

Article

A Computational Approach to Hand Pose Recognition in Early Modern Paintings

Valentine Bernasconi ^{1,*} , Eva Cetinić ² and Leonardo Impett ³¹ Digital Visual Studies, University of Zurich, 8006 Zurich, Switzerland² Digital Society Initiative, University of Zurich, 8001 Zurich, Switzerland; eva.cetinic@uzh.ch³ Cambridge Digital Humanities, University of Cambridge, Cambridge CB2 1RX, UK; li222@cam.ac.uk

* Correspondence: valentine.bernasconi@uzh.ch

Abstract: Hands represent an important aspect of pictorial narration but have rarely been addressed as an object of study in art history and digital humanities. Although hand gestures play a significant role in conveying emotions, narratives, and cultural symbolism in the context of visual art, a comprehensive terminology for the classification of depicted hand poses is still lacking. In this article, we present the process of creating a new annotated dataset of pictorial hand poses. The dataset is based on a collection of European early modern paintings, from which hands are extracted using human pose estimation (HPE) methods. The hand images are then manually annotated based on art historical categorization schemes. From this categorization, we introduce a new classification task and perform a series of experiments using different types of features, including our newly introduced 2D hand keypoint features, as well as existing neural network-based features. This classification task represents a new and complex challenge due to the subtle and contextually dependent differences between depicted hands. The presented computational approach to hand pose recognition in paintings represents an initial attempt to tackle this challenge, which could potentially advance the use of HPE methods on paintings, as well as foster new research on the understanding of hand gestures in art.

Keywords: annotated dataset; hand pose estimation; hand pose; image classification; digital art history; hand gestures; paintings; early modern times; digital humanities



Citation: Bernasconi, V.; Cetinić, E.; Impett, L. A Computational Approach to Hand Pose Recognition in Early Modern Paintings. *J. Imaging* **2023**, *9*, 120. <https://doi.org/10.3390/jimaging9060120>

Academic Editors: Filippo Stanco, Dario Allegra, Mario Molinara and Alessandra Scotto di Freca

Received: 28 April 2023

Revised: 25 May 2023

Accepted: 11 June 2023

Published: 15 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hands represent an important aspect of our mode of communication, and paintings are no exception. Indeed, from medieval representations of symbolic gestures [1] to their more natural depictions in European early modern art [2] (p. 86), hands seem to convey precious information about sociocultural evolution and play a key role in the pictorial narrative system. Although the specific topic of hand gesture analysis has been fairly sparsely addressed in art history and digital humanities, the recent work by the art historian Dimova [3] defines the foundations of a categorization system of painted hand gestures. In total, her system proposes 30 chirograms, each referring to a specific hand pose and its possible significance. This represents a starting point for a new kind of pictorial reading and serves as a basis for the classification problem addressed in this paper. However, there are several important challenges that need to be addressed in order to transform the art historical categorization scheme into a computational workflow. First of all, the art historical categorization of painted hands is very ambiguous and strongly depends on both the context of the depicted scene as well as the context of the creation of the painting. More specifically, in order to correctly classify the hand gesture, one must be familiar with the iconographic theme of the painting as well as the role and the specific position of the figure from which the hand has been extracted. Furthermore, regarding the context of creation of the artwork, in particular, the sociocultural context of the painter itself can be an indication of the gestural convention in place [4]. This indicates the need

for deep contextual knowledge in the data annotation process and hinders the possibility of performing crowdsourcing-based data annotation, making it difficult to work with large-scale datasets.

In the last decade, there has been increased interest in exploring computer vision and machine learning methods for analyzing and categorizing digitized collections of paintings. Most of the work in this field is focused on classification [5–8], pattern recognition and retrieval [9–13], or object detection in paintings [14–17]. Studying bodily depictions in the context of digital art history has so far been mostly exclusively focused on body pose detection and recognition [18–20]. Such works, essentially based on the aforementioned operationalization of theories presented by the famous art historian Aby Warburg [21], ground their methodological approach in human pose estimation (HPE). Based on deep learning models [22,23], trained on large, annotated dataset of real images [24,25], such as OpenPose [22,26], these procedures produce keypoint features that can be used for pose comparison [19,27–29]. However, these models achieve a lower accuracy on artworks [20,30], that integrate visual features other than natural images, such as layouts and brush stroke marks. One of the major current challenges is the lack of proper training datasets that would improve the results of applying HPE techniques to paintings. Moreover, while body pose estimation and its iconographic interpretation has been the subject of recent studies [30–34], the task of hand pose classification in paintings has not yet been computationally addressed.

More specifically, computer-vision-based approaches, such as hand gesture recognition (HGR) and classification, have been extensively addressed and defined within the domain of human–computer interaction (HCI) [35–38]. HCI includes research on the development of sensor-based gesture detection [39], single RGB cameras [40], and depth cameras [41–43], as well as the attribution of hand poses with a recent focus on deep learning methods [44–46]. From a temporal perspective, hand gestures can be divided into *static* or *dynamic* categories. Static hand gestures are commonly called hand poses, and their understanding relies on their orientation, the positions of the fingers, the angles shaped by their articulations, and their relative positions toward the body [36]. In addition to the temporal perspective, communicative and manipulative gestures are also often differentiated without clear taxonomies defined among the field [35]. Ultimately, gesture recognition systems can be sequenced into different processes, such as identification, tracking, and classification or detection, feature extraction, and classification [35,37].

The field of sign language recognition (SLR) also includes notable research on the development of the live detection and attribution of hand poses [47,48]. Methods mostly involve the analysis of RGB and depth images representing the hand, but also the face and body movements of an individual [47]. Although SLR methods are also applied to static hand poses in the case of alphabets for sign language processing [49,50], the methodology mostly applies to dynamic hand gesture recognition. Furthermore, it seems that iconographic categories of painted hands from early modern times significantly differ from the nomenclature of sign language [3]. Indeed, sign language is a living language, defined by grammatical rules that not only involve the hands but also the face and other body parts [47].

Nevertheless, most of the deep learning models used in this context are trained on large annotated datasets of photographic images and achieve lower accuracy on artworks, which presents new challenges for the field of computer vision.

In this work, we propose the exploitation of methods and body characterization proposed in previous research on digital art history and apply it to the hands. Using HPE methods, we explain the process of the creation of a dataset of painted hands. It includes multiple curatorial steps, which reveals gaps between art historical and computational understanding of digitized artistic material. From the dataset produced, we define different features and propose a classification task based on art historical chirograms. In particular, our main contributions are:

- The first annotated dataset of painted hands (Painted Hand Pose dataset), introducing a new standard for hand pose annotation in the context of digital art history;
- New hand pose features based on 2D keypoint information produced by HPE methods;
- The introduction of a novel classification task and a proposed solution method based on keypoint feature extraction.

Our newly introduced Painted Hand Pose dataset and the proposed keypoints feature represent a first attempt to establish a computational framework for the analysis of hand gestures in paintings. This approach can serve as a basis for future research work or different applications, such as browsing data collections of painting through gestures, as presented in [29]. In the context of this work, we additionally define and explore an automated classification task in order to better understand the potential and limitations of the dataset and features. Because of the small size of the dataset and the complex information encoded within the depiction of various hand poses in paintings, this classification reveals a new possible challenge for the field of computer vision, opening new possibilities for future exploration of this subject.

2. Materials and Methods

2.1. Dataset

2.1.1. Data Acquisition and Metadata Processing

The data source consists of digital images and their corresponding metadata from the photographic collection of the Bibliotheca Hertziana—Max Planck Institute for Art History in Rome. Because early modern paintings offer a figurative content and an interesting shift between symbolic and more natural hand gestures, we focused on Western European artworks produced between the 15th and the 17th centuries, thus harvesting a total of 5234 images.

