Saving Cultural Heritage with Digital Make-Believe: Machine Learning and Digital Techniques to the Rescue

A. M. Yasser[1, 2, *], K. Clawson[1, **] C. Bowerman[1, ***] and M. Lévêque[3,****]
[1] Department of Computing, School of Engineering, University of Sunderland, Sunderland, SR1 3SD, UK
[2] Physics Department, Faculty of Science at Qena, South Valley University, 83523, Egypt
[3] Computing Department, IUT de Bordeaux, Université de Bordeaux, 33400 Talence, France
[*] Yasser.mostafa@sci.svu.edu.eg, [**] Kathy.clawson@sunderland.ac.uk ,
[***]chris.bowerman@sunderland.ac.uk, [****]maxence.leveque@etu.u-bordeaux.fr

The application of digital methods for content-based curation and dissemination of cultural heritage data offers unique advantages for physical sites at risk of damage. In areas affected by 2011 Arab spring, digital may be the only approach to create believable cultural experiences. We propose a framework incorporating computational methods such as: digital image processing, multi-lingual text analysis, and 3D modelling, to facilitate enhanced data archive, federated search, and analysis. Potential use cases include experiential search, damage assessment, virtual site reconstruction, and provision of augmented information for education and cultural preservation. This paper presents initial findings from an empirical evaluation of existing scene classification methods, applied to detection of cultural heritage sites in the Palmyra region. Results indicate that deep learning offers an appropriate solution to semantic annotation of publicly available cultural heritage image data.

Keywords: Semantic Archive, Image Retrieval, Cultural Heritage, Data Mining, Deep Learning.

1. INTRODUCTION

In this paper, we consider means of preserving Cultural Heritage by digital means.

Cultural heritage is an umbrella term that encompasses physical and, more recently, intangible attributes deemed as having social or aesthetic interest [1]. Examples of cultural heritage include architecture, landscapes, art, language, ritual and human memory [2]. The proposition that cultural heritage is a fundamental human right has resulted in widespread efforts to ensure preservation of tangible and intangible legacy. However, high costs associated with preservation activities, coupled with expanding definitions of cultural heritage, has posed challenges for facilitating sustainable practice [3].

Since the 2011 Arab spring, a number of middle-eastern countries have suffered damage to their cultural heritage. As a result, there is an increased need to protect those important sites. When heritage exists in regions of conflict, preservation activities must operate within a dynamic environment. The destruction of heritage within conflict regions escalates the need to protect tangible and intangible human memory.

It is conceivable that novel solutions to the problem of sustainable, timely preservation of cultural heritage may lie in adoption of digital technologies. Furthermore, wide scale application of digital techniques that assist with curation, archive and retrieval of heritage information may help increase public awareness, and societal responsibility. This paper explores opportunities for an enhanced digital archive and retrieval of heritage information through application of machine learning. Outputs from the image retrieval system will be multimedia and automated reports of the state of repair of cultural artefacts as well as real-time, elucidating comments for site visitors- hence creating a digital cultural make-believe of damaged cultural artefacts.

We specifically suggest a system architecture which incorporates multimodal data analysis, and content-based augmented data retrieval with the aim of assisting preservation endeavours. To demonstrate the suitability of machine learning and semantic technologies for the documentation of cultural heritage we present preliminary findings from a case study focusing on the Palmyra Region, Syria. The tools developed will also be used in training sessions to increase curators skills and knowledge.

Within existing literature, the applicability of deep learning for digital documentation of cultural heritage has been reported [4]. A protocol to guide the processes of digitization of cultural heritage, respecting needs, requirements

and specificities of cultural assets was introduced [5]. It was found that deep learning offers opportunities for accurate semantic annotation of physical cultural legacy. Using a pre-trained model [6], classification accuracy between 89% and 92% has been reported [4]. This paper extends upon existing research by offering a comparative analysis of deep learning against the bag of features classification framework, and by exploring opportunities for integrating open source tools for enhanced semantic annotation of cultural heritage data. The proposed system is summarised in Section 2. Our methodology and experimental overview is presented in Section 3. Results and discussion are offered in Sections 4. Finally, conclusions are provided in Section 5.

