Pattern Recognition Letters
journal homepage: www.elsevier.com

EGO-CH: Dataset and Challenges for Visitors Behavioral Understanding from Egocentric Vision (Supplementary Material)

Francesco Ragusa[a,b,2], Antonino Furnari[a,2], Sebastiano Battiato[a], Giovanni Signorello[c], Giovanni Maria Farinella[a,c,2,**]

[a] *DMI-IPLab, University of Catania*
[b] *XGD - XENIA s.r.l., Acicastello, Catania, Italy*
[c] *CUTGANA, University of Catania*

ABSTRACT

This supplementary material complements the submitted paper. It reports additional details on the EGO-CH dataset and experiments.

© 2019 Elsevier Ltd. All rights reserved.

1. INTRODUCTION

This document is intended for the convenience of the reader and reports additional information about the proposed dataset and the performed experiments. This supplementary material is related to the following submission:

- F. Ragusa, A. Furnari, S. Battiato, G. Signorello, G. M. Farinella, "EGO-CH: Dataset and Challenges for Visitors Behavioral Understanding from Egocentric Vision", Pattern Recognition Letters, DOI: 10.1016/j.patrec.2019.12.016

The reader is referred to the manuscript and to our web page *http://iplab.dmi.unict.it/EGO-CH/* for further information.

2. THE EGO-CH DATASET

2.1. Data Collection

The dataset has been acquired using a head-mounted Microsoft HoloLens device in two cultural sites located in Sicily, Italy: 1) "Palazzo Bellomo", located in Siracusa [1], and 2) "Monastero dei Benedettini", located in Catania [2].

1) Palazzo Bellomo. This cultural site is composed by 22 environments and contains 191 Points of Interest (e.g., statues, paintings, etc.). Figure 1 and Figure 2 report some frames related to the different environments and some points of interest.

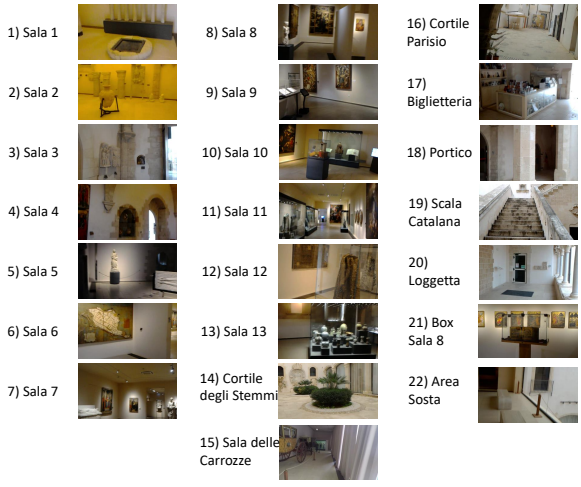

Fig. 1. Sample frames for each of the 22 considered environments of "Palazzo Bellomo".

Table 7 details the list of the acquired training video. Some of the videos are related to the 22 rooms of the cultural site, whereas other are related to specific points of interest. For each video we report its total duration, the amount of required storage, the number of frames, as well as the percentage of frames with respect to the whole training set.

Table 8 reports the list of the 10 test videos acquired by volunteers visiting the cultural site. For each video, we report its total duration, the amount of required storage, the number of frames, the number of environments encountered in the video, as well as the sequence of environments, as visited by the sub-

**Corresponding author: Tel.: +39 095 7337 219; fax: +39 095 330094;
e-mail: gfarinella@dmi.unict.it (Giovanni Maria Farinella)
[2] These authors are co-first authors and contributed equally to this work.

Table 1. FF_1 **Results for Room- Based Localization on Palazzo Bellomo.**

	FF_1 **score**									
Class	Test1	Test2	Test4	Test5	Test6	Test7	Test8	Test9	Test10	AVG
1_Sala1	0,00	0,96	0,96	0,95	0,99	0,94	0,97	0,97	0,91	0,85
2_Sala2	0,82	0,91	0,96	0,95	0,99	0,99	0,99	0,97	0,95	0,95
3_Sala3	0,65	0,84	0,84	0,83	0,94	0,88	0,61	0,85	0,83	0,81
4_Sala4	0,84	0,96	0,98	0,93	/	0,96	0,78	0,97	0,93	0,92
5_Sala5	0,98	0,88	0,96	0,91	0,65	0,91	0,97	0,97	0,99	0,91
6_Sala6	0,84	0,80	0,43	0,00	0,86	0,00	0,18	0,83	0,73	0,52
7_Sala7	0,67	0,93	0,22	0,00	0,38	0,88	0,88	0,40	0,96	0,59
8_Sala8	0,64	0,75	0,60	0,52	0,56	0,56	0,64	0,60	0,77	0,63
9_Sala9	0,90	0,89	0,73	0,88	0,41	0,92	0,98	0,99	0,88	0,84
10_Sala10	0,96	0,98	0,92	0,72	0,00	0,98	0,85	0,97	0,78	0,80
11_Sala11	0,93	0,96	0,96	0,97	0,00	0,97	0,97	0,98	0,95	0,85
12_Sala12	0,82	0,88	0,87	0,90	0,00	0,87	0,85	0,92	0,87	0,78
13_Sala13	0,96	0,95	0,95	0,76	0,00	0,93	0,95	0,94	0,96	0,82
14_CortiledegliStemmi	0,78	0,68	0,74	0,88	0,17	0,90	0,79	0,92	0,00	0,65
15_SalaCarrozze	0,84	0,89	0,95	0,93	0,91	0,84	0,97	0,95	/	0,91
16_CortileParisio	0,33	0,51	0,52	0,45	0,35	0,60	0,81	0,91	0,80	0,59
17_Biglietteria	0,25	0,56	0,00	0,26	0,00	0,00	0,00	0,36	0,00	0,16
18_Portico	0,68	0,67	0,78	0,54	0,69	0,69	0,62	0,78	0,62	0,67
19_ScalaCatalana	0,00	0,00	0,54	0,55	0,78	0,67	0,58	0,85	0,49	0,50
20_Loggetta	0,00	0,50	0,80	0,44	0,72	0,53	0,55	0,83	0,68	0,56
21_BoxSala8	0,96	0,97	0,97	/	0,00	0,97	0,99	0,81	0,94	0,83
22_AreaSosta	0,74	0,85	0,45	0,83	0,00	0,56	0,72	0,85	0,00	0,56
mFF_1	0,66	0,79	0,73	0,68	0,45	0,75	0,76	0,85	0,72	0,71

Table 2. ASF_1 **Results for Room- Based Localization on Palazzo Bellomo.**

	ASF_1 **score**									
Class	Test1	Test2	Test4	Test5	Test6	Test7	Test8	Test9	Test10	AVG
1_Sala1	0,00	0,93	0,93	0,91	0,97	0,89	0,94	0,94	0,84	0,82
2_Sala2	0,64	0,83	0,91	0,90	0,97	0,97	0,97	0,94	0,90	0,89
3_Sala3	0,36	0,53	0,51	0,53	0,61	0,63	0,29	0,59	0,72	0,53
4_Sala4	0,39	0,92	0,96	0,86	/	0,92	0,31	0,94	0,87	0,77
5_Sala5	0,94	0,64	0,66	0,62	0,34	0,63	0,54	0,95	0,97	0,70
6_Sala6	0,46	0,43	0,18	0,00	0,42	0,00	0,09	0,59	0,51	0,30
7_Sala7	0,50	0,86	0,13	0,00	0,10	0,78	0,28	0,25	0,92	0,42
8_Sala8	0,36	0,44	0,42	0,35	0,35	0,41	0,49	0,53	0,57	0,44
9_Sala9	0,81	0,80	0,51	0,78	0,22	0,84	0,95	0,97	0,78	0,74
10_Sala10	0,92	0,96	0,85	0,19	0,00	0,95	0,27	0,94	0,64	0,64
11_Sala11	0,53	0,92	0,91	0,93	0,00	0,93	0,65	0,95	0,91	0,75
12_Sala12	0,38	0,79	0,76	0,81	0,00	0,77	0,30	0,85	0,76	0,60
13_Sala13	0,93	0,90	0,90	0,62	0,00	0,87	0,91	0,89	0,92	0,77
14_CortiledegliStemmi	0,57	0,53	0,54	0,77	0,08	0,81	0,49	0,67	0,00	0,50
15_SalaCarrozze	0,72	0,80	0,91	0,87	0,83	0,72	0,94	0,90	/	0,84
16_CortileParisio	0,30	0,63	0,43	0,65	0,38	0,47	0,59	0,83	0,66	0,55
17_Biglietteria	0,26	0,43	0,00	0,11	0,00	0,00	0,00	0,39	0,00	0,13
18_Portico	0,51	0,47	0,60	0,45	0,48	0,52	0,42	0,64	0,47	0,51
19_ScalaCatalana	0,00	0,00	0,43	0,51	0,61	0,55	0,42	0,75	0,39	0,41
20_Loggetta	0,00	0,31	0,54	0,28	0,46	0,38	0,48	0,72	0,44	0,40
21_BoxSala8	0,92	0,95	0,95	/	0,00	0,93	0,98	0,67	0,88	0,79
22_AreaSosta	0,59	0,73	0,29	0,71	0,00	0,75	0,57	0,74	0,00	0,49
$mASF_1$	0,50	0,67	0,61	0,56	0,32	0,67	0,54	0,76	0,63	0,58

Table 3. FF_1 **Results for Room- Based Localization on Monastero dei Benedettini.**

