized analogue records and born-digital records.
- Identify opportunities to use emerging AI/ML tools for archival processing of digital image-based records.
- Discuss issues associated with authenticity of GenAI-modified or -generated records and issues related to AI/ML tools used in the processing of image-based records (e.g., copyrights, biases in training data, etc.)

Introduction

When it comes to processing image-based records, machine learning (ML) and artificial intelligence (AI) have been implemented into archival workflows since at least the 1990s, although experiments with computer vision started to take off more broadly in the 1970s (Szeliski, 2011). For example, in 1994, UC Davis used a precursor to content-based image retrieval (CBIR), called query-by-image-content (QBIC), to search through a database of over 200,000 image slides based on their appearance, rather than their textual indexes (Holt and Hartwick, 1994). Although their results were variable, particularly when asking the software to identify shapes in fine art images, this experiment illustrates the history of automated image-based records processing in archives and the potential of these technologies to be further integrated into archival workflows (Holt and Hartwick, 1994). Moreover, further advancements in facial identification software, like Meta's DeepFace system, along with Google's reverse image search tool, which is in fact a CBIR system, both highlight potential opportunities for better image-based records searching and processing in archival settings, even if that was not their original application.

In the archival space, recent projects like PergaNet, a deep learning system designed to help create digital reconstructions of medieval parchments based on appearance, demonstrate how improvements in deep learning and neural network ML techniques are increasingly being used in archival settings (Paolanti et al., 2022). Furthermore, improvements in open source AI algorithms like YOLO, which is used for object detection and identification, and other large language models (LLMS) built to work with computer vision, can prove valuable for providing automatically-generated text descriptions of the ingested images. In this sense, there is potential for AI and ML to be used in archival appraisal, arrangement and description workflows. Moreover, CBIR-based systems can help improve record discoverability, searchability, and retrieval.

Nonetheless, while processing image-based records with AI and ML has potential for improving efficiency and reducing archival backlogs, the rapid growth in popularity of generative AI tools, which can be used to create images, can pose challenges in archival contexts. With concerns of bias, hallucinations, copyrights, misappropriation of records as data, and plagiarism, most generative AI models, at this point, are generally inappropriate for use in image-based archives, as their outputs have unclear levels of authenticity and trustworthiness, and thus are unreliable tools for use in archival processing workflows (Bushey, 2023). Therefore, it is important for archivists and records managers to be aware of any potentially

AI-generated content entering the archives, the consequences of training GenAI with local image-based records, and whether other AI tools integrated into their workflows rely on any generative AI models.

History of AI/ML for processing image-based records in archives

As noted in the introduction, computer vision was originally developed in the late 1960s and early 1970s as a field distinct from digital image processing, with computer vision specifically designed to "recover the three-dimensional structure of the world from images and use this as a stepping stone towards full scene understanding" (Szeliski, 2011). As a discipline within the broader field of AI, computer vision uses machine learning (ML) and neural networks to develop models which can derive meaningful information from images, videos, and other visual records (IBM, 2021). While early computer vision in the form of Optical Character Recognition (OCR) was used in archives as early as the 1980s for textual records, image-based processing using Content-Based Image Retrieval (CBIR) does not appear in the archives until the 1990s and early 2000s (see Allen, 1987; Castelli et al., 1998; Holt & Hartwick, 1994). Nonetheless, flat-bed image scanners for digital image processing, like the popular Autokon 8400, were used by archives in the 1980s as they began to digitize their collections. In particular, Ray Kurzweil's pioneering work in developing a system for blind people to read by having a computer read aloud to them using flatbed image scanners and state-of-the-art OCR was essential to the broader development of computer vision workflows (Craine, n.d.). In this sense, while computer vision is distinct from digital image processing, collaboration between the two disciplines is necessary to create effective workflows for digitizing and making image-based collections more accessible.

Historically, Content-Based Image Retrieval (CBIR) is one of the most popular applications of computer vision when it comes to processing image-based records. CIBR, and its precursor query-by-image-content (QBIC), both provide responses to queries based on analyses of images' content, such as colours, shapes and textures, rather than any associated metadata with the record (Date & Allweil, 2022). These analyses are completed using machine learning algorithms which take a digital image to evaluate pixel colours, repetitive patterns, and shapes using edge detection to determine which features define the visual properties of the image (Date & Allweil, 2022). The system can then use these defined features to classify records, aid in descriptions, and respond to user queries by image (e.g., "visually similar images" features, or "more like this" features). For instance, CIBR is the backbone of Google's reverse image search tool, also known as a technique called Query-By-Example, which eliminates the need for users to guess query keywords and instead determines key properties of the image to be cross-referenced across the internet to produce search results. Moreover, convolutional neural networks and deep learning, which will be discussed in further depth below, have also vastly improved reverse image search object detection capabilities in recent years (Singh & Gowdar, 2021). In this sense, archivists using any reverse image searches to identify similar objects or records from other institutions are already engaging with a form of computer vision and AI in their work. Additionally, archives are beginning to implement image match querying in digital archives, in the same way Google images provides similar results through the "More Images Like This" search function. Smeaton et al. illustrate an early experiment with this technology on CCTV video footage using semantic entities identified in the video frames and the video's closed captioning metadata to pull up similar results while users were engaging with an image (Smeaton et al., 2006). This type of image matching search and retrieval system has significant potential for improving image-based digital archives by making it easier for users to find similar records without relying exclusively on metadata.

