Understanding the cultural concerns of libraries based on automatic image analysis

Heng Ding
Central China Normal University, Wuhan, China, and
Wei Lu and Tingting Jiang
School of Information Management, Wuhan University, Wuhan, China

Received 28 November 2018
Revised 10 April 2019
Accepted 30 May 2019

Abstract

Purpose – Photographs are a kind of cultural heritage and very useful for cultural and historical studies. However, traditional or manual research methods are costly and cannot be applied on a large scale. This paper aims to present an exploratory study for understanding the cultural concerns of libraries based on the automatic analysis of large-scale image collections.

Design/methodology/approach – In this work, an image dataset including 85,023 images preserved and shared by 28 libraries is collected from the Flickr Commons project. Then, a method is proposed for representing the culture with a distribution of visual semantic concepts using a state-of-the-art deep learning technique and measuring the cultural concerns of image collections using two metrics. Case studies on this dataset demonstrated the great potential and promise of the method for understanding large-scale image collections from the perspective of cultural concerns.

Findings – The proposed method has the ability to discover important cultural units from large-scale image collections. The proposed two metrics are able to quantify the cultural concerns of libraries from different perspectives.

Originality/value – To the best of the authors' knowledge, this is the first automatic analysis of images for the purpose of understanding cultural concerns of libraries. The significance of this study mainly consists in the proposed method of understanding the cultural concerns of libraries based on the automatic analysis of the visual semantic concepts in image collections. Moreover, this paper has examined the cultural concerns (e.g. important cultural units, cultural focus, trends and volatility of cultural concerns) of 28 libraries.

Keywords Digital humanities, Cultural concerns, Image mining,
Photograph-based culture understanding, Visual semantic concepts

Paper type Research paper

1. Introduction

Cultural heritage witnesses where we come from and where we are (Serageldin *et al.*, 2001). In the digital age, libraries are playing an important role in the long-term preservation, usage and inheritance of cultural heritage in digital forms. Existing studies have discussed the relationships between culture and libraries. Wang and Frank (2002) highlighted the important role of academic libraries in cross-cultural communication. Loach *et al.* (2017) provided a critique of the sustainability policies and research of museums and libraries. They believed that the devotion to cultural sustainability is to the future survival of museums and libraries.

In addition, researchers have been interested in how to use information technologies for preservation, management, and organisation of cultural heritage (Kalfatovic *et al.*, 2008; Rimmer *et al.*, 2008). For example, Russel (1967) proposed the standardisation of mark-ups

0264-0473
DOI 10.1108/EL-11-2018-0229

for the encoding of literary texts in the COCOA programme. In 1998, the famous "Poughkeepsie Principles" was put forward as the text encoding guideline for literary, linguistic and historical research. Then, the extensible mark-up language (XML) became the guideline of text encoding for literary and linguistic computing (Text Encoding Initiative [TEI] Consortium, 2009). To support the preservation of digital objects and ensure long-term usability, the Library of Congress released the PREMIS Data Dictionary and Schema as an international metadata standard (Caplan and Guenther, 2005). Rimmer *et al.* (2008) established the Arts and Humanities Data Service to help researchers with the discovery and preservation of digital resources.

More recently, the Flickr Commons project collected thousands of image collections from many libraries, which might be used to gain a better understanding of the culture and history of libraries (Springer *et al.*, 2008). However, traditional research methods (i.e. image collections analysed by humans) are costly and cannot be applied on a large scale. In this paper, the researchers report an exploratory study to understand the cultural concerns of libraries based on the automatic analysis of large-scale image collections. The value of this study is twofold. On the one hand, the method proposed has the ability to analyse large-scale image collections quantitatively. On the other hand, the study included an examination of the cultural units that libraries were concerned with as well as the cultural concerns of different libraries.

The rest of the paper is organised as follows. In Section 2, the background of this study is introduced, including the quantification of culture and the automatic understanding of images, both of which are related to image mining in the digital humanities. In Section 3, a method is proposed of understanding large-scale image collections from the perspective of cultural concerns. The method first represents culture with a series of visual semantic concepts in the image collections. Second, two metrics are used to quantify the cultural concerns of the image collections. In Section 4, an image dataset collected from the Flickr Commons project, including 85,023 images preserved and shared by 28 libraries, is described to test the method. In Section 5, the effectiveness of the method is demonstrated based on case studies using the collected dataset and the findings about the cultural concerns of libraries.

2. Background

The main idea of this study was to quantify the cultural concerns based on the automatic analysis of visual semantic concepts in the image collections. The related studies are thus reviewed in terms of two topics; the quantification of culture and the automatic understanding of images.

2.1 Quantification of culture

It has been a long time since people started to use quantitative methods to study culture from various aspects. To our best knowledge, Hofstede (1984) was the first who used a quantitative method to study culture from multiple cultural dimensions, such as power distance, individualism versus collectivism, masculinity versus femininity, uncertainty versus avoidance, long-term versus short-term orientation and indulgence versus restraint. Following Hofstede's cultural dimension theory and quantitative method, Stuart-Fox (1986), Ger and Belk (1996) and Schwartz (2006) quantified and compared culture with manual methods (e.g. surveys).

During the past few decades, it was popular to use computational methods to study culture based on special cultural carriers (e.g. books and social media). Stubbs (1996) conducted an analysis of advertisements, newspapers, and scientific research articles in terms of text type and genre with text mining. Michel *et al.* (2011) used a quantitative method and reported a survey of the vast terrain of "culturomics". More specifically, they showed the trends of linguistic and cultural phenomena reflected in the English language between 1,800 and 2,000. Kincl *et al.* (2013) used a keyword analysis method in the examination of the differences in the communication characteristics of international programmes provided by universities from different cultures.

Moreover, researchers have been trying to interpret culture using the symbols of art images since a long time ago, which can be evidenced by a number of studies. Langer (1953) stated that visual symbols are the basis of all human understanding and serve as vehicles of conception for all human knowledge. Images were used to show the culture that people were proud of or that they thought interesting (Jacobs, 1981). Camillo *et al.* (2005) presented a method of analysing "the material culture of happiness" based on photo diaries provided by people from eight countries: Spain, France, England, Germany, Italy, The Netherlands, Finland and Russia.

