

Semantic Object Segmentation in Cultural Sites using Real and Synthetic Data

Francesco Ragusa^{1,2}, Daniele Di Mauro¹, Alfio Palermo¹, Antonino Furnari¹, Giovanni Maria Farinella¹¹FPV@IPLAB - Department of Mathematics and Computer Science, University of Catania, Italy²XDG - XENIA Progetti s.r.l., Acicastello, Catania, Italy

Abstract

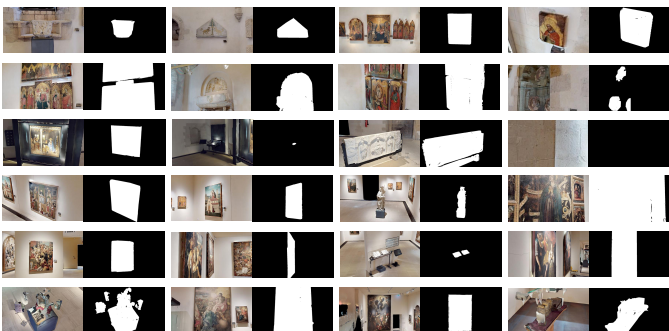
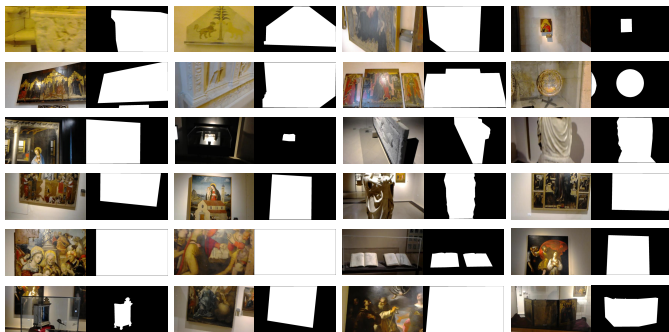
We consider the problem of object segmentation in cultural sites. Since collecting and labeling large datasets of real images is challenging, we investigate whether the use of synthetic images can be useful to achieve good segmentation performance on real data. To perform the study, we collected a new dataset comprising both real and synthetic images of **24 artworks** in a cultural site. The synthetic images have been automatically generated from the 3D model of the considered cultural site using a tool developed for that purpose. Real and synthetic images have been labeled for the task of semantic segmentation of artworks. We compare three different approaches to perform object segmentation exploiting real and synthetic data. The experimental results point out that the use of synthetic data helps to improve the performances of segmentation algorithms when tested on real images. Satisfactory performance is achieved exploiting semantic segmentation together with image-to-image translation and including a small amount of real data during training. To encourage research on the topic, we publicly release the proposed dataset at the following url: <https://iplab.dmi.unict.it/EGO-CH-OBJ-SEG/>

Dataset

DETAILS ABOUT THE PROPOSED DATASET, INCLUDING THE NUMBER OF REAL AND SYNTHETIC IMAGES

	Resolution	Artworks	Environments	Masks	Training Images	Val. Images	Test Images	All Images
Real	1280x720	24	11	5624	4740 (85%)	170 (3%)	678 (12%)	5580
Synthetic	1280x720	24	11	24000	12000 (50%)	1200 (5%)	10800 (45%)	24000

ID	Class	Masks	ID	Class	Masks	ID	Class	Masks
2.1	Acquasantiera	244	5.1	Annunciazione	303	9.1	Adorazione dei Magi	230
2.3	LastraconLeoni	248	5.2	LibroD'OreMin	253	9.2	S.ElenaCost.eMadon.	247
3.1	MadonnainTrono	237	5.3	LastraG.Cabastida	307	9.3	Taccuini di Disegni	212
3.2	FrammentoS.Leo	186	5.4	MadonnadelCard.	223	10.1	MartirioS.Lucia	196
4.1	MadonnainTrono	245	7.1	DisputaS.Tomm.	200	10.2	Volto di Cristo	210
4.2	MonumentoE.d'Aragona	222	7.2	TraslazioneS.Casa	279	11.1	Miracolodi.S.Orsola	250
4.3	Trasf.Cristo	233	7.3	MadonnacoliBam.	231	11.2	Immacolata	219
4.4	Piatti	208	8.1	ImmacolataConc.	245	21.1	Storie della Genesi	196

