Vol.2019-CH-119 No.7
2019/2/16

Faces in an Archive of Buddhism Pictures

Benjamin Renoust[1,a)] Ayaka Uesaka[1] Yuta Nakashima[1] Hajime Nagahara[1] Yutaka Fujioka[2]

概要：We introduce a project aiming at capturing the evolution of Buddhism among time, schools and places, with the support of Computer Vision and Data Analysis. We introduce a first methodology to explore a large archive that includes 50,000 pictures collected by our expert in Art History, raising the different challenges in analyzing this archive.

1. Introduction

The spread and evolution of Buddhism across Asia is the topic of many books [20], [24]. Multiple theories are confronting on which path(s) this spread took across the Asian subcontinent, reaching the coasts of the Japanese archipelago along the Silk Road [20], [22], [23], [24]. Once settled, Buddhism flourished and spread in different schools, with scholars traveling to bring new thought (*e.g.*, Ganjin's travel). Buddhism brought many works of art and their rules (法量) so that local people would craft new artworks by themselves. They would embrace and adapt these rules with their own culture, giving the identity to the resulting style (*e.g.* Fig. 1). Nowadays, only a few experts can identify these works *subjective* to their own knowledge, sometimes disputing explanations [18]. Nonetheless, the evolution of style connects to the evolution of schools [21].

We propose to *objectively* reconstruct this evolution through traces captured in digitalized Buddhism artworks. We wish our approach to be data-oriented, *i.e.* to apply systematic analysis on a large archive of such elements. The topic is very large, but we wish to focus on the representation of Buddha, which is central to Buddhism art. Representations of Buddha, mostly statues can be of many types, and their *canons*[*1] have been normalized over the centuries. These are rather strict and well documented, especially concerning the dimension of faces. This *iconometry* was captured in rule books (法量) to be easily taught across the world. Despite the rules, time and travels allowed for quite an evolution among the style of statues (as represented in Fig. 1), and aligning many of these statues may allow us to capture the traces of this evolution [19].

Since, the Computer Vision community have proposed serious advances in style analysis [16] and face detection and recognition to reach operational maturity [15], we will focus on the faces of Buddha, especially on Buddha statues. Our goal is to capture how artists expressed their freedoms, moving the rules forming the canon, or in between the lines formed by the canons. We believe these are latent features that will allow us to identify many parameters of an artwork, hopefully its area and era of origin, the authors, the school of thoughts, and so on.

図 1 Evolution of Buddha faces: the middle face (Cambodia) is at the intersection of left (Central India) and right (Southern India) archetypes.

The remaining of this paper presents related works in Section 2 before introducing our archive and approach in Section 3. We identify different challenges raised by this archive analysis in Section 4 before concluding.

1 Osaka University
Institute for Datability Science
2 Osaka University
Graduate School of Letters
a) renoust@ids.osaka-u.ac.jp

*1 A *canon of art* refers to a universal set of rules and principles establishing the fundamentals and/or optimal.

2. Related works

Recently, great efforts have been achieved in the domain of style analysis, especially for paintings. The Computer Vision community has focused on this task for a while, proposing solutions using such a technique as wavelet transforms [5], and more recently deep learning-based techniques [12], [13], [14], [16]. By extension, the style analysis gives clues on the shape of style evolution [1], [2].

Analysis of Buddha statues has mainly been tackled from the perspective of 3D data. Scanning statues itself, comes with its load of challenges [4]. Style analysis has been established between Bayon faces [7], contributing in the site understanding. 3D features of Buddha statues from cross-sections have been proven important to identify eras of statues and features of famous artists [19].

3. A Large Archive of Buddhism Pictures

Our archive consists in collection of picture of Buddha statues, accumulating the knowledge of our domain expert over the course of several years.

3.1 Dataset

This hard drive contains about 50,000 pictures of all kinds (about 500GB in total). They have been captured under many conditions, mainly: museum collection pieces acquired with standard methods, on-site captures in museum or temple, carefully captured Buddhist art treasures, outside field trips, and scans of dedicated literature.