In spite of the above efforts devoted to quantifying culture, it is worthwhile to further explore the cultural concerns of libraries based on their image collections using quantitative methods.

2.2 Automatic understanding of images

According to Eakins (2002), people comprehend images based on four types of abstract visual semantic concepts, including object, scene, behaviour and affective. Since then, more and more researchers have tried to develop semantic annotation techniques for automatic image understanding. Early studies mainly focussed on creating manual feature descriptors, such as HOG (Dalal and Triggs, 2005), SIFT (Lowe, 1999), SURF (Bay *et al.*, 2006) and GLOH (Mikolajczyk and Schmid, 2005), for specific image understanding tasks (e.g. face detection and recognition and object detection).

Recently, the deep learning technique has been applied in the automatic understanding of images, which was a major breakthrough. Krizhevsky *et al.* (2012) won the *Large Scale Visual Recognition Challenge 2012* for proposing the first deep learning model AlexNet. AlexNet has a simple architecture with five consecutive convolutional filters, a max-pool layer and three fully connected layers, but received a top-five error rate of 15.3 per cent outperforming the previous best one with an accuracy of 26.2 per cent. Later, a series of studies dived deeper into the sequence of convolutional layers. Simonyan and Zisserman (2014) established the "VGG16" model comprising 16 convolutional layers and introduced ReLU activation functions for nonlinear transformations. In contrast to 11 × 11 filters in AlexNet, VGG16 implements 3 × 3 filters for each convolution layer for the purpose of decreasing the number of parameters in training. He *et al.* (2016) noticed that training and optimising became more and more difficult with the increase of layers in deep models. Therefore, they proposed the ResNet model which tries to learn a residual function for keeping information among convolution layers. ResNet is composed of 152 convolutional layers with 3 × 3 filters using residual learning by block of two layers and obtained a top-five error rate of 4.49 per cent in the *Large Scale Visual Recognition Challenge 2012*. These efforts enable computers to understand the object semantic concepts in images more easily.

Researchers were also interested in automatic understanding of the scene and behaviour semantic concepts in images. Xiao *et al.* (2010) released the SUN dataset and reported some experimental results about scene classification using manual feature descriptors. Their study was able to classify images into 397 scene categories. Zhou *et al.* (2018) released the Place365 dataset for promoting research on scene semantic annotation of images. This dataset was composed of 10 million scene photographs, labelled with scene semantic categories and attributes, including a quasi-exhaustive list of the types of environments encountered in the world. They also created some Place365-CNNs models that have the ability to annotate the scene category of images automatically. Soomro *et al.* (2012) collected a large-scale dataset consisting of 101 action classes and provided the baseline action recognition results using the bag-of-words method. Karpathy *et al.* (2014) implemented convolutional neural networks (CNNs) in recognising 487 sports classes from large-scale videos. Their networks took advantage of local spatio-temporal information and performed significantly better than strong feature-based baselines. Tran *et al.* (2018) discussed several forms of spatio-temporal convolutions for action recognition. They showed that factorising the 3D convolutional filters into separate spatial and temporal components yielded significant improvement in accuracy. Inspired by these studies, the researchers attempted to use deep learning techniques for developing a method of understanding cultural concerns of libraries based on the semantic concepts in their image collections.

3. Methodology

The method proposed is for identifying the characteristics of cultural concerns of libraries from large-scale image collections. Specifically:

- A relational model for mapping visual semantic concepts to cultural units is created (Section 3.1).
- A deep learning technique for detecting visual semantic concepts from images and represented culture with a distribution of these concepts is introduced (Section 3.2).
- Two metrics to measure the cultural concerns from different perspectives are proposed (Section 3.3).

3.1 Mapping visual semantic concepts to cultural units

A study on semantic understanding of images (Eakins, 2002) indicated that people comprehend images based on four types of abstract visual semantic concepts, including object, scene, behaviour and affective. Another study on culture analysis (Stuart-Fox, 1986) suggested that there were three types of cultural units: material, behavioural and mental. The two studies provide the basis for creating the mapping from visual semantic concepts to cultural units. Figure 1 shows the relationships between visual semantic concepts and cultural units. As mental units and affective semantic concepts are difficult to measure and may be easily affected by subjective factors, only material and behavioural units are considered in this study.

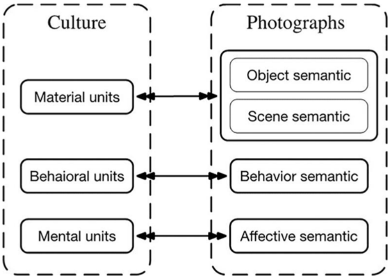


Figure 1.
A relational model for mapping visual semantic concepts to cultural units

3.2 Concept detection and culture representation

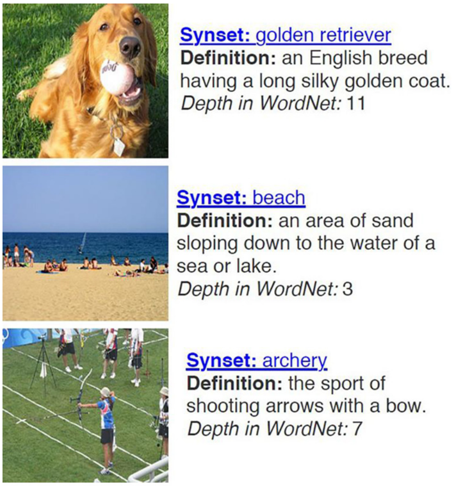
Traditional (manual) methods of analysing the visual semantic concepts in image collections are costly and cannot be applied on a large scale. In this study, a state-of-the-art deep learning technique is adopted to mine visual semantic concepts from image collections automatically. In practice, a CNN is trained on the whole ImageNet collection for detecting semantic concepts in the image collection.