ID_visit	1	2	3	4	Neg.	AVG	ID_visit	1	2	3	4	Neg.	AVG
4805	0,79	0,82	0,74	0,95	0,45	0,75	2043	0,52	0,44	0,62	0,91	0,33	0,564
1804	0,32	0,36	0,79	0,98	0,55	0,6	3996	0,17	0,68	0,57	0,95	0,72	0,618
4377	0,46	0,53	0,8	0,96	0,32	0,614	3455	0,78	0,65	0,8	0,99	0,55	0,754
1669	0,75	0,72	0,9	0,99	0,45	0,762	4785	0,02	0	0,64	0,95	0,33	0,388
1791	0,49	0,5	0,84	0,95	0,43	0,642	2047	0,95	0,84	0,84	0,98	0,38	0,798
3948	0	/	0,59	0,98	0,49	0,515	1912	0,66	0,51	0,69	0,94	0,46	0,652
3152	0,39	0,72	0,87	0,97	0,76	0,742	3232	0,66	0,53	0,85	0,99	0,49	0,704
4361	0,55	0,78	0,82	0,97	0,28	0,68	4442	0,8	0,77	0,82	0,98	0,23	0,72
3976	0,87	0,43	0,81	0,93	0,66	0,74	3646	0,63	0,29	0,66	0,97	0,4	0,59
3527	0,86	0,79	0,85	0,99	0,56	0,81	4833	0,67	0,82	0,67	0,88	0,24	0,656
4105	0,62	0	0,77	0,97	0,09	0,49	3478	0,71	0,75	0,84	0,98	0,37	0,73
1399	0,46	0,23	0,79	0,98	0,43	0,578	4396	0,74	0,69	0,89	0,96	0,26	0,708
3836	0,51	0	0,72	0,99	0,55	0,554	2894	0,83	0,8	0,66	0,92	0,48	0,738
4006	0,57	0,78	0,78	0,99	0,37	0,698	4414	0,75	0,7	0,91	0,95	0,22	0,706
4415	0,86	0,61	0,82	0,97	0,35	0,722	4639	0,57	0,52	0,85	0,99	0,39	0,664
3008	0,69	0	0,7	0,99	0,51	0,578	1004	0,08	0,72	0,49	0,97	0,26	0,504
4660	0,62	0,81	0,83	0,98	0,57	0,762	1917	0,41	0,71	0,36	0,87	0,29	0,528
2826	0,31	0,52	0,76	0,97	0,61	0,634	1153	0,62	0,7	0,71	0,91	0,32	0,652
1099	0,62	0,42	0,85	0,98	0,49	0,672	2244	0,94	0,56	0,74	0,99	0,43	0,732
4391	0,74	0,72	0,72	0,98	0,33	0,698	2614	0,88	0,56	0,83	0,99	0,4	0,732
3929	0,26	0	0,8	0,99	0,43	0,496	1624	0,33	0,8	0,62	0,99	0,33	0,614
3362	0,34	0,68	0,68	0,95	0,21	0,572	3441	0,61	0,25	0,84	0,99	0,41	0,62
1379	0,41	0	0,81	0,96	0,45	0,526	4793	0,52	/	0,68	0,99	0,33	0,63
2600	0,26	0	0,63	0,96	0,3	0,43	4083	0,93	/	0,73	0,99	0,71	0,84
1430	0,94	0	0,69	0,87	0,58	0,616	4906	0,46	0,36	0,59	0,94	0,31	0,532
2956	0,32	0,45	0,33	0,91	0,65	0,532	1160	0,88	0,84	0,72	0,98	0,46	0,776
4742	0,13	0,58	0,59	0,84	0,5	0,528	3416	0,56	/	0,5	0,82	0,2	0,52
3651	0,77	0,41	0,88	0,98	0,41	0,69	1051	0,78	0,76	0,64	0,95	0,57	0,74
1064	0,77	0,23	0,83	0,99	0,22	0,608	2580	0,71	0,45	0,63	0,96	0,44	0,638
3818	0,68	0,64	0,73	0,99	0,47	0,702	1109	0,91	0,28	0,85	0,97	0,36	0,674
							mFF1	**0.59**	**0.51**	**0.73**	**0.96**	**0.42**	**0.64**

Table 4. $AS\,F_1$ **Results for Room- Based Localization on Monastero dei Benedettini.**

ID_visit	1	2	3	4	Neg.	AVG	ID_visit	1	2	3	4	Neg.	AVG
4805	0,55	0,62	0,4	0,12	0,36	0,41	2043	0,4	0,27	0,28	0,13	0,27	0,27
1804	0,32	0,1	0,39	0,05	0,21	0,214	3996	0,23	0,3	0,22	0,05	0,38	0,236
4377	0,31	0,12	0,26	0,08	0,27	0,208	3455	0,63	0,38	0,65	0,99	0,48	0,626
1669	0,68	0,34	0,67	0,9	0,33	0,584	4785	0,06	0	0,25	0,45	0,07	0,166
1791	0,41	0,41	0,25	0,06	0,22	0,27	2047	0,9	0,72	0,72	0,37	0,34	0,61
3948	0	/	0,33	0,26	0,49	0,27	1912	0,42	0,25	0,2	0,08	0,3	0,25
3152	0,39	0,43	0,64	0,44	0,58	0,496	3232	0,52	0,1	0,34	0,13	0,35	0,288
4361	0,26	0,35	0,32	0,08	0,2	0,242	4442	0,61	0,22	0,33	0,06	0,12	0,268
3976	0,68	0,26	0,35	0,03	0,5	0,364	3646	0,62	0,3	0,26	0,19	0,22	0,318
3527	0,65	0,37	0,38	0,28	0,44	0,424	4833	0,54	0,55	0,26	0,1	0,21	0,332
4105	0,49	0	0,49	0,16	0,06	0,24	3478	0,57	0,4	0,47	0,14	0,24	0,364
1399	0,55	0,1	0,33	0,23	0,29	0,3	4396	0,41	0,1	0,2	0,07	0,12	0,18
3836	0,47	0	0,35	0,98	0,34	0,428	2894	0,55	0,67	0,36	0,13	0,31	0,404
4006	0,47	0,49	0,49	0,55	0,26	0,452	4414	0,59	0,26	0,83	0,11	0,08	0,374
4415	0,74	0,07	0,35	0,14	0,23	0,306	4639	0,5	0,35	0,67	0,39	0,27	0,436
3008	0,57	0	0,5	0,99	0,41	0,494	1004	0,22	0,47	0,17	0,11	0,17	0,228
4660	0,45	0,52	0,17	0,05	0,27	0,292	1917	0,39	0,19	0,03	0,03	0,21	0,17
2826	0,31	0,29	0,33	0,06	0,25	0,248	1153	0,47	0,33	0,21	0,14	0,16	0,262
1099	0,52	0,24	0,55	0,96	0,17	0,488	2244	0,77	0,35	0,49	0,35	0,32	0,456
4391	0,54	0,13	0,14	0,2	0,24	0,25	2614	0,43	0,06	0,39	0,99	0,26	0,426
3929	0,15	0	0,31	0,22	0,39	0,214	1624	0,3	0,33	0,24	0,29	0,29	0,29
3362	0,2	0,35	0,38	0,05	0,22	0,24	3441	0,49	0,11	0,44	0,88	0,42	0,468
1379	0,22	0	0,43	0,05	0,35	0,21	4793	0,35	/	0,36	0,28	0,37	0,34
2600	0,32	0	0,39	0,31	0,26	0,256	4083	0,88	/	0,41	0,25	0,48	0,505
1430	0,43	0	0,1	0,02	0,21	0,152	4906	0,36	0,1	0,31	0,06	0,21	0,208
2956	0,41	0,2	0,16	0,29	0,31	0,274	1160	0,79	0,5	0,39	0,14	0,28	0,42
4742	0,15	0,18	0,25	0,11	0,17	0,172	3416	0,32	/	0,25	0,42	0,34	0,3325
3651	0,32	0,25	0,74	0,48	0,28	0,414	1051	0,5	0,31	0,21	0,05	0,4	0,294
1064	0,63	0,09	0,52	0,97	0,19	0,48	2580	0,36	0,29	0,33	0,05	0,28	0,262
3818	0,43	0,49	0,2	0,06	0,29	0,294	1109	0,82	0,17	0,46	0,11	0,28	0,368
							$mAS\,F_1$	**0,46**	**0,26**	**0,37**	**0,28**	**0,28**	**0.40**

Table 5. Point of Interest Retrieval related to Palazzo Bellomo cultural site.

1) Palazzo Bellomo

Variant	K	Precision	Recall	F_1 score
1 - One Shot	1	0.02	0.01	0.00
2 - Many Shots	1	**0.62**	**0.59**	**0.6**
	3	0.62	0.56	0.56
	5	0.62	0.56	0.56
	7	0,61	0,56	0,56
	9	0,61	0,55	0,56
	11	0,61	0,55	0,55

Table 6. Point of Interest Retrieval related to Monastero dei Benedettini cultural site.

2) Monastero dei Benedettini

Variant	K	Precision	Recall	F_1 score
1 - One shot	1	0.38	0.07	0.09
2 - Many Shots	1	0,83	0,83	0,83
	3	0,84	0,83	0,83
	5	**0,84**	**0,84**	**0.83**
	7	0,84	0,83	0,83
	9	0,84	0,83	0,83
	11	0,83	0,83	0,82

ject acquiring the video. All the training/test videos have a resolution of 1280×720 pixels and a frame-rate of 29.97 fps. We also include 191 reference images related to the considered POIs to be used for one-shot image retrieval. The images are akin to pictures generally included in museums catalog. Figure 3 shows some examples of such reference images.

2) Monastero dei Benedettini. This dataset is composed by 4 environments and contains 35 Points of Interest. Figure 4 and Figure 5 report some frames related to the 4 different environments and some of the points of interest. Table 9 reports details on the acquired training videos, highlighting the total duration of the videos, the required storage, the number of frames and the percentage of frames with respect to the whole training set. Training and validation videos have a resolution of 1216×684 pixels and a frame-rate of 24 fps. Five validation videos have been collected by asking volunteers to visit the cultural site with rules similarly to the one used for "Palazzo Bellomo". Table 10 shows the number of frames belonging to each video (left) and the number of frames belonging for each class (right).

Additionally, we collected 60 test videos by asking real visitors to freely visit the cultural site. No specific instructions have been given to the visitors, who were free to explore the 4 environments and the 35 POIs. The 60 test videos have a resolution of 1408×792 pixels and a frame-rate of 30.03 fps. The average number of frames for each video is *39296*. We also include 35 reference images related to the considered POIs to be used for one-shot image retrieval. Figure 6 shows some example of reference images.

Fig. 2. Sample frames of points of interest of "Palazzo Bellomo".

Fig. 3. Sample references images related to the cultural site "Palazzo Bellomo".

Fig. 4. Sample frames from the 4 considered environments of "Monastero dei Benedettini".

2.2. Annotations

Temporal Labels. All test and validation videos have been temporally labeled to indicate in every frame the environment in which the visitor is located and the currently observed point of interest, if any. If the visitor is not located in any of the considered environments or they are not observing any of the considered POIs we mark as that frame as "negative" (Figure 7).

Bounding Box Annotations. A subset of frames from the dataset (sampled at at 1 fps) has been labeled with bounding boxes indicating the presence and locations of all POIs. Figure 5 shows some example of labeled frames from the training set of the "Monastero dei Benedettini".

The EGO-CH dataset is publicy available at our website: *http://iplab.dmi.unict.it/EGO-CH/*. The dataset can be used only for reasearch purposes and is available upon request to the authors.