These images are photographs of paintings taken over the past century. Many of them are in black and white, and early acquisitions correspond to documents holding both an image and written information on the bottom, as shown in Figure 1, traditional material used in the practice of art history. Not all paintings represent bodies and hands, as the collection also holds depictions of landscapes, still-lives, close ups, and portraits, as well as the backs of framed canvases.

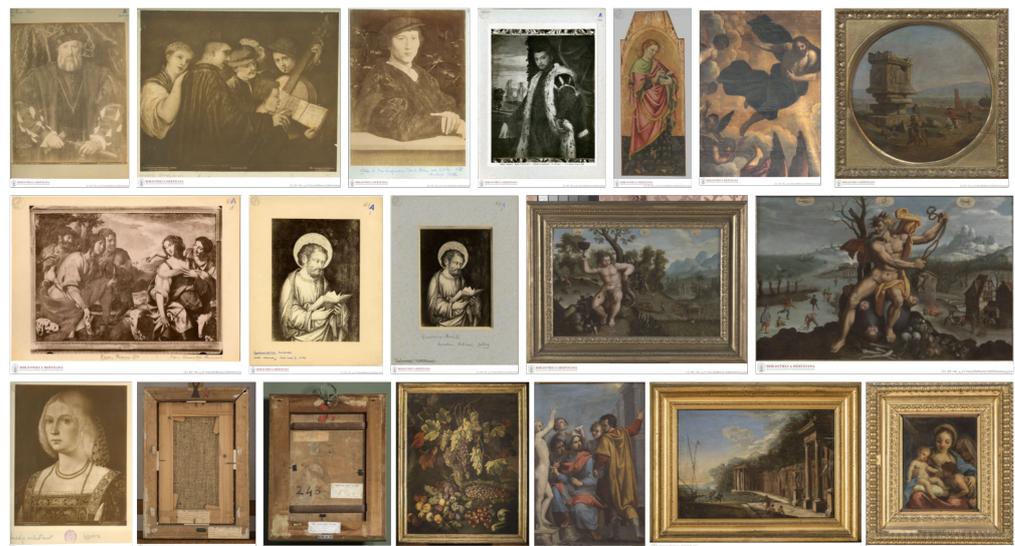


Figure 1. Example images of paintings from the source data collection, the photographic Collection of the Bibliotheca Hertziana, Max Planck Institute for Art History. Reprinted with permission from Bibliotheca Hertziana, Max Planck Institute for Art History in Rome. 2023.

The images are associated with their corresponding metadata, which include the title of the paintings, the author, the estimated year of creation, and the identification number. As the digital photographic collection was created by different art historians and the metadata were produced at different time stages, they often include information in different languages as well as date formats following different art historical conventions. Therefore the first step in the preparation of our dataset included processing, cleaning, and standardizing the available metadata.

2.1.2. Hand Extraction

The second step in the creation of the dataset was extracting hands from the collection of paintings. For this purpose, we compared several different approaches based on pretrained convolutional neural network models, such as the DensePose-RCNN [23], handCNN [51] and OpenPose [22,26]. We found that the OpenPose model yielded the best detection results on our dataset of paintings, and therefore, we used this model to create the final dataset of hands. OpenPose represents a real time multiperson system that can detect human body, hand, facial, and foot keypoints on single images. It detects a total of 135 keypoints, with 21 keypoints assigned to each hand, together with an indication referring to the left or right hand. For the purpose of creating our dataset of painted hands, we first applied the OpenPose model on artwork images for multiperson pose extraction. Each keypoint k_i was defined by its position coordinates (x, y) and a detection confidence value c :

$$k_i = (x_i, y_i, c_i), \text{ where } i \in \{0, 1, \dots, 20\} \quad (1)$$

If no hand was detected for an image, then the coordinates and score value (x, y, c) were set to 0. Out of the 5234 images given as input to the OpenPose model, one or more bodies were estimated on a total of 4256 images. However, the OpenPose model was not able to detect hand keypoints on all of the images that included depictions of bodies. For example, images presenting body depictions within architectural ornamentation, such as walls and ceilings frescos, present a challenge for hand extraction, presumably because of the distortion of the perspective. It is also very difficult to detect the body and hand keypoints in depictions of crowds and in images where the visibility of the body is obscured because the color of clothes or body parts is fused with the color of the background, usually when the image is dark and has low contrast. Figure 2 shows some examples of artwork images on which the OpenPose model failed to detect hands.



Figure 2. Sample of pictures representing an issue for automated hand detection with OpenPose. Reprinted with permission from Bibliotheca Hertziana, Max Planck Institute for Art History in Rome. 2023.

After excluding images without detected hand keypoints, we proceeded to extract the hand images from those artwork images on which the OpenPose model detected

hand keypoints for one or more hands. Based on the coordinates used for the keypoint descriptions, a bounding box was generated around the hand in order to crop the image, as illustrated in Figure 3. The bounding box was defined with the minimum and maximum coordinate values, which were first extracted from all x and y coordinate values of the different hand keypoints. From the minimum and maximum values, the height H of the hand was calculated in order to define a proportional margin m surrounding the hand. This margin was defined by the height value multiplied by a predefined ratio which was set to a fixed value of 0.3 after visually inspecting the results of different values in several experimental setups.

$$\begin{aligned} H &= y_{max} - y_{min} \\ m &= H \times 0.3 \end{aligned} \quad (2)$$

$$(Box_{min}, Box_{max}) = ((x_{min} - m, y_{min} - m), (x_{max} + m), (y_{max} + m))$$

After performing this process on all of the available images, a total of 18,641 cropped images were produced. However, as not all the resulting images actually included hands, manual cleaning of the dataset of cropped images was performed.



Figure 3. Illustration of the hand extraction process using the pretrained OpenPose model on one artwork image example.

During the first visual inspection, cropped images that did not represent a hand were removed from the dataset, which resulted in 15,124 samples. The following step then consisted of checking the quality of the keypoint estimation: the keypoint information was superimposed on each remaining image. This second visual inspection allowed us to remove hands where the keypoints were visibly outside the limits of the depicted hand, as well as left and right hand duplicates, thus resulting in a total of 5993 images of painted hands. Besides the false positives and inaccurate detections, some poses were also more complex to detect and were not recognized by the OpenPose model, which is an issue that has already been documented in related studies [25,30].

2.1.3. Data Categorization and Annotation

The categorization of hands in our dataset was based on an interpretation of the work of the art historian Dimova [3]. In her fundamental work *Le langage des mains dans l'art*, she described a total of 30 hand poses [3] (pp. 320–324), called pictorial chirograms, a notion already introduced by the father of sign language, John Bulwer, in 1644, to illustrate specific hand gestures.

In this work [3], the different hand poses are presented as a lexicon of chirograms, associated with cropped paintings of hands. Although Dimova provided one example

image for each hand pose category, each hand pose can indicate several different meanings. Therefore, the categorization of hand pose images is highly ambiguous, as the interpretation of these hands depends on both the context of the depicted scene as well as the overall context of creation of the artwork itself. Furthermore, the context of creation of the artwork, such as, for example, the sociocultural backgrounds of the artists, can be an indication of the gestural convention in place [4]. In order to correctly classify the different hands, expert-level knowledge is therefore required in the data annotation process.

Another challenge for our data labeling process based on Dimova's categorization is the scarce representativity of specific categories, such as, for example, *thumb up*, *index and auricular up*, or *hand under the chin*. Additionally, the variations between different chirograms can be very subtle. In particular, we noticed the case of the pointing finger, which most often indicates a discursive situation, but which can also refer to a higher authority when pointing up [3] (p. 320). We therefore decided to merge all chirograms representing a pointing finger into one category, the *pointing index*. Finally, after excluding all under-represented and ambiguous categories from the chirograms, we restricted our final data annotation task to nine different categories of hand poses, shown in Figure 4.



