

## Article

# A Systematic Approach for Developing a Robust Artwork Recognition Framework Using Smartphone Cameras

Zenonas Theodosiou <sup>1</sup>, Marios Thoma <sup>1</sup>, Harris Partaourides <sup>1</sup> and Andreas Lanitis <sup>1,2,\*</sup><sup>1</sup> CYENS Center of Excellence, Nicosia 1016, Cyprus<sup>2</sup> Visual Media Computing Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Limassol 3036, Cyprus

\* Correspondence: andreas.lanitis@cut.ac.cy

**Abstract:** The provision of information encourages people to visit cultural sites more often. Exploiting the great potential of using smartphone cameras and egocentric vision, we describe the development of a robust artwork recognition algorithm to assist users when visiting an art space. The algorithm recognizes artworks under any physical museum conditions, as well as camera point of views, making it suitable for different use scenarios towards an enhanced visiting experience. The algorithm was developed following a multiphase approach, including requirements gathering, experimentation in a virtual environment, development of the algorithm in real environment conditions, implementation of a demonstration smartphone app for artwork recognition and provision of assistive information, and its evaluation. During the algorithm development process, a convolutional neural network (CNN) model was trained for automatic artwork recognition using data collected in an art gallery, followed by extensive evaluations related to the parameters that may affect recognition accuracy, while the optimized algorithm was also evaluated through a dedicated app by a group of volunteers with promising results. The overall algorithm design and evaluation adopted for this work can also be applied in numerous applications, especially in cases where the algorithm performance under varying conditions and end-user satisfaction are critical factors.

**Keywords:** egocentric vision; artwork recognition; visiting experience; smartphone app; deep learning



**Citation:** Theodosiou, Z.; Thoma, M.; Partaourides, H.; Lanitis, A. A Systematic Approach for Developing a Robust Artwork Recognition Framework Using Smartphone Cameras. *Algorithms* **2022**, *15*, 305. <https://doi.org/10.3390/a15090305>

Academic Editor: Frank Werner

Received: 12 July 2022

Accepted: 18 August 2022

Published: 27 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

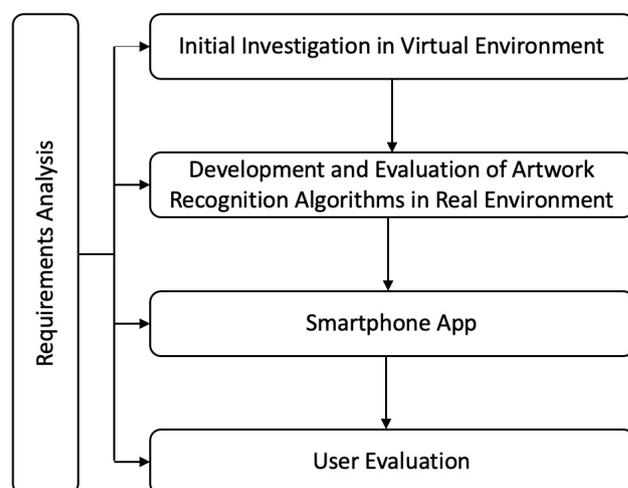
The new era of museums aims to become more interactive, informative and mesmerizing, providing the visitor with an integrated experience that maximizes both knowledge acquisition and entertainment. Emerging technologies play a significant role towards this direction, numbering several applications for analyzing visitors' activities, behaviors and experiences and understanding what is happening in museums. Furthermore, information such as which exhibitions and exhibits visitors attend to, which exhibits visitors spend most time at, and what visitors do when they are in front of each exhibit [1], are very important for curators and museum professionals.

Smartphone apps offer an easy and inexpensive way of handling information, which mainly vary in the type of information content, the way they affect the experience and interact with the user [2]. The majority of the apps developed to service museum visitors are free of use, offer basic types of content in various languages and provide navigation information [3]. The use of smartphone cameras offers considerable potential for knowledge extraction through egocentric vision techniques [4], and hence enables many applications that would not be possible with fixed cameras. However, the framework conditions in which visitors use the apps affect their efficiency and identification accuracy.

In this paper, we present a case study of artwork recognition at the State Gallery of Contemporary Cypriot Art, as part of an effort to utilize mobile smartphones for enhancing the experience of visitors in museums and art galleries. The State Gallery permanently

displays artworks, including paintings, sculptures, ceramics and multimedia art, from the 19th and 20th centuries created by Cypriot artists. The work involves the recognition of the State Gallery's artworks from the visitor's point of view, thus enabling: (a) the provision of additional information about the artwork to the visitor, and (b) the registration of the artworks that the visitor is paying more attention to as a means of automating the process of visitor experience evaluation studies [5].

Going one step further from previous research efforts in egocentric artwork recognition, we concentrated on the development of a robust recognition algorithm able to cover the needs of an art gallery and enhance the overall visiting experience. For this purpose, a systematic approach was followed, which was subdivided into five phases: (i) the collection of requirements through on-site visits and discussions with curators and museum professionals, (ii) an investigation in a virtual environment, (iii) the development of the artwork recognition algorithm and its testing in real setting, (iv) the design and implementation of a smartphone app for testing the developed algorithm and (v) an evaluation concerning both the user experience and recognition accuracy (Figure 1).



**Figure 1.** The systematic approach we followed for the robust artwork recognition.

The artwork recognition in a museum was initially tested in a virtual environment as a proof of concept. Following the preliminary investigation in the virtual environment, an artwork identification algorithm was developed. To train the algorithm, a first-person dataset was created based on the artworks of a specific Gallery's hall, where the rear camera of a midrange smartphone placed on the chest of the wearer with a forward-facing direction to replicate the point of view of a typical visitor of the Gallery. A series of quantitative performance evaluation experiments were run in order to assess the performance of the artwork recognition algorithms. The most efficient recognition model was embedded in a smartphone app that automatically recognized and displayed information about the artwork that the visitor was looking at. The evaluation involved 33 volunteers who tested the app and answered an online questionnaire regarding user experience.

The hierarchical algorithm design and evaluation pipeline adopted in this work, aimed to satisfy the need of implementing a highly accurate algorithm that fulfilled the expectations of the end users. The overall algorithm design and optimization pipeline can also be applied in numerous cases of algorithm formulation, especially in cases where the algorithm performance under varying conditions and the satisfaction of end users are critical factors to the algorithm development process.

The rest of the paper is as follows: In Section 2, a literature review is given, while Section 3 presents the preliminary experimentation in a virtual environment. Section 4 depicts the artwork recognition in a real environment and Section 5 presents the smartphone app and its evaluation regarding the user experience and recognition accuracy. Finally, conclusions and future hints are summarized in Section 6.

## 2. Literature Review

Finding methods to create a pleasant visit has been of great concern to the research community and museum professionals during the last decades. Several studies for enhancing and analyzing the museum experience have been conducted [6]. Professional guides, audiovisual digital material, tablets and smartphones, as well as QR codes have been used to assist the visitors, while questionnaires and interviews, evidence-based feedback, GPS devices and analysis in virtual environments have been studied to get visitors' feedback. The use of these methods has produced fruitful results but is less effective when it is applied on a large number of visitors.

Tesoriero et al. [7] developed mobile technologies aimed at improving museum visitor satisfaction using PDAs and the existing infrastructure of the Cutlery Museum of Albacete such as infrared sensors and radio frequency identification (RFID) tags. Lanir et al. [1] presented an automatic tracking system based on RFID technology for studying the visitors' behavior at the Hecht Museum in Israel. Similarly, Rashed et al. used light detection and ranging (LIDAR) to track and analyze visitor behavior at the Ohara Museum of Modern Arts, Japan [8]. Although these technologies gave useful insights into how people behaved in a museum environment, they required equipment and device installation in both museum space and visitors. More recently, the capabilities of deep learning have been exploited to track visitors by combining badges and fixed cameras placed to cover the space of the museum [9,10].

