A Deep Learning-based Approach to Model Museum Visitors

Alessio Ferrato[a], Carla Limongelli[a], Mauro Mezzini[b] and Giuseppe Sansonetti[a]

[a]*Department of Engineering, Roma Tre University, Via della Vasca Navale 79, 00146 Rome, Italy*
[b]*Department of Education, Roma Tre University, Viale del Castro Pretorio 20, 00185 Rome, Italy*

Abstract

Although ubiquitous and fast access to the Internet allows us to admire objects and artworks exhibited worldwide from the comfort of our home, visiting a museum or an exhibition remains an essential experience today. Current technologies can help make that experience even more satisfying. For instance, they can assist the user during the visit, personalizing her experience by suggesting the artworks of her higher interest and providing her with related textual and multimedia content. To this aim, it is necessary to automatically acquire information relating to the active user. In this paper, we show how a deep neural network-based approach can allow us to obtain accurate information for understanding the behavior of the visitor alone or in a group. This information can also be used to identify users similar to the active one to suggest not only personalized itineraries but also possible visiting companions for promoting the museum as a vehicle for social and cultural inclusion.

Keywords

User interfaces, Computer vision, Deep Learning, Museum visitors

1. Introduction and Background

Recent technological advances have made it possible to significantly improve the experience of citizens when they use public services [1, 2] or when they enjoy points of their interest [3, 4] and itineraries among them [5, 6]. Among the different possible points of interest, there are also museums and exhibits. The first studies concerning the observation and analysis of museum visitor behavior date back to the first half of the twentieth century [7, 8, 9]. Since then, the works that publish studies based on the analysis of visitor tracking have multiplied, namely, on the detailed recording of "not only where visitors go but also what visitors do while inside an exhibition" [10]. In early studies on the subject, the most common method for recording visitor behavior was the paper-and-pencil one. Although this method is simple and low-cost, several aspects limit its validity. Among these, the lack of temporal information, more complicated to collect by the observer, the need to transfer the data collected on paper to the database, and the inability to accurately determine the visitor's real engagement. Fortunately, recent technological advances in Machine Learning [11] have made available to researchers several approaches for automatic visitor tracking [12]. In the research literature, there are several contributions on the technologies adopted for the timing and tracking of museum visitors, highlighting the strengths and weaknesses of each of them (e.g., see [13, 14, 15, 12, 16, 17, 18, 19, 20, 21]). In [22], we have proposed an approach based on Computer Vision technologies [23] that can represent the solution of some of the criticalities shown by the other visitor localization technologies in the museum environment. It takes advantage of image detection and classification techniques through convolutional neural networks (CNNs) capable of providing excellent performance in terms of accuracy [24]. This approach makes use of off-the-shelf cameras and badges such as those provided free to attendees by event and conference organizers [25]. Therefore, the overall cost of the entire instrumentation is reduced, which is certainly a significant advantage over other state-of-the-art technologies. More specifically, this system relies on the Faster Region-based Convolutional Neural Network (Faster R-CNN) model [26]. The architecture of the proposed system is shown in Figure 1. It can be divided into three major parts:

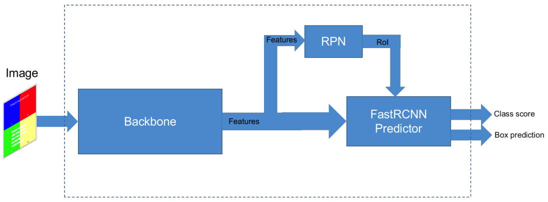

Figure 1: The architecture of the proposed system relies on a Faster Region-based Convolutional Neural Network (Faster R-CNN).

1. A CNN backbone (composed of a ResNet [27] and a Feature Pyramid Network (FPN) [28]) that

Joint Proceedings of the ACM IUI Workshops 2022, March 2022, Helsinki, Finland

ale.ferrato@stud.uniroma3.it (A. Ferrato); limongel@dia.uniroma3.it (C. Limongelli); mauro.mezzini@uniroma3.it (M. Mezzini); gsansone@dia.uniroma3.it (G. Sansonetti)
0000-0003-4953-1390 (G. Sansonetti)

receives in input the image and gives in output a conv feature map;

2. A Region Proposal Network (RPN) that takes as input the conv feature map in output from the backbone and returns a set of rectangular boxes, each of which is associated with a score giving the likelihood that the region contains an object or simple background;
3. The conv feature map given in output from the backbone is also used by a detection system that, given the set of regions from the RPN, determines, using the conv feature map of the backbone, for each distinct class, a score $\in [0, 1]$. This value represents the likelihood that the object belongs to the corresponding class. The detection system also estimates the regression bounding box coordinates x, y, w, h of the object being proposed by the RPN.