This is real data *"in-the-wild"*: the pictures were captured in multiple size and formats; there may be redundancy of a same picture; a same Buddha statue can be taken across many angles and lights; multiple statues in one picture; multiple sub-pictures in one picture; pictures can be a detail of larger artifact; this detail itself can be representation of Buddha; there could be no Buddha representation, or of an explanation text/label or even unrelated pictures can be found.

The pictures are not annotated, but they may be attached to indirect contextual information. They can contain meta-data; they may have a date of creation; they are stored in folders (over 1.7k different folders); they may have a specific filename. We extracted EXIF information but no location data was available.

3.2 Building the database

Our first task is to build a database from these pictures by annotating them and identifying the main artifacts. The textual information from folders and filenames comes actually in 4 different languages, Chinese, English, Japanese, and Korean. This is the richest information we have on the pictures. The filename may label the picture or give orientation/thematic information. Pictures are often grouped in a folder for a reason, and the path to the folder and name of the folder may capture this reason: it is a characteristic shared by all pictures of a folder. The date of creation or the order of pictures added in the folder may give an additional information: a sequence of pictures may be a sequence of a same picture. A folder usually contains a few dozens of pictures.

To identify unique artifacts, first we need to normalize our dataset. Using EXIF information we corrected orientation of pictures and finally manually corrected the remaining. We also normalized the sizes and formats (especially due to many raw pictures). Using image hashing (both differential hashing and perceptual hashing [17]) we mined near duplicates (and manually corrected false positive), reducing the dataset to a subset of about 30k unique instances.

We then use the fact that pictures are stored in a same folder to identify pictures of a unique artifact. We use the prior information that consecutive pictures can be taken of a same object with different angles, or mention the same artifact in a collection. To do so, we want to compare each picture in the order of the folder. We use the order formed by filenames, best suited for collections. We then use an image embedding, the 4096 dimensions of VGG16 trained on ImageNet[11], projected into 50 dimensions with a truncated SVD since the embedding space is very sparse in one folder. We finally compare pairs of consecutive pictures using cosine distance. This is captured in a graph visualization (Fig. 2), which can be annotated by simply cutting links when pictures represent different artifacts, and creating links in the opposite case.

図 2 Example of a similarity sequence visualization generated for the identification of a same Buddha statue.

We finally regrouped the previously matching pictures of different folders and the unique artifacts selection we made to identify unique Buddha statues. We only kept pictures of Buddha statues with faces and multiple pictures, while discarding less informative or unclear collections (about 20k pictures). In the 30k pictures remaining (including 10k duplicates), we collected 3685 unique Buddha statues, 26.2% of which are composed of over 5 different pictures (about 17k pictures in 804 statues). We consider from this arbitrary threshold we consider the most relevant statues, and manually checked them.

We extracted text content from 1.7k folder names and manually curated locations when available, thus enriching the data when possible of the 804 statues. We thus automatically recovered (in terms of annotated statues) 366 statue types, 672 countries with 461 regions and 460 cities in which the statues were taken, 113 countries with 104 regions and 102 cities of origin of the statues, 98 construction eras, and 89 temples.

4. Further analysis

Buddha statues are sculptures, which are by definition 3D objects. The archive is however composed of 2D pictures. Although the volumetric information may precisely capture differences in style, this cannot scale since 3D scans of statues is very expensive in price and time. On the other hand, taking pictures has become cheap and large scale image analysis is now readily available.

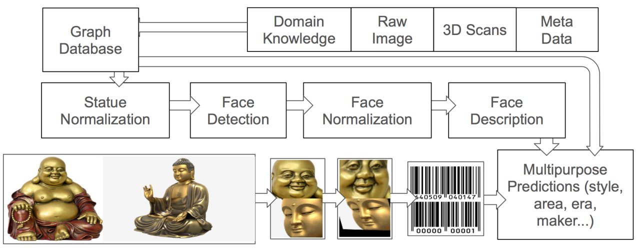

図 3　Pipeline of our proposed analysis.

4.1 Mining faces

We want to deploy face analysis as a support of style identification. The idea is to mine our archive for faces and explore similarity among those faces. This is done in a few steps (Fig. 3): detecting faces in pictures, normalizing them, extracting descriptors for each face, and match them in connection with knowledge data in order to form predictions.

