Imaging Novecento. A Mobile App for Automatic Recognition of Artworks and Transfer of Artistic Styles

Federico Becattini, Andrea Ferracani(✉), Lea Landucci, Daniele Pezzatini, Tiberio Uricchio, and Alberto Del Bimbo

NEMECH - New Media for Cultural Heritage, University of Florence, Viale Morgagni 65, 50124 Firenze, Italy
{federico.becattini,andrea.ferracani,lea.landucci,daniele.pezzatini, tiberio.uricchio,alberto.delbimbo}@unifi.it
http://nemech.unifi.it

Abstract. *Imaging Novecento* is a native mobile application that can be used to get insights on artworks in the "Museo Novecento" in Florence, IT. The App provides smart paradigms of interaction to ease the learning of the Italian art history of the 20^{th} century. *Imaging Novecento* exploits automatic approaches and gamification techniques with recreational and educational purposes. Its main goal is to reduce the cognitive effort of users *versus* the complexity and the numerosity of artworks present in the museum. To achieve this the App provides automatic artwork recognition. It also uses gaming, in terms of a playful user interface which features state-of-the-art algorithms for artistic style transfer. Automated processes are exploited as a mean to attract visitors, approaching them to even lesser known aspects of the history of art.

Keywords: Cultural heritage · Mobile application · Visual recognition · Artistic style transfer · Convolutional neural networks

1 Introduction

Modern museums can provide new paradigms for experiencing artworks. Thanks to the technological development, novel initiatives include pervasive uses of tech to create interactive experiences for visitors throughout a museum. However, making content relevant and appealing through these modern technologies is a difficult problem, requiring more and more interactivity as the audience is shifting towards a 'multimedia point of view'. Moreover, while the massive amount of available artworks constitutes a huge resource for education and recreation purposes, it can also be a cognitive burden for visitors.

The cognitive process related to learning has been an active subject of study in recent decades. According to cognitive load theory, learners must cope with a certain level of cognitive effort to process new information [22]. In this regard, multimedia education, defined as "presenting words and pictures that

M. Ioannides et al. (Eds.): EuroMed 2016, Part I, LNCS 10058, pp. 781–791, 2016.
DOI: 10.1007/978-3-319-48496-9_62

are intended to foster learning" [19], can be an effective remedy because it facilitates the activation of sensory and cognitive perceptions (e.g. visual and notional memory), avoiding visitors from information overloading. This can also be reinforced by gamification, that is the use of playful experience to help a user find personal motivations and engagement with serious content [24]. This combination can enhance the visitor's involvement and further lower its cognitive effort. Using gamified applications, museum visitors have the opportunity to feel the emotion of a game, share results with friends on social networks or become part of a game community [20]. This aspect of learning through gaming is even more valuable in the context of the "Bring Your Own Device" (BYOD) approach [3] that allows on demand access to digital content on personal devices. The BYOD approach and gamification have been identified in the NMC Horizon Report 2015 to be increasingly adopted by museums in one year's time or less for mobile and online engagement [15].

In this paper we report our experience in embedding these concepts into *Imaging Novecento*, a system built around a mobile application developed for the museum "Museo Novecento" in Florence, IT. We aimed at improving the learning process of the visitors by exploiting a simple gamification paradigm, and at reducing visitors cognitive load. To this end, we also developed a state-of-the-art computer vision system that is able to (1) recognize artworks from photos; (2) apply their style to user photos.

1.1 The Museo Novecento in Florence and Innovecento

The "Museo Novecento" in Florence, IT, is a museum opened on June 24, 2014. The museum is dedicated to the Italian art of the 20^{th} century and offers a selection of about 300 artworks distributed in fifteen exhibition halls on two levels. The venue is located in the former hospital of the "Leopoldine" in Piazza Santa Maria Novella. The museum has been an example of innovation since its genesis, thanks to the prompt adoption of the latest multimedia technologies.

In March 2015, in order to improve the visitor experience, the Municipality of Florence has published an open call "INNOVecento - Novecento Museum Innovation Lab" inviting companies and professionals to propose ideas and solutions based on ICT. Five companies specialized in technologies applied to cultural heritage have already responded to the call which, at the time of writing, is still open. As NEMECH, centre of competence of the Tuscany region in Italy, we proposed *Imaging Novecento*. The App features automatic recognition of artworks through the visitor's smartphone and automatic transfer of artistic styles from artworks. These styles can be applied to user images.