In this paper, we show how the gathered information can be used to model museum visitors and identify their nearest neighbors. The aim is to promote the museum as a vehicle to foster the social and cultural inclusion of its visitors.

2. Database with Collected Data

In order to make the analysis of the data collection easy and at the same time effective, we propose the following database implementation and give some sample queries that could cover the most basic and useful needs when a museum staff member wants to extract useful information about visitor behavior from the database. The data collected through the proposed system can be stored in a data structure that supports spatial and temporal analyses of visitor behavior [29]. Let us suppose, for example, that we have m cameras and n badges. Each camera detects, at a generic timestamp, a badge at certain coordinates from the camera. We can store all those detections in a database composed of two tables. The first table, called *positions*, has attributes (TIMESTMP, CAMERA_ID, BADGE_ID, X, Y, Z) and the second table, called *camera*, has attributes (CAMERA_ID, CT, X, Y, Z). A single tuple (t, c_id, b_id, x, y, z) of *position* represents a detection at timestamp t from the camera c_id of the badge b_id at coordinates x, y, z with respect to camera c_id. A single tuple (c_id, ct, x, y, z) of *camera* represents the coordinates x, y, z of the camera c_id in relation to the museum. The value ct is the time period of a frame. If f is the frame rate of the camera, then we have $ct = 1/f$. For the sake of simplicity, hereafter, we suppose that ct assumes the same value for all cameras (i.e., 1/24 s), but all the discussion can be extended with simple and minimal modifications to the general case, in which cameras can have different frame rates. We note that, whilst the table *position* is fed by the detections of the model, the table *camera* is determined and created in advance by the system supervisor. First of all, it can be convenient to create the view *dist_positions* using the SQL Query 1.

Query 1: View creation

```
CREATE VIEW dist_positions AS
SELECT DISTINCT P.TIMESTMP as TIMESTMP, P.BADGE_ID as
      BADGE_ID, C.CT as CT,
    /* Changing the reference system */
    P.X + C.X AS X,
    P.Y + C.Y AS Y,
    P.Z + C.Z AS Z
FROM positions P, camera C
WHERE P.CAMERA_ID = C.CAMERA_ID
```

Furthermore we add another table *artwork* with attributes (ID, AX, AY, AZ, AW, AH). A tuple (id, x, y, z, w, h) of table *artwork* records the *id* of the artwork, the upper right corner coordinates x, y, z (with respect to the museum), the height h and width w of a rectangular box in front of the artwork *id*.

Using this simple database we may easily and effectively retrieve the data described in the previous section. In particular, we may obtain the distances between all pairs of distinct visitors (identified by a badge ID) and at the same choose only those pairs which have a distance less than a prefixed threshold α. In fact, we may first create the SQL view *visit_times* (Query 2) that computes for each artwork the total time a badge has spent nearby the artwork, for all badge ids and for the time interval between t_0 and t_1. In the following, for the sake of simplicity, we assume that every badge has been detected in front of every possible artwork. By this assumption, the query *visit_times* returns for every badge and every artwork at least a value.

Query 2: Badge-Artwork visit time

```
CREATE VIEW visit_times AS
SELECT p.BADGE_ID as BADGE_ID, a.ID as A_ID, SUM(p.CT)
      as V_TIMES
FROM  dist_positions p, artwork a
WHERE p.TIMESTMP BETWEEN t_0 AND t_1 AND
      p.X BETWEEN a.AX AND a.AX + a.AW AND
      p.Y BETWEEN a.AY AND a.AY + a.AH AND
      p.Z BETWEEN a.AZ AND a.AZ + 2.70
GROUP BY p.bid, a.id
```

Using this view, we can compute the total time a single badge spent in front of any artwork (i.e, *total_visit_times* in Query 3).

Query 3: Badge total visit time

```
CREATE VIEW total_visit_times AS
SELECT BADGE_ID,SUM(V_TIMES) as TOTAL_TIMES
FROM  visit_times
GROUP BY BADGE_ID
```

Then, using the two previous views, namely, *visit_times* and *total_visit_times*, we can compute the percentage of time spent by any badge in front of any artwork (i.e., *total_visit_times_perc* in Query 4).