Convolutional neural network. The researchers implemented the Inception V3 CNN for detecting the semantic concepts in each image. The Inception V3 network consists of six convolution layers, two pooling layers, and three inception layers for extracting the visual patterns from the input image. Then a linear layer and a softmax layer were used to output the semantic concepts in the input image. More details of the Inception V3 network can be found in Szegedy *et al.* (2016). Given the Inception V3 network M and a labelled image dataset I[1], the researchers were able to train a neural network model *f*:

$$M(I) \rightarrow f \tag{1}$$

ImageNet Collection. The entire ImageNet collection includes 14,197,122 images with 21,841 general visual semantic concepts. Each image contains object semantic concepts (e.g. tree and golden retriever), scene semantic concepts (e.g. canyon and beach) or behaviour semantic concepts (e.g. riding house and archery). Each concept is represented as a synset in WordNet. Figure 2 shows some examples of the visual semantic concepts in ImageNet. The images on the left are examples of the visual semantic concepts. The texts in blue are WordNet synsets, and the definitions below them are the descriptions of the synsets in WordNet. “Depth in WordNet” refers to the number of nodes between the current synset and the root of WordNet (Dalbello, 2011; Miller, 1995).

Because ImageNet is the largest image collection for image recognition (i.e. detecting visual semantic concepts), the researchers trained the CNN *f* on this collection. That is, in



Synset: golden retriever
Definition: an English breed having a long silky golden coat.
Depth in WordNet: 11

Synset: beach
Definition: an area of sand sloping down to the water of a sea or lake.
Depth in WordNet: 3

Synset: archery
Definition: the sport of shooting arrows with a bow.
Depth in WordNet: 7

Note: Each concept is represented with a WordNet synset with an example image

Figure 2. Visual semantic concepts in the ImageNet collection

equation (1), M denotes the Inception V3 CNN, I denotes the ImageNet collection. In this way, the trained neural model f is able to recognise a variety of visual semantic concepts, and it covers most cultural units.

Culture representation. Let p denote an image and $C = \langle c_1, \ldots, c_n \rangle$ denote the entire set of visual semantic concepts in ImageNet. Given the trained neural model f, we have equation (2):

$$s_i = \mathrm{f}\,(\mathrm{c_i}\ |\mathrm{p})\quad ,\quad \mathrm{i}\ \in\ (1,\mathrm{n}) \tag{2}$$

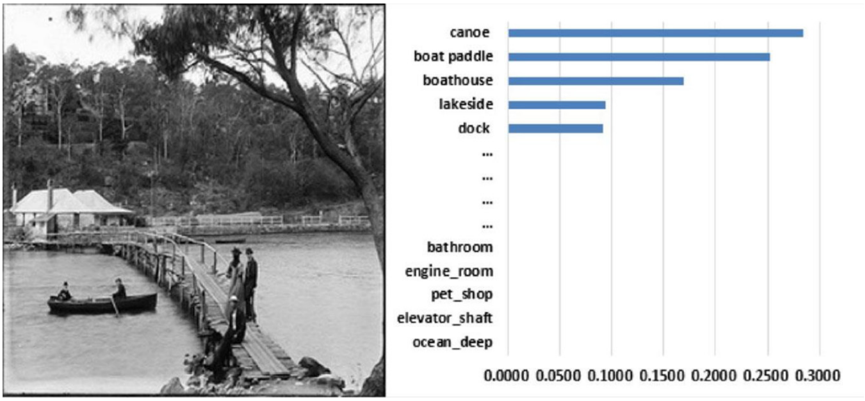
where s_i is the score of the concept c_i in the image p and n denotes the number of visual semantic concepts in ImageNet. Figure 3 shows an example of concept detection. The trained neural model has the ability to capture semantic concepts (i.e. cultural units) in the image, such as "canoe", "boat paddle", "'boathouse', "lakeside", and so on. According to the classic idea that "culture can be understood as a distribution of referenced concepts" (Carley, 1991), the culture implied in each image is represented with a culture vector $\sigma = \langle s_1, \ldots, s_n \rangle$. And the culture of a set of images $P = \langle p_1, \ldots, p_k \rangle$ can be represented with the mean vector $\overline{\sigma}$ as in equation (3):

$$\overline{\sigma} = \frac{\Sigma_{i=1}^{k} \sigma_i}{k} \tag{3}$$

where k is the number of images in set P and σ_i denotes the culture vector of the i-th image in set P.

3.3 Cultural measurement

Inspired by the research on socio-cultural computing (Lietz *et al.*, 2014), two metrics are proposed – that is, cultural focus and cultural difference – for measuring cultural concerns from different perspectives.



Notes: Input image p on the left and the culture vector of p on the right. "Canoe", "boat paddle" and others are visual semantic concepts (i.e. $C = \langle c_1, \ldots, c_n \rangle$) in ImageNet. Blue bars indicate the scores of the concepts (i.e.$\sigma = \langle s_1, \ldots, s_n \rangle$)

Figure 3.
An example of concept detection

Cultural focus is a metric used to quantify the extent to which a group demonstrates a focus F on a few important cultural units. Based on Lietz *et al.* (2014), the researchers used normalised Shannon entropy to measure the cultural focus for each set P [equations (4) and (5)]:

$$F_p = 1 - \frac{-\Sigma_{i=1}^{n}\bar{s}_l \cdot \log_2 \bar{s}_l}{\log_2 n} \tag{4}$$

$$\overline{S}_l = \frac{\Sigma_{i=1}^{k} s_i}{k} \tag{5}$$

where n denotes the number of visual semantic concepts in ImageNet and $\overline{S}_l$ is the score of the i-th semantic concept in set P. F_P falls in the range (0, 1) where 1 means that the content of image set P focuses on a few cultural units and suggests a clear cultural preference. In contrast, 0 means that the content of image set P is distributed uniformly over a lot of cultural units.

Cultural difference is a metric used to quantify the cultural homophily of two sets of images. In this study, the researchers used the cosine distance formula to measure the cultural difference D [equation (6)]:

$$\mathrm{D}(P_x, P_y) = 1 - \frac{\sigma_{P_x} \cdot \sigma_{P_y}}{\|\sigma_{P_x}\| \cdot \|\sigma_{P_y}\|} \tag{6}$$

where σ_{P_x} and σ_{P_y} are the culture vectors of image sets P_x and P_y, respectively. The cultural difference falls in the range (0, 1) where 0 means that the cultural units of P_x are the same as those of P_y, whereas 1 means that the cultural units of P_x is totally different from those of P_y.