1.2 Motivations and Design

The target of the App is rather wide. Although *Imaging Novecento* can be used by anyone (e.g. tourists and residents), during the design process we identified a specific audience. We mainly target the App towards people in a relatively young

age (between 14 and 30 years old), more accustomed to digital technologies, open to technological innovation and to gamification.

One of the main ideas of the App is to exploit the pervasiveness of mobile cameras in modern smartphones to reduce the cognitive effort required to museum visitors. In fact, despite themed rooms and the ubiquitous explanatory cards, users can still be overwhelmed by the great number of artworks present in the museum. Labels in museums can be very concise or, on the contrary, can be filled with lots of explanation, often generic, not highlighting salient features of individual paintings. By using *Imaging Novecento*, the visitor can take a picture of the artwork he is interested in. The App will automatically recognize the painting and provide related information. Another reason for the adoption of this automatic process is the resistance of museums' curators to place or attach additional materials, such as QR codes or BLE iBeacon [12], next to artworks.

Furthermore, tourists and school groups are usually 'hit-and-run' visitors who tend to rapidly forget or do not have the time to process the overload of information. To solve this issue, *Imaging Novecento* leverages a playful feature that employs state-of-the-art algorithms for transferring artistic styles from recognized artworks to user images. This is done using a gamification paradigm at the interface level. Gamification techniques have been proved to be useful in engaging students in the learning process, improving their skills and maximizing their long-term memory [27].

1.3 Previous Work

Several previous works have addressed the problem of providing an engaging experience to museum visitors. Rapid technological development has led to the implementation of a lot of applications. There are several active trends for virtual museums: immersive reality [10,18], natural interaction installations [4,9], mixed reality, mobile applications [5,31]. While they all offer increasing engagement of visitors, only recently studies on the effects of audience have been carried out [15,21]. In particular, a recent audience study has been conducted on the case of the "Keys to Rome" international exhibition, hosted at the "Imperial Fora Museum" in Rome in 2015, to assess the impact of these technologies on cultural heritage. The exhibition was made up of 11 digital installations and applications, installed in the museum [21]. The study highlights some fundamental aspects that must be taken into account when designing applications for virtual museums: (1) the majority of museum visitors are tourists and school groups; (2) visitors generally require applications with an high level of interactivity, particularly on their mobile devices; (3) it is essential for the UX design to use metaphors of informal learning capable to stimulate attention, memory and engagement (e.g. through gamification) in visitors.

Automatic Artwork Recognition. Automatic artwork recognition is a long standing problem in applications for cultural heritage. Descriptors such as SIFT and SURF have been used for years in order to address this task [25,28] due to

their accuracy in recognizing paintings. Crowley and Zisserman [6] retrieve artworks finding object correspondences between photos and paintings by using a deformable part based method. More recent approaches for artwork recognition adopt Convolutional Neural Networks (CNN) as in [2], where a holistic and a part based representation are combined. Peng and Chen [23] exploit CNNs to extract cross-layer features for artist and artistic style classification tasks. Artistic style recognition is also performed in [16] on two novel large scale datasets. Similarly to these works, we explore the use of CNNs features but we aim to obtain a global representation that is semantically meaningful and also capable of retaining low level visual content information. Artwork recognition has also been used with wearable devices, as in [4] where the user's position is jointly estimated with what he is looking at.

Artistic Style Transfer. Regarding the application of artistic style to photos, a lot of research has been done in the past. The problem of rendering a given photo in the style of a particular artwork is known in literature as a branch of non photorealistic rendering [17]. This class of works use texture transfer [7,30] to achieve style transfer. These techniques are non-parametric and directly alter image pixels of the content image into pre-defined styles. Another direction of work focuses on the idea of separating style and content in order to 'remix' them together in different configurations. First works were evaluated on much simpler images such as characters in different handwritings [29] or images representing human body configurations [8]. Only recently, the breakthrough paper from Gatys *et al.* [11] showed the possibility of disentangling the content from the style of natural images by using a convolutional neural network based representation. The advantage of this approach is the capability of performing style transfer from any painting to any kind of content images. The approach was recently extended with a more advanced perceptual loss [14] and also applied to movies [1] by considering the optical flow.