Query 4: Badge-Artwork percentage time

```
CREATE VIEW visit_times_perc AS
SELECT BADGE_ID, A_ID, V_TIMES*100/TOTAL_TIMES as
      V_TIMES_PERC
FROM  visit_times v, total_visit_times t
WHERE v.BADGE_ID = t.BADGE_ID
```

Then, using the two previous views, namely, *visit_times* and *total_visit_times*, we can compute the percentage of time spent by any badge in front of any artwork (i.e., *total_visit_times_perc* in Query 4).

Query 5: Badge-Badge distances

```
SELECT v1.BADGE_ID, v2.BADGE_ID, sum(abs(v1.
      V_TIMES_PERC-v2.V_TIMES_PERC))/200
FROM  visit_times_perc v1, visit_times_perc v2
WHERE v1.A_ID = v2.A_ID AND
      v1.BADGE_ID<v2.BADGE_ID
GROUP BY v1.BADGE_ID, v2.BADGE_ID
HAVING sum(abs(v1.V_TIMES_PERC-v2.V_TIMES_PERC))/200<
      ALPHA
```

3. Distance

We want to define a measure that allows us to compare two visitors according to the time they spend in front on each artwork. Instead of considering absolute values, we reason on a percentage basis. Let us assume that each visitor spends a unit of time at the exhibition and let us calculate the percentage of the time spent observing each artwork. Let us define:

- N: the number of artworks present in the exhibition;
- v_i: the i-th visitor;
- k: the number of visitors we are going to track
- t_i: the overall time that v_i spends to visit the entire exhibition. In this way, the time spent on each artwork is calculated as a percentage and $\sum_{i=1}^{N} t_i = 100$.

We can define a model for the i-th visitor as:

$$um_i = \{t_1^i, \dots, t_N^i\}$$

where $t_h^i, h = 1, \dots, N$ is the percent time that the the visitor i spends in front of the artwork h. The comparison of the previous visitor's model with the the following j-th model:

$$um_j = \{t_1^j, \dots, t_N^j\}$$

is defined by the following measure:

$$\delta(um_i, um_j) = (|t_1^i - t_1^j| + \dots + |t_n^i - t_n^j|)/200 = \frac{1}{200}\sum_{h=1}^{N} |t_h^i - t_h^j| \tag{1}$$

In this way, the measure is a real number in $[0, \dots, 1]$. The definition given in 1 is a measure, as it enjoys properties of *Positiveness*, *Minimality* and *Simmetry*.

Positiveness Formula 1 is a sum of positive values, consequently this property is fulfilled.

Minimality When two *ums* coincide, their distance has to be 0, which means that the times spent for each artwork are the same ($t_h^i = t_h^j, \forall h = 1, \dots, N$) and the overall sum is 0. On the contrary, if two *ums* are completely different, it means that the two visitors have seen different artworks. When $t_n^i \geq 0, t_n^j = 0$ and vice-versa. In this way, the sum is equal to 1.

Simmetry It is given by the absolute value of the observation time difference.

4. Data analysis

Data collection and analysis is useful for museum curators and staff members because it allows them to work on the fruition of the exhibition itself, as well as on the visitor flow. However, it also makes more refined analyzes possible. Our system analyzes the video recorded by the cameras positioned near each point of interest (POI) (i.e., artwork) and from each frame. It can identify the position in pixels of the badge and the distance from the camera for each visitor positioned in front of the POI up to a distance of 6 meters. By collecting and triangulating the data from the different recordings, we can track the path of each visitor in each room of the museum. Some of the analyses that can be performed on the data acquired by this system are:

- Temporal analysis regarding the time spent in front of the artworks;
- Analysis of the trajectory of visitors in front of the artworks;
- Spatial-temporal analyses between viewing distance and time spent in front of an artwork;
- Heatmap of temporal data on visitor positions in the room.

The system also allows us to perform a more refined analysis because it allows facial identification. Therefore, it enables us to collect specific data about the visitor's attention to the artwork. If the face is not identified by the system, we can assume that the visitor is distracted and, therefore, increase the granularity of the analysis.

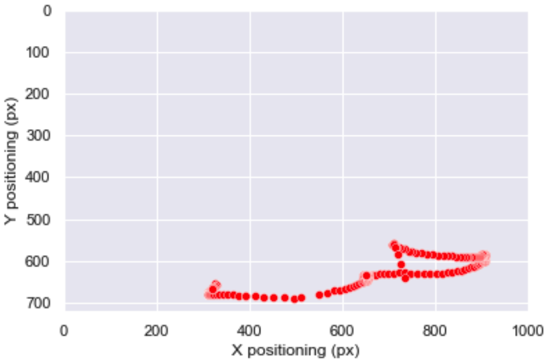

Figure 2: Movements of one visitor.

