Historical and Modern Features for Buddha Statue Classification

Benjamin Renoust, Matheus Oliveira Franca, Jacob Chan, Noa Garcia, Van Le, Ayaka Uesaka, Yuta Nakashima, Hajime Nagahara
Institute for Datability Science, Osaka University
Osaka, Japan

Jueren Wang
Yutaka Fujioka
fujioka@let.osaka-u.ac.jp
Graduate School of Letters, Osaka University
Osaka, Japan

ABSTRACT
While Buddhism has spread along the Silk Roads, many pieces of art have been displaced. Only a few experts may identify these works, subjectively to their experience. The construction of Buddha statues was taught through the definition of canon rules, but the applications of those rules greatly varies across time and space. Automatic art analysis aims at supporting these challenges. We propose to automatically recover the proportions induced by the construction guidelines, in order to use them and compare between different deep learning features for several classification tasks, in a medium size but rich dataset of Buddha statues, collected with experts of Buddhism art history.

CCS CONCEPTS
• Applied computing → Fine arts; • Computing methodologies → Image representations; Interest point and salient region detections.

KEYWORDS
Art History, Buddha statues, classification, face landmarks

1 INTRODUCTION
Started in India, Buddhism spread across all the Asian subcontinent through China reaching the coasts of South-eastern Asia and the Japanese archipelago, benefiting from the travels along the Silk Roads [40, 50]. The story is still subject to many debates as multiple theories are confronting on how this spread and evolution took place [32, 40, 44, 50]. Nonetheless, as Buddhism flourished along the centuries, scholars have exchanged original ideas that further diffused, shaping the different branches of Buddhism and art as we know them today. When Buddhism reached new territories, local people would craft Buddhism art themselves. Not only they observed common rules of crafting, but also they adapted them to express their own culture, giving rise to new styles [35].

With multiple crisis and cultural exchanges, many pieces of art have been displaced, and their sources may remain today uncertain. Only a few experts can identify these works. This is however subject to their own knowledge, and the origin of some statues is still disputed today [25]. However, our decade has seen tremendous progress in machine learning, so that we may harvest these techniques to support art identification [4].

Our work focuses on the representation of Buddha, central to Buddhism art, and more specifically on Buddha statues. Statues are 3D objects by nature. There are many types of Buddha statues, but all of them obey construction rules. These are canons, that make a set of universal rules or principals to establish the very fundamentals of the representation of Buddha. Although the canons have first been taught using language-based description, these rules have been preserved today, and are consigned graphically in rule books (e.g. Tibetan representations [33, 34] as illustrated in Fig. 2a). The study of art pieces measurements, or iconometry, may further be used to investigate the differences between classes of Buddha statues [49].

In this paper, we are interested in understanding how these rules can reflect in a medium size set of Buddha statues (> 1k identified statues in about 7k images). We focus in particular on the faces of the statues, through photographs taken from these statues (each being a 2D projection of the 3D statue). We propose to automatically recover the construction guidelines, and proceed with iconometry in a systematic manner. Taking advantage of the recent advances in image description, we further investigate different deep features and classification tasks of Buddha statues. This paper contributes by setting a baseline for the comparison between “historical” features, the set of canon rules, against “modern” features, on a medium size dataset of Buddha statues and pictures.

The rest of the paper is organized as follows. After discussing the related work, we present our dataset in Section 2. We then introduce the iconometry measurement and application in Section 3. From this point on, we study different embedding techniques and compare them along a larger set of classification tasks (Sec. 4) before concluding.
1.1 Related Work

Automatic art analysis is not a new topic, and early works have focused on handcrafted feature extraction to represent the content typically of paintings [6, 21, 24, 30, 39]. These features were specific to their application, such as the author identification by brushwork decomposition using wavelets [21]. A combination of color, edge, and texture features was used for author/school/style classification [24, 39]. The larger task of painting classification has also been approached in a much more traditional way with SIFT features [6, 30].

This was naturally extended to the use of deep visual features with great effectiveness [1, 13, 14, 17, 23, 26, 28, 37, 46, 47]. The first approaches were using pre-trained networks for automatic classification [1, 23, 37]. Fine-tuned networks have then shown improved performances [7, 28, 37, 46, 47]. Recent approaches [16, 17] introduced the combination of multimedia information in the form of joint visual and textual models [17] or using graph modeling [16] for the semantic analysis of paintings. The analysis of style has also been investigated with relation to time and visual features [13, 14]. Other alternatives are exploring domain transfer for object and face detection and recognition [9–11].