Figure 4. Nine different hand categories based on the chirograms defined by Dimova.

The images were then manually annotated by the first author of this paper. The final annotated dataset, named the Painted Hand Pose (PHP) dataset, represents 23.3% of the entire dataset of the 5993 initially extracted hand images. The distribution of the number of images into each category can be seen in Figure 5. There is significant variation within the distribution of images per category, e.g., two categories, such as the *benedictio* and the *intertwined fingers*, are associated with less than 100 samples, whereas the *pointing index* has almost 250 samples. However, although this represents a challenge for computational classification, it is a property that is inherent to both the dataset and the categorization task itself, as not all hand poses appear equally frequently in paintings from the early modern times.

The resulting PHP dataset has been made publicly available for future research (<https://doi.org/10.5281/zenodo.8069651> (accessed on 24 April 2023)). In order to analyze the potential of this dataset in the context of computational analysis of artwork images, we further explored feature extraction methods and several approaches to the automated classification of the defined hand gestures.

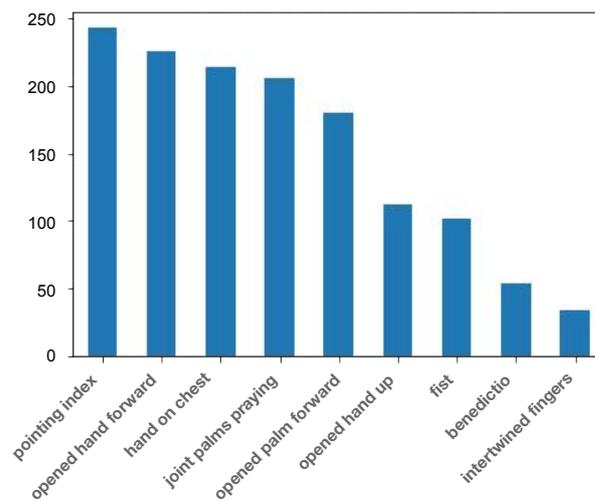


Figure 5. Category-based distribution of images in the Painted Hand Pose dataset.

2.2. Feature Descriptors

2.2.1. Hand Keypoint Features

Based on the hand keypoint coordinates extracted using the OpenPose model, we defined a new set of features for hand pose classification. Figure 6a shows the positions of the 21 hand keypoints together with their defined indices. The OpenPose model outputs a total of 21 keypoints for a single hand, denoted as $k_i, i = [0, \dots, 20]$, from which 1 keypoint corresponds to the base of the palm or the wrist, 5 correspond to the fingertips, and 16 correspond to the articulations of the fingers, also called joints.

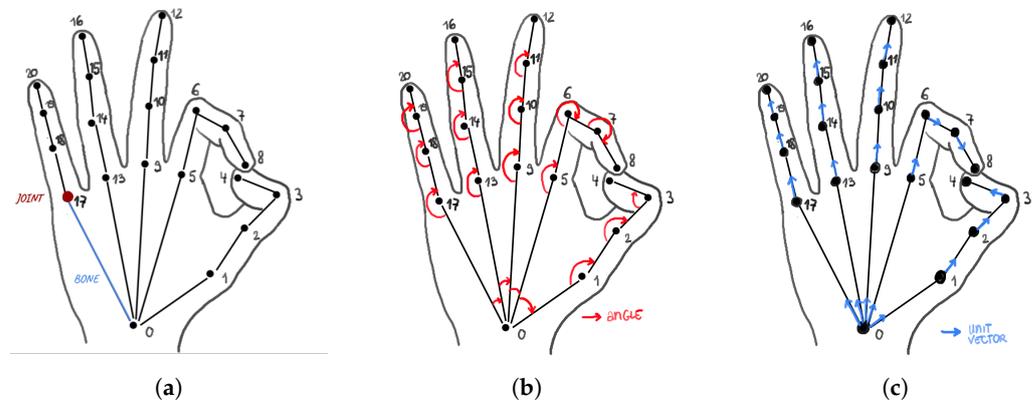


Figure 6. Hand pose representations illustrating the positions of the 21 main OpenPose hand keypoints, which serve as a basis for our hand feature descriptions based on the 19 angles and 20 unit vectors. (a) The 21 hand keypoints, (b) the 19 hand angles in red, and (c) the 20 hand unit vectors in blue.

Inspired by the work of Impett [28] on body pose comparison, we decided to characterize the hand poses by the angles shaped by the joints and use the values of the angles as feature descriptors of the hand images. Each angle was defined with three keypoints, excluding the fingertips, as shown on Figure 6b. A list of 19 possible angles $\theta_{a,b,c}$ was constructed, where a, b, c correspond to the defined keypoint indices, and the angle $\theta_{a,b,c}$ represents the angle between the central keypoint at position b and two neighbouring keypoints a and c :

$$\theta_{a,b,c} \in [\theta_{5,0,1}, \theta_{9,0,5}, \theta_{13,0,9}, \theta_{17,0,13}, \theta_{j-1,j,j+1} \mid j \in \mathbb{Z}, 1 \leq j \leq 19, j \notin \{4, 8, 12, 16\}] \quad (3)$$

As we have the coordinates of every keypoint a, b, c , we can calculate the Euclidean distances between them:

$$\begin{aligned} d_{ab} &= \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} \\ d_{ac} &= \sqrt{(x_c - x_a)^2 + (y_c - y_a)^2} \\ d_{bc} &= \sqrt{(x_c - x_b)^2 + (y_c - y_b)^2} \end{aligned} \tag{4}$$

We can then use the distance values to calculate the angle $\theta_{a,b,c}$ by applying the Law of Cosines:

$$d_{ac}^2 = d_{ab}^2 + d_{bc}^2 - 2d_{ab} \cdot d_{bc} \cdot \cos(\theta_{a,b,c}) \tag{5}$$

where then the angle $\theta_{a,b,c}$ is given by:

$$\theta_{a,b,c} = \arccos\left(\frac{(d_{ab}^2 + d_{bc}^2 - d_{ac}^2)}{(2d_{ab} \cdot d_{bc})}\right) \tag{6}$$

The calculated angle values indicate the relative positions of the fingers and therefore serve as informative feature descriptors for hand gesture recognition. However, besides the angle values between joints, hand gestures are also characterized by the absolute positions and rotation of the palm and fingers. As previously shown in Figure 4, some hand gesture categories present hands that have similar articulations of joints, but more noticeable differences emerge on the level of hand orientation, such as with the categories *hand on chest* and *praying hands*. The orientation of a finger is also instructive regarding the possible space between fingers when, for example, the fingers are spread apart or stuck together. To represent the information related to the direction and orientation of the hand and fingers, we calculated the unit vectors of segments between a pair of keypoints, denoted as *bones* in Figure 6a.

Each bone is defined by two keypoints, excluding the fingertips, as shown on Figure 6c. A list of 20 possible bones $B_{a,b}$ was constructed, where a and b correspond to the defined indices of neighboring keypoints:

$$B_{a,b} \in [B_{0,5}, B_{0,9}, B_{0,13}, B_{0,17}, B_{j,j+1} \mid j \in \mathbb{Z}, 0 \leq j \leq 19, j \notin \{4, 8, 12, 16\}] \tag{7}$$

The unit vector of the bone is represented by a tuple $(u_{x_{ab}}, u_{y_{ab}})$, where

$$\begin{aligned} u_{x_{ab}} &= \frac{x_b - x_a}{d_{ab}} \\ u_{y_{ab}} &= \frac{y_b - y_a}{d_{ab}} \end{aligned} \tag{8}$$

The tuple values of the unit vector for each bone constitutes a total of 40 unit vector features per hand image. These unit vector features are concatenated together with the 19 angle value features, forming the final 59-dimensional feature descriptor, which we refer to as keypoint features (KP features).