Although fixed cameras have significantly improved the process of analyzing museum visits, they can only track the flow of visits without focusing on each visitor individually. The great success of ambient computing in everyday activities has been able to push the research community to experiment with wearable devices in museums and exhibition spaces [11]. Taking advantage of egocentric or first-person vision, which is the domain of computer vision focusing on data resembling the point of view of a user [4], several works have been presented that create new interactive experiences using wearable cameras [12–14] and mobile devices [15]. Ioannakis et al. [15] used different learning combinations based on bags-of-visual-words (BoVWs) and CNN approaches for the recognition of folklore museum exhibits. In [14], a method for localizing visitors in cultural sites was presented, which was tested on a new dataset collected using head- and chest-mounted cameras. A multiclassifier based on the VGG-19 convolutional network with ImageNet for pretraining was utilized to classify the egocentric data into nine location classes.

Along with crowd tracking, visitor localization and museum exhibit recognition, painting recognition has also attracted the interest of the research community. Gultepe et al. [16] presented a system for automatically classifying digitized paintings. Three art features were extracted from the paintings using unsupervised learning method inspired by deep learning and then used to group them into different art styles. A method for identifying copyrighted paintings was presented in [17], which combined the YOLOv2 [18] object detector and a matching approach based on handcrafted visual features. Crowley and Zisserman [19] proposed an approach to find objects in paintings for effective image retrieval. Classifiers were trained using CNN features to retrieve images from a large painting dataset. The authors developed a system for training new classifiers in real time by crawling Google Images for a given object query. More recently Barucci et al. [20] exploited deep learning and proposed the Glyphnet, a CNN-based network, for classifying with high accuracy images of ancient Egyptian hieroglyphs. The authors tested the new network on two different datasets and compared the results with other CNN frameworks indicating the effectiveness of Glyphnet in the specific classification task.

Similarly, egocentric vision techniques have been applied for artwork recognition in data collected through wearable devices. In [21], a system was presented which exploited Google glasses to recognize the painting the wearer was looking at. Different handcrafted features were compared in recognizing paintings from the Musée du Louvre. Other efforts have focused on artwork recognition using deep learning approaches, due to their great ability to learn high-level features directly from the input data, combined with their high

performance when applied in small, labeled datasets. Zhang et al. [13] studied the artwork recognition in egocentric images using deep learning algorithms. Datasets for three different museum art spaces were created using smartphone cameras. The VGG-16 architecture was employed to classify the artworks into paintings, sculptures and clocks classes. The experimental results led to low recognition accuracy, with a maximum of 47.8% in the case of sculptures, highlighting the challenges for successful recognition. Portaz et al. [22] propose a method for museum artwork identification using CNNs. The learning and identification process were based on an image retrieval approach making it suitable for small datasets with low variability. The method was tested on two datasets, created using classic and head-mounted cameras covering both paintings and sculptures in glass boxes.

In [23], painting recognition was presented as part of a unified smart audio guide in the form of a smartphone app for museum visitors. Combining audio and visual information, the app recognized what the visitor was looking at and detected if they walked or talked in a museum hall. The painting classification and localization was based on a CNN algorithm, where further tracking and smoothing techniques were utilized to improve the recognition accuracy. In addition, other applications have been proposed such as digital tour guides that utilize both visual techniques and the available museum databases. Magnus (<http://www.magnus.net/>, accessed on 11 July 2022), (Vizgu <https://vizgu.com/>, accessed on 11 July 2022) and Smartify (<https://smartify.org/>, accessed on 11 July 2022) are apps for museums and art galleries that provide information to visitors for any artwork they choose with their smartphone camera.

In relation to previous articles, in this work, we developed robust artwork recognition algorithms which work under different conditions based on CNN techniques to provide useful information in real time about what a visitor is looking at at an art site. To avoid the discomfort caused by wearable cameras and simplify the process, we incorporated the algorithms in a dedicated smartphone app. Furthermore, the variables which affect the recognition performance were first studied in a virtual environment and taken into consideration in the training stage to develop a robust app. In addition to recognition performance, we evaluated the user experience, enabling the derivation of insights for the app's improvement, to deal effectively with the users' needs.

### 3. Preliminary Investigations in a Virtual Environment

A preliminary investigation aiming to determine the effect of different conditions on the artwork recognition accuracy in a museum/gallery environment was carried out [24]. The use of a virtual environment allowed the controlled introduction of various test scenarios that highlighted the challenges related to artwork recognition, so that efficient ways to deal with the problem could be formulated.

#### 3.1. Virtual Environment

A virtual environment showcasing 10 paintings (see Figure 2) was created. To simulate the points of view observed by a virtual visitor, a virtual camera was placed looking towards the leftmost painting, and gradually moved towards the rightmost painting until it returned to the starting position. A video showing the views of the virtual camera was recorded and all video frames were annotated to indicate the frames where the virtual visitor focused on a specific painting.

During on-site visits to museum/gallery exhibiting artworks, the following potential distractors that may affect the accuracy of automated artwork recognition were identified: (a) gallery wall texture, (b) ambient light intensity, (c) the viewing angle and position while looking at an artwork, (d) the visitor's speed of movement and (e) the introduction of occluding structures in a visitor's point of view due to the existence of other visitors/staff in the area. To assess the effect of the identified factors, appropriate changes were simulated in the virtual environment (see Figure 3) allowing the controlled introduction of the aforementioned distractors.



Figure 2. The setup of the virtual gallery.

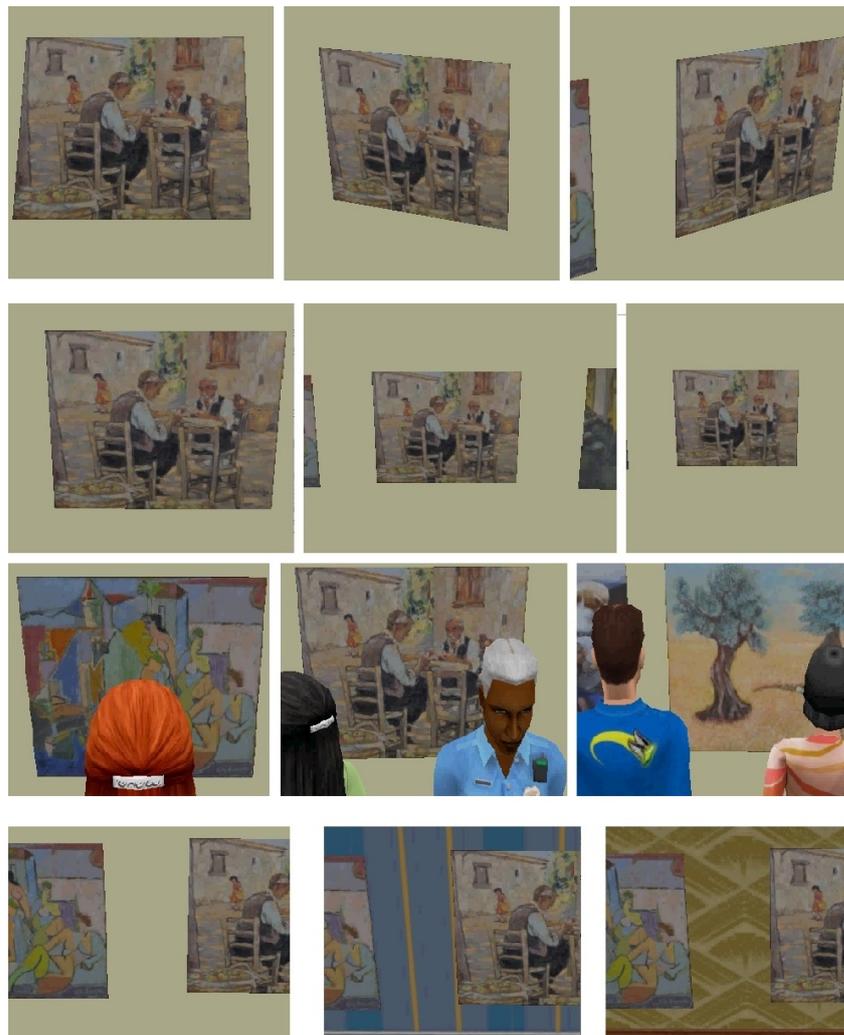


Figure 3. Samples showing examples of different conditions simulated in the virtual environment. Row 1 shows the simulation of changes in camera orientation, row 2 shows changes in distance, row 3 shows the introduction of occlusions in the form of virtual visitors and row 4 shows changes in wall texture.