These methods mostly focus on capturing the visual content of paintings, on very well-curated datasets. However, paintings are very different to Buddha statues, in that sense that statues are 3D objects, created with strict rules. In addition, we are interested by studying the history of art, not limited to the visual appearance, but also about their historical, material, and artistic context. In this work, we explore different embeddings, from ancient Tibetan rules, to modern visual, in addition to face-based, and graph-based, for different classification tasks of Buddha statues.

We can also investigate recent works which are close to our application domain, i.e., the analysis of oriental statues [2, 3, 20, 22, 48, 49]. Although, one previous work has achieved Thai statue recognition by using handcrafted facial features [36]. Other related works focus on the 3D acquisition of statues [20, 22] and their structural analysis [2, 3], with sometimes the goals of classification too [22, 49]. We should also highlight the recent use of inpainting techniques on Buddhism faces for the study and recovery of damaged pieces [48].

Because 3D scanning does not scale to the order of thousands statues, we investigate features of 2D pictures of 3D statues, very close to the spirit of Pornpanomchai et al. [36]. In addition to the study of ancient proportions, we provide modern analysis with visual, face-based (which also implies a 3D analysis), and semantic features for multiple classification tasks, on a very sparse dataset that does not provide information for every class.

2 DATA

This work is led in collaboration with experts who wish to investigate three important styles of Buddha statues. A first style is made of statues from ancient China spreading between the IV and XIII centuries. A second style is made of Japanese statues during the Heian period (794–1185). The last style is also made of Japanese statues, during the Kamakura era (1185–1333).

To do so, our experts have captured (scanned) photos in 4 series of books, resulting in a total of 6811 scanned images, and documented 1393 statues among them. The first series [29] concerns 1076 Chinese statues (1524 pictures). Two book series [41, 42] regroup 132 statues of the Heian period (1847 pictures). The last series [31] collects 185 statues of the Kamakura era (3888 pictures).

To further investigate the statues, our experts have also manually curated extra meta-data information (only when available).
Figure 2: Above: Deriving the Buddha iconometric proportions based on 68 facial landmarks. (a) Proportional measurements on a Tibetan canon of Buddha [33] facial regions and their value (original template © Carmen Mensik, www.tibetanbuddhistart.com). (b) Iconometric proportional guidelines defined from the 68 facial landmark points. (c) Application of landmarks and guidelines detection to the Tibetan model [33]. (d) 3D 68 facial landmarks detected on a Buddha statue image, and its frontal projection (e). Below: Examples of the detected iconometric proportions in three different styles. (f) China. (g) Heian. (h) Kamakura. (i) The combined and superimposed iconometric proportional lines from the examples (f)-(h). Canon image is the courtesy of Carmen Mensink [33].

For the localization, we so far only consider China and Japan. Dimensions are reporting the height of each statue, so we created three classes: small (from 0 to 100 cm), medium (from 100 cm to 250 cm) and big (greater than 250 cm). Many statues also have a specific statue type attributed to them. We threshold them to the most common types, represented by at least 20 pictures, namely Bodhisattva and Buddha.

A temporal information which can be inferred from up to four components: an exact international date, a date or period that may be specific to the Japanese or Chinese traditional dating system, a century information (period), an era that may be specific to Japan or China (period). Because these information may be only periods, we re-align them temporally to give an estimate year in the international system, that is the median year of the intersection of all potential time periods. They all distribute between the V and XIII century.

Material information is also provided but it is made of multiple compounds and/or subdivisions. We observe the following categories: base material can be of wood, wood+lacquer, iron, or brick; color or texture can refer to pigment, lacquered foil, gold leaves, gold paint, plating, dry lacquer finish, or lacquer; type of stone (when applies) may be limestone, sand stone, white marble, or marble; type of wood (also when applies) may be Japanese cypress, Katsura, Japanese Torreya, cherry wood, coniferous, or camphor tree; the material may also imply a construction method among separate pieces, one piece cut, and one piece.

Fig. 1 shows the distribution of all the images across the different classes. Because for each of the statues many information is either uncertain or unavailable, we can note that the data is very sparse, and most of the different classes are balanced unevenly. Note that not all pictures are corresponding to a documented statue, the curated dataset annotates a total of 3065 images in 1393 unique statues. In addition, not the same statues shares the same information, i.e., some statues have color information, but no base material, when others have temporal information only etc. As a consequence, each classification task we describe later in Sec. 4 has a specific subset of images and statues to which it may apply, not necessary overlapping with the subsets of other tasks.