To validate this idea, we demonstrate a proof of concept. We mined faces with the Faster R-CNN [6] in the archive, removing false positive detections, and retaining a maximum of 4 faces per statues (in total 1863 faces). We then detect landmarks in the faces [8] and normalize them (with affine projection). We use a 4096-dimensional embedding space (VGG Faces [10] trained on human faces) for each face. Finally, we project this 4096-dimensional space into a 2D space with cosine similarity by t-SNE [3], [9]. The resulting map of all faces 4 places similar faces together as a proof of concept illustrating the feasibility of our pipeline

The map does not reflect yet the actual proximity between Buddha statues but raises interesting questions. The embedding is purely based on human appearances of statues, ignoring the specific features of Buddha and without integrating any domain knowledge yet.

4.2 Challenges

First is the quality of annotation: together with domain experts, we are harvesting data to complete statue types, location, and so on in order to build a comprehensive database. In the end, we wish to store all the domain information in a graph database, that would represent a knowledge graph.

The second aspect is to fine tune state of the art face detectors and landmarking for annotations. We wish to automatically measure the iconometry and compare with the canon of different periods. We need these detectors to capture the variety within a set of Buddha statues rather than human beings, and even to predict meta information specific to Buddha statues. A knowledge database will give rich ground to train quantities of classifiers.

4.3 Extending to 3D

Our long term goal is to deploy 3D analysis on our large scale archive. While mining the archive, we found that many statues have been captured with multiple views such as illustrated in Fig. 2. To do so, we have started capturing 3D scans of Buddha statues *in situ* to develop a ground truth. It is to be noted that precise reconstruction of statues often comes with many challenges to avoid model deformations [4].

5. Conclusion

In order to achieve our goal, curating proper data and capturing domain knowledge about the statues is critical and we will need to enrich the database with external data (such as *Wikipedia* or *e-museum*). Since scalability is key to our approach, one of our main challenges is to combine both 2D and 3D methods such that it will allow the deployment of iconometry on a large scale.

謝辞　This work was supported by JSPS KAKENHI Grant Number 18H03571.

Vol.2019-CH-119 No.7
2019/2/16

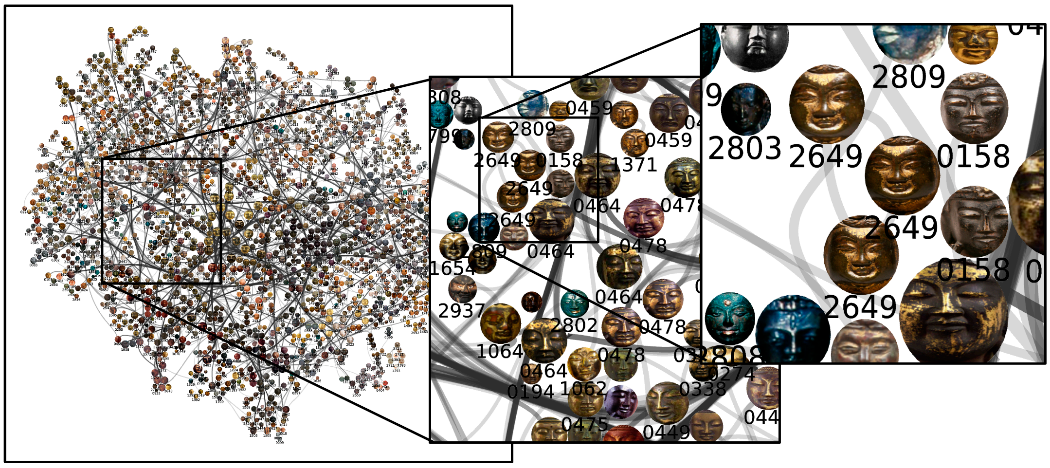

図 4 Visualization of the Buddha faces embedding in the archive after extraction and normalization. Links are connecting faces belonging to the same statue pictures (sometimes multiple statues have been found). Node size depends on the total number of faces found in a statue group.