2.2.2. Neural-Network-Based Image Features

The hand keypoint features were exclusively based on the hand keypoint information obtained using the OpenPose model. In order to understand the descriptive potential of those features for the task of hand gesture recognition, we compared them to other image-based features. In particular, we extracted feature representations of the hand images from two different pretrained neural network models.

In the first case, we used a ResNet50 model pretrained on the well-known ImageNet dataset [52], which has become a standard in many computer vision-related tasks. The model takes a 224×224 resized image as input. We used $2048 \times 7 \times 7$ dimensional outputs

of the last sequential layer as the basis for our feature descriptor. The outputs of this layer were flattened into a vector of size 100,352, on which we performed a Principle Component Analysis (PCA) to reduce the dimensionality to 1000. We refer to this 1000-dimensional image representation as the ImageNet ResNet50 feature descriptor.

Additionally, in order to test whether we can leverage the availability of larger sign language datasets for our task, we fine-tuned the same ImageNet pretrained Resnet50 model on a dataset for sign language recognition, called the American Sign Language (ASL) dataset [53]. The dataset consists of images of hands categorized into 29 different classes: 26 different alphabetic letters from American sign language, two extra classes for hand poses indicating *space* and *delete* signs, and a nonhand class, which includes images without hands. The training dataset contains 87,000 images, where each sample is a 200×200 pixels image. The fine-tuned model requires a 64×64 input shape and the addition of sequential layers with a final softmax activation layer to output a prediction on the 29 possible classes. We fine-tuned the model using the categorical cross entropy loss function and the Adam optimizer for 50 epochs, with a 128 batch size and a learning rate of 0.0001. The model reached 84% accuracy on the American Sign Language (ASL) dataset. After fine-tuning this model, we utilized it to extract features from our Painted Hand Pose dataset. We used the outputs of the last sequential layer, which form a 512-dimensional vector, to refer to this descriptor as the ASL ResNet50FT (fine-tuned) features.

2.3. Classification Settings

The classification of the hands was performed on each type of feature and with different machine learning methods. In total, we compared the results of three different classifiers: a Multilayer Perceptron (MLP), a Support Vector Machine (SVM), and a K-Nearest Neighbor (KNN) classifier. For the SVM, a Radial Basis Function (RBF) kernel was used. The parameters of the MLP setup included a 200 batch size, with a maximum of 10,000 iterations, and a 0.001 learning rate. A ReLu activation layer and a limited-memory BFGS (lbfgs) solver for weight optimization were used. In order to train the different classification models and due to the small size of the dataset, a five-fold cross-validation method was used with an 80% and 20% split between the training and testing datasets over each fold. The reported accuracy of the selected models corresponds to the mean of the five scores obtained for each fold.

3. Results

3.1. Exploratory Analysis

In order to gain a better understanding of the dataset and the defined class categories, we performed an exploratory analysis which included visualizing the separability of different classes within the dataset. We employed the well-known dimensionality reduction method t-distributed Stochastic Neighbor Embedding (t-SNE) to the KP features extracted from each image in our dataset. The visualization presented in Figure 7 shows each image represented with a 2D point, where the color of the point marks the class category and the shape of the point (circle or cross mark) indicates if it is a right or left hand image. The indication of the right or left hand is based on the OpenPose model keypoint prediction.

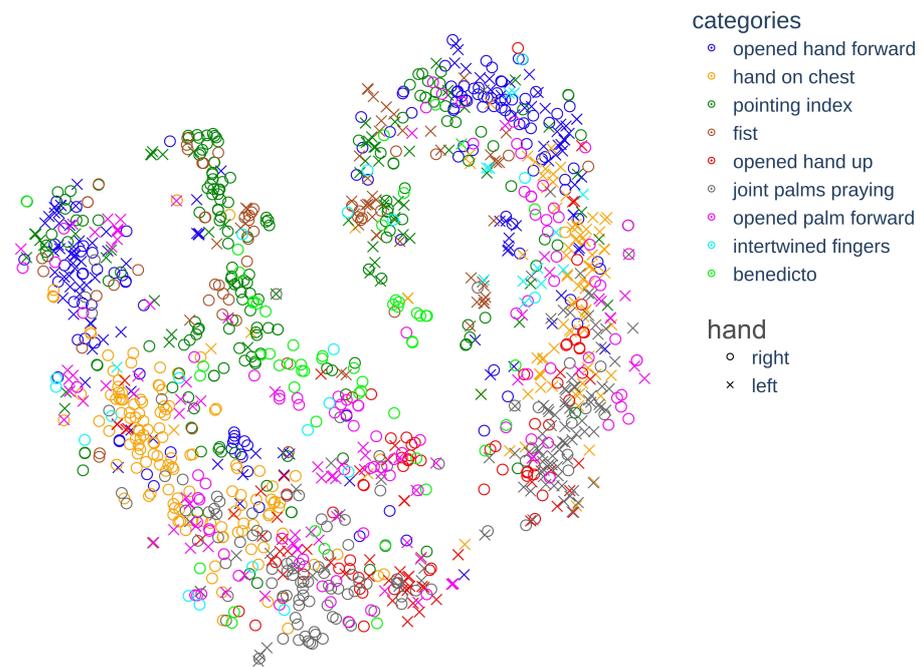


Figure 7. T-SNE projection of the KP features and indication of whether the left or right hand was used.

The overlapping clusters in the t-SNE visualization indicate that certain classes possess similarities, such as the *pointing index* and the *fist*, as well as the *joint palms praying*, the *hand on chest*, and the *opened hand up*.

As well as gaining insight about the underlying data structure in regard to the historically defined art categories, our exploratory analysis also focused on general hand categorization aspects, such the relation of the left and right hands. Based on the indication of the left or right hand keypoints, we found that 58% of the hand images in our dataset were labeled as right hands. A more detailed analysis of the distribution of left and right hands within each class category is shown in Figure 8.

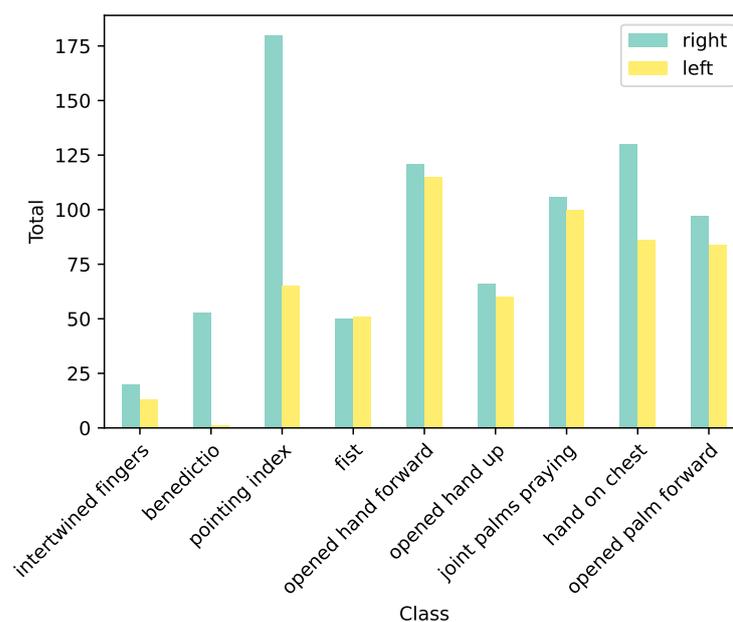


Figure 8. Left and right hand distribution among the categories.