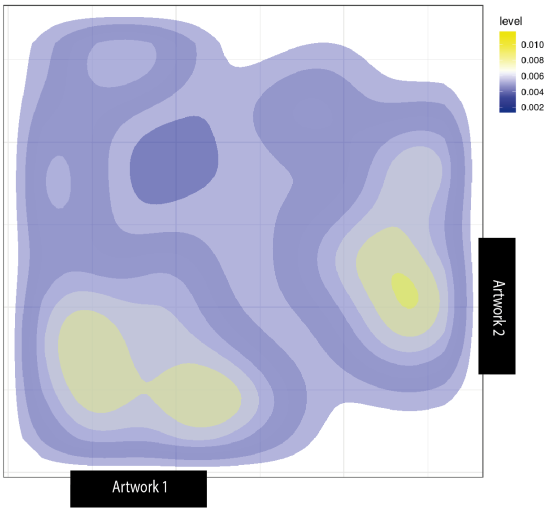

Figure 3: An heatmap of one room with two artworks.

If we shift the focus to the visitor, we can exploit this information to improve the sociality in the museum environment and break down cultural barriers by exploiting the similarity between the behavior of different visitors.

5. Conclusions and Future Works

In this paper, we have presented a deep learning-based approach that allows museum curators and staff members to accurately collect data relating to the visitors' behavior. We have seen how such data can be exploited to identify possible communities [30] of users to transform the museum into a place of social and cultural inclusion. The sharing of interests and preferences can foster the aggregation of individuals, for example, belonging to different cultures.

The possible future developments of the research work presented herein are manifold. First of all, there is the integration of the data collection system within a social recommender system [31, 32, 33]. This would allow us to assess the effective benefits of our system in terms of the inclusion of individuals from different backgrounds [34].

References

[1] G. D'Aniello, M. Gaeta, I. La Rocca, KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis, Artificial Intelligence Review (2022).

[2] G. D'Aniello, M. Gaeta, F. Orciuoli, G. Sansonetti, F. Sorgente, Knowledge-based smart city service system, Electronics (Switzerland) 9 (2020) 1–22.

[3] G. Sansonetti, Point of interest recommendation based on social and linked open data, Personal and Ubiquitous Computing 23 (2019) 199–214.

[4] G. Sansonetti, F. Gasparetti, A. Micarelli, Cross-domain recommendation for enhancing cultural heritage experience, in: Adjunct Publication of the 27th ACM UMAP Conference, ACM, New York, NY, USA, 2019, pp. 413–415.

[5] A. Fogli, G. Sansonetti, Exploiting semantics for context-aware itinerary recommendation, Personal and Ubiquitous Computing 23 (2019) 215–231.

[6] D. D'Agostino, F. Gasparetti, A. Micarelli, G. Sansonetti, A social context-aware recommender of itineraries between relevant points of interest, in: HCI International 2016, volume 618, Springer International Publishing, Cham, 2016, pp. 354–359.

[7] E. S. Robinson, I. C. Sherman, L. E. Curry, H. H. F. Jayne, The behavior of the museum visitor, Publications of the American Association of Museums 1 (1928) 72.

[8] A. W. Melton, Problems of installation in museums of art, Publications of the American Association of Museums 7 (1935) 29–30.

[9] A. W. Melton, Distribution of attention in galleries in a museum of science and industry, Museum News 14 (1935) 6–8.

[10] S. S. Yalowitz, K. Bronnenkant, Timing and tracking: Unlocking visitor behavior, Visitor Studies 12 (2009).

[11] L. Vaccaro, G. Sansonetti, A. Micarelli, An empirical review of automated machine learning, Computers 10 (2021).

[12] F. Zafari, A. Gkelias, K. K. Leung, A survey of indoor localization systems and technologies, IEEE Communications Surveys Tutorials 21 (2019) 2568–2599.

[13] P. Centorrino, A. Corbetta, E. Cristiani, E. Onofri,

Managing crowded museums: Visitors flow measurement, analysis, modeling, and optimization, Journal of Computational Science 53 (2021) 101357.

[14] M. G. Rashed, R. Suzuki, T. Yonezawa, A. Lam, Y. Kobayashi, Y. Kuno, Tracking visitors in a real museum for behavioral analysis, in: Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS), IEEE, 2016, pp. 80–85. Sapporo, Japan, 25–28 August 2016.