3 ICONOMETRY

We begin our analysis with the use of historic iconometry for determining facial proportions in figurative Buddha constructions. For this, we have chosen a model based on a Tibetan-originated 18th century book comprising of precise iconometric guidelines for representing Buddha-related artworks [33, 34]. Although this book primarily encompasses Tibetan-based Buddha drawing guidelines, it gave us insights of how Buddha-artists from different eras and geographical locations proportionate key facial regions in their portrayal of Buddha artworks.

We propose to detect and use these proportions for the analysis and differentiation of Buddha designs from different eras and locations around the world. Fig. 2a depicts the chosen iconometric
proportional measurements of different facial regions that is used in our analysis. The idea is to use automatic landmark detection, so we may infer the iconometry lines from any Buddha face in the dataset. Based on these lines, we can identify and normalize the proportions of each key region of the Buddha faces and compare them together and against the canons.

3.1 Guidelines and Proportions

The guidelines are given for a front facing Buddha statue, but not all pictures are perfectly facing front the camera source point. Finding 3D facial landmarks allows for affine spatial transformation, and for normalizing the statue pose before searching for the iconometric guidelines.

Moreover, we wish to locate the guidelines with relation to important facial points. To do so, we first employ facial landmark detection on the historical Buddha model, and find correspondences between the lines and the model (as detailed in Table 1 and illustrated in Fig. 2a-c). Because the landmark points are defined in a 3-dimensional space, the correspondences are defined on the 2D front-facing orthogonal projection of the landmarks. We employ the Position Map Regression Network (PRN) [15] which identifies 68 3D facial landmarks in faces. Table 1 defines the proportional guidelines that can be drawn from any given 68 facial landmark points (refer to the Fig. 2b for reference to the point numbers).

Once the guidelines are established from the detected 68 landmark points, each key region of the Buddha face is then measured according to the proposed proportions as seen in Fig. 2a. For this analysis we do not make use of the inner diagonal guidelines, but we rather focus on a clear subset of six key facial regions, namely, left forehead (LH), right forehead (RH), eyelids (EL), eyes (E), nose (N), and lower face (LF). Table 2 details how we may derive the proportions from the lines, with their theoretical values, Fig. 2c shows the lines once the whole process is applied to the historical model. Fig. 2d shows the PRN-detected 68 landmark points on a Buddha face and its 2D frontal orthographic projection is presented in Fig. 2e. Results on statues are shown in Fig. 2f-i.

3.2 Analysis

Given our dataset, we apply the above described iconometric proportions for the three main categories of statues. Given that we may have multiple pictures for each statue and that the landmark detection may fail on some pictures, we obtain 179 measurements for statues from China, 894 proportions for Japan Heian statues, and 1994 for Japan Kamakura statues. Results are reported in Fig. 3 against two baselines, the theoretical Tibetan canon baseline, and the actually measured baseline on the same Tibetan model.

Although the proportion differences might be minute, it can be observed that the Buddha designs from China, in general, have much larger noses and shorter eyelids when compared with the other two datasets, while Buddhas from the Kamakura period have their design proportions in-between the other two datasets. Eyelids tend to be slightly smaller for Kamakura designs in comparison to Heian ones. Fig. 2f-i show a sample of the iconometric proportional measurement taken from each of the experimented dataset while Fig. 2i displays a superimposition of the three.

One can also notice some important difference between the theoretical canons of the Tibetan model and their actual measurement in the dataset. Considering the small average distance between the observed model proportions and the different measurements on real statues, we may wonder whether this distance is an artifact due to the measurement methodology – which is trained for human faces – or to an actual approximation of these measures. Even in the original Tibetan model, the proportions of the nose appear to the eye larger than the one originally described.

Although the differences are not striking for the measurements themselves, they do actually differ as the timelines and locations change. This motivates us to further investigate if modern image

Table 1: The proportional guidelines can be drawn from any given 68 facial landmark points as shown in Fig. 2b.

| Line | Description | Point Connections |
|------|-------------|-------------------|
| L1   | Eyebrow Line| Mean of (19,21) to Mean of (24,26) |
| L2   | Top Eye Line| Mean of (38,39) to Mean of (44,45) |
| L3   | Bottom Eye Line| Mean of (41,42) to Mean of (47,48) |
| L4   | Nose Sides Line| 32 to 36 |
| L5   | Jaw Line | 7 to 11 |
| L6   | Center Nose Line| Mean of (22,23) to Mean of (28,29,30,31) |
| L7   | Left Face Line| Line between L1 and L5 through 2 |
| L8   | Left Face Line| Line between L1 and L5 through 16 |
embedding can reveal further differences among different categories of Buddha statues.