3. EXPERIMENTAL DETAILS

In this section, we present additional details on baseline experiments related to the 4 tasks proposed in the paper.

*3.1. **Room-based Localization***

1) Palazzo Bellomo. We split the Training Set into two subsets to train and validate the VGG-19 for the "Discrimination" stage

Table 7. List of training videos of "Palazzo Bellomo".

Name	Time (s)	Storage (MB)	#frame	%frame
1.0_Sala1	124	156.229	3721	3,13%
2.0_Sala2	117	148.480	3525	2,96%
3.0_Sala3	100	125.924	3000	2,52%
3.0_Sala3_S	73	92.589	2200	1,85%
4.0_Sala4	97	122.941	2914	2,45%
5.0_Sala5	99	126.213	2992	2,51%
6.0_Sala6	87	110.451	2630	2,21%
7.0_Sala7	113	143.257	3402	2,86%
8.0_Sala8	147	186.470	4427	3,72%
9.0_Sala9	143	180.971	4298	3,61%
10.0_Sala10	71	90.697	2154	1,81%
11.0_Sala11	104	131.983	3145	2,64%
12.0_Sala12	82	103.785	2463	2,07%
13.0_Sala13	101	128.013	3040	2,55%
14.0_CortiledegliStemmi	104	131.962	3131	2,63%
14.0_CortiledegliStemmi_S	90	113.822	2722	2,29%
15.0_SalaCarrozze	108	136.968	3259	3,12%
16.0_CortileParisio	124	156.605	3718	3,12%
16.0_Cortile_Parisio_S	68	86.646	2062	1,73%
17.0_Biglietteria	83	104.532	2489	2,09%
17.0_Biglietteria_S	53	68.071	1610	1,35%
18.0_Portico	126	159.800	3791	3,18%
18.0_Portico_S	63	80.044	1910	1,60%
19.0_ScalaCatalana	97	123.010	2918	2,45%
19.0_ScalaCatalana_S	116	110.063	3481	2,92%
20.0_Loggetta	80	101.584	2425	2,04%
20.0_Loggetta_S	58	73.722	1744	1,46%
21.0_BoxSala8	85	107.540	2562	2,15%
22.0_AreaSosta	64	81.934	1945	1,63%
22.0_Area_Sosta_S	52	65.340	1560	1,31%
2.1_Sala2_Acquasantiera	54	68.445	1623	2,94%
2.2_Sala2_FrammentiArchitett.	46	58.265	1393	2,53%
2.3_Sala2_LastraconLeoni	47	60.199	1427	2,59%
3.1_Sala3_MadonnainTrono	65	83.198	1972	3,58%
3.2_Sala3_FrammentoS.Leonardo	37	47.061	1113	2,02%
4.1_Sala4_MadonnainTrono	75	94.767	2252	4,08%
4.2_Sala4_MonumentoE.d'Aragona	86	108.296	2580	4,68%
4.3_Sala4_TrasfigurazioneCristo	76	96.106	2277	4,13%
4.4_Sala4_Piatti	49	62.281	1474	2,67%
5.1_Sala5_Annunciazione	76	96.952	2295	4,16%
5.2_Sala5_LibroD'OreMiniato	46	59.011	1406	2,55%
5.3_Sala5_LastraG.Cabastida	100	127.023	3017	5,47%
5.4_Sala5_MadonnadelCardillo	61	77.568	1829	3,32%
7.1_Sala7_DisputaS.Tommaso	74	94.188	2234	4,05%
7.2_Sala7_TraslazioneSantaCasa	76	96.045	2281	4,14%
7.3_Sala7_MadonnacolBambino	90	113.202	2696	4,89%
8.1_Sala8_ImmacolataConcezione	82	104.570	2483	4,50%
9.1_Sala9_AdorazionedeiMagi	60	76.171	1803	3,27%
9.2_Sala9_S.ElenaCostantinoeMadonna	76	96.227	2283	4,14%
9.3_Sala9_TaccuinidiDisegni	70	89.647	2121	3,85%
10.1_Sala10_MartirioS.Lucia	58	74.248	1759	3,19%
10.2_Sala10_VoltodiCristo	64	80.896	1917	3,48%
11.1_Sala11_MiracolodiS.Orsola	66	84.297	2002	3,63%
11.2_Sala11_Immacolata	73	92.424	2196	3,98%
16.1_CortileParisio_LapidiEbraiche	85	108.098	2563	4,65%
16.1_CortileParisio_LapidiEbraiche_S	67	85.173	2031	3,68%
21.1_BoxSala8_StoriedellaGenesi	70	88.300	2099	3,81%
AVG	**81.72**	**103.02**	**2462.53**	**2.07%**

Table 8. List of test videos of "Palazzo Bellomo".

Name	Time (s)	MB	#Frame	%Frame	#Environments	Environments - Temporal sequence
Test1	1906	2.400.360	57123	11,13%	22	16->17->18->1->18->3->2->3->18->4->18->14->15->14->19->20->6->5->6->7->21->8->9->22->10->11->12->13->5->6->20->19->18->17
Test2	1413	1.435.096	42348	8,25%	22	16->17->18->1->18->3->2->3->18->4->18->14->15->14->19->20->6->5->6->7->8->21->8->9->22->10->11->12->13->5->6->20->19->18->17
Test3	1830	2.304.410	54845	10,69%	22	16->17->18->1->18->3->2->3->18->4->18->14->15->14->19->20->6->5->13->12->11->10->22->9->8->21->8->7->6->20->19->18->17
Test4	1542	1.942.200	46214	5,49%	22	16->17->18->1->18->3->2->3->18->4->18->14->15->14->19->20->6->5->6->7->8->21->8->9->22->10->11->12->13->5->6->20->19->18->17
Test5	1034	1.302.612	30989	9,00%	22	16->17->18->1->18->3->2->3->18->4->18->14->15->14->19->20->6->5->6->7->8->9->22->10->11->12->13->5->6->20->19->18->17
Test6	1949	2.273.926	58411	11,38%	22	16->17->8->1->18->3->2->3->18->14->15->14->19->20->6->5->13->5->6->7->8->21->8->9->22->10->11->12->11->10->22->9->8->7->6->20->19->18->17
Test7	1332	1.677.047	39920	7,78%	22	16->17->14->15->14->18->1->18->3->2->3->18->4->18->19->20->6->5->6->7->8->21->8->9->22->10->11->12->13->5->6->20->19->18->17->16
Test8	3023	3.806.383	90599	17,65%	22	16->17->14->5->14->18->4->18->3->2->3->18->1->18->19->20->6->5->13->12->11->10->22->9->8->21->8->7->6->20->19->14
Test9	2236	2.815.878	67013	13,05%	22	16->17->18->1->18->3->2->3->18->4->18->14->15->14->19->20->6->5->13->12->11->10->22->9->8->21->8->7->6->20->19->14->17
Test10	858	1.080.389	25714	5,01%	22	16->17->14->19->20->6->7->8->21->8->9->22->10->11->12->13->5->6->20->19->18->4->18->1->18->3->2->3->18

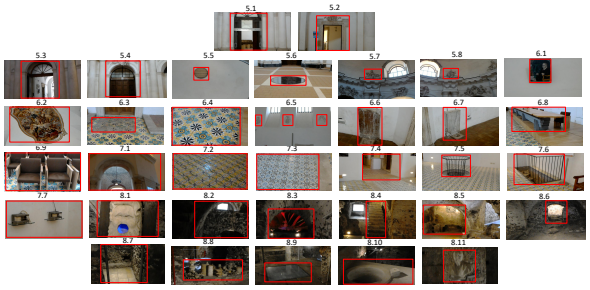

Fig. 5. Sample frames from the 35 considered POIs of "Monastero dei Benedettini", with the related bounding box annotations.

Fig. 6. Sample references images related to the "Monastero dei Benedettini".

(no "negative" frames are used for training). Table 11 reports the number of frames belonging to the two subsets for each of the 22 considered environments.

We report the results obtained by the baseline on the 9 test videos ("Test3" has been used for validation) in Table 12 considering the $FF1$ score metric, and in in Table 13 considering the $ASF1$ score. As example, Figure 8 illustrate qualitatively the segmentation result of the baseline on "Test7". Figure 9 reports the confusion matrix of the baseline on the test set.

2) Monastero dei Benedettini. Similarly to "Palazzo Bellomo", we split the Training Set into two subsets to train and validate the VGG-19 (no "negative" frames are used for training on this cultural site). Table 16 reports the number of frames belonging to the two subsets for each of the 4 considered environments. We report the results obtained by the baseline method over the 60 test videos in Table 17 and Table 18 considering the FF_1 score metric, and in Table 19 and Table 20 considering the ASF_1 score.

3.1.1. *DenseNet Backbone*

We performed experiments using another backbone in the same pipeline to solve the first task. We used DenseNet[3], a densely convolutional connect network which connects each layer to every other layer in a feed-forward fashion. Table 14 and Table 15 report the results obtained with DenseNet at the end of the Sequential Modeling step. We have evaluated the model using F1 score and Asf1 score. As shown the tables we did not obtain an improvement respect the results obtained with the backbone VGG. In particular for the "Palazzo Bellomo" we obtained a FF_1 score of 0.71 and a ASF_1 score of 0.58 which are lower respect the scores obtained with VGG ($FF_1 = 0.82$, $ASF_1 = 0.59$). From Table 17 to Table 20 we report the results obtained in the "Monastero dei Benedettini" using the backbone DenseNet. As shown, neither considering this cultural site we obtained an improvement of FF_1 and ASF_1 scores.