As we can see from the results, all gestures, especially the *benedictio*, the *pointing index*, and the *hand on chest*, tended to be mostly performed by the right hand. This tendency for right-handedness in conventional gestures found in early modern paintings correlates with contemporary research on the perception of the right hand as being the dominant hand [54,55]. Furthermore, the reason for the right hand prevalence is also strongly related to the different symbolic connotations of the right and left hands in the early modern period. Indeed, the *benedictio* was only performed by the right hand, and our findings align with historical art observations on the long-standing gestural convention and the liturgical practices from which it derives [56,57]. However, it is important to note that the accuracy of determining whether a hand is left or right using the OpenPose model can vary significantly, especially in the challenging scenarios that emerged in our dataset. Therefore, this exploratory analysis represents only an initial indication and further research is necessary to derive the quantitative relation between left and right hands in early modern paintings with more certainty.

3.2. Classification Results

The comparison of different feature descriptors based on three different classification methods is shown in Figure 9. The results indicate that our KP features, with a 0.6 accuracy mean obtained using the MLP classifier, outperform the results of the other neural-network-based image features. We also compared the accuracy difference when using all KP features and when using only the KP features related to the angle values. We obtained a significantly lower accuracy when using only the KP angle features, which outlines the importance of the unit vector features for the description of the painted hand gesture as well as confirming the need to encode the direction of the palm and fingers for historical art hand gesture classification.

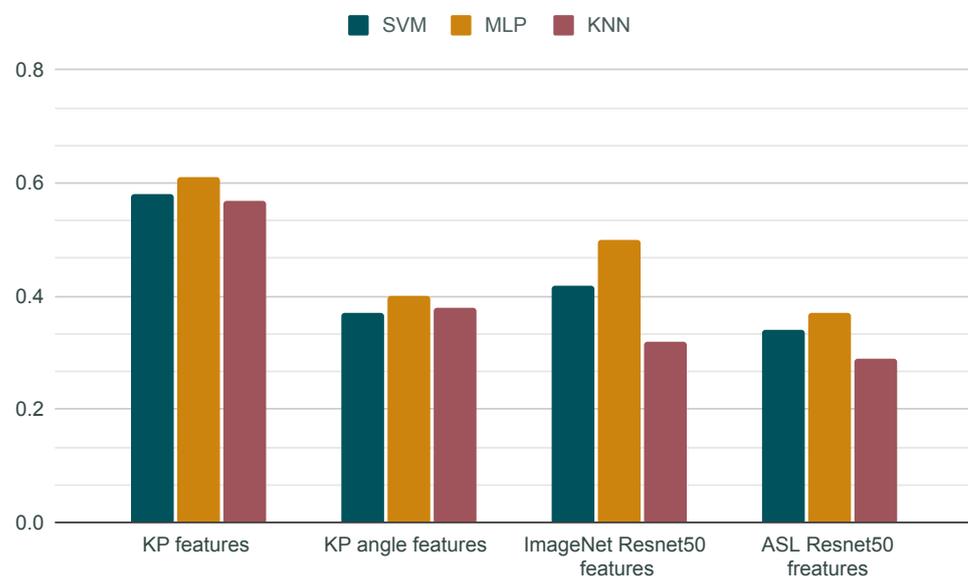


Figure 9. Classification accuracy of different classifiers and features.

The 50% accuracy achieved using the MLP classifier on the ImageNet ResNet50 features also indicates a promising result, particularly considering that the model was trained to classify images from very diverse object categories. These features also significantly outperform features extracted from the ResNet50 model fine-tuned on the sign language datasets, with which we achieved an accuracy of only 37% using the MLP classifier. Our initial intuition that we might achieve better results if we fine-tuned a general object recognition model on a sign language recognition task using a specific dataset of hands was therefore proven wrong. As well as the fact that the dataset of photographic sign language

hand images is very different to our dataset of painted hands, we also assume that the model trained for the very specific task of sign language classification cannot generalize well to different types of hand poses of early modern paintings because those are, in the end, fundamentally different tasks. This outcome also indicates, once again, the very intricate nature of the computational approach towards painted hand gesture classification and demonstrates the limits of leveraging existing pretrained models and available large datasets of real hand images.

In order to better understand the challenges of painted hand gesture classification, we analyzed the per-class classification performance. The detailed per-class classification results of the MLP are presented with the confusion matrix shown in Figure 10.

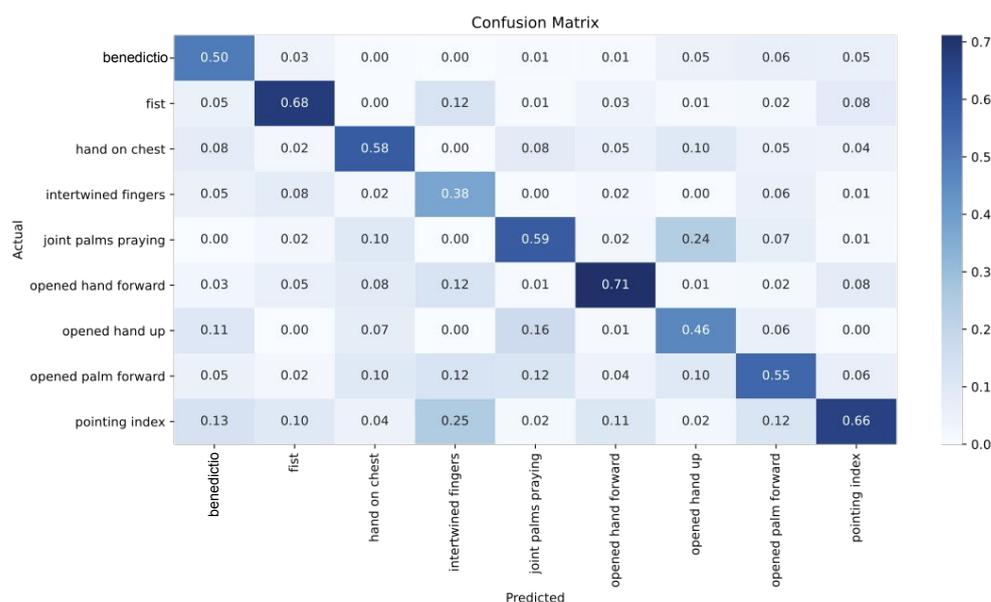


Figure 10. Confusion matrix from the MLP classifier on KP features.

The presented results indicate that the best overall results were achieved for the *fist*, the *opened hand forward*, the *joint palms praying*, and the *pointing index* classes. The confusion matrix demonstrates that classes with the largest available number of images in the training set also achieved the highest per-class accuracy scores. This observation was additionally confirmed by the comparison of the F1 score of each set of features with the number of test set images available in each class, as expressed in Figure 11. We found a tendency for better F1 scores based on the number of images in the test set, with the exception of the *fist* class. The high F1 score for the *fist* class reflects the obvious distinctiveness of this hand gesture in comparison to other gestures, mainly as it usually includes a hand with all fingers folded towards the palm.

Additionally, the comparison of the F1 scores of all the different feature types indicates the strength of each feature type on the different classes. For example, the simplified KP features, which included only the angle value information, successfully described the *fist* and *pointing index*, but generated low results for all other categories. The ImageNet ResNet50 embeddings also showed good descriptive features for the *opened hand up*, and the ASL fine-tuned version offered a slightly better representation for *intertwined fingers*. Overall, it seems that the KP features performed the best for the majority of hand pose classes.

As well as indicating the well-known problem that occurs when classifying datasets with imbalanced numbers of images per class, the results also demonstrate the impact of a low interclass visual variation. For example, the *benedictio* gesture was often misclassified as *pointing index* and *opened palm forward*, while the *hand on chest* was most commonly misclassified as the *praying gesture* and vice versa. Figure 12 illustrates the complexity of the

hand gesture classification tasks, as the differences between hands belonging to different categories are often very subtle. Our in-depth review of the misclassified samples also showed that an issue often emerges when hands present an interplay with an object or with another hand.