2. PROPOSED SYSTEM

We propose to develop a digital heritage search platform (ICARE) that will allow users to archive digital heritage content and perform semantic queries over multimodal cultural heritage data archives. To achieve this, the system will utilise current tools in natural language processing, image processing, and semantic classification. ICARE facilitates interactions between three main entities, specifically:

- Requesters: individuals and organisations who wish to search the data archives.
- Curators: individuals and organisations who wish to submit data to the archives.
- ICARE: the system that analyses multimedia, associated metadata, and text archives, and offers search services to requesters.

We perceive the intended system will:

- Analyse publicly available heritage data, for example social media multimedia and associated metadata.
- Integrate multiple, geographically disparate archives through either manual upload (curator initiated) or automatic public data collection and analysis.
- Support content-based augmented data retrieval via semantic queries, including ‘query-by-location’ and ‘query-by-image’.
- Support future digital reconstruction of damaged and degraded sites of interest.
- Offer opportunities for enhanced public awareness / education regarding heritage training and associated issues.

A reference architecture for ICARE is presented in Figure 1. Core functionality of the ICARE Platform includes:

Data Curation: Data curation functionality is initiated by entities who generate and manage cultural heritage data, for example the Egyptian Supreme Council of Antiquities (SCA) [7]. Curation activities can be automatic (web crawl for relevant public content) or manual (upload specific content).

Data Modelling: This module will perform natural language processing or multimedia data processing to: generate feature vectors from cultural heritage data inputs, and; generate semantic annotations describing the nature of input data, for example objects, buildings, landscapes, and regions. The goal is to apply machine learning to generate search labels interpretable by the Search Module to return rich information to requesters. Natural language processing will also be used to generate automatic reports about cultural artefacts and providing shorter elucidating comments for visitors

Interface: This module exposes the ICARE user interface and associated functionality to end users.

Search: Search functionality involves the cultural heritage data archive and search engine. The search engine will match search queries with archived data and metadata to identify and return relevant content.

3. METHODOLOGY: MACHINE LEARNING APPLIED TO CULTURAL HERITAGE.

The ICARE system is intended to be used to curate, document, retrieve, recreate cultural heritage artefacts, and assist in training. The current system makes use of both cultural heritage data and machine learning- as detailed below.

3.1 Data Collection

Cultural heritage images were collected from Flickr without known restrictions on copyright. The collection process was depending on using text search related to tags of images. After keyword search, returned results were evaluated manually to ensure that only true positive samples of each search term were included. Our final data set comprised 432 images corresponding to 4 regions of cultural interest, specifically Baalshamin (50 samples), Temple of Bel (171 samples), Tetrapylon (113 samples) and Roman Theatre at Palmyra (99 samples). Example images are provided in Figure 2.

3.2 Machine Learning

Semantic annotations are generated using two popular machine learning methods, specifically the Bag of Features (BOF) framework [8], and deep learning using Convolution Neural Networks (CNN) [9, 10]. Image data is also evaluated using the Google Cloud Vision REST API [11].