3.2. *Points of Interest Recognition*

1) Palazzo Bellomo. We used the subset of the Training set considered in the first task annotated with bounding box for a

Table 9. List of training videos of "Monastero dei Benedettini"

Name	Time (s)	Storage (MB)	#frame	%frame
5.0_Antirefettorio	247	244.500	5933	1,21%
5.0_Antirefettorio1	263	260.395	6315	1,29%
6.0_SantoMazzarino	241	239.007	5800	1,18%
7.0_Cucina	239	237.268	5753	1,17%
8.0_Ventre	679	832.844	20385	4,15%
5.1_Antirefettorio_PortaA.S.Mazz.Ap.	67	67.001	1630	0,33%
5.1_Antirefettorio_PortaA.S.Mazz.Ch.	74	73.589	1785	0,36%
5.2_Antirefettorio_PortaMuseoFab.Ap.	50	49.840	1211	0,25%
5.2_Antirefettorio_PortaMuseoFab.Ch.	62	61.846	1503	0,31%
5.3_Antirefettorio_PortaAntiref.	51	58.557	1537	0,31%
5.4_Antirefettorio_PortaRef.Piccolo	54	53.767	1306	0,27%
5.5_Antirefettorio_Cupola	57	56.089	1377	0,28%
5.6_Antirefettorio_AperturaPavimento	55	54.586	1322	0,27%
5.7_Antirefettorio_S.Agata	48	47.820	1165	0,24%
5.8_Antirefettorio_S.Scolastica	58	57.919	1407	0,29%
5.9_Antirefettorio_ArcoconFirma	62	76.700	1864	0,38%
5.10_Antirefettorio_BustoVaccarini	65	64.298	1563	0,32%
6.1_SantoMazzarino_QuadroS.Mazz.	71	70.401	1716	0,35%
6.2_SantoMazzarino_Affresco	213	211.424	5124	1,04%
6.3_SantoMazzarino_PavimentoOr.	99	98.691	2397	0,49%
6.4_SantoMazzarino_PavimentoRes.	69	69.148	1675	0,34%
6.5_SantoMazzarino_BassorilieviManc.	117	115.348	2823	0,57%
6.6_SantoMazzarino_LavamaniSx	151	149.882	3637	0,74%
6.7_SantoMazzarino_LavamaniDx	93	92.928	2256	0,46%
6.8_SantoMazzarino_TavoloRelatori	150	148.661	3603	0,73%
6.9_SantoMazzarino_Poltrone	108	107.374	2604	0,53%
7.1_Cucina_Edicola	369	437.219	11086	2,26%
7.2_Cucina_PavimentoA	52	52.163	1268	0,26%
7.3_Cucina_PavimentoB	52	52.244	1266	0,26%
7.4_Cucina_PassavivandePavim.Orig.	81	80.733	1961	0,40%
7.5_Cucina_AperturaPav.1	57	57.170	1385	0,28%
7.5_Cucina_AperturaPav.2	53	52.875	1280	0,26%
7.5_Cucina_AperturaPav.3	62	62.320	1509	0,31%
7.6_Cucina_Scala	77	76.587	1856	0,38%
7.7_Cucina_SalaMetereologica	156	154.394	3748	0,76%
8.1_Ventre_Doccione	103	102.683	2492	0,51%
8.2_Ventre_VanoRacc.Cenere	126	124.837	3026	0,62%
8.3_Ventre_SalaRossa	300	296.792	7202	1,47%
8.4_Ventre_ScalaCucina	214	212.372	5152	1,05%
8.5_Ventre_CucinaProvv.	148	146.379	3553	0,72%
8.6_Ventre_Ghiacciaia	69	68.562	1668	0,34%
8.6_Ventre_Ghiacciaia1	266	263.817	6398	1,30%
8.7_Ventre_Latrina	102	100.542	2468	0,50%
8.8_Ventre_OssaScarti	154	152.861	3713	0,76%
8.8_Ventre_OssaScarti1	36	36.345	886	0,18%
8.9_Ventre_Pozzo	300	297.167	7209	1,47%
8.10_Ventre_Cisterna	57	56.572	1384	0,28%
8.11_Ventre_BustoP.Tacchini	110	109.520	2662	0,54%
AVG	**133.06**	**137.38**	**3351.31**	**1.04%**

Table 10. List of validation videos of "Monastero dei Benedettini".

Name	#frame	Class	#frame
Test1	4141	1 Antirefettorio	88613
Test3	18678	2 Aula S. Mazzarino	8395
Test4	13731	3 Cucina	9712
Test5	15958	4 Ventre	20513
Test7	1124	Negatives	6399
Total	**53632**	**Total**	**53632**

Fig. 7. Sample frames from "Monastero dei Benedettigni" marked as "negative locations".

total of *56686* frames. We split this subset in training/validation sets to train and validate the object detector. In particular, we used *41111* frames as training set and *15575* as validation set. To find the optimal detection threshold, we used "Test3" as validation video. We obtaining the value of *0.05* maximizing mAP over the validations et. In Table 25 we report the results on the 9 test videos (excluding "Test3") of the object detector when using the mean Average Precision (mAP) as evaluation metric. The table also reports the number of frames annotated with bounding box for each test video. Per-class AP values are reported in Table 26.

2) Monastero dei Benedettini. A subset of frames from the dataset (sampled at at 1 fps) has been labeled with bounding boxes. The annotated frames with bounding box of the Training set used in the first task are *33366*. We split this set into training/validation sets to train and validate the object detector.

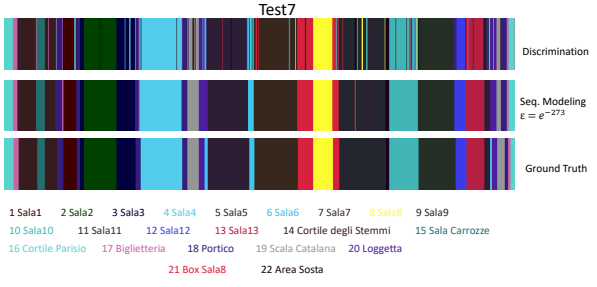

Fig. 8. Color-coded segmentations for the test video "Test7" of "Palazzo Bellomo".

Table 11. Number of frames belonging to the two subsets (Training/Validation) to train the CNN for "Palazzo Bellomo".

	Training	Validation
1 Sala1	2605	1116
2 Sala2	5578	2390
3 Sala3	5800	2486
4 Sala4	8048	3449
5 Sala5	8023	3438
6 Sala6	1841	789
7 Sala7	7429	3184
8 Sala8	4837	2073
9 Sala9	7354	3152
10 Sala10	4081	1749
11 Sala11	5140	2203
12 Sala12	1724	739
13 Sala13	2128	912
14 Cortile degli Stemmi	4097	1756
15 Sala delle Carrozze	2281	978
16 Cortile Parisio	7262	3112
17 Biglietteria	2869	1230
18 Portico	3991	1710
19 Scala Catalana	4479	1920
20 Loggetta	2918	1251
21 Box Sala8	3263	1398
22 Area Sosta	2454	1052
Total	**98202**	**42084**

In particular, we used *23363* frames as training set and *10003* as validation set. We used the 10 validation videos (*2235* frames) to find the optimal threshold of the object detector. Table 27 reports the AP values obtained for each of the 35 considered points of interest belonging to the validation set, using the optimal threshold found through validation (0.001). We annotated the 60 real visits with bounding boxes for a total of *71310* images. We tested the object detector on these frames. The results are shown in Table 28. Also, for each test video, we report the number of frames annotated with bounding boxes. Per-class AP values are reported in Table 29.

3.3. ***Object Retrieval***

To address this task, we extracted image patches from the bounding box annotations of the dataset.

1) Palazzo Bellomo. In Table 30, we report the number of image patches which have been extracted for each test video. For one-shot learning, we have used reference images for training and all the image patches for testing, whereas to perform many-shot learning, we used the patches belonging to test videos $1-7$ for training (*15185* patches) and the others to test (*8542* patches).

2) Monastero dei Benedettini. In Table 31, we report the number of image patches extracted from the 60 test videos. One-shot learning has been performed using reference images from training and all extracted patches for testing. For many-shot learning, we used *30497* image patches images belonging to the

Table 12. Detailed results of the 9 test videos of "Palazzo Bellomo" using the FF_1 score. The "/" sign indicates that no samples from that class was present in the test video.

	FF_1 score									
Class	Test1	Test2	Test4	Test5	Test6	Test7	Test8	Test9	Test10	AVG
1_Sala1	0,16	0,00	0,81	0,96	0,86	0,96	0,90	0,85	0,92	0,71
2_Sala2	0,78	0,67	0,96	0,96	0,99	0,99	0,97	0,97	0,96	0,92
3_Sala3	0,68	0,75	0,97	0,87	0,72	0,96	0,83	0,89	0,91	0,84
4_Sala4	0,89	0,96	0,93	0,91	/	0,98	0,86	0,91	0,95	0,92
5_Sala5	0,90	0,94	0,98	0,95	0,89	0,95	0,95	0,95	0,97	0,94
6_Sala6	0,84	0,86	0,80	0,59	0,76	0,57	0,93	0,93	0,68	0,77
7_Sala7	0,99	0,95	0,99	0,88	0,84	0,99	0,92	0,93	0,97	0,94
8_Sala8	0,85	0,95	0,96	0,91	0,67	0,90	0,84	0,95	0,93	0,89
9_Sala9	0,86	0,97	0,95	0,90	0,76	0,93	0,94	0,94	0,90	0,91
10_Sala10	0,87	0,96	0,96	0,97	0,00	0,98	0,95	0,97	0,90	0,84
11_Sala11	0,86	0,96	0,97	0,97	0,00	0,97	0,94	0,96	0,96	0,84
12_Sala12	0,82	0,91	0,86	0,94	0,00	0,88	0,96	0,94	0,88	0,80
13_Sala13	0,74	0,94	0,91	0,85	0,00	0,95	0,97	0,92	0,94	0,80
14_CortiledegliStemmi	0,92	0,80	0,93	0,89	0,73	0,92	0,97	0,94	0,57	0,85
15_SalaCarrozze	0,91	0,89	0,96	0,93	0,90	0,90	0,85	0,96		0,91
16_CortileParisio	0,74	0,50	0,92	0,71	0,48	0,91	0,99	0,74	0,72	0,75
17_Biglietteria	0,64	0,79	0,81	0,49	0,74	0,55	0,66	0,61	0,53	0,65
18_Portico	0,71	0,42	0,77	0,70	0,70	0,73	0,71	0,75	0,72	0,69
19_ScalaCatalana	0,70	0,78	0,80	0,77	0,39	0,86	0,83	0,95	0,76	0,76
20_Loggetta	0,62	0,39	0,75	0,58	0,67	0,77	0,84	0,94	0,81	0,71
21_BoxSala8	0,97	0,97	0,98	/	0,79	0,97	0,99	0,94	0,94	0,94
22_AreaSosta	0,24	0,81	0,56	0,46	0,00	0,87	0,56	0,78	0,77	0,56
mFF1	0,76	0,78	0,89	0,82	0,57	0,89	0,88	0,90	0,84	0,81

Table 13. Detailed results of the 9 test videos of "Palazzo Bellomo" using the ASF_1 score. The "/" sign indicates that no samples from that class was present in the test video.