[15] R. Angeloni, R. Pierdicca, A. Mancini, M. Paolanti, A. Tonelli, Measuring and evaluating visitors' behaviors inside museums: the co. me. project, SCIRES-IT-SCIentific RESearch and Information Technology 11 (2021) 167–178.

[16] M. G. Rashed, R. Suzuki, T. Yonezawa, A. Lam, Y. Kobayashi, Y. Kuno, Robustly tracking people with lidars in a crowded museum for behavioral analysis, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E100A (2017) 2458–2469.

[17] A. Kontarinis, C. Marinica, D. Vodislav, K. Zeitouni, A. Krebs, D. Kotzinos, Towards a better understanding of museum visitors' behavior through indoor trajectory analysis, in: Digital Presentation and Preservation of Cultural and Scientific Heritage, volume 7, 2017, pp. 19–30.

[18] R. Giuliano, G. C. Cardarilli, C. Cesarini, L. Di Nunzio, F. Fallucchi, R. Fazzolari, F. Mazzenga, M. Re, A. Vizzarri, Indoor localization system based on bluetooth low energy for museum applications, Electronics 9 (2020).

[19] P. Spachos, K. N. Plataniotis, Ble beacons for indoor positioning at an interactive iot-based smart museum, IEEE Systems Journal 14 (2020) 3483–3493.

[20] C. Martella, A. Miraglia, J. Frost, M. Cattani, M. V. Steen, Visualizing, clustering, and predicting the behavior of museum visitors, Pervasive Mob. Comput. 38 (2017) 430–443.

[21] A. Emerson, N. Henderson, J. Rowe, W. Min, S. Lee, J. Minogue, J. Lester, Early prediction of visitor engagement in science museums with multimodal learning analytics, in: Proceedings of the 2020 International Conference on Multimodal Interaction, 2020, pp. 107–116.

[22] A. Ferrato, C. Limongelli, M. Mezzini, G. Sansonetti, Using deep learning for collecting data about museum visitor behavior, Applied Sciences 12 (2022).

[23] A. Micarelli, A. Neri, G. Sansonetti, A case-based approach to image recognition, in: Proceedings of the 5th European Workshop on Advances in Case-Based Reasoning, EWCBR '00, Springer-Verlag, Berlin, Heidelberg, 2000, pp. 443–454.

[24] G. Sansonetti, F. Gasparetti, G. D'Aniello, A. Micarelli, Unreliable users detection in social media: Deep learning techniques for automatic detection, IEEE Access 8 (2020) 213154–213167.

[25] M. Mezzini, C. Limongelli, G. Sansonetti, C. De Medio, Tracking museum visitors through convolutional object detectors, in: Adjunct Publication of the 28th ACM UMAP Conference, ACM, New York, NY, USA, 2020, pp. 352–355.

[26] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Proceedings of the 28th NIPS Conference - Volume 1, MIT Press, Cambridge, MA, US, 2015, pp. 91––99.

[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2016, pp. 770–778.

[28] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: 2017 IEEE CVPR, IEEE Computer Society, Los Alamitos, CA, USA, 2017, pp. 936–944.

[29] F. Gasparetti, A. Micarelli, G. Sansonetti, Exploiting web browsing activities for user needs identification, in: 2014 International Conference on Computational Science and Computational Intelligence, volume 2, 2014, pp. 86–89.

[30] F. Gasparetti, G. Sansonetti, A. Micarelli, Community detection in social recommender systems: a survey, Applied Intelligence 51 (2021) 3975–3995.

[31] S. Caldarelli, D. F. Gurini, A. Micarelli, G. Sansonetti, A signal-based approach to news recommendation, in: CEUR Workshop Proceedings, volume 1618, CEUR-WS.org, Aachen, Germany, 2016, pp. 1–4.

[32] H. A. M. Hassan, G. Sansonetti, F. Gasparetti, A. Micarelli, J. Beel, Bert, elmo, USE and infersent sentence encoders: The panacea for research-paper recommendation?, in: M. Tkalcic, S. Pera (Eds.), Proceedings of ACM RecSys 2019 Late-Breaking Results, volume 2431, CEUR-WS.org, 2019, pp. 6–10.

[33] I. Guy, Social recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer US, Boston, MA, 2015, pp. 511–543.

[34] R. Cordier, B. T. Milbourn, R. Martin, A. Buchanan, D. Chung, R. Speyer, A systematic review evaluating the psychometric properties of measures of social inclusion, PLoS ONE 12 (2017).