### 3.2. Paintings Classification

Given the success of CNN for similar applications, as exemplified in Section 2, in our approach, the process of artwork recognition was carried out using the convolutional neural network “SqueezeNet” [25]. SqueezeNet is a deep network that uses design strategies to decrease the parameters number, particularly by using fire modules which “squeeze” parameters. SqueezeNet has 26 convolution layers with the ReLU activation function and

max pooling in each of those layers followed by a global average pooling layer. During the training stage, data augmentation, which included reflection, rotation, scale and translation of all training images, was used for increasing the training set. Transfer learning was utilized to adapt a version of the network trained on more than one million images from the ImageNet dataset [26,27] to the artwork classification problem. During the transfer learning process, the last convolution layer and the output layer from the pretrained SqueezeNet were tuned to the artwork recognition problem. To train the network, the Adam optimizer [28] was used, which is a method for efficient stochastic optimization. The learning rate was set to 0.0001, and the training process was run for 15 epochs, where each epoch consisted of 55 iterations. During the training process, 70% of the training samples were used for training and the remaining 30% of samples constituted the validation set.

During the classification phase, first-person images captured by virtual visitors were classified using the trained network.

### 3.3. Experimental Evaluation

During the performance evaluation stage, all frames in a video were classified and the percentage of correct classification was estimated. Since at this stage, the aim was to assess the effect of different conditions rather than producing the final classification system, the experiment did not consider the frames where the view was split among two paintings to an extent that it was not possible to indicate the painting which the visitor focused on. Two experiments were run as described below: for the first experiment a single image per painting was used as part of the training procedure, whereas for the second, multiple images for each painting were used in the training set. In this case, a training set with approximately 100 images per painting was generated using a virtual environment setup. For this purpose, each painting was placed in a virtual wall and several images of each painting were captured from different angles, different distances and different lighting (see Figure 4). In effect, the attempt here was to introduce types of variations present in real conditions in the training set, which were not introduced through the usual data augmentation methods.



**Figure 4.** Typical examples of simulated training images for a particular painting.

It should be noted that in the pilot experiments carried out in a virtual environment setting, the aim was to determine the factors affecting the paintings identification accuracy, rather than obtaining optimized classification rates. For this reason, the standardized cross-validation methodology usually adopted in similar problems, where a data set is divided into training, validation and test subsets [20], was not used in this case. During the training stage, the training set was split into training and validation sets. However, in our experiments, the training and test sets were completely disjointed sets, where the training set was based either on a single image per painting (Experiment 1) or multiple images showing each painting under different conditions (Experiment 2), whereas the test set was a video showing the image frames as captured by a virtual camera in a way that simulated images captured by a wearable camera attached on visitors.

### 3.4. Experimental Results and Conclusions

Table 1 shows how the recognition performance was affected by different simulated conditions. In the case of a single training image per painting, the recognition performance was considerably affected by reduced lighting, but other factors such as distance, occlusions, viewing angle and wall texture also affected the recognition performance. In the case of using multiple images per painting for training (Experiment 2), the recognition results

improved considerably. The only cases where a noticeable performance deterioration was observed were the cases of a highly reduced lighting (this is an extreme case that is not likely to be encountered in a real gallery) and the addition of multiple occlusions. As shown in Table 1, when using multiple images per painting in the training set, a significant performance improvement was also recorded in the case where the camera was moved randomly in the virtual space, capturing images displaying combined modes of variation in terms of distance, viewing angle and occlusions.

The results provided a useful insight into issues related to the efficiency of artwork recognition. The most important conclusion was the need for using multiple images per painting during the training stage, rather than relying on a single image per painting. Furthermore, museum curators that may adopt such technologies should ensure that a plain wall texture is used, a strong ambient lighting exists in all areas, and limit the number of visitors per room to avoid excessive occlusions.

The findings of this preliminary investigation were used to guide the further development of the proposed artwork recognition app, as described in subsequent sections.

**Table 1.** Recognition rates for different simulated conditions.

Condition	Parameters	Experiment 1: Correct Classification Rate	Experiment 2: Correct Classification Rate
Wall texture	No texture (baseline)	99%	100%
	Green stripe	95%	100%
	Blue stripe	94%	98%
	Gold fan	88%	100%
Number of visitors	0 (baseline)	99%	100%
	5	92%	100%
	10	71%	98%
	15	83%	93%
Distance	150 cm (baseline)	99%	100%
	200 cm	88%	100%
	250 cm	78%	100%
	300 cm	76%	100%
Camera point of view	Forward (baseline)	99%	100%
	Upwards	98%	100%
	Downwards	77%	100%
	Left	87%	100%
	Right	86%	100%
Ambient light intensity	255 (baseline)	99%	100%
	200	96%	100%
	150	63%	99%
	100	27%	46%
Camera speed	10 s	100%	100%
	20 s (baseline)	99%	100%
	30 s	99%	100%
	40 s	99%	100%
	Random camera movements	N/A	72%

#### 4. Artwork Recognition in a Real Environment

This section deals with artwork recognition experiments in a real environment, and more specifically in the State Gallery of Contemporary Art.

##### 4.1. Dataset

For our experiments, we chose a specific hall in the State Gallery which hosts 16 artworks in total, including 13 paintings and 3 sculptures (shown in Figure 5). A first-person

dataset consisting of two subsets was created, using a midrange smartphone camera, to be realistic, on regular visitors.



**Figure 5.** The artworks used in the experimental setup (artworks marked with an \* are 3D objects).

Following the outcomes of the experiments in the virtual environment, two subsets of training images were created. The first subset contained 500 images for each artwork captured from different points of view and distances. In addition, it included 500 images for the “No Artwork” category covering the floor, ceiling and hall’s walls. In total, the first subset contained 8500 images and was used for the training of different CNN models capable of recognizing the hall’s artworks.

The second subset was used for further testing the recognition accuracy of the best performing CNN model of the previous step. It contained 750 images per artwork captured from three different distances (1 m, 1.5 m and 2.5 m), using five different camera points of view (forward, downwards, upwards, left and right), representing 15 combinations in total. Fifty images were collected for each combination, resulting in 12,000 images (15 combinations  $\times$  16 artworks  $\times$  50 images) (720  $\times$  1280 resolution). The subset was expanded to include changes in ambient light intensity and occlusions after processing at image level using the Python Pillow library [29]. In the case of light intensity, the brightness level of all images was modified from 255 to 200, 150, 100 and 50 (Figure 6). Similarly, human silhouettes were added as occlusions between the visitor and artwork at various percentages (e.g., 50–90%) of the silhouette’s height in relation to the height of the image, as shown in Figure 7.