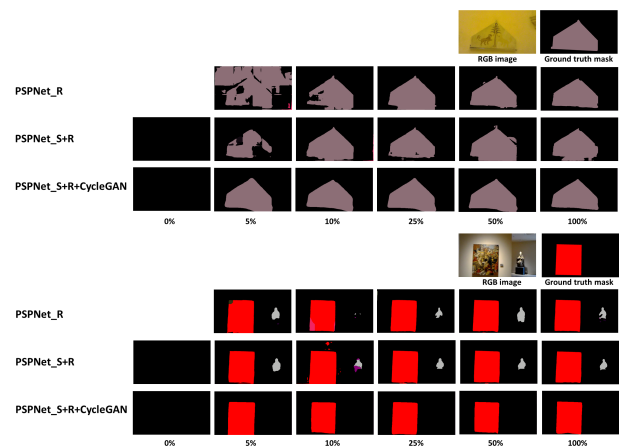


Baselines

In order to show the goodness of the dataset we present some results on it using several baselines:

- **PSPNet on real data (PSPNet_R)**. To assess the amount of labeled real data needed to obtain reasonable performance, we train the model with 5%, 10%, 25%, 50% and 100% of training data.
- **PSPNet on real and synthetic data (PSPNet_S+R)** it uses both synthetic and real data during training. Two stages training phase: in the first stage, PSPNet is trained using only synthetic data, in the second stage, we fine-tune using real data.
- **PSPNet on real and translated synthetic data (PSPNet_S+R+CycleGAN)** This approach uses CycleGAN to reduce the domain shift between real and synthetic images.

Results



Results of the compared methods on real test data

	Real Training	Accuracy %	Class Accuracy %	MeanIoU %	FWAVACC %
PSPNet_R	5%	71.10	43.73	29.47	56.96
	10%	76.28	46.66	31.88	62.49
	25%	80.95	58.54	43.47	68.86
	50%	82.47	59.40	44.37	70.86
	100%	83.51	63.15	47.15	72.76
PSPNet_S+R	0%	58.32	8.45	5.50	35.60
	5%	70.18	42.38	27.06	56.54
	10%	80.23	57.87	40.87	67.71
	25%	82.14	58.55	45.03	69.90
	50%	83.07	65.09	47.80	72.51
PSPNet_S+R+CycleGAN	100%	83.70	59.02	47.06	72.00
	0%	80.52	53.93	39.43	67.77
	5%	87.82	77.90	59.49	79.85
	10%	88.58	81.67	66.19	80.45
	25%	88.62	79.91	60.93	80.57
	50%	90.23	78.72	68.25	82.44
	100%	90.23	81.22	68.20	82.77

Results of PSPNet_S+R on the synthetic data

	Chunk	Accuracy %	Class Accuracy %	MeanIoU %	FWAVACC %
PSPNet_S+R	0%	95.31	88.10	81.48	91.20
	5%	77.88	58.39	39.14	69.28
	10%	80.32	60.94	43.44	71.44
	25%	86.55	63.76	50.15	77.48
	50%	83.15	62.93	47.08	74.33
	100%	86.63	59.64	49.35	76.90

Conclusions

We have considered the problem of object segmentation in cultural sites. Starting from the assumption that manually labeling images with semantic masks is expensive and time-consuming, we have studied whether the availability of large amounts of synthetic images can allow to improve performance on real images. We have hence compared three approaches to semantic segmentation which use both real and synthetic images. Results highlight that synthetic images can be beneficial to improve performance on real data, especially when coupled with image-to-image translation techniques, to reduce the domain shift arising from the two different data sources.

Contact

Daniele Di Mauro, PhD

dimauro@dm.unict.itFPV@IPLAB - Department of Mathematics and Computer Science,
University of Catania, Italy

Acknowledgment

This research is supported by MIUR - Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 - Dottorati Innovativi a Caratterizzazione Industriale XXXIII CICLO, by the project VALUE - Visual Analysis for Localization and Understanding of Environments (N. 08CT6209090207, CUP G69J18001060007) granted by PO FESR 2014/2020 - Azione~1.1.5, and by 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'Identità Siciliana - Dipartimento dei Beni Culturali e dell'Identità Siciliana and Polo regionale di Siracusa per i siti culturali - Galleria Regionale di Palazzo Bellomo.