4. Data

A large-scale image collection with at least 10,000 images is needed to test the method. Given the important roles of libraries in the preservation of cultural materials, we collected an image dataset shared by libraries from the Flickr Commons[2] project which had accumulated 6,800,000 images shared by 109 institutions, including national libraries, public libraries, university libraries, museums, and others. The researchers manually examined the type of each institution and identified all the libraries. Then the images uploaded and shared by these libraries and their metadata were collected using the Flickr API (www.flickr.com/services/api/). The metadata of each image includes five fields, that is, "owner id", "image id", "title of image", "description" and "taken time" (Figure 4). In total, we collected 85,023 images shared by 28 libraries from Flickr Commons. Table I shows the metadata of an example image. All the images were taken from 1000 to May 2016, with only seven taken between 1000 and 1799. We eliminated the seven images, and the rest of the images can be traced back to as early as 1800 (Figure 5). It can be easily detected that there are two peak periods of photo taking; that is, 1900-1950 and 2000-2016. That is, this collection mainly reflects the cultural concerns of the libraries during these two periods.

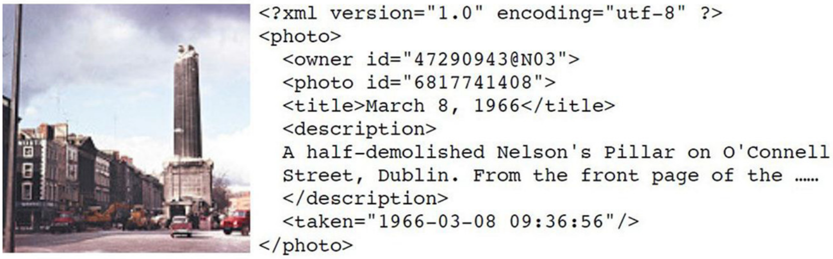
5. Analysis

The effectiveness of the method was tested based on the case studies on the above image collection. Specifically speaking, the intent is to answer the following research questions:

RQ1. Is the method able to discover effectively the cultural units that the libraries are concerned with?

RQ2. Which libraries have clear cultural preferences? What are these preferences?

RQ3. What are the patterns of the cultural concerns of the libraries?



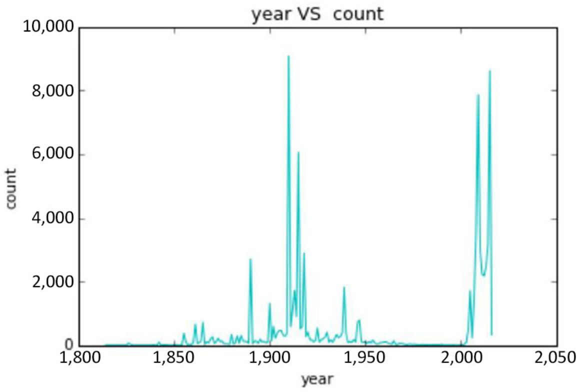
```
<?xml version="1.0" encoding="utf-8" ?>
<photo>
  <owner id="47290943@N03">
  <photo id="6817741408">
  <title>March 8, 1966</title>
  <description>
  A half-demolished Nelson's Pillar on O'Connell
  Street, Dublin. From the front page of the ……
  </description>
  <taken="1966-03-08 09:36:56"/>
</photo>
```

Note: The image is on the left and its metadata on the right

Figure 4.
The metadata of an image collected from Flickr Commons

ID	Library name	Country	Count of images	Library type
L1	National Library of Medicine	The USA	584	National
L2	National Library of Norway	Norway	3,316	National
L3	National Library of Sweden	Sweden	165	National
L4	National Library of Australia	Australia	968	National
L5	The Royal Library, Denmark	Denmark	188	National
L6	National Library of Ireland	Ireland	1,752	National
L7	National Library of Scotland	Scotland	2,313	National
L8	National Library of Wales	Wales	2,117	National
L9	National Library of New Zealand	New Zealand	4,304	National
L10	Library of Congress	The USA	26,140	National
L11	Camden Public Library (Maine)	The USA	283	Public
L12	Tyrrell Historical Library	The USA	156	Public
L13	Hamilton Public Library	The USA	403	Public
L14	Library Company of Philadelphia	The USA	1,021	Public
L15	Vancouver Public Library	Canada	537	Public
L16	Bergen Public Library	Norway	982	Public
L17	District of Columbia Public Library	The USA	230	Public
L18	State Library of Queensland	Australia	3,452	Public
L19	New York Public Library	The USA	2,525	Public
L20	Library of Virginia	The USA	986	Public
L21	State Library of New South Wales	Australia	2,911	Public
L22	Library of Texas State University	The USA	5,254	University
L23	Miami University Libraries	The USA	9,440	University
L24	Glucksman Library, University of Limerick	Ireland	610	University
L25	Cornell University Library	The USA	3,971	University
L26	Central University Libraries	Mexico	7,441	University
L27	Library of University of Washington	The USA	802	University
L28	London School of Economics Library	The UK	2,172	University

Table I.
Details of the image collection



Note: The x-axis indicates the year, and the y-axis is the number of images taken in each year

Figure 5. Temporal distribution of the images

To answer these research questions, several case studies were conducted on the collected dataset.