4 MODERN EMBEDDINGS

Since the previous method based on a historical description of facial landmarks does not give a clear cut between classes, we also explore modern types of embeddings designed for classification, namely, image embeddings that take full image for description, face embeddings trained for facial recognition, and graph embeddings purely built on the semantics of the statues.

4.1 Classification Tasks

Our initial research question is a style classification (T1), i.e. the comparison of three different styles: China, Kamakura period, and Heian period. Given the rich dataset we have been offered to explore, we also approach four additional classification tasks.

We conduct a statue type classification (T2) which guesses the type of Buddha represented, and dimension classification (T3) which classifies the dimension of a statue across the three classes determined in Sec. 2.

We continue with a century classification (T4), given the temporal alignment of our statues, each could be assigned to a different century (we are covering a total of nine centuries in our dataset).

We conclude with the material classifications (T5), which comprises: base material (T5.1), color/texture (T5.2), type of stone (T5.3), type of wood (T5.4), and construction method (T5.5). Note that all material classifications except for task T5.5 are actually multilabel classification tasks, indeed a statue can combine different materials, colors, etc. Only the construction method is unique, thus single label classification.

To evaluate classification across each of these tasks, we limit our dataset to the 1393 annotated and cleaned statues, covering a total of 3315 images. To compare the different methods on the same dataset, we further limit our evaluation to the 2508 pictures with a detectable face as searched during Sec. 3, using PRN [15]. Due to the limited size of the dataset, we train our classifiers using a 5-fold cross-validation.

4.2 Image Embeddings

To describe our Buddha statue 2D pictures, we propose to study existing neural network architectures which already have proven great success in many classification tasks, namely VGG [16] and ResNet50 [19].

For the classification of Buddha statues from the global aspect of their image, we use each of these networks with their standard pre-trained weights (from ImageNet [12]).

To study the classification performances of statues with regards to their face, we first restrain the face region using PRN [15]. To compare the relevance of the facial region for classification, we evaluate against two datasets. The first one evaluates ImageNet-trained embeddings on the full images (referred to as VGG16full and ResNet50full), the second one evaluates the same features, but only on the cropped region of the face (VGG16cropped and ResNet50cropped). In addition, each of the networks is also fine-tuned using VGGFace2 [5], a large-scale dataset designed for the face recognition task (on cropped faces), hereafter VGG16vggface2 and ResNet50vggface2.

Whichever the method described above, the size of the resulting embedding space is of 2048 dimensions.

4.3 Semantic Embedding

Given the rich data we are provided, and inspired by the work of Garcia et al. [16], we may also explore semantic embedding in the form of an artistic knowledge graph.

Instead of traditional homophily relationships, our artistic knowledge graph KG = (V, E) is composed of multiple types of node: first of all, each statue picture is a node (e.g. the Great Buddha of Kamakura). Then, each value of each family of attributes also has a node, connected to the nodes of the statues they qualify (for example, the Great Buddha of Kamakura node will be connected to the Bronze node).

From the metadata provided, we construct two knowledge graphs. A first knowledge graph KG only uses the following families of attributes: Dimensions, Materials, and Statue type. Because we are curious in testing the impact of time as a determinant of style, we also add the Century attributes in a more complete graph KGtime. In total, the resulting KG presents 3389 nodes and 16756 edges, and KGtime presents 3401 nodes and 20120 edges. An illustrative representation of our artistic knowledge graph is shown in Fig. 4.

Considering the sparsity of all our data, due to noisy and/or missing values, this graph definition suits very well our case. However,
4.4 Evaluation

We use two types of classifiers for each task.

- We use a classical Support Vector Machine (SVM) classifier [8]. To improve the quality of the classifier given imbalanced data, we adjust the following parameters: \( y = 1/|M| \) \(|M|\), the number of classes), \( \text{penalty} = 1 \), linear kernel, and adjusted class weights \( w_m = N/k \cdot n_m \) (inversely proportional to class frequency in the input data: for a class \( m \) among \( k \) classes, having \( n_m \) observations among \( N \) observations).