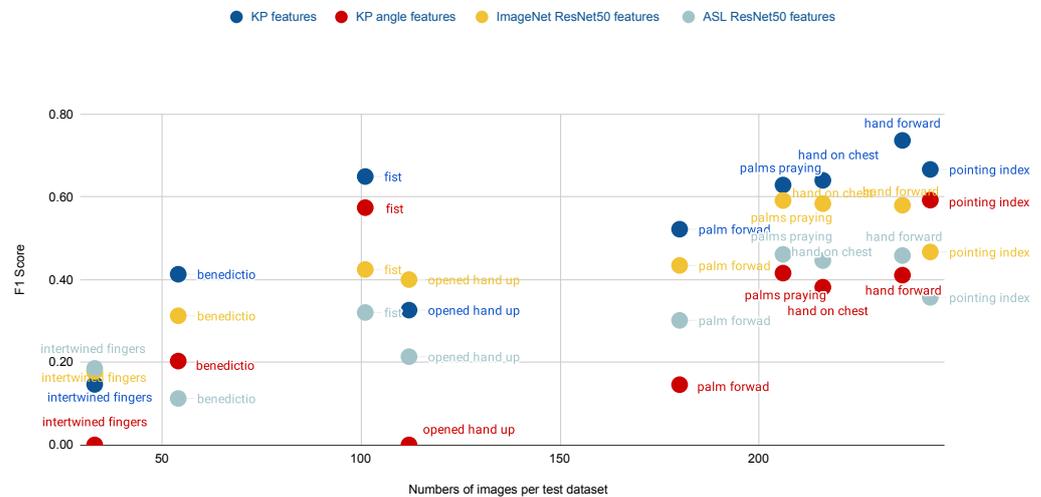


Figure 11. The F1 score value in relation to the number of images in the per-class test set and the different feature types obtained when using the MLP classifier.

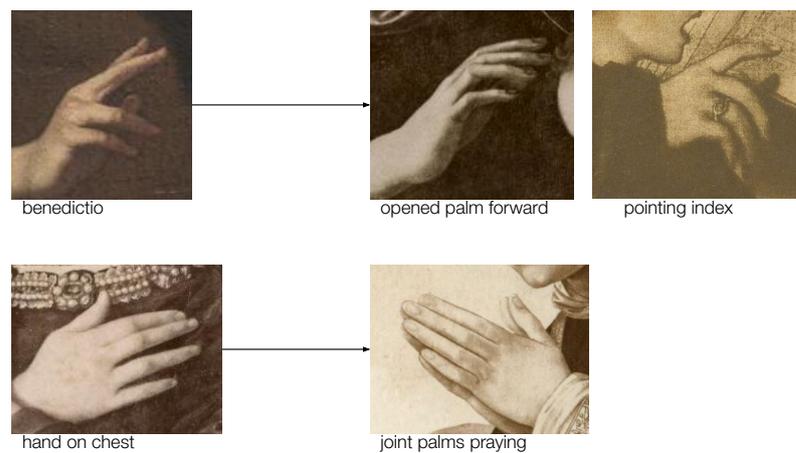


Figure 12. Example images that belong to different classes but depict hands in similar poses.

3.3. Application for Gesture-Based Image Search

Finally, it is important to indicate that our dataset and the defined hand keypoint features can be utilized beyond the scope of the presented classification task. More specifically, the defined hand keypoint features serve as a basis for the gesture-based image search engine Gestures for Artworks Browsing (GAB), which is presented and described in more detail in [29]. This application represents an unique framework in the context of digital art history, as it enables the exploration of collections of digitized artworks without using the conventional methods of word- or image-based searches. Figure 13 illustrates the main functionalities of GAB. The code of the GAB implementation is publicly available (https://github.com/VBernasconi/GAB_project (accessed on 22 March 2022)).

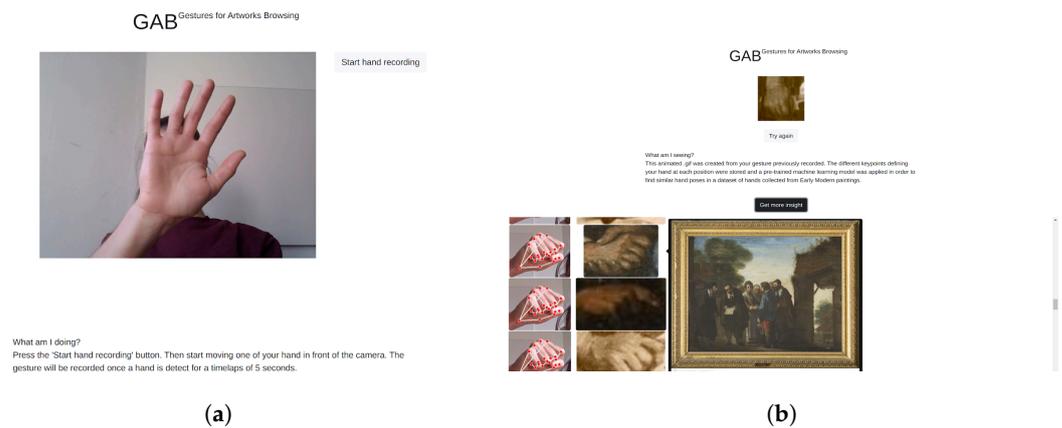


Figure 13. Gestures for Artworks Browsing application. (a) Input: real-time recording of the hand, (b) Output: similarly painted hand images and their source artworks. Screenshot of the Gestures for Artworks Browsing application.

As shown in Figure 13, the hand pose of the user is used to retrieve similar hands in the database of paintings and access their corresponding original images. The implementation of the browsing tool uses the KP features and the k-NN algorithm to find the most similar painted hands in relation to the image taken from the user’s hand. This application demonstrates the efficiency of the KP features as a pose descriptor in a context-independent environment. Furthermore, KP features have also been used for real-time interactions in the digital work of art called *La main baladeuse* [58] presented in the digital exhibition of the museum Le Jeu de Paume from May to December 2022. The concept of the work was based on the simultaneous interplay of similarly drawn and painted hands from different time periods and the continuous gestural hand expression of the user facing the camera, as illustrated in Figure 14.



Figure 14. Interactive interface from the art installation *La main baladeuse*, with the hand of the user represented as a skeleton in the center.

An application such as the Gesture-Based Browsing Interface can be used in a research context to easily access specific painted hands, but it can also be used as a novel and unconventional way of exploring art collections in museum spaces, offering a new experience of the artworks. Combined with a classification model trained for the recognition of pictorial early modern hand gestures, the tool could potentially become a didactic instrument, em-

ployed to better understand the diversity of depicted hands and their various connotations in early modern paintings.

4. Discussion

To summarize, in this paper, we introduced a new dataset of hand images extracted from paintings from the European early modern period, named the Painted Hand Pose dataset. We presented, in detail, the various challenging steps in the dataset creation process, which highlight the complexity of working with digitized collections of artworks. The dataset creation process included significant manual cleaning effort as well as laborious categorization and annotation work that required historical art expertise.

As well as introducing a novel dataset, we proposed a new feature extraction method based on the use of deep learning methods for human pose estimation. More specifically, we focused on the hand keypoint coordinates obtained using the OpenPose model. These keypoints were used as the basis for our hand keypoint feature descriptor which integrates information about the absolute and relative positions of the palm and fingers by combining the angle values between the different joints in the hand and unit vector values of the different bones of the hand.

Those features serve as the basis for additional computational exploration. For this purpose, in the context of our paper, we introduced a new challenging and interdisciplinary classification task. The work of the art historian Dimova on the categorization of hand gestures in paintings served as a basis for defining our classification task and identifying the different hand pose classes. In this paper, we propose a classification setting in order to better understand the complexity of hand pose classification in the context of artworks as well as to compare our keypoint features with other neural-network-based image features. The features created from the HPE 2D keypoint information showed promising results for this classification task and demonstrated the importance of the angles shaped by the articulation of the fingers and the orientation of the hand as descriptors.