More recent experiments in archives using AI/ML tools for processing image-based records have focused on detecting and defining semantic objects (i.e., conceptual entities) depicted on images rather than their primitive visual properties (i.e., hue, shape, texture, etc.), as was done in the Finnish Wartime Photograph Project from the Finnish Defense Forces' photographic archive. Using the YOLO algorithm, which will be explained in further depth below, the project team was able to identify semantic objects in over 160,000 digitized Finnish photographs from World War II (Chumachenko et al., 2020). Another prominent example is the Perganet project, briefly mentioned in the introduction of this module. Perganet relies on deep learning to automatically analyze and process digitized medieval parchments in order to classify them (Paolanti et al., 2022). In particular, while Perganet technically does some textual processing by recognizing and identifying medieval scripts, it is also concerned with the notary symbols (i.e., signa) and other images present in the document (Paolanti et al., 2022). In this sense, Perganet does not only focus on recognizing and classifying the objects detected in the image, but also their location on the layout of record itself, which is often very relevant for studying early manuscripts from a diplomatics perspective (Paolanti et al., 2022).

More broadly, there is also a significant amount of work being done by the InterPARES Trust AI project (2021-2026) in exploring the use of AI tools in archives for image-based records from all angles. For instance, the RA01 study, led by Adam Jensen, is exploring how AI can be used to generate more meaningful descriptive metadata for photographs and multimedia-based records. On the other hand, the CU08 study, led by Dr. Jessica Bushey, explores how AI-generated images are becoming an emerging record format, and investigates how archivists can determine the reliability and trustworthiness of these records. In this sense, generative AI, although potentially not yet implemented into archival record processing workflows, is still an emerging concern, as AI generated content is increasingly making it into archival settings. As such, it is also worth considering how new types of image-based records will be processed in the archive, and whether existing archival approaches to protecting records' authenticity and trustworthiness will be sufficient for AI-generated images (Bushey, 2023).

ACTIVITY #1: Experience CBIR on Special Collections

iART is a proof-of-concept for GLAM institutions of AI-powered CIBR and clustering on visual materials. In this activity, you will experience CBIR on iART. More information about this German project here and here.

1. Go to iART and find an image by providing some keywords on their search engine. For example, run a query search with the word "horse." Pick one image from the result set to use for CBIR. Click on the image. On the pop-up window showing the item's digital record for the selected image, click on the red search icon at the bottom right of the image (magnifying glass icon). An "Append to Search" option will show. Click on it. This action selects this specific image to be used for a visual query by the search engine to find visually similar images. Make sure you delete the query work "horse" from the search field to focus only on the CBIR. Students can also upload an image, similar to what is done with Google Images, by clicking on the image icon at the end of the search box and run a CBIR search based on the uploaded image.
2. Use "Search with Google Lens," if you are using a Google Chrome browser with this functionality, or Reverse Image search on Google Image, to determine if you can find any additional provenance about the image or more relevant metadata than the one found on iART.

3. Reflect and share your experience using CBIR/reverse image search, what worked well, what didn't, and how applicable you think this technology is, or could be to archival workflows.

How does CBIR work?

As briefly mentioned above, computer vision applications like CBIR rely on machine learning algorithms to accurately and efficiently analyze images and produce results. When it comes to processing image-based records with AI, ML is essential for extracting key features and labeling objects for detection and recognition by subsequent systems or algorithms (Khan et al., 2018). The two main types of training for ML models: supervised and unsupervised learning, are also found in CBIR applications. Supervised learning uses labelled data, which means the data has already been classified, usually by a human, to train the model, while unsupervised learning does not. Both methods have benefits, as using labelled data allows for the model's inputs and outputs to be monitored for accuracy and learning over time, whereas unlabeled data allows the models to discover patterns hidden in the data without human intervention (Delua, 2021). When it comes to processing image-based records in archives archives, supervised learning models can be helpful for identification and classification, while unsupervised learning models are best suited for processing large amounts of data for creating generic descriptions, and highlighting possible connections between records. For archivists to use these tools, whether supervised or unsupervised, they first need to acquire the data to train the algorithm, either by digitizing analogue records or by putting together a collection of born-digital records. Then, archivists will need to clean up the training data set of images to be fed to the algorithm by identifying, correcting or removing any inaccurate, corrupt, or irrelevant data (commonly known as the data cleaning process).

If archivists are working with supervised learning models in particular, they also need to label the data before ingesting it into the model. For image-based records, this process requires using a labeling tool like AnyLabeling to draw boxes around important objects on the images and then assigning these boxes to semantic categories. As such, when choosing whether to use a supervised or unsupervised model, archivists need to consider their training data, and how much extra time and resources would be required to prepare it ingestion. Furthermore, archivists should also consider what type of problems they are trying to solve, and what their ultimate goal is with using AI/ML.

Although it can vary, CIBR and other computer vision applications are increasingly relying on deep learning and neural networks, which are computational models based on the structure and functions of biological neural networks like the human brain, to interpret queries and improve feature recognition accuracy, particularly pattern recognition (Ashraf, 2015). Neural networks, rather than relying on pre-defined rules for inputs and outputs, instead use mathematical models to map inputs and outputs together (Date & Allweil, 2021).