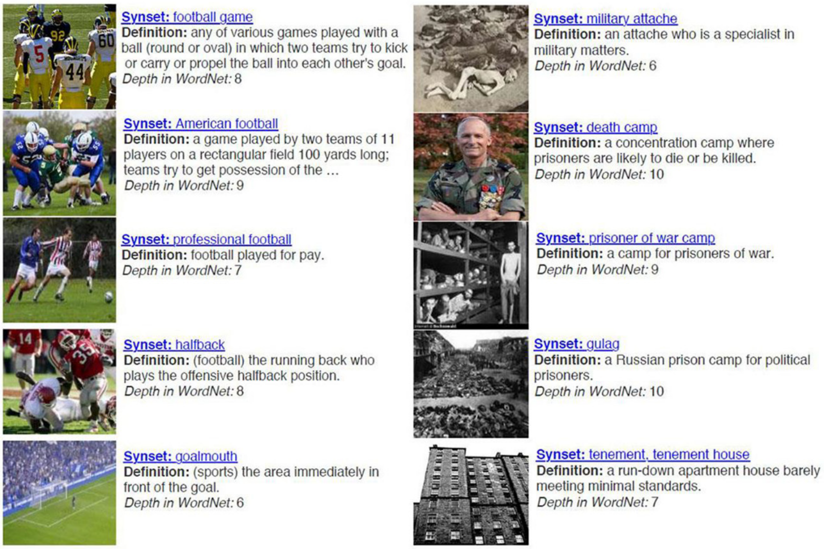
5.1 Answer to RQ1

In the method, the culture vector $\overline{\varpi}$ [equation (3)] represents the cultural units that libraries were proud of, thought interesting, and wanted to show to others. To verify this idea, two of the libraries, TX State University Library (*L22*) and National Library of Scotland (*L7*), were chosen for further study because the sizes of their image sets are moderate. *L22* shared 5,254 images and *L7* 2,313 images. Both image sets are sufficiently large for reflecting the advantage of the automatic method but not too large to impede subsequent manual examination.

The pre-trained CNN was used to detect visual semantic concepts in both image sets and the culture vector $\overline{\varpi}$ for each library was calculated using equation (3). Also examined were the cultural units of both libraries by visiting their home pages, Wikipedia pages and other related online resources.

It is very interesting to see that the method has the ability to discover important cultural units from large-scale image collections. For example, the top ten visual semantic concepts with the highest scores in the culture vector $\overline{\varpi}$ of *L22* are “football game”, “American football”, “professional football”, “halfback”, “goalmouth”, “drum majorette”, “basketball player”, “tennis”, “pushball”, and “pitching coach” (Figure 6), which suggests a strong preference for sports and athletics. On the Wikipedia page of Texas State University (https://en.wikipedia.org/wiki/Texas_State_University), one can find that “athletics”, especially “football bowl subdivision”, is deemed a core cultural unit in the history of Texas State University (Figure 7).

The top ten visual semantic concepts of *L7* are “death camp”, “military attaché”, “prisoner of war camp”, “Gulag”, “tenement house”, “orthochromatic film”, “passe-partout”, “colliery”, “slit trench” and “foxhole” (Figure 6), which indicates that “the war” is an important cultural unit of *L7*. After viewing the library’s web site (http://



Synset: football game
Definition: any of various games played with a ball (round or oval) in which two teams try to kick or carry or propel the ball into each other's goal.
Depth in WordNet: 8

Synset: American football
Definition: a game played by two teams of 11 players on a rectangular field 100 yards long; teams try to get possession of the ...
Depth in WordNet: 9

Synset: professional football
Definition: football played for pay.
Depth in WordNet: 7

Synset: halfback
Definition: (football) the running back who plays the offensive halfback position.
Depth in WordNet: 8

Synset: goalmouth
Definition: (sports) the area immediately in front of the goal.
Depth in WordNet: 6

Synset: military attache
Definition: an attache who is a specialist in military matters.
Depth in WordNet: 6

Synset: death camp
Definition: a concentration camp where prisoners are likely to die or be killed.
Depth in WordNet: 10

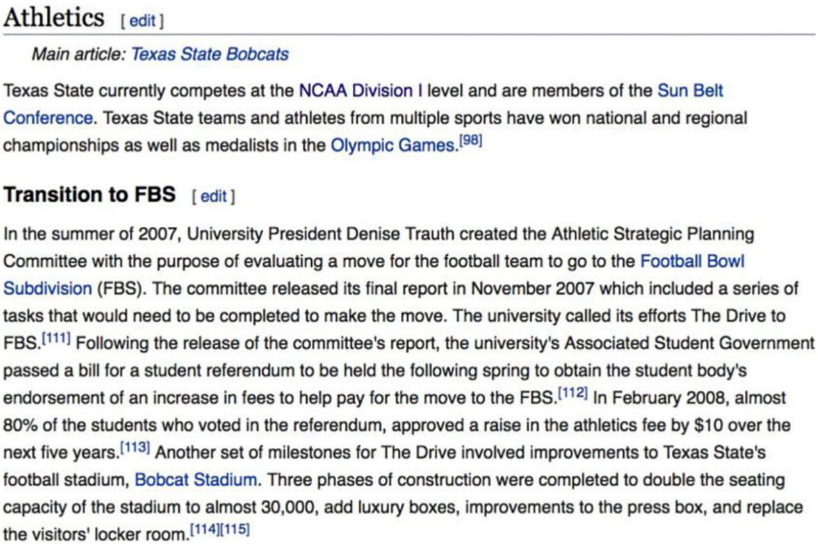
Synset: prisoner of war camp
Definition: a camp for prisoners of war.
Depth in WordNet: 9

Synset: gulag
Definition: a Russian prison camp for political prisoners.
Depth in WordNet: 10

Synset: tenement, tenement house
Definition: a run-down apartment house barely meeting minimal standards.
Depth in WordNet: 7

Note: Only the top five concepts are shown for each library for lack of space

Figure 6. The significant visual semantic concepts detected from the image sets of *L22* (left) and *L7* (right)



Athletics [edit]

Main article: Texas State Bobcats

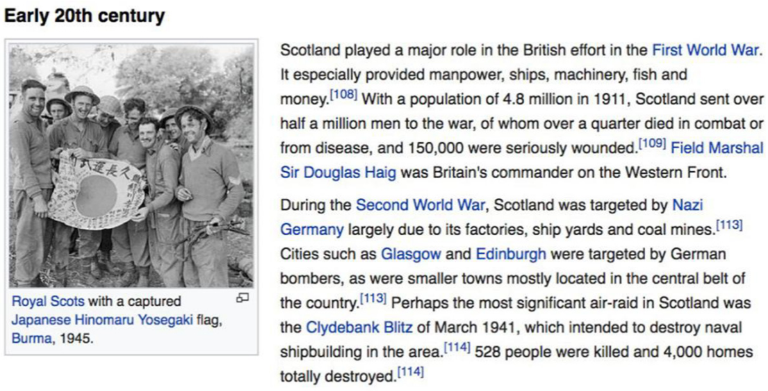
Texas State currently competes at the NCAA Division I level and are members of the Sun Belt Conference. Texas State teams and athletes from multiple sports have won national and regional championships as well as medalists in the Olympic Games.[98]