The presented classification results primarily serve to illustrate the complexity of adopting a computational approach towards hand pose recognition in early modern paintings. There are various possibilities for improving the classification results that might be explored in the future, such as exploring different feature ensemble techniques or including features extracted from deep neural network models fine-tuned on different tasks. However, the goal of this work was not to find the best possible classification method but to introduce a novel and challenging interdisciplinary task. The subtle differences between hand poses belonging to different historical art gestural categories outline the importance of integrating additional contextual information. Therefore, another possible future extension of this work includes combining hand pose image features with other global image scene features or existing metadata information (e.g., the title of the painting) in a multimodal setting.

Possible future improvements of the classification results also include the augmentation of the dataset, although this would require manual human effort and expert-level knowledge on hand gestures in the early modern times. There is, however, the possibility of augmenting the existing dataset with synthetically generated images by using generative models such as variational autoencoders, generative adversarial networks, or diffusion models. As well as the issue of homogenization, where not enough diverse data are produced, it seems that hands also represent a great challenge for contemporary text-to-image generative diffusion models, as reported in various media [59,60]. These examples also highlight our own conclusion that hand gestures represent a particularly difficult subject for computational image understanding, even for the most advanced contemporary deep neural network models.

Finally, our newly introduced classification task represents an interesting challenge in the context of the emerging discipline of digital art history. It also contributes to existing challenges in the field of computer vision and hand gesture recognition. Our research represents the first step in the direction of computational understanding of hand poses

in paintings. It will, therefore, hopefully foster not only new methods in the context of computer vision, but also new research work on art history that will lead to a better understanding of the language of hands in art, the comprehension of various gestural patterns, and their evolution in early modern times.

Author Contributions: Conceptualization, V.B. and L.I.; methodology, V.B.; validation, V.B., L.I. and E.C.; formal analysis, V.B.; investigation, V.B.; data curation, V.B.; visualization, V.B.; writing—original draft preparation, V.B. and E.C.; writing—review and editing, E.C. and L.I.; supervision, L.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Painted Hand Pose dataset is publicly available at <https://doi.org/10.5281/zenodo.8069651> (accessed on 24 April 2023).

Acknowledgments: We would like to thank the Bibliotheca Herziana—Max Planck Institute for Art History in Rome for providing the digital photographic collection for the purpose of scientific research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schmitt, J.C. *La Raison des Gestes Dans L'Occident Médiéval*; Editions Gallimard: Paris, France, 1990.
2. Wittkower, R. *La Migration des Symboles*; Iconologia, Thames & Hudson: Paris, France, 1992.
3. Dimova, T. *Le Langage des Mains Dans L'art: Histoire, Significations et Usages des Chirogrammes Picturaux aux XVIIe et XVIIIe Siecles*; Brepols Publishers: Turnhout, Belgium, 2020.
4. Spicer, J. The Renaissance elbow. In *A Cultural History of Gesture. From Antiquity to the Present Day*; Bremmer, J., Roodenburg, H., Eds.; Polity Press: Cambridge, UK, 1991.
5. Agarwal, S.; Karnick, H.; Pant, N.; Patel, U. Genre and Style Based Painting Classification. In Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 5–9 January 2015; pp. 588–594. [CrossRef]
6. Arora, R.S.; Elgammal, A. Towards automated classification of fine-art painting style: A comparative study. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan, 11–15 November 2012; pp. 3541–3544.
7. Cetinic, E.; Lipic, T.; Grgic, S. Fine-tuning convolutional neural networks for fine art classification. *Expert Syst. Appl.* **2018**, *114*, 107–118. [CrossRef]
8. Tan, W.R.; Chan, C.S.; Aguirre, H.E.; Tanaka, K. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3703–3707. [CrossRef]
9. Seguin, B.; Striolo, C.; diLenardo, I.; Kaplan, F. Visual Link Retrieval in a Database of Paintings. In Proceedings of the Computer Vision—ECCV 2016 Workshops, Amsterdam, The Netherlands, 11–14 October 2016; Hua, G., Jégou, H., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; pp. 753–767. [CrossRef]
10. Shen, X.; Efros, A.A.; Aubry, M. Discovering Visual Patterns in Art Collections With Spatially-Consistent Feature Learning. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9270–9279. [CrossRef]
11. Ufer, N.; Simon, M.; Lang, S.; Ommer, B. Large-scale interactive retrieval in art collections using multi-style feature aggregation. *PLoS ONE* **2021**, *16*, e0259718. [CrossRef]
12. Shen, X.; Champenois, R.; Ginosar, S.; Pastrolin, I.; Rousselot, M.; Bounou, O.; Monnier, T.; Gidaris, S.; Bougard, F.; Raverdy, P.G.; et al. Spatially-consistent Feature Matching and Learning for Heritage Image Analysis. *Int. J. Comput. Vis.* **2022**, *130*, 1325–1339. [CrossRef]
13. Bell, P.; Schlecht, J.; Ommer, B. Nonverbal Communication in Medieval Illustrations Revisited by Computer Vision and Art History. *Vis. Resour.* **2013**, *29*, 26–37. [CrossRef]
14. Thomas, C.; Kovashka, A. Artistic Object Recognition by Unsupervised Style Adaptation. In *Computer Vision—ACCV 2018*; Lecture Notes in Computer Science; Jawahar, C.V., Li, H., Mori, G., Schindler, K., Eds.; Springer International Publishing: Cham, Switzerland, 2019; Volume 11363, pp. 460–476. [CrossRef]
15. Yin, R.; Monson, E.; Honig, E.; Daubechies, I.; Maggioni, M. Object recognition in art drawings: Transfer of a neural network. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 2299–2303. [CrossRef]
16. Smirnov, S.; Eguizabal, A. Deep learning for object detection in fine-art paintings. In Proceedings of the 2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo), Cassino, Italy, 22–24 October 2018; pp. 45–49. [CrossRef]