2 The System

The system is composed by two main components: a mobile App and a computer vision system responsible to address the two tasks of automatically recognize artworks and apply artwork styles to user photos. The mobile App is used by the visitor in the museum and is the *fulcrum* of the user interaction. Once installed by the user in his mobile phone, it allows to take pictures, deliver artwork information and request the style transfer to new photos. Due to the limited amount of computational power available on most mobile devices, the computer vision system is deployed on a scalable web server system that processes requests from the mobile App. Since the two tasks use quite different technologies, we discuss them separately in the following sections.

2.1 The Mobile App

Imaging Novecento has been developed as an Android application using Ionic[1]. Ionic is a framework, based on Sass and AngularJS, for building highly interactive native web apps through mobile-optimized HTML, CSS and JS components and tools. *Imaging Novecento* is a contextual App that can be used exclusively inside the Museo Novecento in Florence. An information flyer of the App is delivered to the visitor at the ticket office. In the flyer there are a QR code, through which the visitor can download the App from the Google Play store, and the list of the artworks on which the App can perform the automatic recognition and style transfer processes. The list comprises a selection of twenty artworks for which the museum's curators have provided multimedia materials. The App interface (Fig. 1) is quite simple and is organized in two main views: (1) the Camera View and (2) the Artwork Details view.

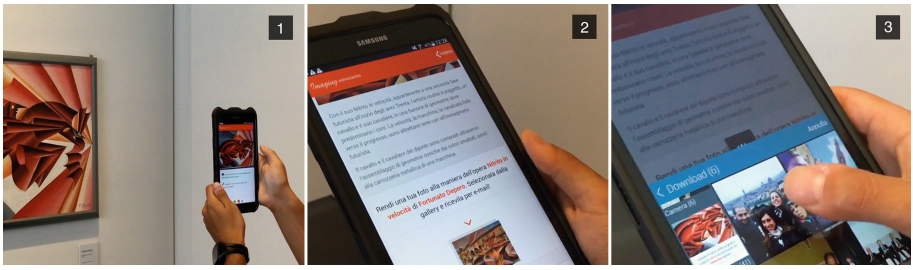

Fig. 1. *Imaging Novecento* in action: (1) the user takes a picture of an artwork; (2) the artwork is recognized and insights are shown; (3) the user selects a photo from his own gallery in order to apply that artwork style and to share the results on social networks.

The Camera View allows the visitor to frame one of the artworks on the list in order to have it immediately recognized by the automatic system. Proper feedback is given in case the recognition is not successful. Once the artwork is recognized, the Artwork Details view is activated. In this view, exhaustive but concise information about the author, the history of the artwork and its artistic style are given. An infographic is presented to the user. It works as a "call to action" for enabling the transfer of the recognized painting style to a photo from the user's device gallery. The infographic provides an animated preview that shows the result of the artistic style transfer on a predefined picture. After the image has been successfully uploaded, the remote process for style transfer is performed. The result of the elaboration is then sent to the user in a few minutes by email. The image has a resolution of 900 px wide preserving the original image aspect ratio and can be shared on the most popular social networks (e.g. Facebook).

[1] http://ionicframework.com/.

2.2 Automatic Artwork Recognition

Artwork recognition is performed through a Python web server with a REST interface. The server processes the image and returns the ID of the recognized painting. The recognition step combines modern deep features with classical Support Vector Machines (SVM) in order to classify photos of paintings. Image features are extracted using a deep convolutional neural network (CNN), and are then evaluated using a set of classifiers, one for each recognizable artwork. The neural network we adopted is the Caffe reference model [13], fine-tuned for style recognition using the FlickrStyle dataset [16][2].

In order to obtain a representation which is at the same time semantically meaningful and capable of retaining low level visual content information, we extract image features from an intermediate level of the network. In particular, we adopt the *pool5* feature map, the latest one before the fully connected (FC) layers of the CNN. In fact, FC layers trade spatial information for a more semantic representation, which is highly coupled with the task and with the visual domain on which the network has been trained. This choice is therefore motivated by the fact that our visual domain, while being quite close, is different from the one of FlickrStyle. Moreover, since a sufficiently large dataset was not available to perform a further fine-tuning step, SVM classifiers have been trained to adapt the framework to the App's domain and be able to classify artworks correctly. For training the classifiers we used approximately 1,800 images, gathered at the museum using different smartphones and tablets, namely Galaxy S4, Galaxy Tab, iPhone 6, iPad Mini and OnePlus One. These images represent all of the twenty artworks plus a 'negative' set of images containing other scenes and paintings inside of the museum. They are used to reduce the false positive rate when the user accidentally attempts to recognize other paintings. All the classifiers are One vs All SVMs. During the evaluation phase, the ID of the highest scoring one is returned to the mobile App, if it scores above a cross validated threshold. Details about the recognized artwork are then provided to the user, who can upload a personal photo to get the style of the painting transferred on to it. Calls to the webserver are handled asynchronously and each request takes approximately 300 ms on a CPU.