As such, with the growing popularity of deep learning techniques, convolutional neural networks are also increasingly being implemented into computer vision applications for image processing (Khan et al., 2018). Convolutional neural networks (CNN) are not fed data beforehand to map together using mathematical models like traditional neural networks, but instead process the data as it is ingested to produce an internal representation of the data most suitable to the required task at hand (Piras & Giacinto, 2017). Although first used successfully to identify handwritten numbers and characters, CNNs now have significant potential for enabling semantic segmentation of images, which associates each pixel with a semantic class rather than a geometric one (Date & Allweil, 2021). In the context of computer vision, CNNs can identify features of interest that are not explicitly modeled or connected to a geometric property of the image (Date & Allweil, 2021). This could be of particular use to archivists and records managers who are interested in identifying, describing or tagging semantically-meaningful image features. Moreover, CNNs may be more flexible in recognizing semantic features in images as opposed to traditional CBIR systems which base feature detection on the geometry detected on the rastered images (Date & Allweil, 2021).

As an example, facial recognition systems are increasingly making use of deep learning techniques like CNNs to identify individuals. Looking at the Spellman College Photograph Collection at the Advanced Information Collaboratory at University of Maryland, Proctor and Marciano demonstrated how computer vision, along with natural language processing (NLP) can be used to transform and connect image-based metadata (Proctor & Marciano, 2021). Working with a curated selection of 40 photographs from a collection of over 1200 images, they first established a baseline accuracy measures for the facial recognition tools used as the photographs were primarily of people of colour, which facial recognition tools have traditionally struggled to recognize due to bias (Proctor & Marciano, 2021). Once the accuracy measures were established, they ran the photographs through the models and determined that the tools were able to reliably recognize the same individuals across images. Subsequently, this information was used to improve the image descriptions and create linked data visualizations between the records that was published online (Proctor & Marciano, 2021). This illustrates how computer vision has potential for not only improving image-based records' discoverability, but also for improving descriptions and making connections between records using deep learning algorithms and techniques. Moreover, although the study was initially conceived as a proof of concept rather than a practical and applicable tool, it also highlights how AI/ML tools can be scalable to actual archival practice.

ACTIVITY #2: Labeling an Image Dataset
Students should create a small dataset of archival images collected from the Public Domain Image Archive (All Images | Public Domain Image Archive) containing common objects (e.g., people, flowers, animals, nature, cultural objects, etc.). Then, using AnyLabeling (AnyLabeling), or another similar tool, students should label their dataset into a few distinct semantic categories (5-6) for such objects. Afterwards, in groups students can discuss the types of datasets they created, how they chose to label them, and consider how different data labelling strategies can influence the models' final outputs.

Architectures and algorithms for processing image-based records with AI/ML

Many algorithms which rely on complicated CNNs make use of residual neural networks, a type of deep learning architecture. Residual deep learning is a method of solving the degradation problem of neural networks, where the more layers are added to a model for data to be processed through, the less accurate it becomes (He et al., 2015). Residual deep learning tackles this degradation problem by creating connections between the model's layers and their outputs, which essentially allows each layer to build upon the work of the previous layer, rather than having to generate the output entirely from scratch at every level (Shafiq & Gu, 2022). This allows for models to be built with many tens or hundreds of layers, with each one fine-tuning the results of the previous one, leading to improved learning within the network and thus, better performance outputs. As this architecture was originally created for image recognition, residual deep learning models have been used quite successfully in tasks like image classification, object detection, and semantic segmentation (Shafiq & Gu, 2022). While residual deep learning models can be expensive to train and maintain given the amount of data and computing power required to run them, they nonetheless still have potential applications in archival work. For instance, Mechi et al. proposed the implementation of a residual neural network framework for identifying handwritten text in historical images (2022). In particular, this application of residual deep learning is interesting because it focuses on interpreting the whole text image as input, rather than using a segmentation process which is common in most handwritten text recognition (HTR) and object detection frameworks (Mechi et al., 2022). Using a gated residual neural network, which allows for certain information to be retained by the model for a short period of time, it is possible to prevent information loss from the source data throughout the network while also prioritizing the most relevant features to be identified from the image (Mechi et al., 2022). In other words, the model attempts to reduce the complexity of deep neural networks, while still remaining efficient and providing accurate outputs. Residual deep learning models can serve as valuable tools for archives looking to use more complex image recognition algorithms without sacrificing accuracy; however, they may require additional computer expertise and related-cost for computer processing, which at the moment could be too expensive and out-of-reach for most archives. Either way, it is important for archivists to know these more complex architectures exist, and how they could be utilized for archival workflows in the future when they become cheaper and more commonplace.

As computer vision and other AI tools continue to progress, there is a burgeoning potential for these tools to be combined to develop even larger, and more sophisticated tools. In particular, there has been a growing interest in combining computer vision algorithms with large language models (LLMs) to produce text- and image-outputs based on both computational understandings of text and interpretations of visual data (Hamadi, 2023). Called visual language models (VLMs), these types of tools can create visual outputs based on input text, or textual outputs based on images or video

(IBM, 2025). Built from two main components, VLMs usually include some type of language encoder, which captures the semantic meanings and contexts connected using natural language processing (NLP), and a vision encoder, which extracts key visual properties from images or videos usually using CNNs or deep learning techniques (IBM, 2025).

