The Application of Machine Learning to At-Risk Cultural Heritage Image Data

ROBERTS, MATTHEW,IAN

The Application of Machine Learning to At-Risk Cultural Heritage Image Data

Matthew Roberts

A Thesis presented for the degree of
Master by Research

Department of Archaeology
University of Durham
England
November 2019

The Application of Machine Learning to At-Risk Cultural Heritage Image Data

Matthew Roberts

Submitted for the degree of Master by Research
November 2019

Abstract

Abstract

This project investigates the application of Convolutional Neural Network (CNN) methods and technologies to problems related to At-Risk cultural heritage object recognition. The primary aim for this work is the use of developmental software combining the disciplines of computer vision and artefact studies, developing applications in the field of heritage protection specifically related to the illegal antiquities market. To accomplish this digital image data provided by the Durham University Oriental Museum was used in conjunction with several different implementations of pre-trained CNN software models, for the purposes of artefact Classification and Identification. Testing focused on data capture using a variety of digital recording devices, guided by the developmental needs of a heritage programme seeking to create software solutions to heritage threats in the Middle East and North Africa (MENA) region. Quantitative data results using information retrieval metrics is reported for all model and test sets, and has been used to evaluate the models predictive results.

Table of Contents

List of Figures

ACRONYMS/ABBREVIATIONS

CIDOC	International Committee for Documentation
CMS	Collection Management System
CNN	Convolutional Neural Networks
COTS	Custom-Off-The-Shelf
CRISP-DM	Cross-industry standard process for data mining
GPU	Graphics Processing Units
ICOM	International Council of Museums#
JPL	Jet Propulsion Laboratory
MEC	Model Export Certificate
MIT	Massachusetts Institute of Technology
ML	Machine Learning
SQL	Structured Query Language
UNESCO	United Nations Educational, Scientific and Cultural Organization (UNESCO)
UNIDROIT	International Institute for the Unification of Private Law
VASARI	Visual Arts System for Archiving and Retrieval of Images
WCO	World Customs Organization

DECLARATION OF AUTHORSHIP

I, Matthew Roberts, declare that this thesis titled, "The Application of Machine Learning to At-Risk Cultural Heritage Image Data" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____

Date: ____

Acknowledgements/Dedication

This thesis would not have been possible without the resources from the Durham University Oriental Museum, and I would like to give my most sincere gratitude to Helen Armstrong, Rachel Barclay, Craig Barclay, and Charlotte Spink for all their support these past several years.

Thanks to Dr. Kamal Badreshany and Dr. Marco Nebbia for their guidance in beginning this work, and to my supervisor Professor Anna Leone for the knowledge and encouragement necessary to see it to its end. Thanks to the Training in Action and the Heritage Documentation and Protection projects for support material and equipment.

Thanks to Dr. Boguslaw Obara, Dr. Chris Willcocks, Dr. Noura Al Moubayed, Stephen Bonner, and John Brennan from the Computer Science department for all their technical reviews and insights. Thanks to the CS postgraduate researcher crew for advice, support, and late-night Karaoke.

Thanks to Ustinov College for providing friends and family far from home.

Thanks to Mom; not all who wander are lost.

Thanks to Dad; no matter where you go, there you are.

Glossary[1]

Activation Function - A function which takes in the weighted sum of all of the inputs from the previous layer and then generates and passes an output value to the next layer.

Backpropagation - The primary algorithm for performing gradient descent for a neural network. First, the output values of each node are calculated in a forward pass. Then, the partial derivative of the error with respect to each parameter is calculated in a backward pass.

Batch - The set of examples used in one iteration (that is, one gradient update) of model training.

Bias - An intercept or offset from an origin. Bias (also known as the bias term) is referred to as *b* or w_0 in machine learning models.

Class - One of a set of enumerated target values for a label. For example, in a binary classification model that detects spam, the two classes are spam and not spam. In a multi-class classification model that identifies dog breeds, the classes would be poodle, beagle, pug, and so on.

Convergence - Informally, often refers to a state reached during training in which training loss and validation loss change very little or not at all with each iteration after a certain number of iterations. In other words, a model reaches convergence when additional training on the current data will not improve the model.