17. Lin, H.; Van Zuijlen, M.; Wijntjes, M.W.A.; Pont, S.C.; Bala, K. Insights from a Large-Scale Database of Material Depictions in Paintings. *arXiv* **2020**. [[CrossRef](#)]
18. Impett, L.; Süssstrunk, S. Pose and Pathosformel in Aby Warburg's Bilderatlas. In Proceedings of the Computer Vision—ECCV 2016 Workshops, Amsterdam, The Netherlands, 8–10 October 2016; Hua, G., Jégou, H., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; pp. 888–902. [[CrossRef](#)]
19. Marsocci, V.; Lastilla, L. POSE-ID-on—A Novel Framework for Artwork Pose Clustering. *ISPRS Int. J.-Geo-Inf.* **2021**, *10*, 257. [[CrossRef](#)]
20. Madhu, P.; Villar-Corrales, A.; Kosti, R.; Bendschus, T.; Reinhardt, C.; Bell, P.; Maier, A.; Christlein, V. Enhancing Human Pose Estimation in Ancient Vase Paintings via Perceptually-grounded Style Transfer Learning. *J. Comput. Cult. Herit.* **2022**, *16*, 1–17. [[CrossRef](#)]
21. Ohrt, R.; Ohrt, R. *Aby Warburg: Bilderatlas Mnemosyne: The Original*; Kulturgeschichte; Hatje Cantz Verlag: Berlin, Germany, 2020.
22. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [[CrossRef](#)]
23. Guler, R.A.; Neverova, N.; Kokkinos, I. DensePose: Dense Human Pose Estimation in the Wild. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7297–7306. [[CrossRef](#)]
24. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In Proceedings of the Computer Vision—ECCV, Zurich, Switzerland, 6–12 September 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; pp. 740–755. [[CrossRef](#)]
25. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3686–3693. [[CrossRef](#)]
26. Simon, T.; Joo, H.; Matthews, I.; Sheikh, Y. Hand Keypoint Detection in Single Images using Multiview Bootstrapping. In Proceedings of the CVPR, Honolulu, HI, USA, 21–26 July 2017.
27. Impett, L.; Bell, P. Ikonographie Und Interaktion. Computergestützte Analyse von Posen in Bildern der Heilsgeschichte. *Das Mittelalt.* **2019**, *24*, 31–53. [[CrossRef](#)]
28. Impett, L. Analyzing Gesture in Digital Art History. In *The Routledge Companion to Digital Humanities and Art History*; Routledge: London, UK, 2020; pp. 386–407.
29. Bernasconi, V. GAB—Gestures for Artworks Browsing. In Proceedings of the 27th International Conference on Intelligent User Interfaces, Online, 22–25 March 2022; Association for Computing Machinery: New York, NY, USA, 2022; IUI '22 Companion; pp. 50–53. [[CrossRef](#)]
30. Springstein, M.; Schneider, S.; Althaus, C.; Ewerth, R. Semi-Supervised Human Pose Estimation in Art-Historical Images. In Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 10–14 October 2022; Association for Computing Machinery: New York, NY, USA, 2022; MM '22; pp. 1107–1116. [[CrossRef](#)]
31. Jenicek, T.; Chum, O. Linking Art through Human Poses. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, Australia, 20–25 September 2019; pp. 1338–1345.
32. Zhao, S.; Akdağ Salah, A.; Salah, A.A. Automatic Analysis of Human Body Representations in Western Art. In Proceedings of the Computer Vision—ECCV 2022 Workshops, Tel Aviv, Israel, 23–27 October 2022; Proceedings, Part I; Springer: Berlin/Heidelberg, Germany, 2023; pp. 282–297.
33. Milani, F.; Fraternali, P. A Dataset and a Convolutional Model for Iconography Classification in Paintings. *J. Comput. Cult. Herit.* **2021**, *14*, 1–18. [[CrossRef](#)]
34. Cetinic, E. Towards Generating and Evaluating Iconographic Image Captions of Artworks. *J. Imaging* **2021**, *7*, 123. [[CrossRef](#)] [[PubMed](#)]
35. Carfi, A.; Mastrogiovanni, F. Gesture-Based Human–Machine Interaction: Taxonomy, Problem Definition, and Analysis. *IEEE Trans. Cybern.* **2023**, *53*, 497–513. [[CrossRef](#)] [[PubMed](#)]
36. Pisharady, P.K.; Saerbeck, M. Recent methods and databases in vision-based hand gesture recognition: A review. *Comput. Vis. Image Underst.* **2015**, *141*, 152–165. [[CrossRef](#)]
37. Chakraborty, B.K.; Sarma, D.; Bhuyan, M.; MacDorman, K.F. Review of constraints on vision-based gesture recognition for human–computer interaction. *IET Computer Vision* **2018**, *12*, 3–15. [[CrossRef](#)]
38. Oudah, M.; Al-Naji, A.; Chahl, J. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. *J. Imaging* **2020**, *6*, 73. [[CrossRef](#)]
39. Ahmed, S.; Kallu, K.D.; Ahmed, S.; Cho, S.H. Hand Gestures Recognition Using Radar Sensors for Human–Computer-Interaction: A Review. *Remote Sens.* **2021**, *13*, 527. [[CrossRef](#)]
40. Zhang, F.; Bazarevsky, V.; Vaku nov, A.; Tkachenka, A.; Sung, G.; Chang, C.L.; Grundmann, M. MediaPipe Hands: On-device Real-time Hand Tracking. *arXiv* **2020**. [[CrossRef](#)]
41. M, S.; Rakesh, S.; Gupta, S.; Biswas, S.; Das, P.P. Real-time hands-free immersive image navigation system using Microsoft Kinect 2.0 and Leap Motion Controller. In Proceedings of the 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna, Bihar, 16–19 December 2015; pp. 1–4. [[CrossRef](#)]

42. Ren, Z.; Yuan, J.; Meng, J.; Zhang, Z. Robust Part-Based Hand Gesture Recognition Using Kinect Sensor. *IEEE Trans. Multimed.* **2013**, *15*, 1110–1120. [[CrossRef](#)]
43. Marin, G.; Dominio, F.; Zanuttigh, P. Hand gesture recognition with leap motion and kinect devices. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 1565–1569. [[CrossRef](#)]
44. Núñez, J.C.; Cabido, R.; Pantrigo, J.J.; Montemayor, A.S.; Vélez, J.F. Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition. *Pattern Recognit.* **2018**, *76*, 80–94. [[CrossRef](#)]
45. Köpüklü, O.; Gunduz, A.; Kose, N.; Rigoll, G. Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–8. [[CrossRef](#)]
46. Sung, G.; Sokal, K.; Uboweja, E.; Bazarevsky, V.; Baccash, J.; Bazavan, E.G.; Chang, C.L.; Grundmann, M. On-device Real-time Hand Gesture Recognition. *arXiv* **2021**. [[CrossRef](#)]
47. Rastgoo, R.; Kiani, K.; Escalera, S. Sign Language Recognition: A Deep Survey. *Expert Syst. Appl.* **2021**, *164*, 113794. [[CrossRef](#)]
48. Cheok, M.J.; Omar, Z.; Jaward, M.H. A review of hand gesture and sign language recognition techniques. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 131–153. [[CrossRef](#)]
49. Kumar, M.; Gupta, P.; Jha, R.K.; Bhatia, A.; Jha, K.; Shah, B.K. Sign Language Alphabet Recognition Using Convolution Neural Network. In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 6–8 May 2021; pp. 1859–1865. [[CrossRef](#)]
50. Shin, J.; Matsuoka, A.; Hasan, M.A.M.; Srizon, A.Y. American Sign Language Alphabet Recognition by Extracting Feature from Hand Pose Estimation. *Sensors* **2021**, *21*, 5856. [[CrossRef](#)] [[PubMed](#)]
51. Zhang, X.; Huang, H.; Tan, J.; Xu, H.; Yang, C.; Peng, G.; Wang, L.; Liu, J. Hand Image Understanding via Deep Multi-Task Learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 11281–11292.
52. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [[CrossRef](#)]
53. Nagaraj, A. ASL Alphabet. 2018. Available online: <https://www.kaggle.com/datasets/grassknoted/asl-alphabet> (accessed on 21 February 2023)
54. Lucafò, C.; Marzoli, D.; Zdybek, P.; Malatesta, G.; Smerilli, F.; Ferrara, C.; Tommasi, L. The Bias toward the Right Side of Others Is Stronger for Hands than for Feet. *Symmetry* **2021**, *13*, 146. [[CrossRef](#)]
55. Marzoli, D.; Lucafò, C.; Pagliara, A.; Cappuccio, R.; Brancucci, A.; Tommasi, L. Both right- and left-handers show a bias to attend others' right arm. *Exp. Brain Res.* **2015**, *233*, 415–424. [[CrossRef](#)]
56. Hertz, R. La prééminence de la main droite: Étude sur la polarité religieuse. *Revue Philosophique de la France et de L'Étranger* **1909**, *68*, 553–580.
57. Barasch, M. *Giotto and the Language of Gesture*; Cambridge studies in the history of art; University Press: Cambridge, UK, 1987.
58. Bernasconi, V. La main baladeuse. *Jeu de Paume en ligne* 2022. as part of the online exhibition Contagions visuelles. Available online: <https://jdp.visualcontagions.net/nautilus> (accessed on 7 June 2022).
59. Hughes, A. Why AI-generated hands are the stuff of nightmares, explained by a scientist. *BBC Science Focus Magazine*, 4 February 2023.
60. Chayka, K. The Uncanny Failure of A.I.-Generated Hands. *The New Yorker*, 10 March 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.