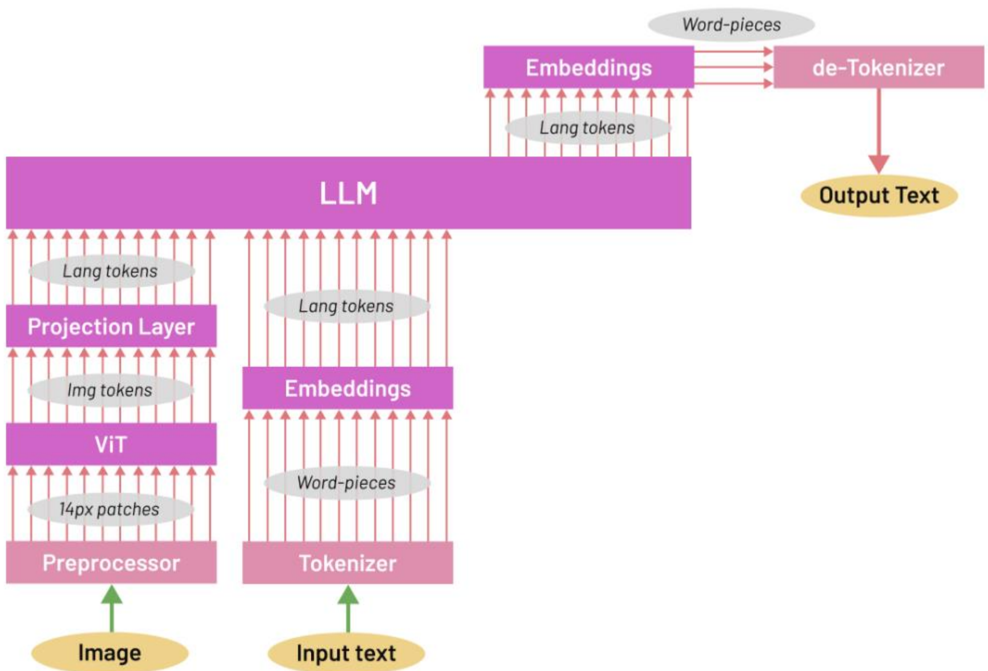

Figure 1: Vision Language Model Architecture (Dirac et al., 2024)

As a result, VLMs are able to do a variety of vision language tasks like captioning and summarization, image generation, search and retrieval, and object detection and segmentation (IBM, 2025). However, there are still challenges with these emerging models like biases, hallucinations, and over-generalizations as well as concerns of cost and environmental impact considering that VLMs combine two already complex AI architectures into one even more complicated tool (IBM, 2025).

As CBIR and other image-based AI tools are becoming increasingly complex with CNNs, deep learning and other innovations in machine learning, there are many different algorithms available for processing images. A key type of algorithm for image recognition are object detection algorithms, which uses computer vision to detect specific instances of semantic objects. There are two main types of object detection algorithms: single-shot and double shot-object detectors. Single shot detectors only review ingested images once to make predictions about semantic objects present (Bushey, 2024). For example, the YOLO (You Only Look Once) algorithm uses CNNs to build boxes around objects and classify them into categories based on a single review of the entire photo (Kundu, 2023). On the other hand, two-shot detectors, as evidenced by the name, review the image twice, first to identify semantic objects and then second to classify those objects into categories. Two-shot detection can often be more accurate because of the multiple image reviews, but single-shot detectors are much more computationally efficient (Kundu, 2023). An example of a two-shot detector is the Mask Region-based CNN (R-CNN) algorithm, which uses deep neural networks to identify regions of interest in the photograph and build boxes around objects found throughout the image (Bushey, 2024). This process is called image segmentation, of which there are two types: semantic segmentation and instance segmentation. Semantic segmentation identifies semantic objects without specifying how many instances of that object are present in the image (Boesch, 2022). So for example, if there are five people in a photograph, semantic segmentation will identify them as a single semantic object, rather than five instances of the same semantic object. Comparatively, instance segmentation identifies all the semantic objects in an image and segments them into specific instances to accurately distinguish between each object of a similar semantic category (Boesch, 2022). Mask R-CNN relies on image segmentation to first identify semantic objects before taking a second pass of the image to classify the objects and refine their boundaries using both semantic and instance segmentation.

A good example of a CBIR system built on object detection algorithms in an archival setting comes out of the Galt Museum and Archives in Lethbridge, Alberta. Working in collaboration with the Swiss software company "4eyes GmbH" the Galt Archives has implemented the Archipanion search tool into their access services, which is an AI model designed to understand and interpret image-based records without any additional written information attached to the record like metadata or descriptions (Tah, 2024). The model is purpose-built for archives to make their collections more immediately searchable and accessible, without the time-consuming process of manually filling out metadata and image descriptions before users can access the records. At the Galt Archives, Archipanion has two main search options available for users to browse the collection and find similar images. First, the 'scene search' method allows users to describe an image using keywords or ideas, with the subsequent results based on the system's interpretation of the prompt. These keywords can even include more abstract concepts like 'dangerous activities', for instance, as opposed to solely tangible objects visible within a record (Tah, 2024). Second, the 'search by text' method pulls together records which have similar text present within the images. For example, searching 'store' with the 'search by text' method would pull up all the records which have the word 'store' visible within the image itself, while doing the same search using the 'scene search' would pull up results which match the system's interpretation of a store. In this sense, Archipanion not only provides new ways to search records, but also makes records with no descriptions or metadata attached visible to users when they would otherwise not be available due to archival backlogs. As such, the model represents a big step forward in terms of meaningful applications of AI to archival workflows which lessen the burden of backlogs on archivists and makes collections more accessible for users.