	ASF_1									
Class	Test1	Test2	Test4	Test5	Test6	Test7	Test8	Test9	Test10	AVG
1_Sala1	0,07	0,00	0,25	0,92	0,15	0,92	0,62	0,55	0,84	0,48
2_Sala2	0,32	0,46	0,92	0,92	0,97	0,98	0,65	0,94	0,92	0,79
3_Sala3	0,18	0,34	0,66	0,63	0,19	0,88	0,16	0,66	0,83	0,50
4_Sala4	0,58	0,92	0,41	0,45	/	0,95	0,21	0,28	0,89	0,59
5_Sala5	0,57	0,56	0,84	0,77	0,27	0,72	0,49	0,64	0,94	0,64
6_Sala6	0,47	0,61	0,44	0,35	0,32	0,40	0,87	0,80	0,44	0,52
7_Sala7	0,97	0,91	0,97	0,28	0,07	0,97	0,21	0,16	0,94	0,61
8_Sala8	0,40	0,86	0,84	0,59	0,13	0,68	0,70	0,74	0,85	0,64
9_Sala9	0,19	0,93	0,90	0,45	0,32	0,18	0,64	0,14	0,49	0,47
10_Sala10	0,28	0,92	0,92	0,93	0,00	0,96	0,42	0,94	0,81	0,69
11_Sala11	0,21	0,92	0,94	0,94	0,00	0,94	0,13	0,21	0,92	0,58
12_Sala12	0,37	0,82	0,76	0,89	0,00	0,79	0,91	0,63	0,78	0,66
13_Sala13	0,32	0,88	0,83	0,74	0,00	0,90	0,94	0,40	0,89	0,66
14_CortiledegliStemmi	0,78	0,61	0,83	0,79	0,16	0,84	0,72	0,69	0,39	0,64
15_SalaCarrozze	0,49	0,80	0,92	0,87	0,37	0,82	0,17	0,91	/	0,67
16_CortileParisio	0,31	0,22	0,65	0,41	0,16	0,80	0,98	0,47	0,54	0,50
17_Biglietteria	0,41	0,67	0,69	0,32	0,45	0,29	0,38	0,37	0,36	0,44
18_Portico	0,54	0,28	0,60	0,43	0,51	0,58	0,56	0,49	0,58	0,51
19_ScalaCatalana	0,61	0,48	0,66	0,64	0,41	0,72	0,65	0,91	0,62	0,63
20_Loggetta	0,43	0,31	0,37	0,46	0,23	0,63	0,70	0,87	0,64	0,51
21_BoxSala8	0,65	0,94	0,96	/	0,34	0,94	0,98	0,64	0,88	0,79
22_AreaSosta	0,24	0,69	0,42	0,47	0,00	0,76	0,41	0,64	0,62	0,47
mASF1	0,43	0,64	0,72	0,63	0,24	0,76	0,57	0,59	0,72	0,59

Table 14. Detailed results of the 9 test videos of "Palazzo Bellomo" using the FF_1 score. The backbone used is DenseNet. The "/" sign indicates that no samples from that class was present in the test video.

FF_1 score

Class	Test1	Test2	Test4	Test5	Test6	Test7	Test8	Test9	Test10	AVG
1_Sala1	0,00	0,96	0,96	0,95	0,99	0,94	0,97	0,97	0,91	0,85
2_Sala2	0,82	0,91	0,96	0,95	0,99	0,99	0,99	0,97	0,95	0,95
3_Sala3	0,65	0,84	0,84	0,83	0,94	0,88	0,61	0,85	0,83	0,81
4_Sala4	0,84	0,96	0,98	0,93	/	0,96	0,78	0,97	0,93	0,92
5_Sala5	0,98	0,88	0,96	0,91	0,65	0,91	0,97	0,97	0,99	0,91
6_Sala6	0,84	0,80	0,43	0,00	0,86	0,00	0,18	0,83	0,73	0,52
7_Sala7	0,67	0,93	0,22	0,00	0,38	0,88	0,88	0,40	0,96	0,59
8_Sala8	0,64	0,75	0,60	0,52	0,56	0,56	0,64	0,60	0,77	0,63
9_Sala9	0,90	0,89	0,73	0,88	0,41	0,92	0,98	0,99	0,88	0,84
10_Sala10	0,96	0,98	0,92	0,72	0,00	0,98	0,85	0,97	0,78	0,80
11_Sala11	0,93	0,96	0,96	0,97	0,00	0,97	0,97	0,98	0,95	0,85
12_Sala12	0,82	0,88	0,87	0,90	0,00	0,87	0,85	0,92	0,87	0,78
13_Sala13	0,96	0,95	0,95	0,76	0,00	0,93	0,95	0,94	0,96	0,82
14_CortiledegliStemmi	0,78	0,68	0,74	0,88	0,17	0,90	0,79	0,92	0,00	0,65
15_SalaCarrozze	0,84	0,89	0,95	0,93	0,91	0,84	0,97	0,95	/	0,91
16_CortileParisio	0,33	0,51	0,52	0,45	0,35	0,60	0,81	0,91	0,80	0,59
17_Biglietteria	0,25	0,56	0,00	0,26	0,00	0,00	0,00	0,36	0,00	0,16
18_Portico	0,68	0,67	0,78	0,54	0,69	0,69	0,62	0,78	0,62	0,67
19_ScalaCatalana	0,00	0,00	0,54	0,55	0,78	0,67	0,58	0,85	0,49	0,50
20_Loggetta	0,00	0,50	0,80	0,44	0,72	0,53	0,55	0,83	0,68	0,56
21_BoxSala8	0,96	0,97	0,97	/	0,00	0,97	0,99	0,81	0,94	0,83
22_AreaSosta	0,74	0,85	0,45	0,83	0,00	0,56	0,72	0,85	0,00	0,56
mFF1	0,66	0,79	0,73	0,68	0,45	0,75	0,76	0,85	0,72	0,71

Table 15. Detailed results of the 9 test videos of "Palazzo Bellomo" using the ASF_1 score. The backbone used is DenseNet. The "/" sign indicates that no samples from that class was present in the test video.

ASF_1 score

Class	Test1	Test2	Test4	Test5	Test6	Test7	Test8	Test9	Test10	AVG
1_Sala1	0,00	0,93	0,93	0,91	0,97	0,89	0,94	0,94	0,84	0,82
2_Sala2	0,64	0,83	0,91	0,90	0,97	0,97	0,97	0,94	0,90	0,89
3_Sala3	0,36	0,53	0,51	0,53	0,61	0,63	0,29	0,59	0,72	0,53
4_Sala4	0,39	0,92	0,96	0,86	/	0,92	0,31	0,94	0,87	0,77
5_Sala5	0,94	0,64	0,66	0,62	0,34	0,63	0,54	0,95	0,97	0,70
6_Sala6	0,46	0,43	0,18	0,00	0,42	0,00	0,09	0,59	0,51	0,30
7_Sala7	0,50	0,86	0,13	0,00	0,10	0,78	0,28	0,25	0,92	0,42
8_Sala8	0,36	0,44	0,42	0,35	0,35	0,41	0,49	0,53	0,57	0,44
9_Sala9	0,81	0,80	0,51	0,78	0,22	0,84	0,95	0,97	0,78	0,74
10_Sala10	0,92	0,96	0,85	0,19	0,00	0,95	0,27	0,94	0,64	0,64
11_Sala11	0,53	0,92	0,91	0,93	0,00	0,93	0,65	0,95	0,91	0,75
12_Sala12	0,38	0,79	0,76	0,81	0,00	0,77	0,30	0,85	0,76	0,60
13_Sala13	0,93	0,90	0,90	0,62	0,00	0,87	0,91	0,89	0,92	0,77
14_CortiledegliStemmi	0,57	0,53	0,54	0,77	0,08	0,81	0,49	0,67	0,00	0,50
15_SalaCarrozze	0,72	0,80	0,91	0,87	0,83	0,72	0,94	0,90	/	0,84
16_CortileParisio	0,30	0,63	0,43	0,65	0,38	0,47	0,59	0,83	0,66	0,55
17_Biglietteria	0,26	0,43	0,00	0,11	0,00	0,00	0,00	0,39	0,00	0,13
18_Portico	0,51	0,47	0,60	0,45	0,48	0,52	0,42	0,64	0,47	0,51
19_ScalaCatalana	0,00	0,00	0,43	0,51	0,61	0,55	0,42	0,75	0,39	0,41
20_Loggetta	0,00	0,31	0,54	0,28	0,46	0,38	0,48	0,72	0,44	0,40
21_BoxSala8	0,92	0,95	0,95	/	0,00	0,93	0,98	0,67	0,88	0,79
22_AreaSosta	0,59	0,73	0,29	0,71	0,00	0,75	0,57	0,74	0,00	0,49
mFF1	0,50	0,67	0,61	0,56	0,32	0,67	0,54	0,76	0,63	0,58

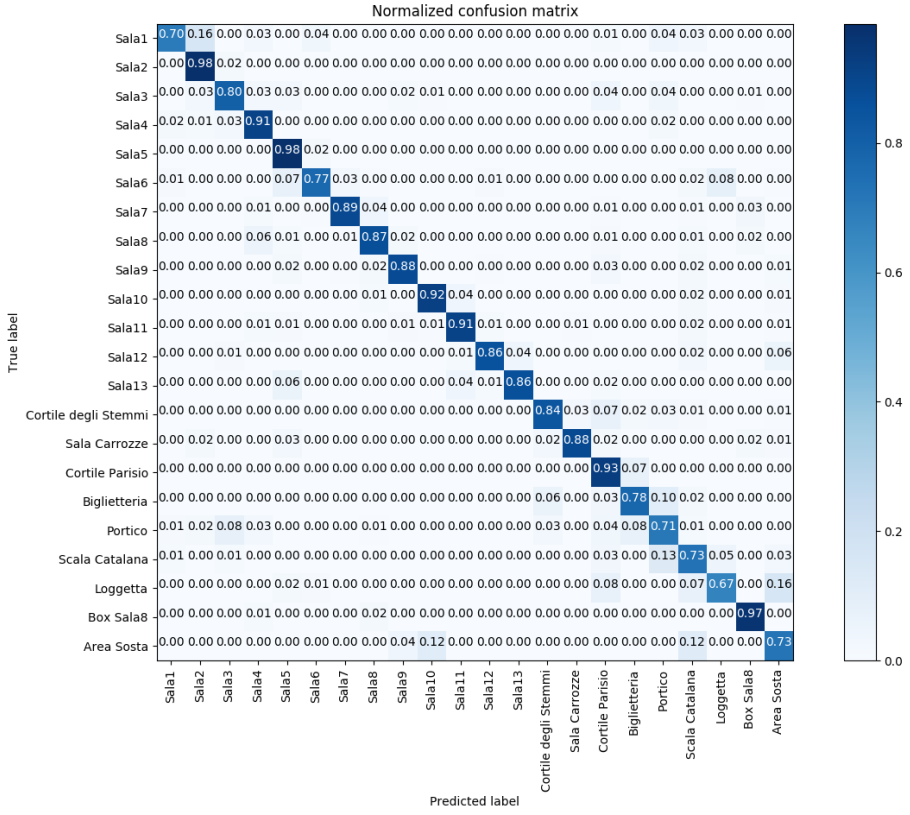

Fig. 9. Confusion matrix for localization in "Palazzo Bellomo".