Epoch - A full training pass over the entire dataset such that each example has been seen once. Thus, an epoch represents N/batch size training iterations, where N is the total number of examples.

Feature - An input variable used in making predictions.

Gradient Descent - A technique to minimize loss by computing the gradients of loss (a vector of partial derivatives of the model function, pointing in the direction of steepest ascent) with respect to the model's parameters, conditioned on training data. Informally, gradient descent iteratively adjusts parameters, gradually finding the best combination of weights and biases to minimize loss.

Learning Rate - A scalar used to train a model via gradient descent. During each iteration, the gradient descent algorithm multiplies the learning rate by the gradient. The resulting product is called the gradient step.

Model - The representation of what a machine learning system has learned from the training data.

Neural Network - A model that, taking inspiration from the brain, is composed of layers (at least one of which is hidden) consisting of simple connected units or neurons followed by nonlinearities (i.e. calculations which do not vary outputs in proportion to their inputs).

Neuron - A node (connected unit) in a neural network, typically taking in multiple input values and generating one output value. The neuron calculates the output value by applying an activation function to a weighted sum of input values.

[1] Definitions taken from the Google Developers Machine Learning Glossary https://developers.google.com/machine-learning/glossary

Normalization - The process of converting an actual range of values into a standard range of values, typically -1 to +1 or 0 to 1. For example, suppose the natural range of a certain feature is 800 to 6,000. Through subtraction and division, you can normalize those values into the range -1 to +1.

Overfitting - Creating a model that matches the training data so closely that the model fails to make correct predictions on new data.

Partial Derivative - A derivative in which all but one of the variables is considered a constant. For example, the partial derivative of $f(x, y)$ with respect to x is the derivative of f considered as a function of x alone (that is, keeping y constant). The partial derivative of f with respect to x focuses only on how x is changing and ignores all other variables in the equation.

Test Set - The subset of the dataset that you use to test your model after the model has gone through initial vetting by the validation set.

Training - The process of determining the ideal parameters comprising a model.

Training Set - The subset of the dataset used to train a model.

Validation - A process used, as part of training, to evaluate the quality of a machine learning model using the validation set. Because the validation set is disjoint from the training set, validation helps ensure that the model's performance generalizes beyond the training set.

Validation Set - A subset of the dataset—disjoint from the training set—used in validation.

Weight - A coefficient for a feature in a linear model, or an edge in a deep network. The goal of training a linear model is to determine the ideal weight for each feature.

1 INTRODUCTION

In the field of archaeological preservation, digital image data has become increasingly common in heritage documentation. Preserving heritage assets relies on recording as much accurate information as possible, both in service to the values of heritage custodianship (i.e. that the act of capturing a recording by itself has value) as well as a way to safeguard against potential threats to that heritage (vandalism, war, theft, natural degradation and many other potential destructive hazards). Accomplishing this is heavily reliant upon the tools available at the time of the heritage effort; nowadays, those toolsets might include digital cameras and tablets, remote sensing drones and many other methods of “born-digital” recordings. These tools in turn generate new types of documentation information, such as images, 3D models, and multispectral recordings, along with all the associated metadata. Their adoption has the potential to improve the work of heritage professionals by increasing the volume and ease with which asset data can be collected. However, it is this very potential which can also serve as the greatest challenge to heritage professionals, as the expertise necessary to properly utilize the tools and manage the information they produce can in turn overwhelm the resources of the project if not planned for appropriately. Vast amounts of data are only useful if a method exists to elicit meaningful information from that collection. The goal of automating the search of digital images in particular has been a cross-disciplinary research goal for many diverse fields, all with the ultimate intent of achieving better methods of data retrieval; in Computer Science and Data Engineering, this has traditionally fallen under the category of Computer Vision and Image Processing (da Silva Torres and Falcao 2006) (Liu, et al. 2007).