In terms of specific algorithms being applied to archival work, the YOLO algorithm in particular, has been extensively implemented when it comes to object detection in historical image-based collections, as mentioned above in the discussion of the Finnish Wartime Photograph Project and the Perganet Project. There have been major improvements on the initial YOLO algorithm since its release in 2015 including refining its feature extraction capabilities and accuracy, as well as supporting a broader range of tasks beyond object detection like image classification, instance segmentation, and pose estimation, which involves identifying specific points of interest within processed images (Ultralytics, n.d.). As a one-shot detector, the YOLO algorithm processes the entire image through a neural network at once rather than segmenting it sequentially, while identifying objects and predicting categories for them simultaneously (Yildiz & Rukanci, 2025). This macro-perspective of object recognition makes the YOLO algorithm well-suited to efficiently complete tasks like facial recognition, scene analysis and object tracking without sacrificing accuracy or precision. The use of computer vision and the YOLO algorithm in particular is increasingly popular in archives, as it allows for more efficient organization and retrieval of image-based records and enables users to access the materials more easily and analyze them deeper (Yildiz & Rukanci, 2025). For instance, using a Google Colab workspace and a YOLOv9 algorithm, Yildiz and Rukanci developed a facial recognition model for prominent 20th century figures in Turkish cinema (Yildiz & Rukanci, 2025). Working with a curated dataset of 1638 images, which was expanded to just under 4000 images using data augmentation (i.e., synthetic data), the model was trained to recognize figures with an accuracy rate of 91.8% and a recall rate of 85.2% (Yildiz & Rukanci, 2025). This experiment demonstrates the extent to which YOLO-based models are becoming increasingly robust, in particular when used with historic visual collections. In this sense, the YOLO algorithm is a valuable tool for archivists for improving the organization, classification, and accessibility of large image-based collections which could not be processed as efficiently manually.

ACTIVITY #3: Using YOLO for object detection

This activity allows students to experiment with AI object detection using a pre-built YOLO algorithm available on Google Colab here. Students should make a copy of the code and save it to their drive. Then, following the instructions in the notebook, students should run through the code and experiment with the object detection algorithm by choosing a custom dataset from the Roboflow 100 Universe.

Archival Processes for image-based records using AI/ML

As discussed throughout this module so far, there are some key archival processes which can be enhanced using AI and machine learning. In particular, there has been significant work done in the sphere of using AI for improving archival metadata. For example, the Computer–Aided Metadata generation for Photo Archives Initiative (CAMPI) was developed in 2020 at the Carnegie Mellon University Library and designed to help improve limited or inaccurate tags on archival images (Lincoln et al., 2020). CAMPI, as a web-based platform, allows archivists to tag images and it recommends other similar images based on their visual content to be tagged with the same label. It allows archivists to find related materials more easily and enrich their metadata, especially those with limited or inaccurate descriptions (Lincoln et al., 2020). CAMPI can also be used to retroactively tag images already uploaded online, which is helpful for finding potentially culturally inappropriate descriptions that need to be corrected. In this sense, CAMPI illustrates how AI can improve archival descriptions, not by directly writing them using a LLM or generative AI, but instead through using computer vision to garner more information from the image itself for improving search terms and other metadata associated with the record.

AI and machine learning have also been key tools for improving diplomatic analysis of medieval documents, as explained earlier in this module with the case of the Perganet project. This project is particularly concerned with identifying and classifying notary signatures and seals, which are essential in a diplomatic analysis of archival records (Paolanti et al., 2022). Moreover, Perganet is designed to work with large corpuses of scanned medieval parchments, which helps make information about the records more accessible without contributing to their deterioration through handling during manual analysis (Paolanti et al., 2022). As such, the Perganet project demonstrates how computer vision applications, when developed with archival diplomatic principles in mind, can have major impacts on reducing backlogs and making records more promptly accessible to users.

Nonetheless, while there is considerable potential in integrating AI/ML into metadata creation and diplomatic analysis workflows, other aspects of archival work have yet to see the implementation of these tools. Digital preservation workflows, as an example, have yet to see any significant applications of AI or machine learning technologies. The digital preservation process relies primarily on the Open Archival Information System or OAIS model, whose implementations, as of yet, predate recent applications of AI/ML tools. Additionally, the OAIS model and digital preservation processes more broadly have been primarily developed with the long-term preservation of records in mind and rely on the use of trustworthy digital repositories (TDR). TDRs are unlikely to integrate technologies that are consistently changing, uncertain, and evolving and therefore, it will be sometime until AI/ML tools make their way into TDRs (Oliviera et al., 2023). Even so, it is still relevant for archivists and records managers to consider what aspects of the digital preservation process could be improved with AI/ML tools in the hopes of encouraging further research into the integration of these applications into digital preservation workflows in the future.

Exploring GenAI or AI-modified images as records

With the advent of generative AI in particular, there has also been growing discussions around how archivists and records managers can detect and identify potentially AI-generated images coming into the archives. Paradata, as a method for documenting how AI tools are being used and the outputs they produce, can be a useful tool for investigating the origin of born-digital images coming into collections. As AI presents a challenge to archivists and records managers aiming to protect the authenticity and trustworthiness of records held in their collections, paradata can serve as a tool to identify how AI was used in the context of creation of the records, and document the changing provenance of records modified or created with AI (Cameron et al., 2023). Paradata is necessary for combatting the often opaque, 'black box,' AI models, which as they become increasingly sophisticated, become less interrogatable. In this sense, as GenAI models, especially those with image generation capabilities, are often unexplainable, they increase the likelihood of manipulation, misinformation, inaccurate attribution, and bias (Bushey, 2023). As a result, documenting the entirety of the scope, context, and use of AI for processing or creating records is essential for minimizing risk and ensuring the trustworthiness of the record. Paradata is helpful in this scenario for going beyond explainable AI and extending the question to how, why, and to what extent AI tools were used in a given context of creation of a visual record, including documenting how human actors were (or not) involved in the process. As such, as AI-generated or modified materials continue to emerge as a potential record format, archivists and records managers should consider using paradata as a method for ensuring the authenticity, transparency and impartiality of records accepted into archival collections.