Transition to FBS [edit]

In the summer of 2007, University President Denise Trauth created the Athletic Strategic Planning Committee with the purpose of evaluating a move for the football team to go to the Football Bowl Subdivision (FBS). The committee released its final report in November 2007 which included a series of tasks that would need to be completed to make the move. The university called its efforts The Drive to FBS.[111] Following the release of the committee's report, the university's Associated Student Government passed a bill for a student referendum to be held the following spring to obtain the student body's endorsement of an increase in fees to help pay for the move to the FBS.[112] In February 2008, almost 80% of the students who voted in the referendum, approved a raise in the athletics fee by $10 over the next five years.[113] Another set of milestones for The Drive involved improvements to Texas State's football stadium, Bobcat Stadium. Three phases of construction were completed to double the seating capacity of the stadium to almost 30,000, add luxury boxes, improvements to the press box, and replace the visitors' locker room.[114][115]

Figure 7. The Wikipedia page of Texas State University, screen captured at 1:42 PM, 29 May 2017

digital.nls.uk/gallery/category/war) and the Wikipedia of Scotland (Figure 8), we found that Scotland played an important role in the two World Wars and a lot of people died during the wars. This in a degree explains why the wars are an unforgettable unit in their culture and history.



Early 20th century

Royal Scots with a captured Japanese Hinomaru Yosegaki flag, Burma, 1945.

Scotland played a major role in the British effort in the First World War. It especially provided manpower, ships, machinery, fish and money.[108] With a population of 4.8 million in 1911, Scotland sent over half a million men to the war, of whom over a quarter died in combat or from disease, and 150,000 were seriously wounded.[109] Field Marshal Sir Douglas Haig was Britain's commander on the Western Front.

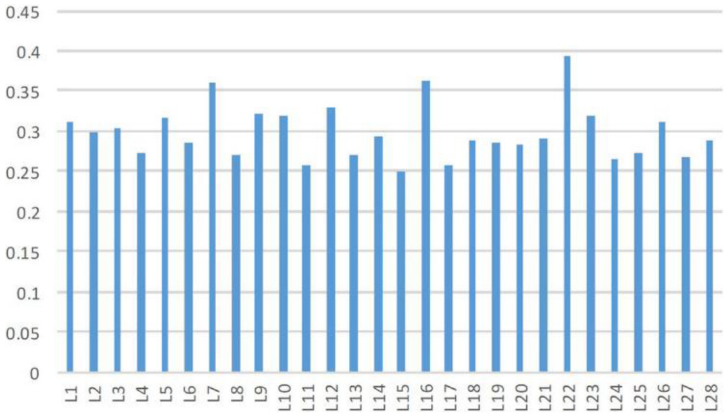
During the Second World War, Scotland was targeted by Nazi Germany largely due to its factories, ship yards and coal mines.[113] Cities such as Glasgow and Edinburgh were targeted by German bombers, as were smaller towns mostly located in the central belt of the country.[113] Perhaps the most significant air-raid in Scotland was the Clydebank Blitz of March 1941, which intended to destroy naval shipbuilding in the area.[114] 528 people were killed and 4,000 homes totally destroyed.[114]

Figure 8. The Wikipedia page of Scotland, screen captured at 1:51 PM, 29 May 2018

5.2 Answer to RQ2

The above case studies show that the culture vector is able to capture the cultural concerns (important cultural units) of the libraries. Then the libraries which have clear culture preferences were identified using the proposed metric Cultural Focus (Section 2.4). This metric is a quantitative indicator that describes cultural concerns. Generally speaking, if the cultural focus of a library is close to 1, then the library is concerned with very few cultural units. When the cultural focus is close to 0, the library has a variety of cultural concerns.

The cultural focus for the 28 libraries was calculated based on their image sets using equations (4) and (5). Figure 9 shows the results. It can be seen that *L22* (Library of Texas State University) has the highest cultural focus score ($F = 0.39272$). Most images shared by *L22* are about football series, tennis series, basketball series, track meets, and famous athletes in the history of Texas State University. One can infer that the library was very proud of the university's sports culture, and sports and athletics have played an important role in the life of the people at Texas State University. Such finding is consistent with that of



Note: The x-axis indicates the library ID and the y-axis is the score of cultural focus

Figure 9. The cultural focus of each library in the collection

the automatic analysis in Section 4.1. Additionally, *L16* (Bergen Public Library) and *L7* (National Library of Scotland) also have high cultural focuses ($F > 0.35$), indicating that the cultural units they want to share with others are very focussed. In all, 80 per cent (787/982) [3] of the images from *L16* are the portraits of famous historical figures or photos of people and 86 per cent (1991/2313) of the images from *L7* are related to World War I.

The cultural focus of *L15* (Vancouver Public Library) is the lowest ($F = 0.25034$). The researchers also manually examined the image set of *L15* and found that the library has shared a lot of historical images that cover a diversity of cultural units, including buildings, persons, wars, sports, leisure life and so forth. These images do not have a clear cultural concern, perhaps because *L15* just wanted to show everything they had.

5.3 Answer to RQ3

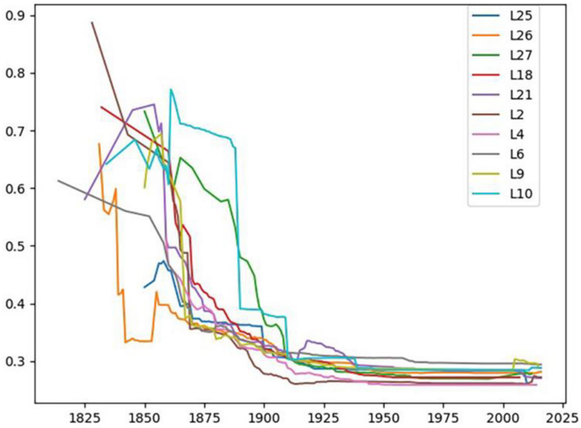
The researchers described the patterns of the cultural concerns of the libraries from two aspects:

(1) the trend of cultural concerns; and

(2) the volatility of cultural concerns.