- We additionally train a Neural Network classifier (NN), in form of a fully connected layer followed by softmax activation with categorical crossentropy loss \( L(y, \hat{y}) \) if only one category is applicable:

\[
L(y, \hat{y}) = -\sum_{j=0}^{M} \sum_{i=0}^{N} (y_{ij} \cdot \log(\hat{y}_{ij}))
\]

with \( M \) categories. Otherwise, we use a binary crossentropy \( H_p(q) \) for multi-label classification, as follows:

\[
H_p(q) = \frac{1}{N} \sum_{i=1}^{N} y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i))
\]

With \( y \) the embedding vector, \( N \) is the training set. Both cases use Adam optimizer, and the size of the output layer is then matched to the number of possible classes.

For each of the task we report the weighted average, more adapted for classification with unbalanced labels, of precision and recall under the form of F1-score (precision and recall values are very comparable across all our classifiers, so F1-score works very well in our case). Classification results are presented in Table 3.

5 DISCUSSION

The first point we may notice from the classification results is that using iconometry does not perform very well in comparison to neural networks. In average, we obtain with iconometry a precision score of 0.49 for a recall score of 0.67, whereas scores are very similar for all the other methods. Deep learning based methods perform better, but not equally on all tasks. Style (T1), base material (T5.1) and type of wood (T5.4) are the tasks that are the best classified. Type of stone (T5.3) is by far the most difficult to classify, probably due to the imbalance of its labels. Iconometry is significantly worse than the other methods in guessing the construction method (T5.5) and the color/texture (T5.2). The neural network classifier (NN) usually perform better than SVM, except in the multilabel classification tasks (T5.1–T5.4). It suggests that those classes have a more linear distribution.

VGGFace2 [5] trained methods perform a little worse than their counterpart trained with ImageNet [12] on the cropped faces, which in turn perform slightly worse than the full images. This may happen because VGGFace2 dataset takes into account variations between ages, contrarily to the shape of Buddha faces which are more rigid (from the point of view of the construction guidelines). It might also suggest that faces of Buddha statues differ fundamentally from a standard human face, which makes the transfer learning not so effective from the networks pretrained in VGGFace2. This proposition agrees with the fact that iconometry did not perform well. The differences are not significant though, giving us great hope for fine tuning models directly from Buddha faces.

In addition, ResNet50 [19] appears to show the best results overall. Remarkably, when we focus only on the face region, ResNet50 performs even better than the others when classifying type of wood and construction method, which encourages the idea of using face region for a good discriminator, specially for classification related to the material.

The semantic embeddings based on artistic knowledge graph perform as well as the best of image-based embeddings for style classification (T1), a result consistent with Garcia et al.’s observations [16]. This is probably due to the contextual information carried by the graph. However, if the century is not present in the KG, ResNet50 still shows better results than node2vec. We can additionally underline that temporal information is a good predictor.
of style, since the classification performance is slightly improved after adding the centuries information in the knowledge graph.

We may further investigate the space defined by those embeddings as illustrated in Fig. 5. It is interesting to see the similarities in the space between node2vec and the iconometry used as embeddings. However, their classification performances are very different. The iconometry embeddings do not look like to well separate the three styles, but there seems to be quite notable clusters forming that should be interesting to investigate further. The advantage of iconometry over other embeddings is its explainability. Integrating time into the KG clearly shows a better separability of the three styles.

By looking at the spread of the face-based ResNet50\textsubscript{cropped} ResNet50\textsubscript{face} embeddings, we may also notice that the different classes are much more diffused than ResNet50\textsubscript{full}. The face region is very specialized in the space of all shapes. Although the vgg face2 embeddings are trained for a different task, facial recognition, i.e. to identify similar faces with quite some variability, we do not see a clear difference between the separation of style from the face regions. To show the effectiveness of facial analysis against whole picture analysis, we will need to proceed with further experiments, including increasing the variety of Buddha statues in our dataset in order to train specific models designed for Buddha faces.

6 CONCLUSION

We have presented a method for acquisition of iconometric guidelines in Buddha faces and use them for different classification tasks. We have compared them with different modern embeddings, which have demonstrated much higher classification performances. Still there is one advantage of the iconometric guidelines from their simplicity and ease of understanding. To further understand what makes a style, we would like to investigate visualization and parameter regressions in future works, and identify salient areas that are specific to a class.

We have presented one straightforward method for the identification of iconometric landmarks in Buddha statue, but many statues did not show good enough landmarks to be measured at all. We could extend our landmark analysis, and boost the discrimination power of landmarks by designing a specific landmark detector for Buddha statues.