**Figure 6.** Example of the ambient light intensity variations in the hall.



**Figure 7.** Example of figures of virtual visitors superimposed on images showing paintings, in order to simulate occlusions (the human silhouettes used originated from a freely available dataset [30]).

Qualitative and quantitative training data significantly improve the performance of deep learning algorithms. The availability of annotated datasets is an important step in the development of first-person algorithms. Although there are several ways that can facilitate the collection and annotation process, creating a first-person dataset remains a difficult and time-consuming process due to the large amount of data and rich visual content. The created dataset will be made available to the research community for assisting researchers involved in the development of first-person technologies, artwork recognition and applications in museums and art galleries (due to copyright issues, the dataset is only available for downloading after sending a request to the corresponding author).

#### 4.2. CNN Artwork Recognition Models

Three different CNN architectures were used to train models for artwork recognition using the TensorFlow framework, including the VGG19 [31], InceptionV3 [32] and MobileNetV2 [33] architectures.

The first subset was used to train the 3 CNNs architectures. The subset was randomly divided into 3 sets, a 70% training set (5950 images), a 20% validation set (1700 images, used to fine-tune the model and improve its performance after each epoch) and a 10% test set (850 frames, used to test and compare the accuracy of the created models). The frames were subsampled to a size of  $224 \times 224$  pixels and normalized to the range of  $[-1, 1]$  for the InceptionV3 and MobileNetV2 models, and to the range of  $[0, 1]$  for the VGG19 model. Random transformations such as random crops, variations in brightness and horizontal flips were applied to the images of the training subset, prior to the subsampling and normalization step. Random transformations resulted in a model with rotation-invariant convolutional filters without explicitly enforcing it on the model architecture. The rotation invariance was required in the proposed framework since a smartphone is a constantly moving device and the user might not take the capture completely leveled.

The networks were trained using the layer configurations shown in Table 2 and were initialized with the pretrained weights of ImageNet. Each CNN was trained for 20 epochs and a batch size of 128, using a Google Colab environment configured with a GPU runtime and 13 GB of RAM.

**Table 2.** The layer configuration used for the training of the CNNs.

InceptionV3	MobileNetV2	VGG19
InceptionV3 original layers (excl. top layer, initialized with ImageNet weights)	MobileNetV2 original layers (excl. top layer, initialized with ImageNet weights)	VGG19 original layers (excl. top layer, initialized with ImageNet weights)
Global average pooling 2D layer	Flatten layer	Flatten layer
Dense layer (1024, activation = 'ReLU')	Dropout layer (rate = 0.5)	Dropout layer (rate = 0.5)
Dense layer (17 classes, softmax)	Dense layer (17 classes, softmax)	Dense layer (17 classes, softmax)

Since the purpose of the trained networks was to be incorporated into a smartphone app, the models were converted to the app-compatible TensorFlow Lite format (.tflite). Additionally, by taking the advantage of the post-training quantization technique ([https://www.tensorflow.org/lite/performance/post\\_training\\_quantization](https://www.tensorflow.org/lite/performance/post_training_quantization)), accessed on 11 July 2022, which reduces model size and improves CPU performance with a slight reduction in model accuracy, quantized TensorFlow Lite versions of the models were developed. The characteristics of the resulting CNNs are listed in Table 3.

While the accuracy for all architectures when tested with the 10% test subset was quite similar, the models differed in their file size and the inference time for each frame. The ideal CNN architecture for the app would lead to a model file with the smallest size, the lowest inference time, and the highest accuracy. Although the quantized MobileNetV2 model led to the smallest file size and lowest inference time, it failed to achieve a high accuracy, making the MobileNetV2 model the preferable choice for the development of the app.

**Table 3.** Comparison of the trained CNNs.

Architecture	Model Size (.tflite)	Inference Time/Frame	Accuracy
VGG19	78 MB	4000 ms+	0.9739
VGG19 (quantized)	20 MB	1800–2000 ms	0.9749
InceptionV3	91 MB	600–800 ms	0.9902
InceptionV3 (quantized)	23 MB	400–550 ms	0.9869
MobileNetV2	15 MB	200–350 ms	0.9967
MobileNetV2 (quantized)	4 MB	180–300 ms	0.8693

#### 4.3. Experimental Results

The MobileNetV2 recognition model was tested using the second subset, in order to evaluate its performance in real conditions and under the different factors affecting the recognition accuracy. The model was loaded into memory and fed with the images in the dataset one by one, recording its prediction for each case to calculate the total recognition accuracy. The confusion matrix depicted in Figure 8 shows the high recognition accuracy achieved before the changes in brightness intensity and introduction of occlusions. The increased model misclassifications between artworks 5 and 7 indicate a need to increase the examples of these classes during training. The increased misclassifications between artworks 5 and 7 arise because the two artworks are located next to each other in the gallery. Furthermore, artwork 5 is a 3D object (a sculpture) placed in a corner of the gallery hall while artwork 7 is a painting placed close to artwork 5, on the adjacent wall. As a result, in numerous frames showing artwork 5 (depending on the angle of view), parts of artwork 7 are also visible (see Figure 9). Due to the fact that the pattern of artwork 7 is distinct, the chance for misclassifications increases when parts of artwork 7 are shown in combination

with artwork 5. The use of additional training samples of 3D objects, where other artworks also appear in the scene, may enhance the ability of the classifier to classify correctly 3D objects in the presence of visual stimuli from other artworks.

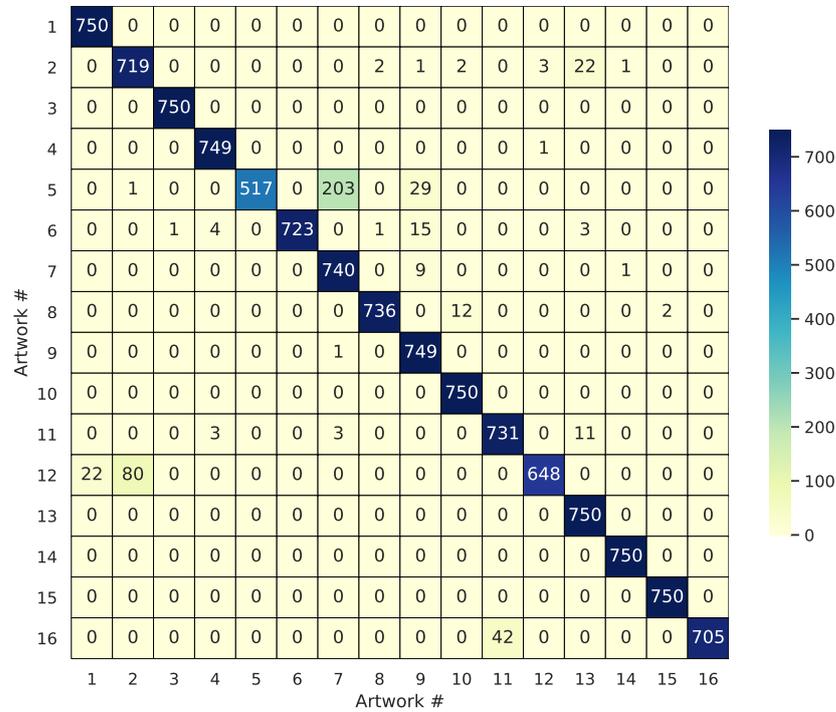
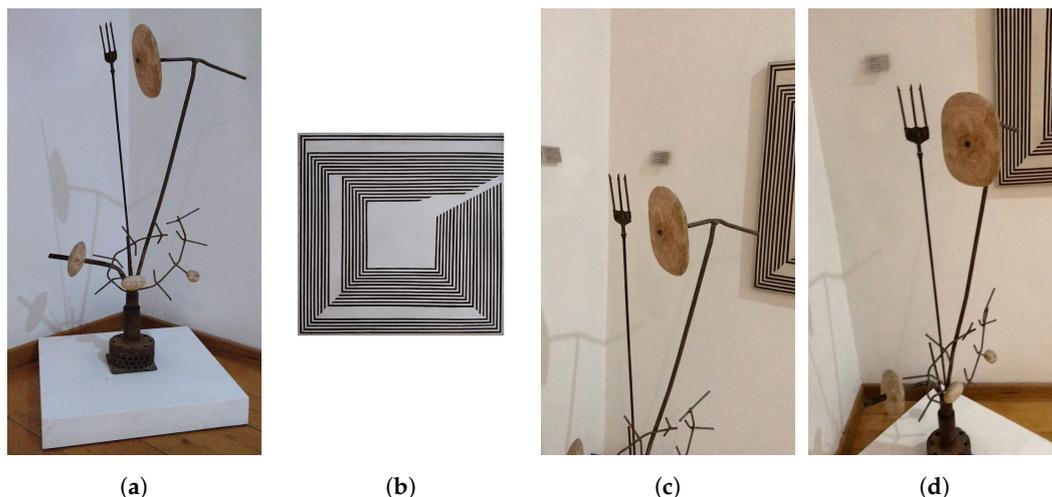