In order to test the recognition accuracy we collected an additional set of photos which were not used for training. For each one of the twenty artworks in our system, we collected approximately 30 photos taken from different viewpoints, with different scales and degrees of occlusion. Figure 2 shows some of the photos from the test set. Some of them are "difficult" in a sense that might be blurred or taken from challenging viewpoints and artworks may be partially occluded by other visitors. Despite these difficulties our system achieves an overall good performance with a mean accuracy of 94.01 %. In detail, in Fig. 3 we report the confusion matrix for the twenty artworks in the test set, showing how often each painting is correctly classified or confused with other artworks. As can be seen,

[2] The network is available online at http://caffe.berkeleyvision.org/gathered/examples/finetune_flickr_style.html.

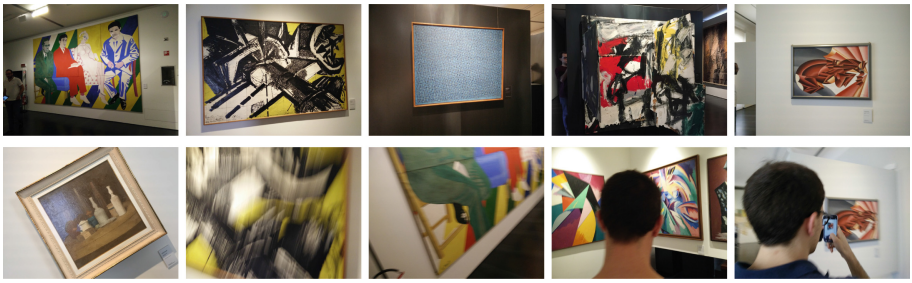

Fig. 2. Samples from the dataset collected at the museum. On the first row standard pictures are shown, depicting the painting in their entirety. On the second row instead, are reported more challenging photos, due to blur, occlusion or rotation.

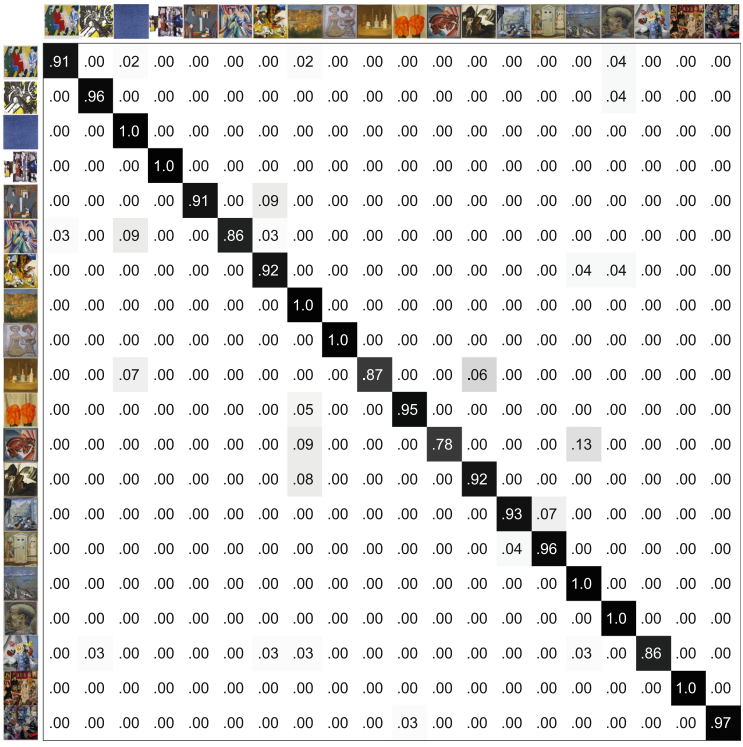

Fig. 3. Confusion matrix for the artwork recognition module. Each row indicates the percentages of correct and incorrect classifications for a given artwork.

the majority of artworks are perfectly recognized. Only four artworks have performance slightly inferior to 0.9, due to the difficult lightening conditions present in their specific locations at the museum.