To examine the trend of cultural concerns, the accumulative cultural focus for each library was calculated. Let $P_j = \langle p_i, \ldots, p_l \rangle$ denotes all the images from a library that were taken before the year $j \in (1800, 2016)$ (includes the year j), the accumulative cultural focus of the library in year j is calculated with equations (2) to (5). The accumulative cultural focus of the libraries year by year from 1800 to 2016 was plotted. To see the long-term trend of cultural focus, the researchers focussed on the libraries whose earliest images were taken more than 40 years ago. Figure 10 shows the plot. It is not surprising that the accumulative cultural focus shows a decreasing trend. However, a sharp decreasing mostly happened before 1950, followed by flat curves from 1950 to 2016.

To examine the volatility of cultural concerns, the adjacent cultural difference for each library was calculated. Let $P_j = \langle p_i, \ldots, p_m \rangle$ denote all the images from a library that were taken in the year $t \in (1800, 2016)$, the adjacent cultural difference of the year t is calculated as:



Note: The x-axis indicates the year and the y-axis is the score of cultural focus

Figure 10.
The long-term trend of the cultural focus of the libraries in the collection

$$D(P_t, P_{t-1}) = 1 - \frac{\sigma_{P_t} \cdot \sigma_{P_{t-1}}}{\|\sigma_{P_t}\| \cdot \|\sigma_{P_{t-1}}\|} \tag{7}$$

where σ_{P_t} and $\sigma_{P_{t-1}}$ are the culture vectors of image sets P_t and P_{t-1}, respectively, and σ_{P_t} and $\sigma_{P_{t-1}}$ are computed using equation (3). The long-term trend of the adjacent cultural difference can be found in Figure 11. Here, we also focussed on those libraries whose earliest images were taken more than 40 years ago. There are some void years during which no photo was taken. As a result, only considered were the years with non-zero numbers of images.

The peaks in the plots reflect that the cultural concerns of the library changed from some cultural units to others between two adjacent years. In other words, a peak represents a dramatic transition of cultural concerns. In contrast, the troughs in the plots reflect the cultural concerns of the library were stable between two adjacent years.

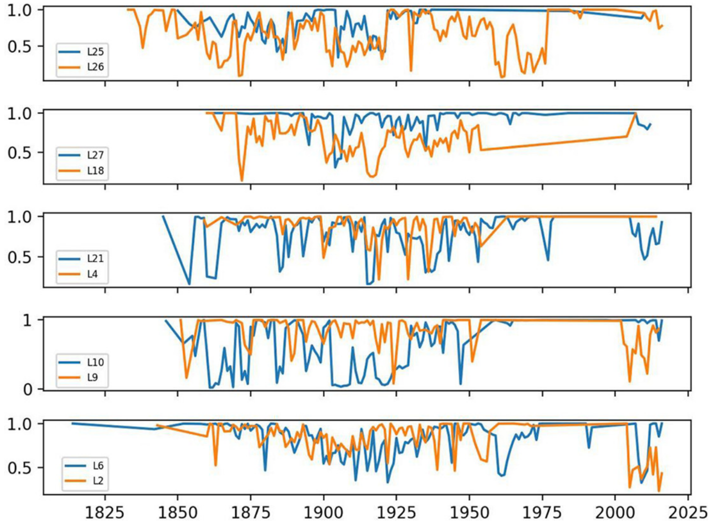
One may find that the adjacent cultural difference of *L27* (Library of University of Washington) and that of *L9* (National Library of New Zealand) is close to 1 in most years, which suggests that both libraries have a high volatility in their cultural concerns.

6. Conclusions

In this paper, an exploratory study of understanding the cultural concerns of libraries based on automatic analysis of image collections is presented.

To the best of our knowledge, this is the first time that automatic image mining was applied in cultural concerns discovery. Additionally, a method was proposed that quantifies the cultural concerns based on visual semantic concept detection using a deep learning approach and two proposed metrics.

The case studies demonstrated the effectiveness of the method from several aspects:



Note: The x-axis denotes the year and the y-axis is cultural difference between P_t and P_{t-1}, where P_t denotes all the images of a library that were taken in the year t

Figure 11. The long-term trend of adjacent cultural difference of the libraries in the collection

- The method has the ability to highlight the noticeable cultural units in image collections without human efforts.
- The proposed metric, cultural focus, is able to measure the diversity of cultural concerns (i.e. cultural preferences).

Unlike traditional methods that depend mainly on humans to analyse the content of images, the method uses the deep learning approach to automatically analyse the visual semantic concepts in images and could be used on large-scale image collections. Case studies were conducted to show the great potential and promise of the method for understanding cultural concerns. This idea may be applied to other objects than image collections of libraries in the future.

Notes

1. Each image in dataset I has been annotated with semantic concept labels (e.g. "beach", "golden retriever", and so on).
2. The goal of the Flickr Commons Project is to share hidden treasures from the world's public image archives.
3. Here the former is the number of images that are related to people, the latter is the number of images the library owns.

References

Bay, H., Tuytelaars, T. and Van Gool, L. (2006), "Surf: speeded up robust features", *2006 European Conference on Computer Vision, Graz, Austria*, *Springer*, pp. 404-417.

Camillo, F., Tosi, M. and Traldi, T. (2005), "Semiometric approach, qualitative research and text mining techniques for modelling the material culture of happiness", *Knowledge Mining*, Springer, Berlin, Heidelberg, pp. 79-92.

Caplan, P. and Guenther, R.S. (2005), "Practical preservation: the PREMIS experience", *Library Trends*, Vol. 54 No. 1, pp. 111-124.