1.1 COMPUTER VISION AND IMAGE PROCESSING

The beginnings of meaningful digital image processing and computer vision tasks took place in the 1960's and 70's, driven in large part by the development of advanced computer systems by both the space program (such as the Jet Propulsion Lab Ranger-7 transmitting images of the moon in 1964) as well as new techniques and technologies in the medical imaging profession (the development of computerized axial tomography for medical diagnosis, for example) (Gonzalez and Woods 2006, 3-7). Spurred forward by this progression of technology, academic research began to explore the potential applications and capabilities of computer vision. These initial efforts were in many cases more ambitious rather than achievable, given the technology of the time; a prime example of this is the Massachusetts Institute of Technology (MIT) Summer Vision project (Papert 1966), whose stated purpose was "an attempt to use our summer workers effectively in the construction of a significant part of a visual system" and whose final goal was "OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects." Even at this early stage of development researchers were considering the possibilities for computer enhanced photography to support artefact analysis (Burton, et al. 1970). Archaeologists and mathematicians with an interest in the use of computer applications began to pursue common cause and organization, with the first Computer Applications in Archaeology conference being held in Birmingham in 1973 (J. D. Wilcock 1973). Broadly speaking, much of the published work from this period focused on using computer imaging in relation to aerial photography, artefact shape imaging (usually for pottery profiling), and the replacement of traditional hand-drawn site and artefact mapping with software drafting tools (Ryan 1988, 16-18). Moreover, this type of work was constrained by researchers with access to mainframe computer facilities, and as such was primarily limited to western institutions. It wasn't until the 1980's that archaeologists began to methodically explore the possibilities of computer graphics and their potential application to archaeological science, in large part due to the availability of microcomputers to individual archaeologists (J. D. Wilcock 1999). However, the technology necessary for the implementation and support of digital archaeology was still not widespread in terms of availability and scope. The catalyst necessary to see wider adoption only came about when digital cameras and scanners became commercial commodities.

1.1.1 Digitization and Reconstruction of Heritage

The early 1990s marked the point at which digital image capture began to replace traditional photographing techniques. In the wider business environment, steady effort was being made to convert physical paper records to digital as part of cost saving measures; digital imaging of the time was focused on how to then create systems capable of providing recognition and retrieval of these document’s textual and image components (Cullen, Hull and Hart 1997). Given this business-oriented focus, it is unsurprising that the first regular usage of digital imaging related to heritage came about from the commercial art world, rather than in museums and the field (Hamber 2006). At the beginning of the decade digital camera technology was still in its infancy and had not yet become a commonly held item of personal use. The Auction House Christie's was one of the first institutions concerned with photography of heritage objects to devote research and development towards implementing a fully digital replacement of traditional photography. Post-processing was the main developments here, balancing out image capture against the choice of color profiles. At the time commercial and industrial standards were only beginning to be discussed, most notably the specification produced by the International Color Consortium (a group formed by Adobe, Agfa, Apple, Kodak, Microsoft, Silicon Graphics, Sun Microsystems, and Taligent in 1993) (International Color Consortium 1996). While there were numerous challenges faced by the implementation of the new technology, the project nevertheless showed that the new born-digital workflow sped up process flow, improved quality, reduced costs, and created the opportunity for new services.

1.1.1.1 Antique Books and Manuscripts

The use of computer vision to help with reading faded or damaged historical manuscripts is one of the earliest examples of digital imaging being used in heritage work; the research performed by historians at the California Technical Institute (Dr. John Benton, J. M. Soha, and A. R. Gillespie) are an early demonstration of this development (Benton, Gillespie and Soha 1979). Utilizing image enhancing techniques developed at JPL to restore and clean up images received from Mars (including contrast enhancement and spatial filtering), they managed to restore previously illegible medieval documents and advance the state of discussion for sections previously in dispute. However, after a decade of evaluation, the team’s conclusion in the 1980’s was that digital imaging was not yet commercially viable for manuscript studies (Kiernan 1991). Nevertheless, other heritage experts inspired by their work continued to explore the usage of digital imaging, testing it against known problems in manuscript visualization. Again, this early work was in many ways limited by expensive equipment found in specialized research facilities (Caltech and the British Library, for example). Dedicated digital cameras were being commercially produced (primarily for medical imaging), and financing initiatives by top archive institutions were beginning to see the value in supporting digitization efforts (such as the British Library’s Initiative for Access programme, producing the Electronic Beowulf (Prescott 1997)). The landmark effort to digitize the damaged and fragile Beowulf manuscript demonstrated several advantages that would become the focus of heritage efforts for the next decade; primarily this included increasing accessibility for research (combining efforts from the British Museum, Harvard, and the University of Kansas), improving workflows (UV and infrared lighting of the document which had been time and cost prohibitive previously suddenly became technically practical for all elements of the folio), and providing the raw digital material for image processing (restoring known missing letters lost to deterioration or hidden by previous conservation efforts, as well as removing discolorations and improving legibility through thresholding techniques). The project was, in fact, the foundational example used by the British Library to describe their new strategic direction to support digital collaborations in research and scholarship (Brindley 2002).