Table 16. Number of frames belonging to the two subsets (Training/Validation) to train the CNN for "Monastero dei Benedettini".

	Training	**Validation**
1 Antirefettorio	7000	3000
2 Aula Santo Mazzarino	7000	3000
3 Cucina	7000	3000
4 Ventre	7000	3000
Total	28000	12000

visits with IDs from 100 to 147 for training, and *14551* patches belonging to the visits with ID from 148 to 166 for testing.

3.3.1. Object Retrieval with DenseNet

We tried to extract features using a different backbone respect to the proposed baseline based on VGG-19. We extracted the features from the FC7 layer of DenseNet[3]. In this way, we obtained for each image a fixed-size vector of 1024 values. We have followed the same pipeline used with VGG to perform object retrieval. Table 32 shows the results obtained using DenseNet as backbone. The results show that using DenseNet we didn't obtained an improvement considering the F_1 score as evaluation measure in both variants (One-shot and Many-shots) for both cultural sites.

3.4. Survey Prediction

Table 33 shows the results obtained on binary classification using a KNN with different values of k. The method is evaluated using precision, recall and F_1 score. $K = 9$ gives the best results. Table 34 shows the results of multi-class classification using a KNN classifier with different values of K. We obtained the best results for $K = 9$.

ACKNOWLEDGMENTS

This research is part of the project VALUE - Visual Analysis for Localization and Understanding of Environments (N. 08CT6209090207, CUP G69J18001060007) supported by PO FESR 2014/2020 - Azione 1.1.5. - "Sostegno allavanzamento tecnologico delle imprese attraverso il finanziamento di linee pilota e azioni di validazione precoce dei prodotti e di dimostrazioni su larga scala" del PO FESR Sicilia 2014/2020, and Piano della Ricerca 2016-2018 linea di Intervento 2 of DMI, University of Catania. The authors would like to thank Regione Siciliana Assessorato dei Beni Culturali dell'Identit Siciliana - Dipartimento dei Beni Culturali e dell'Identit Siciliana and Polo regionale di Siracusa per i siti culturali - Galleria Regionale di Palazzo Bellomo.

References

[1] Galleria regionale di palazzo bellomo (2007).
URL `http://www.regione.sicilia.it/beniculturali/palazzobellomo/`

[2] Monastero dei benedettini.
URL `http://www.monasterodeibenedettini.it/`

Table 17. Detailed results on the 60 test videos of "Monastero dei Benedettini", considering the evaluation measure FF_1 score. The "/" sign indicates that no samples from that class were present in the test video. The four classes are: 1) Antirefettorio, 2) Aula S. Mazzarino, 3) Cucina, 3) Ventre, whereas Neg. represents the negatives.

ID_Visit	1	2	3	4	Neg.	AVG
4805	0,91	0,33	0,75	0,99	0,43	0,682
1804	0,69	0,43	0,8	0,99	0,43	0,668
4377	0,77	0,47	0,84	0,99	0,5	0,714
1669	0,8	0,51	0,92	0,99	0,67	0,778
1791	0,59	0,22	0,78	0,98	0,35	0,584
3948	0,8	/	0,66	0,99	0,61	0,765
3152	0,71	0,25	0,86	0,98	0,75	0,71
4361	0,89	0,32	0,81	0,099	0,62	0,5478
3976	0,97	0,62	0,92	0,98	0,68	0,834
3527	0,85	0	0,82	0,99	0,63	0,658
4105	0,81	0	0,77	0,99	0,66	0,646
1399	0,65	0	0,76	0,99	0,62	0,604
3836	0,65	0	0,76	0,99	0,62	0,604
4006	0,81	0,82	0,92	0,99	0,75	0,858
4415	0,87	0,49	0,77	0,98	0,73	0,768
3008	0,82	0,2	0,63	0,99	0,23	0,574
4660	0,82	0,2	0,63	0,99	0,23	0,574
2826	0,79	0,57	0,81	1	0,41	0,716
1099	0,77	0,27	0,64	0,98	0,5	0,632
4391	0,8	0,03	0,79	0,98	0,55	0,63
3929	0,94	0	0,84	0,99	0,67	0,688
3362	0,46	0,25	0,76	0,99	0,46	0,584
1379	0,84	0	0,75	0,98	0,33	0,58
2600	0,74	0	0,78	0,99	0,59	0,62
1430	0,57	0,22	0,72	0,96	0,47	0,588
2956	0,53	0,33	0,73	0,96	0,8	0,67
4742	0,14	0,43	0,92	0,99	0,67	0,63
3651	0,94	0,66	0,82	0,99	0,49	0,78
1064	0,77	0,12	0,88	0,99	0,31	0,614
3818	0,93	0,14	0,71	0,99	0,51	0,656

[3] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.

Table 18. Continued from Table 17

ID_Visit	1	2	3	4	Neg.	AVG
2043	0,72	0,47	0,79	0,98	0,66	0,724
3996	0,47	0,4	0,72	0,98	0,76	0,666
3455	0,85	0,33	0,88	0,99	0,61	0,732
4785	0,03	0	0,73	0,96	0,65	0,474
2047	0,95	0,55	0,86	1	0,6	0,792
1912	0,79	0,44	0,67	0,96	0,62	0,696
3232	0,89	0,33	0,82	0,99	0,73	0,752
4442	0,9	0,37	0,82	0,99	0,54	0,724
3646	0,67	0,09	0,8	1	0,34	0,58
4833	0,78	0,59	0,86	0,95	0,66	0,768
3478	0,82	0,51	0,94	1	0,67	0,788
4396	0,81	0,53	0,89	0,99	0,38	0,72
2894	0,67	0	0,87	0,97	0,55	0,612
4414	0,88	0,51	0,9	0,98	0,58	0,77
4639	0,73	0,21	0,83	0,99	0,61	0,674
1004	0,15	0,44	0,6	0,98	0,48	0,53
1917	0,44	0,48	0,51	1	0,42	0,57
1153	0,78	0,54	0,81	0,98	0,57	0,736
2244	0,86	0,3	0,77	0,99	0,32	0,648
2614	0,97	0,59	0,84	0,99	0,39	0,756
1624	0,91	0,71	0,69	0,99	0,48	0,756
3441	0,82	0,45	0,79	0,99	0,46	0,702
4793	0,82	/	0,74	0,99	0,42	0,7425
4083	0,84	/	0,72	0,99	0,73	0,82
4906	0,77	0,2	0,74	0,99	0,42	0,624
1160	0,84	0,66	0,74	1	0,42	0,732
3416	0,77	/	0,82	0,99	0,68	0,815
1051	0,79	0,43	0,78	0,98	0,4	0,676
2580	0,73	0,18	0,89	0,99	0,48	0,654
1109	0,81	0,28	0,89	0,99	0,43	0,68
AVG	**0,75**	**0,33**	**0,79**	**0,99**	**0,54**	**0,68**

Table 19. Detailed results on the 60 test videos of "Monastero dei Benedettini", considering the evaluation measure $AS\,F_1$ score. The "/" sign indicates that no samples from that class were present in the test video. The four classes are: 1) Antirefettorio, 2) Aula S. Mazzarino, 3) Cucina, 3) Ventre, whereas Neg. represents the negatives.

ID_Visit	1	2	3	4	Neg.	AVG
4805	0,71	0,06	0,2	0,31	0,22	0,3
1804	0,55	0,06	0,26	0,15	0,14	0,232
4377	0,55	0,05	0,18	0,98	0,25	0,402
1669	0,4	0,22	0,27	0,93	0,49	0,462
1791	0,46	0,14	0,21	0,25	0,19	0,25
3948	0,67	/	0,26	0,98	0,51	0,605
3152	0,54	0,07	0,66	0,88	0,47	0,524
4361	0,47	0,12	0,17	0,27	0,29	0,264
3976	0,94	0,24	0,45	0,22	0,56	0,482
3527	0,72	0	0,44	0,46	0,55	0,434
4105	0,54	0	0,49	0,39	0,51	0,386
1399	0,15	0	0,44	0,98	0,26	0,366
3836	0,15	0	0,44	0,98	0,26	0,366
4006	0,56	0,69	0,53	0,98	0,5	0,652
4415	0,48	0,15	0,24	0,49	0,48	0,368
3008	0,68	0,1	0,26	0,98	0,1	0,424
4660	0,68	0,09	0,26	0,98	0,09	0,42
2826	0,71	0,14	0,24	0,99	0,41	0,498
1099	0,54	0,15	0,16	0,49	0,31	0,33
4391	0,67	0,05	0,32	0,96	0,53	0,506
3929	0,91	0	0,43	0,98	0,46	0,556
3362	0,36	0,07	0,45	0,57	0,36	0,362
1379	0,49	0	0,21	0,18	0,23	0,222
2600	0,64	0	0,41	0,98	0,38	0,482
1430	0,17	0,09	0,03	0,08	0,1	0,094
2956	0,33	0,06	0,46	0,37	0,48	0,34
4742	0,1	0,09	0,39	0,62	0,17	0,274
3651	0,63	0,19	0,54	0,65	0,28	0,458
1064	0,63	0,05	0,3	0,65	0,13	0,352
3818	0,78	0,11	0,17	0,25	0,35	0,332

Table 20. Continued from Table 19

ID_Visit	1	2	3	4	Neg.	AVG
2043	0,4	0,22	0,2	0,27	0,25	0,268
3996	0,45	0,12	0,39	0,4	0,49	0,37
3455	0,69	0,16	0,39	0,96	0,5	0,54
4785	0,05	0,11	0,53	0,13	0.33	0,205
2047	0,9	0,38	0,22	0,99	0,36	0,57
1912	0,54	0,1	0,16	0,31	0,39	0,3
3232	0,73	0,04	0,49	0,98	0,53	0,554
4442	0,59	0,285	0,22	0,27	0,25	0,323
3646	0,4	0,13	0,2	0,66	0,34	0,346
4833	0,54	0,15	0,53	0,19	0,36	0,354
3478	0,71	0,29	0,54	0,99	0,46	0,598
4396	0,35	0,02	0,05	0,18	0,08	0,136
2894	0,43	0	0,46	0,35	0,18	0,284
4414	0,75	0,11	0,31	0,24	0,3	0,342
4639	0,61	0,12	0,44	0,39	0,43	0,398
1004	0,22	0,11	0,27	0,21	0,38	0,238
1917	0,41	0,11	0,15	0,31	0,26	0,248
1153	0,54	0,11	0,33	0,59	0,26	0,366
2244	0,58	0,16	0,45	0,97	0,26	0,484
2614	0,65	0,05	0,39	0,98	0,3	0,474
1624	0,71	0,29	0,35	0,39	0,23	0,394
3441	0,59	0,13	0,26	0,88	0,25	0,422
4793	0,68	/	0,52	0,98	0,38	0,64
4083	0,65	/	0,31	0,35	0,59	0,475
4906	0,57	0,06	0,33	0,59	0,33	0,376
1160	0,7	0,15	0,34	0,99	0,29	0,494
3416	0,45	/	0,24	0,97	0,19	0,4625
1051	0,55	0,28	0,33	0,4	0,27	0,366
2580	0,37	0,1	0,68	0,21	0,37	0,346
1109	0,66	0,16	0,63	0,98	0,36	0,558
AVG	**0,54**	**0,12**	**0,34**	**0,60**	**0,33**	**0,40**

Table 21. Detailed results on the 60 test videos of "Monastero dei Benedettini", considering the evaluation measure FF_1 score. We used the backbone DenseNet. The "/" sign indicates that no samples from that class were present in the test video. The four classes are: 1) Antirefettorio, 2) Aula S. Mazzarino, 3) Cucina, 3) Ventre, whereas Neg. represents the negatives.