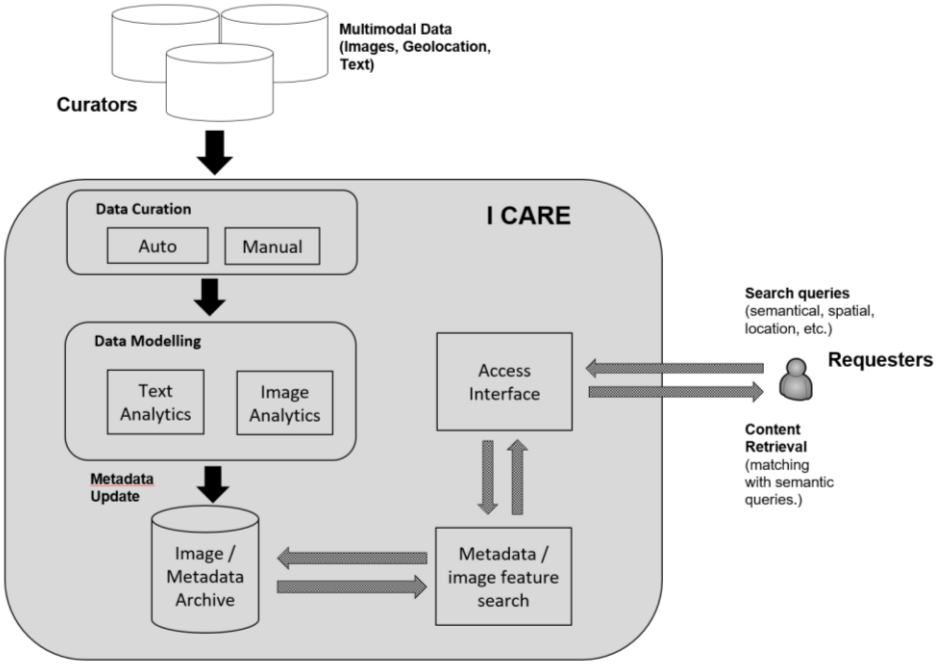

Figure 1: *Proposed System Architecture*

We utilise SIFT features to create a visual vocabulary and BOF representation, and train a multi class support vector machine with Gaussian radial basis function. This methodology is a popular approach to object recognition [8], having high reported accuracy for both image and video scene analysis. Advantages of the BOF methodology include its simplicity, efficiency, and invariance to scale, rotation, and illumination.

Deep learning, in opposition to traditional machine learning methodologies, replaces feature extraction algorithms such as SIFT with large, carefully curated sets of training data and a training algorithm. This unsupervised approach enables features to learned from lower level features such as edges and blobs. Deep learning has been found to achieve state of the art performance across a variety of recognition tasks but a need for very large datasets hence we shall use episodic memory techniques to improve results [13, 14].

We incorporate pre-trained models, specifically AlexNet within a Convolution Neural Network framework [10] and replace the final layer in the network via new training using the categories defined in Section 3.1. This method of transfer learning is applicable for smaller datasets. To exemplify the need for large datasets when training deep networks, we also establish classification models from scratch using Python's Keras package [12] (number of

epochs = 200), and compare results. Across all experiments, data is partitioned into 80% training and 20% test sets and average accuracy over all classes is reported as the performance metric.

4. RESULTS AND DISCUSSION

Mean classification accuracy across all methods is summarised in Table 1. It can be seen from Table 1 that highest classification accuracy is achieved using the transfer learning approach (93%, 1000 iterations). Furthermore, the requirement for large datasets when developing generalised models within deep learning frameworks is exemplified- SVM achieved 11% higher mean classification accuracy than the full CNN (63%, Table 2) when applied to our small dataset. When using full CNN, classification accuracy is a function of the number of epochs, as illustrated in Figure 3. As in general, precision and accuracy increase as number of epochs increase. We envisage future activities to expand the size of our cultural heritage dataset will enable development of more powerful full CNN classification models and increase overall classification accuracy.

Table 1. Mean Classification Accuracy

Method	Accuracy %
SVM	74
Transfer CNN	93
Full CNN	63
Cloud Vision	62

Figure 2. *Sample Images (from top, per row): Baalshamin, Roman Theatre at Palmyra, Temple of Bel and Tetrapylon.*

The mean classification matrix for SVM is illustrated in Figure 4. It is apparent from Figure 4 that lowest classification accuracy is achieved within the Temple of Bel image class. Classification success for Temple of

Bell is lower than other classes across all experiments, with mean accuracy of 0.4% (SVM), 0.6% (Transfer CNN), 0.26% (full CNN) and 0.4% (Cloud Vision API), respectively.

***Figure 3**: The relationship between number of epochs, precision, test accuracy and f1 score.*

Baalshamin	**0.8**	0.2	0	0
Palmyra Theater	0	**1**	0	0
Temple of Bel	0	0.2	**0.4**	0.4
Tetrapylon	0.2	0	0	**0.8**

***Figure 4**: Classification Matrix, SVM.*

In cases where classification likelihoods are low, the opportunity exists to semantically annotate images with lower-level labels generated from existing REST APIs. The Cloud Vision API can detect broad sets of categories within images, ranging from landscape scenes and buildings, to animals. Each returned label is assigned a confidence level. Examples of labels generated using test images are provided in Figure 5. It can be seen in Figure 5 that the API can correctly identify features such as 'Column', 'Building', 'Temple' and 'Archaeological site.' Combining high level classifier outputs with lower level semantic labels and geolocation data may enhance ICARE archive, search and retrieval capabilities. Furthermore, the exclusion of images containing multiple non-relevant labels may assist with identifying and eliminating false-positive samples during data curation.