Carley, K. (1991), "A theory of group stability", *American Sociological Review*, Vol. 56 No. 3, pp. 331-354.

Dalal, N. and Triggs, B. (2005), "Histograms of oriented gradients for human detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA*, *IEEE Computer Society*, pp. 886-893.

Dalbello, M. (2011), "A genealogy of digital humanities", *Journal of Documentation*, Vol. 67 No. 3, pp. 480-506.

Eakins, J.P. (2002), "Towards intelligent image retrieval", *Pattern Recognition*, Vol. 35 No. 1, pp. 3-14.

Ger, G. and Belk, R.W. (1996), "Cross-cultural differences in materialism", *Journal of Economic Psychology*, Vol. 17 No. 1, pp. 55-77.

He, K., Zhang, X., Ren, S. and Sun, J. (2016), ""Deep residual learning for image recognition", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *IEEE Computer Society*, *Las Vegas, NV*, pp. 770-778.

Hofstede, G. (1984), "Cultural dimensions in management and planning", *Asia Pacific Journal of Management*, Vol. 1 No. 2, pp. 81-99.

Jacobs, D.L. (1981), "Domestic snapshots: toward a grammar of motives", *The Journal of American Culture*, Vol. 4 No. 1, pp. 93-105.

Kalfatovic, M.R., Kapsalis, E., Spiess, K.P., Van Camp, A. and Edson, M. (2008), "Smithsonian team Flickr: a library, archives, and museums collaboration in web 2.0 space", *Archival Science*, Vol. 8 No. 4, p. 267.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L. (2014), "Large-scale video classification with convolutional neural networks", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH*, *IEEE Computer Society*, pp. 1725-1732.

Kincl, T., Novák, M. and Štrach, P. (2013), "5A cross-cultural study of online marketing in international higher education: a keyword analysis", *New Educational Review*, Vol. 32 No. 2, pp. 49-65.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012), "Imagenet classification with deep convolutional neural networks", in *Advances in Neural Information Processing Systems*, Neural Information Processing Systems Foundation, Lake Tahoe, NV, pp. 1097-1105.

Langer, S.K. (1953), *Feeling and Form: A Theory of Art Developed from Philosophy in a New Key*, Charles Scribner's Sons, New York, NY.

Lietz, H., Wagner, C., Bleier, A. and Strohmaier, M. (2014), "When politicians talk: assessing online conversational practices of political parties on Twitter", *Eighth International AAAI Conference on Weblogs and Social Media*, *Association for the Advancement of Artificial Intelligence*, *Ann Arbor, MI*, pp. 285-294.

Loach, K., Rowley, J. and Griffiths, J. (2017), "Cultural sustainability as a strategy for the survival of museums and libraries", *International Journal of Cultural Policy*, Vol. 23 No. 2, pp. 186-198.

Lowe, D.G. (1999), ""Object recognition from local scale-invariant features", *Proceedings of the International Conference on Computer Vision, Kerkyra, Corfu, Greece*, *IEEE Computer Society*, pp. 1150-1157.

Michel, J.B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J. and Pinker, S. (2011), "Quantitative analysis of culture using millions of digitized books", *Science*, Vol. 331 No. 6014, pp. 176-182.

Mikolajczyk, K. and Schmid, C. (2005), "A performance evaluation of local descriptors", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27 No. 10, pp. 1615-1630.

Miller, G.A. (1995), "WordNet: a lexical database for English", *Communications of the ACM*, Vol. 38 No. 11, pp. 39-41.

Rimmer, J., Warwick, C., Blandford, A., Gow, J. and Buchanan, G. (2008), "An examination of the physical and the digital qualities of humanities research", *Information Processing and Management*, Vol. 44 No. 3, pp. 1374-1392.

Russel, D.B. (1967), *Cocoa: A Word Count and Concordance Generator for Atlas*, Atlas Computer Laboratory, Chilton.

Schwartz, S. (2006), "A theory of cultural value orientations: explication an applications", *Comparative Sociology*, Vol. 5 Nos 2/3, pp. 137-182.

Serageldin, I., Shluger, E. and Martin-Brown, J. (Eds) (2001), *Historic Cities and Sacred Sites: Cultural Roots for Urban Futures*, World Bank Publications, Washington, DC.

Simonyan, K. and Zisserman, A. (2014), "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409.1556*.

Soomro, K. Zamir, A.R. and Shah, M. (2012), "UCF101: a dataset of 101 human actions classes from videos in the wild", *arXiv preprint arXiv:1212.0402*.

Springer, M., Dulabahn, B., Michel, P., Natanson, B., Reser, D., Woodward, D. and Zinkham, H. (2008), *For the Common Good: The Library of Congress Flickr Pilot Project*, Library of Congress, Washington, DC.

Stuart-Fox, M. (1986), "The unit of replication in socio-cultural evolution", *Journal of Social and Biological Structures*, Vol. 9 No. 1, pp. 67-89.

Stubbs, M. (1996), *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*, Blackwell, Oxford.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016), "Rethinking the inception architecture for computer vision", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV*, *IEEE Computer Society*, pp. 2818-2826.

Text Encoding Initiative (TEI) Consortium (2009), "TEI P5: guidelines for electronic text encoding and interchange", available at: www.tei-c.org/Vault/P5/1.4.0/doc/tei-p5-doc/en/html/ (accessed 12 November 2018).

Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y. and Paluri, M. (2018), "A closer look at spatiotemporal convolutions for action recognition", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT*, *IEEE Computer Society*, pp. 6450-6459.

Wang, J. and Frank, D.G. (2002), "Cross-cultural communication: implications for effective information services in academic libraries", *Portal: Libraries and the Academy*, Vol. 2 No. 2, pp. 207-216.

Xiao, J., Hays, J., Ehinger, K.A., Oliva, A. and Torralba, A. (2010), "Sun database: large-scale scene recognition from abbey to zoo", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA*, *IEEE Computer Society*, pp. 3485-3492.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A. (2018), "Places: a 10 million image database for scene recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40 No. 6, pp. 1452-1464.

Corresponding author
Wei Lu can be contacted at: weilu@whu.edu.cn