ID_Visit	1	2	3	4	Neg.	AVG
4805	0,79	0,82	0,74	0,95	0,45	0,75
1804	0,32	0,36	0,79	0,98	0,55	0,6
4377	0,46	0,53	0,8	0,96	0,32	0,614
1669	0,75	0,72	0,9	0,99	0,45	0,762
1791	0,49	0,5	0,84	0,95	0,43	0,642
3948	0	/	0,59	0,98	0,49	0,515
3152	0,39	0,72	0,87	0,97	0,76	0,742
4361	0,55	0,78	0,82	0,97	0,28	0,68
3976	0,87	0,43	0,81	0,93	0,66	0,74
3527	0,86	0,79	0,85	0,99	0,56	0,81
4105	0,62	0	0,77	0,97	0,09	0,49
1399	0,46	0,23	0,79	0,98	0,43	0,578
3836	0,51	0	0,72	0,99	0,55	0,554
4006	0,57	0,78	0,78	0,99	0,37	0,698
4415	0,86	0,61	0,82	0,97	0,35	0,722
3008	0,69	0	0,7	0,99	0,51	0,578
4660	0,62	0,81	0,83	0,98	0,57	0,762
2826	0,31	0,52	0,76	0,97	0,61	0,634
1099	0,62	0,42	0,85	0,98	0,49	0,672
4391	0,74	0,72	0,72	0,98	0,33	0,698
3929	0,26	0	0,8	0,99	0,43	0,496
3362	0,34	0,68	0,68	0,95	0,21	0,572
1379	0,41	0	0,81	0,96	0,45	0,526
2600	0,26	0	0,63	0,96	0,3	0,43
1430	0,94	0	0,69	0,87	0,58	0,616
2956	0,32	0,45	0,33	0,91	0,65	0,532
4742	0,13	0,58	0,59	0,84	0,5	0,528
3651	0,77	0,41	0,88	0,98	0,41	0,69
1064	0,77	0,23	0,83	0,99	0,22	0,608
3818	0,68	0,64	0,73	0,99	0,47	0,702

Table 22. Continued from Table 21

ID_Visit	1	2	3	4	Neg.	AVG
2043	0,52	0,44	0,62	0,91	0,33	0,564
3996	0,17	0,68	0,57	0,95	0,72	0,618
3455	0,78	0,65	0,8	0,99	0,55	0,754
4785	0,02	0	0,64	0,95	0,33	0,388
2047	0,95	0,84	0,84	0,98	0,38	0,798
1912	0,66	0,51	0,69	0,94	0,46	0,652
3232	0,66	0,53	0,85	0,99	0,49	0,704
4442	0,8	0,77	0,82	0,98	0,23	0,72
3646	0,63	0,29	0,66	0,97	0,4	0,59
4833	0,67	0,82	0,67	0,88	0,24	0,656
3478	0,71	0,75	0,84	0,98	0,37	0,73
4396	0,74	0,69	0,89	0,96	0,26	0,708
2894	0,83	0,8	0,66	0,92	0,48	0,738
4414	0,75	0,7	0,91	0,95	0,22	0,706
4639	0,57	0,52	0,85	0,99	0,39	0,664
1004	0,08	0,72	0,49	0,97	0,26	0,504
1917	0,41	0,71	0,36	0,87	0,29	0,528
1153	0,62	0,7	0,71	0,91	0,32	0,652
2244	0,94	0,56	0,74	0,99	0,43	0,732
2614	0,88	0,56	0,83	0,99	0,4	0,732
1624	0,33	0,8	0,62	0,99	0,33	0,614
3441	0,61	0,25	0,84	0,99	0,41	0,62
4793	0,52	/	0,68	0,99	0,33	0,63
4083	0,93	/	0,73	0,99	0,71	0,84
4906	0,46	0,36	0,59	0,94	0,31	0,532
1160	0,88	0,84	0,72	0,98	0,46	0,776
3416	0,56	/	0,5	0,82	0,2	0,52
1051	0,78	0,76	0,64	0,95	0,57	0,74
2580	0,71	0,45	0,63	0,96	0,44	0,638
1109	0,91	0,28	0,85	0,97	0,36	0,674
mFF1	**0,59**	**0,51**	**0,73**	**0,96**	**0,42**	**0,64**

Table 23. Detailed results on the 60 test videos of "Monastero dei Benedettini", considering the evaluation measure $AS F_1$ score. We used the backbone DenseNet. The "/" sign indicates that no samples from that class were present in the test video. The four classes are: 1) Antirefettorio, 2) Aula S. Mazzarino, 3) Cucina, 3) Ventre, whereas Neg. represents the negatives.

ID_visit	1	2	3	4	Neg.	AVG
4805	0,55	0,62	0,4	0,12	0,36	0,41
1804	0,32	0,1	0,39	0,05	0,21	0,214
4377	0,31	0,12	0,26	0,08	0,27	0,208
1669	0,68	0,34	0,67	0,9	0,33	0,584
1791	0,41	0,41	0,25	0,06	0,22	0,27
3948	0	/	0,33	0,26	0,49	0,27
3152	0,39	0,43	0,64	0,44	0,58	0,496
4361	0,26	0,35	0,32	0,08	0,2	0,242
3976	0,68	0,26	0,35	0,03	0,5	0,364
3527	0,65	0,37	0,38	0,28	0,44	0,424
4105	0,49	0	0,49	0,16	0,06	0,24
1399	0,55	0,1	0,33	0,23	0,29	0,3
3836	0,47	0	0,35	0,98	0,34	0,428
4006	0,47	0,49	0,49	0,55	0,26	0,452
4415	0,74	0,07	0,35	0,14	0,23	0,306
3008	0,57	0	0,5	0,99	0,41	0,494
4660	0,45	0,52	0,17	0,05	0,27	0,292
2826	0,31	0,29	0,33	0,06	0,25	0,248
1099	0,52	0,24	0,55	0,96	0,17	0,488
4391	0,54	0,13	0,14	0,2	0,24	0,25
3929	0,15	0	0,31	0,22	0,39	0,214
3362	0,2	0,35	0,38	0,05	0,22	0,24
1379	0,22	0	0,43	0,05	0,35	0,21
2600	0,32	0	0,39	0,31	0,26	0,256
1430	0,43	0	0,1	0,02	0,21	0,152
2956	0,41	0,2	0,16	0,29	0,31	0,274
4742	0,15	0,18	0,25	0,11	0,17	0,172
3651	0,32	0,25	0,74	0,48	0,28	0,414
1064	0,63	0,09	0,52	0,97	0,19	0,48
3818	0,43	0,49	0,2	0,06	0,29	0,294

Table 24. Continued from Table 23

ID_Visit	1	2	3	4	Neg.	AVG
2043	0,4	0,27	0,28	0,13	0,27	0,27
3996	0,23	0,3	0,22	0,05	0,38	0,236
3455	0,63	0,38	0,65	0,99	0,48	0,626
4785	0,06	0	0,25	0,45	0,07	0,166
2047	0,9	0,72	0,72	0,37	0,34	0,61
1912	0,42	0,25	0,2	0,08	0,3	0,25
3232	0,52	0,1	0,34	0,13	0,35	0,288
4442	0,61	0,22	0,33	0,06	0,12	0,268
3646	0,62	0,3	0,26	0,19	0,22	0,318
4833	0,54	0,55	0,26	0,1	0,21	0,332
3478	0,57	0,4	0,47	0,14	0,24	0,364
4396	0,41	0,1	0,2	0,07	0,12	0,18
2894	0,55	0,67	0,36	0,13	0,31	0,404
4414	0,59	0,26	0,83	0,11	0,08	0,374
4639	0,5	0,35	0,67	0,39	0,27	0,436
1004	0,22	0,47	0,17	0,11	0,17	0,228
1917	0,39	0,19	0,03	0,03	0,21	0,17
1153	0,47	0,33	0,21	0,14	0,16	0,262
2244	0,77	0,35	0,49	0,35	0,32	0,456
2614	0,43	0,06	0,39	0,99	0,26	0,426
1624	0,3	0,33	0,24	0,29	0,29	0,29
3441	0,49	0,11	0,44	0,88	0,42	0,468
4793	0,35	/	0,36	0,28	0,37	0,34
4083	0,88	/	0,41	0,25	0,48	0,505
4906	0,36	0,1	0,31	0,06	0,21	0,208
1160	0,79	0,5	0,39	0,14	0,28	0,42
3416	0,32	/	0,25	0,42	0,34	0,3325
1051	0,5	0,31	0,21	0,05	0,4	0,294
2580	0,36	0,29	0,33	0,05	0,28	0,262
1109	0,82	0,17	0,46	0,11	0,28	0,368
mASF1	0,46	0,26	0,37	0,28	0,28	**0,40**

Table 25. Detailed results obtained using the YoloV3 object detector Yolov3 on the 9 test videos of "Palazzo Bellomo". The last column reports the number of frames belonging to each test video. The last row indicates the average of mAP score obtained for each test video.

	mAP	#images
Test1	12,72	1644
Test2	13,61	1238
Test4	12,31	1398
Test5	8,65	848
Test6	8,9	1453
Test7	10,29	1200
Test8	10,98	2826
Test9	9,97	2004
Test10	7,85	791
AVG	**10,59**	**13402**

Table 26. Per-class AP values obtained on the 9 test videos. The "/" sign indicates that no samples from that class were present in the test videos.