Figure 8. Confusion matrix for the predictions in data without occlusions and brightness intensity modifications.

As shown in Table 4, the recognition model performs very well in strong ambient light condition and without occlusions between the visitor and artwork with almost 100% accuracy at 1 m and 1.5 m, and 88.5% accuracy at 2.5 m. The majority of the misclassified predictions at 2.5 m are caused by the appearance of more than one artwork in the image which sometimes leads to the classification of an artwork in the adjacent class. The camera point of view does not play a significant role in the recognition with a prediction accuracy ranging between 91% and 98% for all distances. The predictions verify the experiments conducted in the virtual environment in regard to the effects of brightness and the number of visitors who are present at the same time in the hall. As the brightness level changes from 255 to 50, the recognition accuracy decreases for most combinations with the smallest accuracy observed in the case of the left (−0.8 %) and the highest in the case of the upwards camera view (−13.9%). Similarly, the accuracy decreases while the height of the human silhouettes increases, achieving 72.6% as the highest value when the height is equal to 50% of the image.

The results confirmed the superiority of deep neural networks in egocentric vision and culture heritage, achieving a high recognition accuracy especially when a satisfactory number of training data were available. The MobileNetV2 performed almost perfectly in artwork recognition for most of the combinations. The good quality of the input image in CNN model positively affected the output result. The hall’s strong ambient lighting, limited occlusions and reasonable capturing distance played an important role in recognition accuracy, which was in full agreement with the findings of the experimentation in virtual environment.



**Figure 9.** Misclassifications between artworks 5 and 7: (a) artwork 5; (b) artwork 7; (c,d) taking a photo of artwork 5 under certain angles of view.

**Table 4.** Prediction accuracy (%) for all brightness variations and visitor occlusions at 50% height.

View \ Distance	1 m	1.5 m	2.5 m	Avg	1 m	1.5 m	2.5 m	Avg	
<b>Brightness = 255/255 (Visitor height = 0%)</b>					<b>Brightness = 200/255 (Visitor height = 0%)</b>				
Forward	100.0%	100.0%	94.2%	98.1%	100.0%	100.0%	93.2%	97.8%	
Downwards	100.0%	99.9%	92.1%	97.3%	100.0%	100.0%	91.4%	97.1%	
Upwards	100.0%	96.1%	78.9%	91.7%	98.1%	98.8%	77.8%	91.5%	
Left	100.0%	100.0%	90.6%	96.9%	100.0%	100.0%	85.4%	95.1%	
Right	100.0%	100.0%	87.8%	95.9%	100.0%	100.0%	91.4%	97.1%	
Avg	100.0%	99.2%	88.7%	96.0%	99.6%	99.8%	87.8%	95.7%	
<b>Brightness = 150/255 (Visitor height = 0%)</b>					<b>Brightness = 100/255 (Visitor height = 0%)</b>				
Forward	100.0%	100.0%	93.0%	97.7%	100.0%	100.0%	93.4%	97.8%	
Downwards	100.0%	100.0%	91.0%	97.0%	100.0%	100.0%	88.1%	96.0%	
Upwards	97.5%	97.8%	78.1%	91.1%	99.2%	94.2%	78.5%	90.7%	
Left	100.0%	100.0%	86.6%	95.5%	100.0%	100.0%	91.1%	97.0%	
Right	100.0%	100.0%	93.1%	97.7%	100.0%	99.8%	89.8%	96.5%	
Avg	99.5%	99.5%	88.4%	95.8%	99.8%	98.8%	88.2%	95.6%	
<b>Brightness = 50/255 (Visitor height = 0%)</b>					<b>Visitor height = 50% (Brightness = 255/255)</b>				
Forward	89.2%	92.6%	86.1%	89.3%	68.1%	70.4%	49.9%	62.8%	
Downwards	98.0%	98.8%	78.0%	91.6%	47.6%	54.2%	25.1%	42.3%	
Upwards	87.2%	89.1%	56.9%	77.8%	60.5%	61.0%	33.4%	51.6%	
Left	99.9%	99.0%	85.4%	94.8%	72.0%	72.6%	46.4%	63.7%	
Right	97.4%	98.0%	65.2%	86.9%	56.9%	64.4%	28.5%	49.9%	
Avg	94.4%	95.5%	74.3%	88.1%	61.0%	64.5%	36.7%	54.1%	

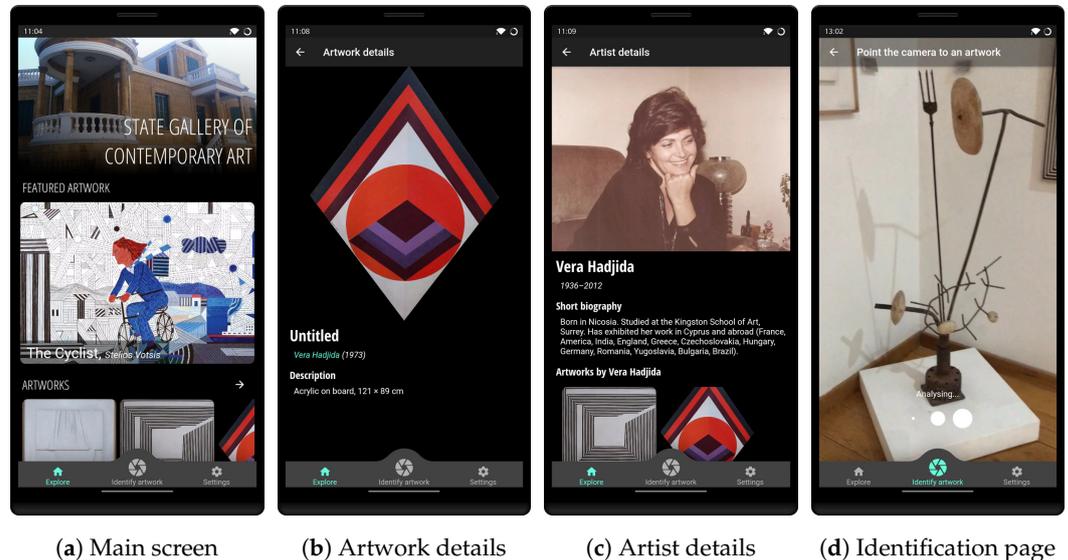
### 5. Smartphone App Development and Evaluation

In this section, we describe the app development and the app’s preliminary testing and evaluation focusing on the user experience.