1.1.1.2 Easel Paintings and Photographs

Given the enthusiasm for the potential benefits of digital imaging, it is unsurprising that parallel efforts were also being developed for paintings, photographs, and other similar two-dimensional heritage objects. One of the earliest initiatives to replace traditional photography with a digital system was the European visual arts system for archiving and retrieval of images (VASARI) project (Martinez, Cupitt, et al. 2002), initially funded between 1989 and 1992. Installed both in the National Gallery in London as well as the Doerner Institute in Munich, its purpose was to demonstrate the feasibility of high-resolution digital image processing techniques (Martinez 1991). As with the Initiative for Access programme, many of the key advantages of digital restoration were demonstrated by this project: experiments with a virtual copy do not damage the original, can be more easily supplied to researchers, and have the potential of automation for algorithmic corrections (Stanco, Restrepo and Ramponi 2011). Moreover, these types of objects suffer similar degradation issues as manuscripts, including aging discolouration (chemical, fungal, metal-induced, etc) or staining. Over the course of a decade, the project was used to aid several areas of conservation research, including detection of colour change over a period of time, damage resulting from transportation, simulation of conservation efforts, and superposition of multispectral images for comparative studies. Further research extended upon those initial ideas. Colour pigments that had faded could be returned to their former appearance through pixel hue transformations (Dik, et al. 2002), while crack development (craquelure) from improper storage/transport or internal stress could be automatically detected and removed (Giakoumis and Pitas 1998).

1.1.2 Computer Vision and Museum Collections

The move towards digital brought with it an explosion in the field of image processing techniques, as researchers sought new ways to organize and classify these new archives of digital content (Smeulders, et al. 2000). Colour, shape, and texture all became an area of focus, and effort was made to create methods of object recognition in digital images using local feature description and detection algorithms. Computer vision sought new approaches to region and patch analysis, focused on identifying key points for object recognition which were invariant to changes in scale, viewpoint, and obfuscation (due to clutter/noise). When detecting point or feature (one which persisted over multiple scales), the patch or neighbourhood surrounding it could then be extracted and composed into a feature vector. This in turn could be used for various methods of comparison, often as a histogram of gradients or features. These types of traditional feature engineering algorithms were in turn adapted to the heritage sector, most notably as visual guides for museums. Research which followed these initial efforts has primarily followed the same pattern of implementation, using some variety of feature extraction and engineering software based on interest point detection (Föckler, et al. 2005) (Bay, Fasel and Gool 2006) (Ruf, Kokiopoulou and Detyniecki 2008) (Blighe, et al. 2008). Others have sought to use further specialized methods of image processing to classify according to art type or cultural provenance, including combinations of image segmentation techniques (Haladova 2010) and statistical analysis of edge orientations (Redies, Brachmann and Wagemans 2017). Nevertheless, half a century after MIT's announcement that they would solve Object Identification over the course of the summer, that task remains an aspirational goal within the computer vision field. The fundamental difficulty continues to be the inability for algorithmic processing of raw digital image data to accurately achieve the sort of high level (or semantic) descriptions which a human observer would comprehend and label a visual object (Hare, et al. 2006). Digital image data is stored as a numerical representation, and even slight variations in the recording of an object can change those representational numbers entirely. Such challenges in the domain of image recognition include viewpoint variation, illumination, deformation, occlusion, and background clutter. Traditional image processing initially sought to address