Even with the advent of paradata as a tool for documenting AI processes, there are still concerns for archivists around handling AI-generated or modified image-based records. As Bushey highlights, current efforts to explore AI-generated images as an emerging record format have not meaningfully engaged with archival and recordkeeping knowledge (Bushey, 2023). In particular, there are three main concerns around how AI-generated or modified images are processed in archival settings. First of all, current approaches to metadata capture and creation, which are crucial aspects of archival work, are still being developed for AI-generated or modified records, and may not meet the archival standards for properly describing and contextualizing an AI record throughout its lifecycle (Bushey, 2023). Second, there is a need to preserve the original data used to create or modify an AI image, including things like training data, original and unaltered images, algorithms, and prompts, to ensure the authenticity of the images in an archival sense (Bushey, 2023). Finally, archivists need to learn about and be prepared for acquiring and appraising large volumes of born-digital records, especially those which contain AI-generated or modified images, to assess current preservation strategies and determine whether new approaches are needed for extensive born-digital collections (Bushey, 2023).

When it comes to metadata creation and capture, the transdisciplinary field of computational archival science (CAS), developed primarily by Dr. Victoria Lemieux and Dr. Richard Marciano, proposes using computational workflows to rapidly convert descriptive image metadata into linked data for use in reference and access work (Proctor and Marciano, 2021). Computational processing of AI-generated or modified images could provide archivists with a solution to the time and labour-intensive work of creating and capturing metadata by leveraging metadata captured by the software, systems and platforms used to create or modify AI images. Moreover, several large technology companies including Microsoft, Intel, Arm, TruePic and Adobe have joined together in the Coalition for Content Provenance and Authenticity (C2PA) to develop "technical standards for certifying the source and history (referred to as provenance) of media content" (Coalition for Content Provenance and Authenticity, n.d.). Formally introduced in 2022, the C2PA has presented a technical standard for cryptographically binding provenance data to born-digital records, including AI-generated or modified images, which can be applied to records using C2PA-enabled software and platforms, or through certified digital signatures of actors involved in the creation of modification of the records (Bushey, 2023). For instance, the development of the Content Credentials system by Adobe's Content Authority Initiative (CAI) has implemented the C2PA technical standard in partnership with Leica cameras to capture technical metadata about camera makes and models, along with additional metadata about the images' content, and whether it has been edited or altered, which is essential information for providing an "unbroken chain of authenticity from the time a picture is taken to the time it is published" (Bushey, 2023). While this is a new approach and still undergoing development, Content Credentials are a valuable tool for archivists processing large collections of born-digital records for evaluating the authenticity and provenance of the records and it presents a starting point for the capture and creation of metadata about AI records to improve their accessibility, transparency, and usability.

Similarly, archival diplomatics has also been investigated as a method of assessing the authenticity of AI records. Bushey highlights that applying diplomatics to AI records requires identifying key attributes in the records themselves which must be preserved long-term in order to demonstrate and maintain the record's authenticity, reliability, and accuracy over time (Bushey, 2023). In this sense, archival diplomatics can be used as a conceptual framework for defining the necessary requirements for creating, managing, and preserving trustworthy image-based records in trusted digital repositories (Bushey, 2023). Furthermore, the capture, creation, management, and preservation of born-digital records' technical, descriptive and administrative metadata must also undergo diplomatic analyses to ensure they meet the baseline requirements for verifying a record's authenticity, especially concerning AI-generated or modified images. As an example, law enforcement, journalism, and even wildfire services, particularly in Canada and in the US, but also more broadly around the world, have faced challenges in regulating AI-generated images produced to spread disinformation, called deepfakes. Deepfakes, or AI-generated synthetic media, rely on computer vision, facial recognition software and deep learning to "manipulate existing images and produce realistic-looking images of people that are false" (Bushey, 2023). Deepfakes have been used to respond to political and cultural events in real-time, spread mis- and dis-information, as well as facilitate identity theft and manipulate evidence submitted for legal proceedings, as there are few mechanisms in place to meaningfully assess and verify the authenticity of an image in law enforcement, government services, and journalistic contexts (Bushey, 2023). In this sense, diplomatic analysis can help identify the baseline requirements of authenticity for born-digital image-based records, which could be valuable in discerning real images from deepfakes, especially in legal settings. Despite further research being required to concretely establish the key attributes required for authenticity in AI-generated or modified images, the conceptual diplomatic framework is still valuable for archivists and records managers considering how to process these types of records into their collections and maintain the full context around the image.