#### 5.1. Smartphone App Development

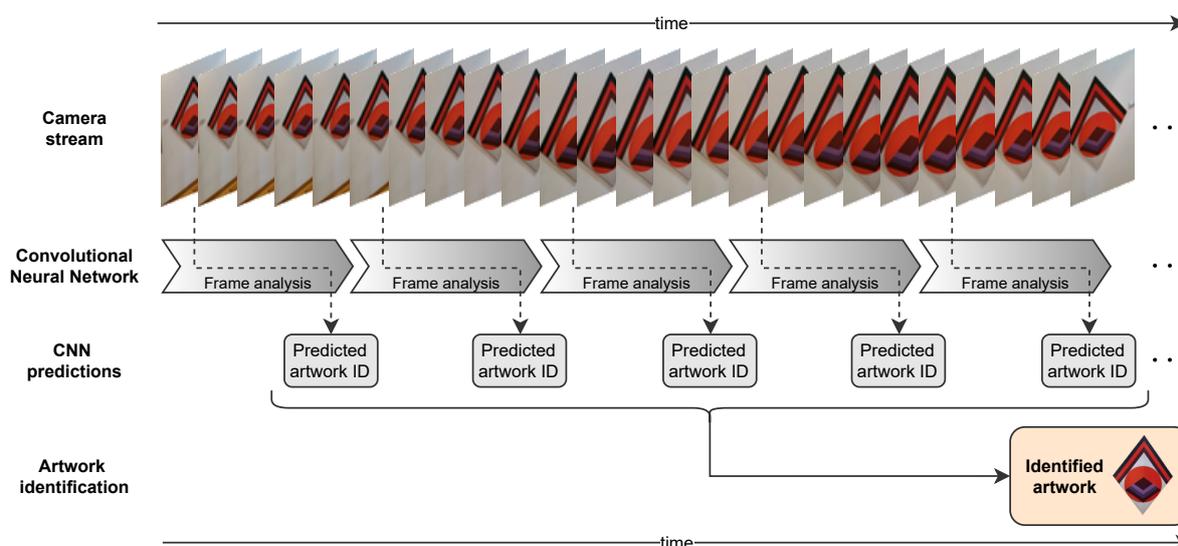
The app was developed using the Flutter application framework (<https://flutter.dev/>, accessed on 11 July 2022), which allows cross-platform development, and so it can be compiled to run in both major mobile operating systems, Android and iOS. The app is supported by an SQLite database containing information about the artworks and their creators, as well as high-resolution images of each artwork, and thus it does not require any network connection to function. It provides the user with the opportunity to browse

and view information about artworks and artists, as depicted in Figure 10. The artist page includes links to all the artist’s works included in the app. The app’s UI is currently available in both Greek and English and can be readily internationalized to any other language if deemed necessary.



**Figure 10.** Screenshots taken from an Android device running the app prototype.

The MobileNetV2 model can recognize artworks in visual data collected by a camera mounted on the body or held in the user’s hands. However, for the app’s development, we chose to continue with the latter approach providing the user with the opportunity to select the artwork of their preference for recognition as follows: As soon as the user clicks the “Identify artwork” button, the app instructs the user to point their smartphone camera to the artwork they want to identify (Figure 10d) and begins sending individual frames from the camera stream to the CNN for analysis. Once the analysis of a frame is finished (inference takes 200–350 ms, see Table 3), the currently available frame in the camera stream is sent for analysis, i.e., the intervening frames during the duration of each analysis are not used (see Figure 11). The result of each frame analysis by the CNN is a predicted artwork label and a probability for the prediction (a number in the range  $[0, 1]$ ). To ensure accurate artwork identifications, the app considers a sliding window of 5 consecutive analyses (as shown in Figure 11) and calculates the average probability for each predicted artwork label in the window (the sliding window used proceeded in steps of 1). The artwork with the greatest average probability is chosen as the winner, whereupon the app takes the user to the identified artwork’s details page (Figure 10b). In the case of a tie between the average probabilities of two or more artworks, the app proceeds to the next window, until a winner emerges. If the label with the greatest average probability is “No artwork” (see Section 4.1), the app displays a relevant warning to the user and continues analyzing frames until the prediction is an artwork.



**Figure 11.** The process used for artwork identification: while the user is pointing at an artwork, the embedded CNN is used to continuously analyze individual frames from the phone’s camera stream. The app considers a sliding window of 5 CNN predictions and picks the artwork with the highest mean probability in the window as the correct artwork.

### 5.2. Smartphone App Evaluation

Due to the COVID-19 lockdown measures and difficulties to test the proposed smartphone app in the wild, a preliminary remote testing and evaluation regarding the user experience (UX) was carried out. A total of 33 volunteer users participated in the evaluation process, including ordinary visitors, two of the State Gallery’s personnel, as well as other professionals in the field of cultural heritage. Users were asked to download the app on their personal smartphones, use it and familiarize themselves with it, and then follow an online questionnaire created for the evaluation.

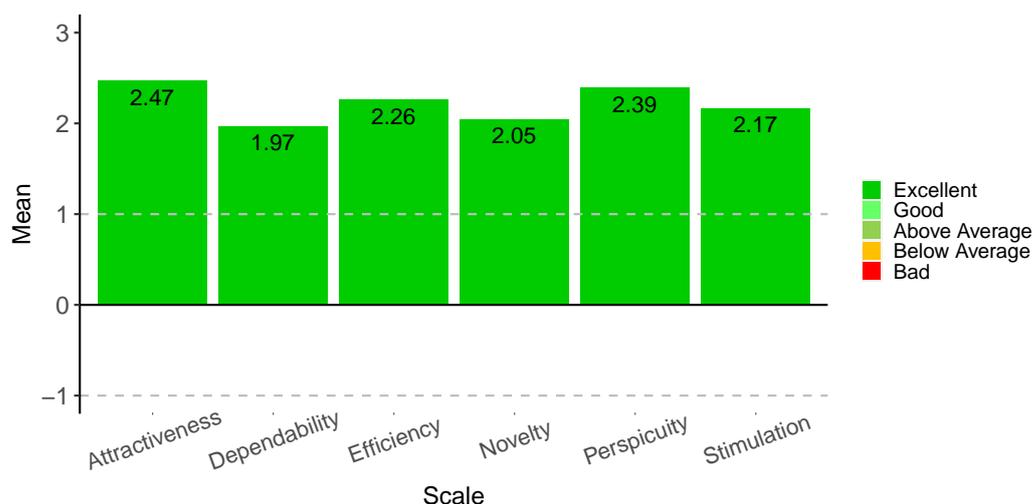
The online questionnaire was structured into four parts: (A) demographics, (B) using the app, (C) evaluating the app’s recognition accuracy and (D) UX evaluation. Five demographic questions were included in Part A concerning the gender, age, familiarity with smartphones, frequency of museum visits and educational background and/or work experience in museology and cultural heritage. Fifteen female and eighteen male participants aged between 16 and 60 years old, who were moderately to extremely familiar with smartphone devices, were involved in the evaluation procedure. One participant stated that they visited a museum one or more times per month, five participants stated that they visited once every 3 months, seven participants stated that they visited once every 6 months, twelve stated that they visited once a year, while the remaining eight stated that they visited a museum less than once a year. Finally, six participants answered positively to the question of whether they had relevant experience or were considered as experts in the evaluation process.

The second part of the questionnaire (Part B) asked the participants to explore the app and use it to identify artworks as they would use it during their visit to the State Gallery. More specifically, they were able to select any artwork, from the eight randomly selected, for display on their computer screen and then use the app for its automatic identification and provision of auxiliary information. Participants could repeat the process without restrictions before proceeding to Part C for evaluating recognition accuracy. At this part, the participants were asked to select the name given by the app during automatic recognition in three randomly selected artworks. The app recognized artworks with a high accuracy giving 32 correct answers to the first and second artwork (96.73%), and 31 correct answers to the third one (93.43%).