Class	AP	Class	AP	Class	AP	Class	AP
0	10,53	50	6,67	100	5,34	150	0,07
1	41,45	51	0,00	101	9,45	151	2,18
2	49,60	52	0,60	102	5,44	152	17,34
3	41,79	53	0,00	103	10,69	153	15,24
4	13,27	54	0,00	104	10,77	154	35,10
5	66,73	55	1,85	105	2,71	155	0,00
6	66,97	56	0,00	106	2,67	156	0,71
7	72,62	57	0,00	107	0,00	157	3,90
8	52,41	58	0,00	108	3,80	158	1,92
9	68,21	59	0,00	109	6,11	159	0,00
10	2,69	60	1,85	110	16,50	160	0,63
11	14,79	61	0,13	111	0,00	161	18,92
12	2,19	62	6,05	112	12,32	162	11,44
13	44,65	63	0,00	113	0,00	163	18,31
14	35,27	64	2,34	114	0,00	164	12,00
15	16,58	65	2,22	115	21,75	165	26,78
16	61,05	66	0,74	116	9,50	166	11,97
17	28,68	67	18,36	117	4,98	167	11,18
18	46,37	68	9,19	118	1,97	168	1,04
19	9,68	69	5,70	119	1,64	169	23,41
20	51,04	70	1,14	120	33,36	170	11,95
21	11,11	71	2,64	121	0,39	171	0,52
22	45,00	72	9,19	122	9,74	172	2,82
23	48,80	73	11,26	123	3,62	173	5,12
24	10,40	74	0,11	124	17,70	174	37,13
25	0,00	75	34,85	125	0,00	175	30,37
26	17,47	76	6,56	126	0,96	176	18,87
27	0,00	77	0,75	127	1,05	177	/
28	14,01	78	10,02	128	4,66	178	0,00
29	0,00	79	5,16	129	12,58	179	0,00
30	3,11	80	16,57	130	15,71	180	3,43
31	16,71	81	17,89	131	6,43	181	0,00
32	0,61	82	/	132	2,74	182	0,43
33	2,76	83	25,18	133	0,00	183	0,10
34	0,99	84	0,86	134	0,00	184	0,00
35	0,00	85	1,00	135	0,00	185	0,04
36	0,56	86	0,00	136	5,02	186	0,00
37	4,60	87	15,17	137	4,49	187	0,00
38	16,01	88	12,64	138	0,30	188	3,86
39	3,38	89	8,99	139	0,00	189	0,00
40	14,20	90	25,49	140	0,83	190	10,42
41	0,00	91	0,29	141	11,66	**mAP**	**10.66**
42	0,00	92	0,00	142	0,00		
43	25,33	93	8,84	143	1,25		
44	0,11	94	11,98	144	6,56		
45	0,00	95	4,79	145	25,27		
46	0,00	96	0,00	146	2,01		
47	0,00	97	0,00	147	31,66		
48	20,07	98	0,00	148	2,15		
49	12,90	99	0,00	149	0,23		

Table 27. Per-class AP values obtained on the validation set using the optimal threshold of 0.001**.**

Class	AP
5.1 PortaAulaS.Mazzarino	36,05
5.2 PortaIngressoMuseoFabbrica	37,99
5.3 PortaAntirefettorio	20,44
5.4 PortaIngressoRef.Piccolo	26,07
5.5 Cupola	73,98
5.6 AperturaPavimento	80,52
5.7 S.Agata	74,89
5.8 S.Scolastica	66,84
6.1 QuadroSantoMazzarino	76,89
6.2 Affresco	60,39
6.3 PavimentoOriginale	37,41
6.4 PavimentoRestaurato	13,11
6.5 BassorilieviMancanti	25,7
6.6 LavamaniSx	41,55
6.7 LavamaniDx	25,8
6.8 TavoloRelatori	14,59
6.9 Poltrone	23,48
7.1 Edicola	42,57
7.2 PavimentoA	6,03
7.3 PavimentoB	0,93
7.4 PassavivandePavimentoOriginale	44,44
7.5 AperturaPavimento	33,12
7.6 Scala	46,49
7.7 SalaMetereologica	30,83
8.1 Doccione	46,65
8.2 VanoRaccoltaCenere	75,6
8.3 SalaRossa	18,52
8.4 ScalaCucina	42,48
8.5 CucinaProvv.	48,26
8.6 Ghiacciaia	47,7
8.7 Latrina	72,86
8.8 OssaeScarti	50,36
8.9 Pozzo	51,74
8.10 Cisterna	17,03
8.11 BustoPietroTacchini	41,15
Negatives	5,8
mAP	**40.51**

Table 28. Detailed results of the YoloV3 object detector on the the 60 **real visits. The second column reports the number of frames annotated with bounding box contained in each visit.**

ID_Visit	#images	mAP	ID_Visit	#images	mAP
156	1288	19,89	117	1349	14,11
154	2443	18,73	115	1511	20,64
153	1620	20,34	135	1396	14,73
155	786	18,58	137	2484	12,04
110	1671	18,36	136	566	16,99
109	679	19,96	132	1233	17,74
108	1065	13,84	134	1177	18,41
107	1728	15,24	130	1401	15,54
158	1660	13,37	105	1434	15,04
157	874	12,40	124	936	20,68
160	660	19,20	123	1571	15,15
159	654	16,08	103	2411	11,65
129	751	14,88	104	1492	10,43
125	544	19,51	122	1794	14,36
126	892	21,75	120	939	15,88
163	683	12,75	140	1050	10,68
165	1689	17,91	139	1454	16,62
161	1563	22,52	138	1736	13,27
162	979	20,02	145	1343	11,25
166	597	10,80	146	1370	15,12
164	1197	13,98	114	868	11,45
142	1161	17,27	112	840	9,49
144	868	11,61	111	726	13,79
143	824	9,64	113	851	11,29
101	1894	13,52	149	1612	11,69
102	824	10,50	148	847	8,84
100	1343	19,07	147	450	19,92
119	564	17,43	152	1519	18,71
118	740	16,74	150	1437	13,82
116	1618	16,55	151	942	16,71
			Tot./AVG	**71310**	**15,45**

Table 29. Per-class AP values obtained on the 60 real visits.

Class	AP
5.1 PortaAulaS.Mazzarino	37,86
5.2 PortaIngressoMuseoFabbrica	27,18
5.3 PortaAntirefettorio	2,23
5.4 PortaIngressoRef.Piccolo	15,44
5.5 Cupola	65,80
5.6 AperturaPavimento	0,89
5.7 S.Agata	41,82
5.8 S.Scolastica	31,64
6.1 QuadroSantoMazzarino	13,22
6.2 Affresco	53,69
6.3 PavimentoOriginale	6,95
6.4 PavimentoRestaurato	4,35
6.5 BassorilieviMancanti	16,57
6.6 LavamaniSx	0,83
6.7 LavamaniDx	0,58
6.8 TavoloRelatori	4,99
6.9 Poltrone	9,39
7.1 Edicola	34,01
7.2 PavimentoA	1,59
7.3 PavimentoB	3,26
7.4 PassavivandePavimentoOriginale	9,79
7.5 AperturaPavimento	22,77
7.6 Scala	20,44
7.7 SalaMetereologica	11,71
8.1 Doccione	13,78
8.2 VanoRaccoltaCenere	16,47
8.3 SalaRossa	17,35
8.4 ScalaCucina	13,36
8.5 CucinaProvv.	16,33
8.6 Ghiacciaia	3,98
8.7 Latrina	22,29
8.8 OssaeScarti	29,39
8.9 Pozzo	13,43
8.10 Cisterna	5,45
8.11 BustoPietroTacchini	23,13
Negatives	16,81
mAP	**17,47**

Table 30. Number of patches extracted from each of the 10 test videos of "Palazzo Bellomo".

Test video	#images
Test1	2568
Test2	2048
Test3	2672
Test4	2224
Test5	1439
Test6	2086
Test7	2148
Test8	4108
Test9	2914
Test10	1520
Total	**23727**

Table 31. Number of image patches extracted from the 60 test videos of "Monastero dei Benedettini".

ID_Visit	#images	ID_Visit	#images
100	770	135	765
101	696	136	414
102	613	137	354
103	1733	138	824
104	768	139	707
105	929	140	494
107	1011	142	770
108	659	143	536
109	234	144	598
110	918	145	897
111	365	146	1307
112	727	147	173
113	288	148	692
114	561	149	954
115	623	150	609
116	968	151	652
117	810	152	699
118	907	153	1244
119	545	154	1691
120	669	155	666
121	774	156	709
122	1156	157	515
123	957	158	846
124	652	159	544
125	491	160	327
126	702	161	985
129	587	162	902
130	820	164	693
132	771	165	1133
134	884	166	690
		Total	**44978**

Table 32. Results using Densenet.
Points of Interest Retrieval

1) Palazzo Bellomo

Variant	K	Precision	Recall	F1 score
1 - One Shot	1	0.02	0.01	0.00
2 - Many Shots	1	**0.62**	**0.59**	**0.6**
	3	0.62	0.56	0.56
	5	0.62	0.56	0.56
	7	0,61	0,56	0,56
	9	0,61	0,55	0,56
	11	0,61	0,55	0,55

2) Monastero dei Benedettini

Variant	K	Precision	Recall	F1 score
1 - One shot	1	0.38	0.07	0.09
2 - Many Shots	1	0,83	0,83	0,83
	3	0,84	0,83	0,83
	5	**0,84**	**0,84**	**0,83**
	7	0,84	0,83	0,83
	9	0,84	0,83	0,83
	11	0,83	0,83	0,82

Table 33. Results of the binary classifier obtained using a KNN with different values of K.

K	Precision	Recall	F1 score	Support
1	0,62	0,61	0,62	
3	0,62	0,65	0,63	
5	0,63	0,67	0,64	1980
7	0,64	0,69	0,65	
9	**0,65**	**0,7**	**0,66**	

Table 34. Results of the multi-class classifier obtained using a KNN with different values of K.

K	Precision	Recall	F1 score	Support
1	0,2	0,2	0,2	
3	0,2	0,23	0,19	
5	0,2	0,24	0,21	1980
7	0,22	0,24	0,22	
9	**0,23**	**0,27**	**0,23**	