Label	Likelihood
Structure	0.91
Building	0.82
Archaelogical Site	0.82
Column	0.72
Temple	0.69

Label	Likelihood
Ampitheatre	0.88
Archaelogical Site	0.88
Building	0.81
Column	0.8
Egyptial Temple	0.53

***Figure 5**: Example labels, Google Vision API. Left: Temple of Bel. Right: Palmyra Theatre.*

4. CONCLUSIONS

We are at the early stage of development of the digital heritage platform (ICARE) to protect and preserve cultural heritage sites digitally. We seek to use innovative techniques from computing, computer vision, image and natural language processing to analyse images and enable semantic labelling and retrieval for varied user groups. Preliminary work indicates Transfer CNNs perform well though further experiments need to improve training and optimise the network [4] and compare other algorithms and ensembles.

Table 2. Full CNN Classification (epochs = 200)

Category	precision	accuracy	f1-score	support
Baalshamin	1.00	0.23	0.38	13
Palmyra Theatre	0.71	0.91	0.80	35
Temple of Bel	0.43	0.26	0.32	23
Tetrapylon	0.44	0.69	0.54	16
avg / total	0.63	0.60	0.56	87

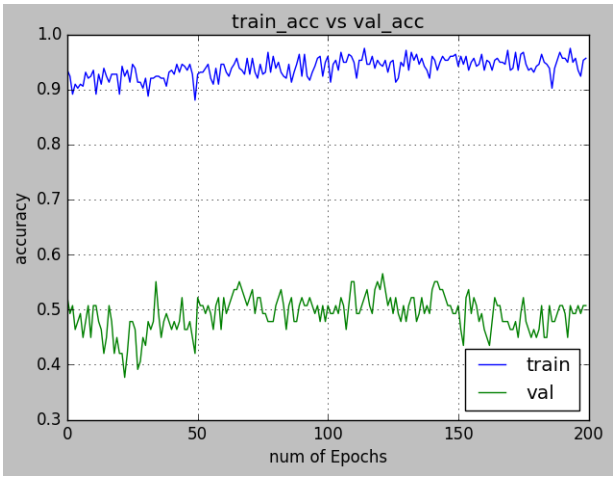

Figure 6: *The relation between number of epochs and training and validation data sets (full CNN).*

6. REFERENCES

[1] Vecco, M. (2010). A definition of cultural heritage: From the tangible to the intangible. Journal of Cultural Heritage, 11(3), 321-324.

[2] Silverman, H., & Ruggles, D. F. (2007). Cultural heritage and human rights. In Cultural heritage and human rights (pp. 3-29). Springer New York.

[3] Barthel-Bouchier, D. (2016). Cultural heritage and the challenge of sustainability. Routledge.

[4] Llamas J., Lerones P.M., Zalama E., Gómez-García-Bermejo J. (2016). Applying Deep Learning Techniques to Cultural Heritage Images Within the INCEPTION Project. In: Ioannides M. et al. (eds) Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. EuroMed 2016. Lecture Notes in Computer Science, vol 10059. Springer, Cham.

[5] Di Giulio, R., Maietti, F., Piaia, E., Medici, M., Ferrari, F., and Turillazzi, B. (2017). Integrated Data Capturing Requirements For 3d Semantic Modelling of Cultural Heritage: The Inception Protocol, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W3, 251-257, doi:10.5194/isprs-archives-XLII-2-W3-251-2017.

[6] Shankar S, Robertson D, Ioannou Y, A Criminisi, Cipolla R, (2016). Refining Architectures of Deep Convolutional Neural Networks. Conference on Computer Vision and Pattern Recognition.

[7] SCA http://www.sca-egypt.org/eng/main.htm , accessed on 12/5/2017

[8] Csurka, G., C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual Categorization with Bags of Keypoints. Workshop on Statistical Learning in Computer Vision. ECCV 1 (1–22), 1–2.

[9] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[11] Cloud Vision API

https://cloud.google.com/vision/

[12] Chollet, F and others, (2015). Keras,

https://github.com/fchollet/keras, GitHub.

[13] Charles Blundell, Benigno Uria, Alexander Pritzel, Yazhe Li, Avraham Ruderman, Joel Z Leibo, Jack Rae, Daan Wierstra, Demis Hassabis (2016) Model-Free Episodic Control]

[14] Blundell, Uria, Pritzel, Li, Ruderman, Leibo, Rae, Wiestra, Hassabid (2017) Mode Free Episodic Control