Part D was concerned with the UX evaluation by utilizing a shortened version of the User Experience Questionnaire (UEQ) [34]. Two UEQ questions for each one of the following six scales, attractiveness, dependability, efficiency, novelty, perspicuity and stimulation, were included in the questionnaire (Table 5). The scores depicted in Figure 12 show that the app was positively evaluated, achieving an *Excellent* evaluation for all scales. Concerning the differences between demographic factors, the results reveal only a significant difference ( $p$ -value = 0.02) between experts and non-experts in the way they evaluate the efficiency of the app. All experts evaluated the app as very fast and practical to very practical.

**Table 5.** UEQ scales used in the app evaluation and their individual components.

Scale	Component 1	Component 2
Attractiveness	Attractive–unattractive	Friendly–unfriendly
Dependability	Meets expectations–does not meet expectations	Obstructive–supportive
Efficiency	Fast–slow	Impractical–practical
Novelty	Conservative–innovative	Inventive–conventional
Perspicuity	Clear–confusing	Complicated–easy
Stimulation	Not interesting–interesting	Valuable–inferior



**Figure 12.** Results of the app evaluation using the UEQ questionnaire.

The last part of the questionnaire focused on collecting users’ feedback on their overall experience and suggestions for possible improvements. All participants were asked: “What is the likelihood that the app will positively enhance the experience of a visitor at the Gallery?”. About 44% of participants answered “Very likely”, 35.5% answered “Likely”, 17.5% answered “Moderately likely”, and 3% answered “Neither likely nor unlikely”. According to the demographics, participants who visit museums or galleries once every 3 months and once every 6 months answered significantly different compared to the participants with more or less visits. They answered “Likely” to “Very likely” with a  $p$ -value equal to 0.015 for those with a visit once every 3 months and equal to 0.038 for those with a visit once every 6 months, respectively. When asked how likely they would recommend the app to others, 96% of the participants answered, “Moderately likely” to “Very likely”, with the remaining 4% answering “Neither likely nor unlikely”. It is worth mentioning here that the only participant with more than one visit to museums and galleries per month noted that they very likely would recommend the app to other users.

No major problems were reported during the testing of the app, except an identification issue that occurred when the testing ran on low resolution screen. Based on the participants’ feedback, the issue was resolved by placing their smartphone closer to the screen during the

tests. The problem concerned only the use of the app for the remote online testing and did not concern the identification in real museum conditions, on which we relied for the training of algorithms. In regard to possible improvements, the participants suggested: (i) providing audio information about the recognized artwork, (ii) adding additional information about the Gallery and the artists, as well as descriptive details regarding the artworks, giving a historical background, conceptual and morphological information, (iii) using a frame window to guide the user to scan the artwork of their preference and (iv) improved access to the artworks' list by giving their location on the Gallery's map and providing a thematic search based on year of creation, artist, etc.

## 6. Conclusions

In this work, we presented the robust artwork recognition algorithm we developed for enhancing the visiting experience at museums and art galleries. The ultimate aim of our work was to formulate and adopt a comprehensive algorithm design hierarchical process, that enabled the testing of the algorithm developments at different levels, starting from a virtual space and ending at an end-user evaluation process. As part of the process to improve recognition accuracy in any context, we studied how different shooting conditions affected accuracy in a virtual environment and included the main findings in the proposed algorithm. It was then embedded in a dedicated smartphone app which was designed and implemented to provide visitors with assistive information about each artwork during their visit. Due to COVID-19 pandemic restrictions, the app evaluation was carried out remotely regarding the user experience, achieving an *Excellent* score in the UEQ sections. The results revealed that there were differences in the way that experts and non-experts evaluated the efficiency section with the former giving higher scores. Most participants believed that the app would greatly enhance the visiting experience as they were also very positive in recommending it to other users. In addition, the volunteers gave useful feedback and hints which will be considered for the improvement and finalization of the artwork recognition method to be ready to use in real museum conditions. The developed machine learning models and smartphone app source code are publicly available (Github Repository: <https://github.com/msthoma/Robust-Artwork-Recognition-using-Smartphone-Cameras>, accessed on 11 July 2022) for further research experiments, benchmark tests, etc.

Our future plans will first include the finalization of the app's functionality, including scanning window and artwork searching. Second, it will be enhanced with further artwork documentation and the algorithm will be extended to include more artworks of the State Gallery. Third, the app will be extended to record the tour route, the time spent observing each artwork and other useful information that will help the museum experts to study and analyze the visit. Fourth, the opportunity to use a mounted smartphone camera on users' body and provide audio information during the visit without their involvement in the recognition process will be studied and developed. Fifth, the app will be tested and evaluated for different scenarios in real conditions by more volunteers, including ordinary visitors and museum professionals, before being made available to State Gallery visitors. Finally, the issue of the scalability of the proposed approach will be addressed to accommodate the inclusion of new artworks in the classification algorithm by exploring dedicated methods that allow the adaptation of the model to deal with new classes without the need for retraining the model from scratch [35,36]. For this purpose, continuous learning approaches [37] will be investigated. From a practical point of view, to support the operation of adding new artworks, a dedicated extension of the app will be added to allow the gallery staff to capture videos and images of new artworks, that can be used as training data for tuning the classification model to deal with the new artworks. In addition, we plan to use the hierarchical algorithm design process proposed in this work for the development of algorithms for other egocentric vision applications.

**Author Contributions:** Conceptualization, Z.T. and A.L.; methodology, Z.T. and H.P.; software, M.T.; validation, Z.T., M.T. and H.P.; formal analysis, M.T.; investigation, Z.T. and M.T.; data curation, Z.T., M.T. and H.P.; writing—original draft preparation, Z.T.; writing—review and editing, A.L.; supervision, A.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under gGrant agreement no. 739578 and from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset used in this study is available on request from the corresponding author. The dataset is not publicly available due to copyright issues. The developed machine learning models and smartphone app source code are openly available in Github Repository: <https://github.com/msthoma/Robust-Artwork-Recognition-using-Smartphone-Cameras> (accessed on 11 July 2022).

**Acknowledgments:** This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under gGrant agreement no. 739578 and from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