参考文献

[1] Elgammal, A., Mazzone, M., Liu, B., Kim, D. and Elhoseiny, M.: The Shape of Art History in the Eyes of the Machine, *arXiv preprint arXiv:1801.07729* (2018).

[2] Elgammal, A. and Saleh, B.: Quantifying creativity in art networks, *arXiv preprint arXiv:1506.00711* (2015).

[3] Gisbrecht, A., Schulz, A. and Hammer, B.: Parametric nonlinear dimensionality reduction using kernel t-SNE, *Neurocomputing*, Vol. 147, pp. 71–82 (2015).

[4] Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M. and Okamoto, Y.: The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects, *International Journal of Computer Vision*, Vol. 75, No. 1, pp. 189–208 (2007).

[5] Jafarpour, S., Polatkan, G., Brevdo, E., Hughes, S., Brasoveanu, A. and Daubechies, I.: Stylistic analysis of paintings using wavelets and machine learning, *Signal Processing Conference, 2009 17th European*, IEEE, pp. 1220–1224 (2009).

[6] Jiang, H. and Learned-Miller, E.: Face detection with the faster R-CNN, *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, IEEE, pp. 650–657 (2017).

[7] Kamakura, M., Oishi, T., Takamatsu, J. and Ikeuchi, K.: Classification of Bayon faces using 3D models, *Virtual Systems and Multimedia*, pp. 751–760 (2005).

[8] Kazemi, V. and Sullivan, J.: One millisecond face alignment with an ensemble of regression trees, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1867–1874 (2014).

[9] Maaten, L. v. d. and Hinton, G.: Visualizing data using t-SNE, *Journal of machine learning research*, Vol. 9, No. Nov, pp. 2579–2605 (2008).

[10] Parkhi, O. M., Vedaldi, A., Zisserman, A. et al.: Deep face recognition., *BMVC*, Vol. 1, No. 3, p. 6 (2015).

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

[12] Sirirattanapol, C., Matsui, Y., Satoh, S., Matsuda, K. and Yamamoto, K.: Deep Image Retrieval Applied on Kotenseki Ancient Japanese Literature, *Multimedia (ISM), 2017 IEEE International Symposium on*, IEEE, pp. 495–499 (2017).

[13] Strezoski, G. and Worring, M.: OmniArt: Multi-task Deep Learning for Artistic Data Analysis, *arXiv preprint arXiv:1708.00684* (2017).

[14] Tan, W. R., Chan, C. S., Aguirre, H. E. and Tanaka, K.: Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification, *Image Processing (ICIP), 2016 IEEE International Conference on*, IEEE, pp. 3703–3707 (2016).

[15] Wang, N., Gao, X., Tao, D., Yang, H. and Li, X.: Facial feature point detection: A comprehensive survey, *Neurocomputing*, Vol. 275, pp. 50–65 (2018).

[16] Wynen, D., Schmid, C. and Mairal, J.: Unsupervised Learning of Artistic Styles with Archetypal Style Analysis, *arXiv preprint arXiv:1805.11155* (2018).

[17] Zauner, C.: Implementation and benchmarking of perceptual image hash functions (2010).

[18] 久保文乃，村上征勝：同時代仏師との比較による快慶作品の特徴について—様式と法量からみる—，研究報告人文科学とコンピュータ (CH)， Vol. 2011, No. 1, pp. 1–6 (2011).

[19] 修　山田：彫刻文化財にみられる図学的解釈，大会学術講演論文集，pp. 23–28 (2014).

[20] 　山本勉：仏像のひみつ，朝日出版社 (2006).

[21] 西村公朝，小川光三：仏像の見分け方 (1987).

[22] Felice Fischer (監修) 副島弘道 (監修)：英訳付 仏像鑑賞ガイド—A Guide to Japanese Buddhist Sculpture，池田書店 (2008).

[23] 　名畑崇：仏教伝来と仏像の伝説–光るほとけ，大谷学報，Vol. 65, No. 4, pp. p1–16 (1986).

[24] 清水眞澄：仏像の顔——形と表情をよむ, 岩波新書 (2013).