2.3 Artistic Style Transfer

From the Artwork Detail view of the mobile App, the user has the possibility to upload a personal image on which the style of the artwork will be applied. In this way, entertaining personal pictures that share similarities with the artworks can be obtained and shared on social networks. As a result, a visit at the museum can become a playful experience, combining gaming and learning aspects for young visitors. We base our approach on that of Gatys *et al.* [11], that is capable of freely mixing style and content of two different photos. The main advantage of this approach is its broad applicability to different styles, in contrast to fixed handcrafted styles [7,30]. This allows a museum curator to easily add new artworks in the system without requiring the development of a new transfer style algorithm. Following [11], our approach uses a CNN to derive a neural representation of content and style. The feature responses of a pre-trained network on object recognition (VGG-19 [26]) are used to capture the appearance of an artwork image and the content of a user photo under the form of texture information. We start from a blank novel image that is altered with back-propagation until its neural representation is similar in terms of euclidean distance to the style and content representations (Fig. 4).

Fig. 4. Two examples of image stylization: (1) Baccio Maria Bacci, "Il tram di Fiesole", applied to a picture of the Battistero in Florence, IT; (2) Alberto Moretti, "Malcom X ed altri", applied to a picture of Piazza della Repubblica, also in Florence.

Unfortunately, the generation of the image is quite computational intensive. For an image of 900 pixel large, it takes about ∼90 seconds on a K80 NVIDIA GPU. As a result, the requests have to be handled offline since it is not possibile to obtain the output image in few seconds. Considering also that multiple requests can be made at the same time from multiple users, we implemented a scalable web server that is able to be easily deployed on several interconnected

nodes. Web requests are handled in Python and enqueued to a distributed queue run by a Celery[3] server. By treating each request as a single unit task, it allows to process the images in a distributed batch fashion on several GPUs and several servers if available. After completing the computation, each output image is sent to the user via email, together with a description of the artwork. We also include links to share the image to several social media, with the aim of enabling viral publicity of the museum.

3 Conclusion

We presented the *Imaging Novecento* App, recently developed for the "Museo Novecento" in Florence, IT. Following previous studies on cultural heritage audience and applications, the App aims at enhancing the experience in the museum reducing cognitive load and exploiting gamification. The App automatically recognizes a selection of paintings and provides insights on artworks and their authors. The user can upload a personal picture with his smartphone to get it stylized with the recognized artwork style. He also has the possibility of sharing it on social networks. In the paper we show how computer vision technologies can be exploited to increase interactivity and reduce cognitive load. This can attract the targeted audience to the museum and further engage people with content.

Acknowledgments. We acknowledge the support of the "Museo Novecento" in Florence, IT, and the Municipality of Florence, IT. This research was supported by "THE SOCIAL MUSEUM AND SMART TOURISM", MIUR project no. CTN01_00034_23154_SMST.

References

1. Anderson, A.G., Berg, C.P., Mossing, D.P., Olshausen, B.A.: Deepmovie: using optical flow and deep neural networks to stylize movies. arXiv preprint arXiv:1605.08153 (2016)
2. Anwer, R.M., Khan, F.S., van de Weijer, J., Laaksonen, J.: Combining holistic and part-based deep representations for computational painting categorization. In: Proceedings of the 2016 ACM ICMR, pp. 339–342. ACM (2016)
3. Ballagas, R., Rohs, M., Sheridan, J.G., Borchers, J.: Byod: bring your own device. In: Proceedings of the Workshop on Ubiquitous Display Environments, Ubicomp, vol. 2004 (2004)
4. Baraldi, L., Paci, F., Serra, G., Benini, L., Cucchiara, R.: Gesture recognition using wearable vision sensors to enhance visitors Museum experiences. IEEE Sens. J. **15**(5), 2705–2714 (2015)
5. Chen, C.Y., Chang, B.R., Huang, P.S.: Multimedia augmented reality information system for Museum guidance. Pers. Ubiquit. Comput. **18**(2), 315–322 (2014)
6. Crowley, E., Zisserman, A.: The state of the art: object retrieval in paintings using discriminative regions. In: BMVC (2014)

[3] http://www.celeryproject.org.

7. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 341–346. ACM (2001)
8. Elgammal, A., Lee, C.S.: Separating style and content on a nonlinear manifold. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 1, pp. I–478. IEEE (2004)
9. Fanini, B., d'Annibale, E., Demetrescu, E., Ferdani, D., Pagano, A.: Engaging and shared gesture-based interaction for museums the case study of K2R international expo in rome. In: 2015 Digital Heritage, vol. 1, pp. 263–270. IEEE (2015)
10. Fineschi, A., Pozzebon, A.: A 3D virtual tour of the Santa Maria della Scala Museum Complex in Siena, Italy, based on the use of Oculus RIFT HMD. In: 2015 International Conference on 3D Imaging (IC3D), pp. 1–5. IEEE (2015)
11. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015)
12. He, Z., Cui, B., Zhou, W., Yokoi, S.: A proposal of interaction system between visitor and collection in Museum hall by iBeacon. In: 2015 10th International Conference on Computer Science & Education (ICCSE), pp. 427–430. IEEE (2015)
13. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM MM, pp. 675–678. ACM (2014)
14. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. arXiv preprint arXiv:1603.08155 (2016)
15. Johnson, L., Adams Becker, S., Estrada, V., Freeman, A.: The NMC Horizon Report: 2015 Museum Edition. ERIC (2015)
16. Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., Winnemoeller, H.: Recognizing image style. arXiv preprint arXiv:1311.3715 (2013)
17. Kyprianidis, J.E., Collomosse, J., Wang, T., Isenberg, T.: State of the "Art": a taxonomy of artistic stylization techniques for images and video. IEEE Trans. Vis. Comput. Graph. **19**(5), 866–885 (2013)
18. Malomo, L., Banterle, F., Pingi, P., Gabellone, F., Scopigno, R.: Virtualtour: a system for exploring cultural heritage sites in an immersive way. In: 2015 Digital Heritage, vol. 1, pp. 309–312. IEEE (2015)
19. Mayer, R.E., Moreno, R.: Nine ways to reduce cognitive load in multimedia learning. Educ. Psychol. **38**(1), 43–52 (2003)
20. Negruşa, A.L., Toader, V., Sofică, A., Tutunea, M.F., Rus, R.V.: Exploring gamification techniques and applications for sustainable tourism. Sustainability **7**(8), 11160–11189 (2015)
21. Pagano, A., Armone, G., De Sanctis, E.: Virtual Museums and audience studies: the case of "Keys to Rome" exhibition. In: 2015 Digital Heritage, vol. 1, pp. 373–376. IEEE (2015)
22. Park, S.: The effects of social cue principles on cognitive load, situational interest, motivation, and achievement in pedagogical agent multimedia learning. J. Educ. Technol. Soc. **4**, 211–229 (2015)
23. Peng, K.C., Chen, T.: Cross-layer features in convolutional neural networks for generic classification tasks. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 3057–3061. IEEE (2015)
24. Reiners, T., Wood, L.C. (eds.): Gamification in Education and Business. Springer International Publishing, Switzerland (2015)

25. Ruf, B., Kokiopoulou, E., Detyniecki, M.: Mobile Museum guide based on fast SIFT recognition. In: Detyniecki, M., Leiner, U., Nürnberger, A. (eds.) AMR 2008. LNCS, vol. 5811, pp. 170–183. Springer, Heidelberg (2010). doi:10.1007/978-3-642-14758-6_14
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
27. de Sousa Borges, S., Durelli, V.H., Reis, H.M., Isotani, S.: A systematic mapping on gamification applied to education. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing, pp. 216–222. ACM (2014)
28. Temmermans, F., Jansen, B., Deklerck, R., Schelkens, P., Cornelis, J.: The mobile Museum guide: artwork recognition with eigenpaintings and SURF. In: WIAMIS 2011, Delft, The Netherlands, April 13–15, 2011. TU Delft; EWI; MM; PRB (2011)
29. Tenenbaum, J.B., Freeman, W.T.: Separating style and content with bilinear models. Neural Comput. **12**(6), 1247–1283 (2000)
30. Xie, X., Tian, F., Seah, H.S.: Feature guided texture synthesis (FGTS) for artistic style transfer. In: Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts, pp. 44–49. ACM (2007)
31. Yoon, S.A., Wang, J.: Making the invisible visible in science museums through augmented reality devices. TechTrends **58**(1), 49–55 (2014)