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

## References

1. Lanir, J.; Kuflik, T.; Sheidin, J.; Yavin, N.; Leiderman, K.; Segal, M. Visualizing Museum Visitors’ Behavior: Where Do They Go and What Do They Do There? *Pers. Ubiquitous Comput.* **2017**, *21*, 313–326. [[CrossRef](#)]
2. Tomiuc, A. Navigating Culture. Enhancing Visitor Museum Experience through Mobile Technologies. From Smartphone to Google Glass. *J. Media Res.* **2014**, *7*, 33–47.
3. Miluniec, A.; Swacha, J. Museum Apps Investigated: Availability, Content and Popularity. *e-Rev. Tour. Res.* **2020**, *17*, pp. 764–776.
4. Theodosiou, Z.; Lanitis, A. Visual Lifelogs Retrieval: State of the Art and Future Challenges. In Proceedings of the 2019 14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), Larnaca, Cyprus, 9–10 June 2019; pp. 1–5. [[CrossRef](#)]
5. Loizides, F.; El Kater, A.; Terlikas, C.; Lanitis, A.; Michael, D. Presenting Cypriot Cultural Heritage in Virtual Reality: A User Evaluation. In Proceedings of the Euro-Mediterranean Conference, Sousse, Tunisia, 1–4 November 2022; Springer: Berlin/Heidelberg, Germany, 2014; pp. 572–579. [[CrossRef](#)]
6. Hooper-Greenhill, E. Studying Visitors. In *A Companion to Museum Studies*; Macdonald, S., Ed.; Blackwell Publishing Ltd.: Malden, MA, USA, 2006; Chapter 22, pp. 362–376. [[CrossRef](#)]
7. Tesoriero, R.; Gallud, J.A.; Lozano, M.; Penichet, V.M.R. Enhancing visitors’ experience in art museums using mobile technologies. *Inf. Syst. Front.* **2014**, *16*, 303–327. [[CrossRef](#)]
8. Rashed, M.G.; Suzuki, R.; Yonezawa, T.; Lam, A.; Kobayashi, Y.; Kuno, Y. Tracking Visitors in a Real Museum for Behavioral Analysis. In Proceedings of the 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS), Sapporo, Japan, 25–28 August 2016; pp. 80–85. [[CrossRef](#)]
9. Mezzini, M.; Limongelli, C.; Sansonetti, G.; De Medio, C. Tracking Museum Visitors through Convolutional Object Detectors. In Proceedings of the Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization, Genoa, Italy, 14–17 July 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 352–355. [[CrossRef](#)]
10. Ferrato, A.; Limongelli, C.; Mezzini, M.; Sansonetti, G. Using Deep Learning for Collecting Data about Museum Visitor Behavior. *Appl. Sci.* **2022**, *12*, 533. [[CrossRef](#)]
11. Shapiro, B.R.; Hall, R.P.; Owens, D.A. Developing & Using Interaction Geography in a Museum. *Int. J. -Comput.-Support. Collab. Learn.* **2017**, *12*, 377–399. [[CrossRef](#)]
12. Mason, M. The MIT Museum Glassware Prototype: Visitor Experience Exploration for Designing Smart Glasses. *J. Comput. Cult. Herit.* **2016**, *9*, 1–28. [[CrossRef](#)]
13. Zhang, R.; Tas, Y.; Koniusz, P. Artwork Identification from Wearable Camera Images for Enhancing Experience of Museum Audiences. In Proceedings of the MW17: Museums and the Web, Cleveland, Ohio, 19–22 April, 2017; Springer: Berlin/Heidelberg, Germany, pp. 240–247.
14. Ragusa, F.; Furnari, A.; Battiato, S.; Signorello, G.; Farinella, G.M. Egocentric Visitors Localization in Cultural Sites. *J. Comput. Cult. Herit.* **2019**, *12*, 1–19. [[CrossRef](#)]
15. Ioannakis, G.; Bampis, L.; Koutsoudis, A. Exploiting artificial intelligence for digitally enriched museum visits. *J. Cult. Herit.* **2020**, *42*, 171–180. [[CrossRef](#)]
16. Gulpepe, E.; Edward. Conturo, T.; Makrehchi, M. Predicting and grouping digitized paintings by style using unsupervised feature learning. *J. Cult. Herit.* **2018**, *31*, 13–23. [[CrossRef](#)] [[PubMed](#)]

17. Hong, Y.; Kim, J. Art Painting Detection and Identification Based on Deep Learning and Image Local Features. *Multimed. Tools Appl.* **2019**, *78*, 6513–6528. [[CrossRef](#)]
18. Nakahara, H.; Yonekawa, H.; Fujii, T.; Sato, S. A Lightweight YOLOv2: A Binarized CNN with A Parallel Support Vector Regression for an FPGA. In Proceedings of the 2018 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, FPGA '18, Monterey, CA, USA, 25–27 February 2018; Association for Computing Machinery: New York, NY, USA, 2018; p. 31–40. [[CrossRef](#)]
19. Crowley, E.J.; Zisserman, A. In Search of Art. In Proceedings of the Computer Vision—ECCV 2014 Workshops, Zurich, Switzerland, 6–12 September 2014; Agapito, L., Bronstein, M.M., Rother, C., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 54–70.
20. Barucci, A.; Cucci, C.; Franci, M.; Loschiavo, M.; Argenti, F. A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification. *IEEE Access* **2021**, *9*, 123438–123447. [[CrossRef](#)]
21. Dalens, T.; Sivic, J.; Laptev, I.; Campedel, M. Painting recognition from wearable cameras. Technical Report hal-01062126, INRIA, 2014. Available online: <https://www.di.ens.fr/willow/research/glasspainting/> (accessed on 11 July 2022)
22. Portaz, M.; Kohl, M.; Quénot, G.; Chevallet, J. Fully Convolutional Network and Region Proposal for Instance Identification with Egocentric Vision. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 2383–2391. [[CrossRef](#)]
23. Seidenari, L.; Baccchi, C.; Uricchio, T.; Ferracani, A.; Bertini, M.; Bimbo, A.D. Deep Artwork Detection and Retrieval for Automatic Context-Aware Audio Guides. *ACM Trans. Multimed. Comput. Commun. Appl.* **2017**, *13*, 1–21. [[CrossRef](#)]
24. Lanitis, A.; Theodosiou, Z.; Partaourides, H. Artwork Identification in a Museum Environment: A Quantitative Evaluation of Factors Affecting Identification Accuracy. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*; Ioannides, M., Fink, E., Cantoni, L., Champion, E., Eds.; EuroMed 2020. Lecture Notes in Computer Science; Springer, Cham, Switzerland, 2021; Volume 12642, pp. 588–595. [[CrossRef](#)]
25. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5 MB Model Size. *arXiv* **2016**, arXiv:1602.07360.
26. Deng, J.; Dong, W.; Socher, R.; Li, L.; Kai, L.; Li, F. 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](#)]
27. ImageNet. Online. 2020. Available online: <http://image-net.org> (accessed on 17 March 2021).
28. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
29. Clark, A. Pillow (PIL Fork) Documentation, 2020. Version 8.0.0. Available online: <https://pillow.readthedocs.io/en/stable/releasesnotes/8.0.0.html> (accessed on 11 July 2022) .
30. XOIO-AIR. Cutout People—Greenscreen Volume 1, 2012. Available online: [https://xoio-air.de/2012/greenscreen\\_people\\_01/](https://xoio-air.de/2012/greenscreen_people_01/) (accessed on 11 July 2022).
31. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
32. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826. [[CrossRef](#)]
33. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520. [[CrossRef](#)]
34. Schrepp, M.; Hinderks, A.; Thomaschewski, J. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In *Design, User Experience, and Usability, Theories, Methods, and Tools for Designing the User Experience, Proceedings of the Third International Conference, Herakleion, Greece, 22–27 June 2014*; Marcus, A., Ed.; Springer International Publishing: Cham, Switzerland, 2014; pp. 383–392. [[CrossRef](#)]
35. Schulz, J.; Veal, C.; Buck, A.; Anderson, D.; Keller, J.; Popescu, M.; Scott, G.; Ho, D.K.C.; Wilkin, T. Extending deep learning to new classes without retraining. In Proceedings of the Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXV, Online, 27 April–8 May 2020; Bishop, S.S., Isaacs, J.C., Eds.; International Society for Optics and Photonics, SPIE: Bellingham, WA, USA, 2020; Volume 11418, pp. 13–26. [[CrossRef](#)]
36. Draelos, T.J.; Miner, N.E.; Lamb, C.C.; Cox, J.A.; Vineyard, C.M.; Carlson, K.D.; Severa, W.M.; James, C.D.; Aimone, J.B. Neurogenesis deep learning: Extending deep networks to accommodate new classes. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 526–533. [[CrossRef](#)]
37. Parisi, G.I.; Kemker, R.; Part, J.L.; Kanan, C.; Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Netw.* **2019**, *113*, 54–71. [[CrossRef](#)]