Scanning books is definitely more scalable today than 3D-captures of the statues. However, with the high results of the deep-learning
methods for style classification, we could question how influential was the data acquisition method on the classification. Each book paper may have a slightly different grain that deep neural networks may have captured. Nonetheless, the different classification tasks are relatively independent from the book source while still showing quite high results. One of our goals is to continue develop this dataset and multiply our data sources, so it would further diminish the influence of data acquisition over analysis.

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

REFERENCES

[1] Yaniv Bar, Noga Levy, and Lior Wolf. 2014. Classification of Artistic Styles Using Binarized Features Derived from a Deep Neural Network. In ECCV Workshops.

[2] Andrew Bevan, Xiaohen Li, Marcos Martín-Torres, Susan Green, Yin Xia, Kun Zhao, Zhen Zhao, Shengtao Ma, Wei Cao, and Thilo Rehren. 2014. Computer vision, archaeological classification and China’s terracotta warriors. Journal of Archaeological Science 49 (2014), 249–254.

[3] Gopa Bhaumik, Shefaliika Ghosh Samaddar, and Arun Baran Samaddar. 2018. Recognition Techniques in Buddhist Iconography and Challenges. In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 1285–1289.

[4] Alexander Blessing and Kai Wen. 2010. Using machine learning for identification of art paintings. Technical report (2010).

[5] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. 2018. VGGFace2: A dataset for recognising faces across pose and age. In International Conference on Automatic Face and Gesture Recognition.

[6] Gustavo Carneiro, Nuno Pinho da Silva, Alessio Del Bue, and João Paulo Costeira. 2012. Artistic image classification: An analysis on the printart database. In ECCV.

[7] Wei Ta Chu and Yi-Ling Wu. 2018. Image style classification based on learnt deep correlation features. IEEE Transactions on Multimedia 20, 9 (2018), 2491–2502.

[8] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. Machine learning 20, 3 (1995), 273–297.

[9] Elliot Crowley and Andrew Zisserman. 2014. The State of the Art: Object Retrieval in Paintings using Discriminative Regions. In BMVC.

[10] Elliot J Crowley, Omkar M Parkhi, and Andrew Zisserman. 2015. Face Painting: querying art with photos. In BMVC. 65–6.

[11] Elliot J Crowley and Andrew Zisserman. 2016. The art of detection. In ECCV Springer.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. IEEE, 248–255.

[13] Ahmed Elgammal, Marian Mazzone, Bingchen Liu, Diana Kim, and Mohamed Elhoseiny. 2018. The Shape of Art History in the Eyes of the Machine. arXiv preprint arXiv:1803.07729 (2018).

[14] Ahmed Elgammal and Babak Saleh. 2015. Quantifying creativity in art networks. arXiv preprint arXiv:1506.00711 (2015).

[15] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. 2018. Joint 3d face reconstruction and dense alignment with position map regression network. In Proceedings of the European Conference on Computer Vision (ECCV). 534–551.

[16] Noa Garcia, Benjamin Renoust, and Yuta Nakashima. 2014. Context-Aware Embeddings for Automatic Art Analysis. In Proceedings of the 2014 on International Conference on Multimedia Retrieval. ACM, 25–33.

[17] Noa Garcia and George Vogiatzis. 2018. How to Read Paintings: Semantic Art Understanding with Multi-Modal Retrieval. In ECCV Workshops.

[18] Aditya Grover and Jure Leskovec. 2016. Node2Vec: Scalable Feature Learning for Networks. In Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16). ACM, New York, NY, USA, 855–864. https://doi.org/10.1145/2939672.2939754

[19] Kaining He, Xiangyu Zhang, Shaoming Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778.

[20] Katsushi Ikeuchi, Takeshi Oishi, Jun Takamatsu, Ryusuke Sagawa, Atsushi Nakazawa, Ryosuke Kurazume, Ko Nishino, Mawo Kamakura, and Yasuhide Okamoto. 2007. The great baubles project: Digitally archiving, restoring, and analyzing cultural heritage objects. International Journal of Computer Vision 75, 1 (2007), 189–208.

[21] C Richard Johnson, Ella Hendries, Igor J Berenzohny, Eugene Brevdo, Shannon M Hughes, Ingrid Daubechies, Jia Li, Eruc Postma, and James Z Wang. 2008. Image processing for artist identification. IEEE Signal Processing Magazine 25, 4 (2008).

[22] Mawo Kamakura, Takeshi Oishi, Jun Takamatsu, and Katsushi Ikeuchi. 2005. Classification of Bayon faces using 3D models. In Virtual Systems and Multimedia. 751–760.