Limitations to processing image-based records with AI/ML

Although there is significant potential in using AI and machine learning technologies for processing image-based records, there are still limitations and challenges associated with using these tools. Most prominently, as with most AI tools, is the concern with bias. Like other AI models, computer vision tools' outputs are highly dependent on the training data they receive, which can often reflect existing social inequalities and entrench them within the algorithms. For instance, racial bias has been an ongoing challenge with computer vision tools, particularly with facial recognition software. A study published by MIT Media Lab in 2018 found that facial recognition software had an average error rate of 0.8% for light-skinned men, versus a 34.7% error rate for darker-skinned women (Fergus, 2024). Similarly, a 2023 experiment with StableDiffusions text-to-image GenAI found that prompts about high-paying white-collar jobs yielded images of men with lighter skin, while prompts about criminal activities like drug dealing and terrorism primarily resulted in images of men with darker skin (Nicoletti & Bass, 2023). Both of these examples illustrate how societal biases are often engrained in AI model training datasets, which lead to the models reproducing those biases in their outputs. In addition to this, the training data of computer vision models has relied mostly on visual content available to be harvested from the Internet, which in its majority corresponds to content developed during the last three decades. These models would underperform on images that are of a more historical character, commonly found in visual archives, since they are underrepresented in the training data (Stančić, 2024). As a result, it is important for archivists and records managers to consider how the models they are using were trained, what types of biases may subsequently surface in the model's outputs, and how to adjust for them. In an ideal world, archivists choosing to use AI in their workflows would use models purposefully-trained for archival contexts using relevant training data (Stančić, 2024). However, custom-built AI models are often costly and labour-intensive to develop, and are likely not realistic for most institutions, particularly smaller communities.

Another challenge when it comes to using computer vision tools, and particularly GenAI, is the question of copyright. The Copyright Office of the United States has ruled copyright can only apply to materials with human creativity, and AI-generated images, although potentially promoted by a human actor, do not reasonably contain human authorship (Bushey, 2023). Moreover, there have been concerns around the uncompensated use of copyrighted works in the development or training of AI tools using data mining techniques, as well as the potential for AI to infringe on existing copyrighted works (Bushey, 2023). Discussions about how to approach attribution for AI-generated or modified content are ongoing, yet there has still been minimal enforcement remedies for rights holders whose work has already been fed to AI algorithms and/or replicated using GenAI. In this sense, archivists should always check the copyright restrictions of materials in their collections before using them to train AI/ML models, especially when these models are not hosted locally. More broadly, archivists need to seriously consider the costs, benefits, and risks associated with ingesting image-based records into AI models, particularly generative ones, which could use the records as training data at a later date, especially if there is a donor agreement, legal restrictions (e.g., privacy, jurisdiction-bound records, etc.), or a community agreement in place (e.g. stewardship of records from Indigenous communities) (See Module 2). While the potential of AI models to improve archival workflow efficiency and reduce backlogs is high, care should always be taken to ensure ethical archival practices as well as the integrity and authenticity of the records are maintained.

SUMMARY

In summary, while processing image-based records with AI is not necessarily new to archives, as computer vision tools like content-based image retrieval (CBIR) and query-by-image-content (QBIC) have been used in these spaces for over thirty years, the rapid development of tools like reverse image-searching, facial recognition software, and deep learning architectures has opened up new possibilities for processing image-based collections more efficiently and effectively. In particular, tools like YOLO algorithms, convolutional neural networks (CNNs) and vision-language models have the potential to aid in metadata creation, object detection, and even diplomatic analysis. Furthermore, projects like Perganet, Archipanion, and CAMPI illustrate how practical applications of AI tools to image-based records can improve their discoverability and accessibility to users in archival contexts. Nonetheless, archivists must remain mindful of the limitations of these tools when it comes to bias and copyright issues, particularly when working with generative AI, as well as the emerging concerns around protecting archival authenticity when working with AI-generated or modified images as records. Still, by combining technical awareness with critical reflection and using strategies such as paradata and archival diplomatics, archivists and records managers can ensure that the integration of AI into archival workflows improves accessibility without compromising trustworthiness or ethical responsibility. Ultimately, while the thoughtful integration of AI/ML has the potential to reduce backlogs, improve discoverability, and expand the possibilities for engaging with visual records, its use must always be guided with care by archival principles.

REFERENCES

Allen, M. (1987). Optical Character Recognition: Technology with New Relevance for Archival Automation Projects. *The American Archivist*, *50*(1), 88–99.

Ashraf, R., Bashir, K., Irtaza, A., & Mahmood, M. T. (2015). Content Based Image Retrieval Using Embedded Neural Networks with Bandletized Regions. *Entropy*, *17*(6), Article 6. https://doi.org/10.3390/e17063552

Boesch, G. (2022, March 19). Everything About Mask R-CNN: A Beginner's Guide. *Viso.Ai*. https://viso.ai/deep-learning/mask-r-cnn/

Bushey, J. (2023). AI-Generated Images as an Emergent Record Format. *2023 IEEE International Conference on Big Data (BigData)*, 2020–2031. https://doi.org/10.1109/BigData59044.2023.10386946

Caballar, R. D., & Stryker, C. (2025, February 25). *What Are Vision Language Models (VLMs)?* https://www.ibm.com/think/topics/vision-language-models

Castelli, V., Bergman, L. D., Kontoyiannis, I., Li, C.-S., Robinson, J. T., & Turek, J. J. (1998). Progressive search and retrieval in large image archives. *IBM Journal of Research and Development*, *42*(2), 253–268. https://doi.org/10.1147/rd.422.0253

Chumachenko, K., Mannisto, A., Iosifidis, A., & Raitoharju, J. (2020). Machine Learning Based Analysis of Finnish World War II Photographers. *IEEE Access*, *8*, 144184–144196. https://doi.org/10.1109/ACCESS.2020.3014458

Craine, A. G. (n.d.). *Ray Kurzweil*. Britannica. Retrieved May 1, 2025, from https://www.britannica.com/biography/Raymond-Kurzweil

Date, K., & Allweil, Y. (2022). Towards a new image archive for the built environment. *Environment and Planning B*, *49*(2), 519–534. https://doi.org/10.1177/23998083211011474

Delua, J. (2021, March 12). Supervised vs. Unsupervised Learning: What's the Difference? *IBM Blog*. https://admin01.prod.blogs.cis.ibm.net/blog/supervised-vs-unsupervised-learning/

Dirac, L., Kumar, S., & Wright, F. (2024, July 16). *How does a Vision-Language-Model (VLM) work?* Groundlight.Ai. https://www.groundlight.ai/blog/how-vlm-works-tokens

Fergus, R. (2024, February 29). *Biased Technology: The Automated Discrimination of Facial Recognition*. ACLU of Minnesota. https://www.aclu-mn.org/en/news/biased-technology-automated-discrimination-facial-recognition

Hamadi, R. (2023). *Large Language Models Meet Computer Vision: A Brief Survey* (arXiv:2311.16673). arXiv. https://doi.org/10.48550/arXiv.2311.16673

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. https://doi.org/10.48550/arXiv.1512.03385

Holt, B., & Hartwick, L. (1994). Retrieving art images by image content: The UC Davis QBIC project. *Aslib Proceedings*, *46*(10), 243–248. https://doi.org/10.1108/eb051371

IBM. (2021, July 27). *What is Computer Vision?* https://www.ibm.com/think/topics/computer-vision

Khan, A., Laghari, A., & Awan, S. (2018). Machine Learning in Computer Vision: A Review. *ICST Transactions on Scalable Information Systems*, 169418. https://doi.org/10.4108/eai.21-4-2021.169418

Kundu, R. (2023, January 17). *YOLO Algorithm for Object Detection Explained*. https://www.v7labs.com/blog/yolo-object-detection

Lincoln, M., Corrin, J., Davis, E., & Weingart, S. B. (2020). *CAMPI: Computer-Aided Metadata Generation for Photo archives Initiative*. Carnegie Mellon University. https://doi.org/10.1184/R1/12791807.v2

Mechi, O., Mehri, M., & Essoukri Ben Amara, N. (2022). Recognizing Handwritten Text Lines in Ancient Document Images Based on a Gated Residual Recurrent Neural Network. In C. Bădică, J. Treur, D. Benslimane, B. Hnatkowska, & M. Krótkiewicz (Eds.), *Communications in Computer and Information Science* (pp. 250–263). Springer International Publishing. https://doi.org/10.1007/978-3-031-16210-7_20

Nicoletti, L., & Bass, D. (2023, June 9). Humans Are Biased. Generative AI Is Even Worse. *Bloomberg.Com*. https://www.bloomberg.com/graphics/2023-generative-ai-bias/

Oliveira, A., Moura, L., & Terra, A. L. (2021). Trusted Digital Repositories for Document Preservation: A Tools' Evaluation. *Bobcatsss*, 280–287. https://doi.org/10.34630/bobcatsss.vi.4986

Paolanti, M., Pietrini, R., Della Sciucca, L., Balloni, E., Compagnoni, B. L., Cesarini, A., Fois, L., Feliciati, P., & Frontoni, E. (2022). PergaNet: A Deep Learning Framework for Automatic Appearance-Based Analysis of Ancient Parchment Collections. In P. L. Mazzeo, E. Frontoni, S. Sclaroff, & C. Distante (Eds.), *Image Analysis and Processing. ICIAP 2022 Workshops* (Vol. 13374, pp. 290–301). Springer International Publishing. https://doi.org/10.1007/978-3-031-13324-4_25

Piras, L., & Giacinto, G. (2017). Information fusion in content based image retrieval: A comprehensive overview. *Information Fusion*, *37*, 50–60. https://doi.org/10.1016/j.inffus.2017.01.003

Proctor, J., & Marciano, R. (2021, December 15). *An AI-Assisted Framework for Rapid Conversion of Descriptive Photo Metadata into Linked Data*. IEEE International Conference on Big Data. https://doi.org/10.1109/BigData52589.2021.9671715

Shafiq, M., & Gu, Z. (2022). Deep Residual Learning for Image Recognition: A Survey. *Applied Sciences*, *12*(18). https://doi.org/10.3390/app12188972

Singh, P. N., & Gowdar, T. P. (2021). Reverse Image Search Improved by Deep Learning. *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, 596–600. https://doi.org/10.1109/MysuruCon52639.2021.9641572

Smeaton, A. F., Gurrin, C., & Lee, H. (2006). Interactive Searching and Browsing of Video Archives: Using Text and Using Image Matching. In R. I. Hammoud (Ed.), *Interactive Video: Algorithms and Technologies* (pp. 189–206). Springer. https://doi.org/10.1007/978-3-540-33215-2_9

Stančić, H. (2024). Training AI on digitised portraits. 5th International Symposium, ITrustAI. Honolulu, HI, February 23, 2024. https://interparestrustai.org/assets/public/dissemination/3-Stancic-TrainingAIondigitisedportraits_pptx.pdf

Szeliski, R. (2011). Introduction. In R. Szeliski (Ed.), *Computer Vision: Algorithms and Applications* (pp. 1–25). Springer. https://doi.org/10.1007/978-1-84882-935-0_1

Tah, P. (2024, February 5). *Galt Museum Introduces a Revolutionary Image Search AI Model*. Galt Museum & Archives. https://www.galtmuseum.com/news/galt-museum-introduces-a-revolutionary-image-search-ai-model

Yıldız, M., & Rukancı, F. (2025). AI-powered visual classification in archives: A computer vision approach to facial recognition in historical archives. *Archival Science*, *25*(2), 18. https://doi.org/10.1007/s10502-025-09486-w