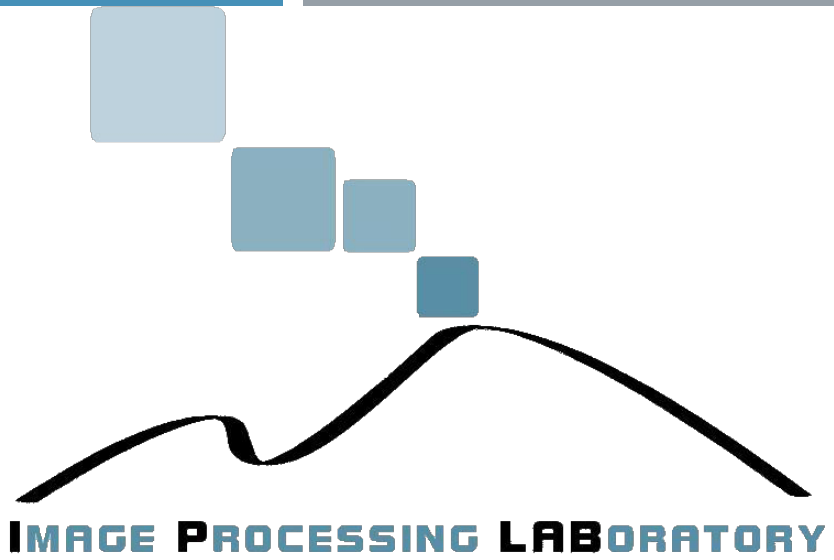
UNIVERSITÀ
degli STUDI
di CATANIA



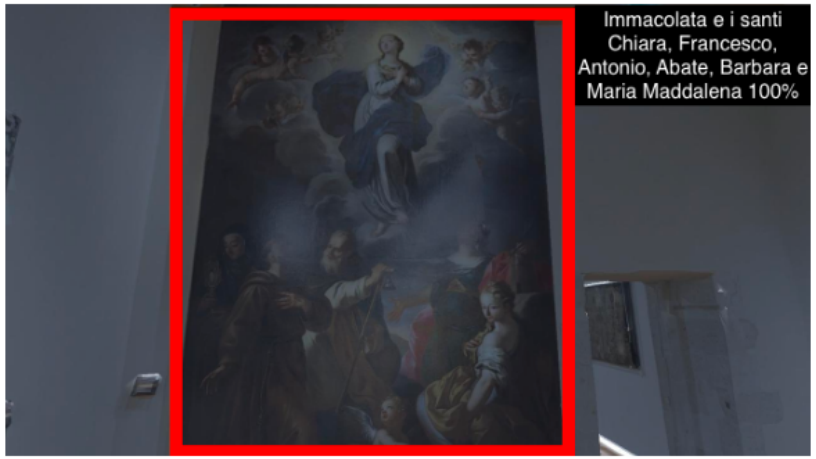
Unsupervised Domain Adaptation for Object Detection in Cultural Sites

G. Pasqualino, A. Furnari, G. M. Farinella

https://iplab.dmi.unict.it/EGO-CH-OBJ-ADAPT/

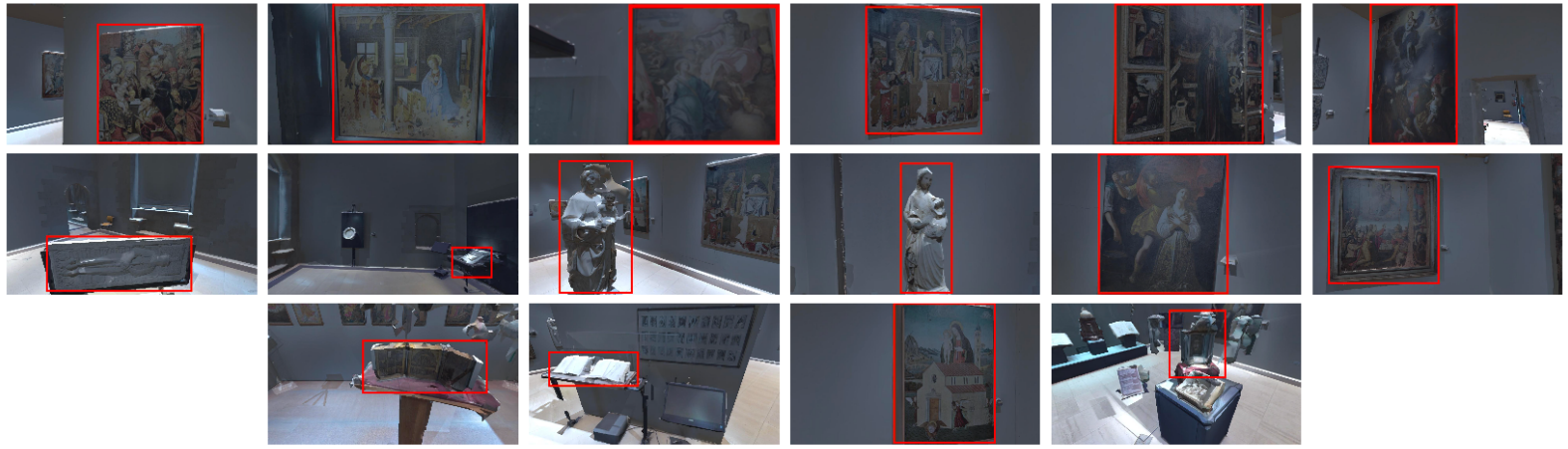
MOTIVATIONS

- Recognize artworks in cultural sites;
- Expensive labeling process;
- Automatic labeling tool for synthetic images;
- Training on synthetic data and testing on real data produces poor performance;
- Domain difference between synthetic and real images.

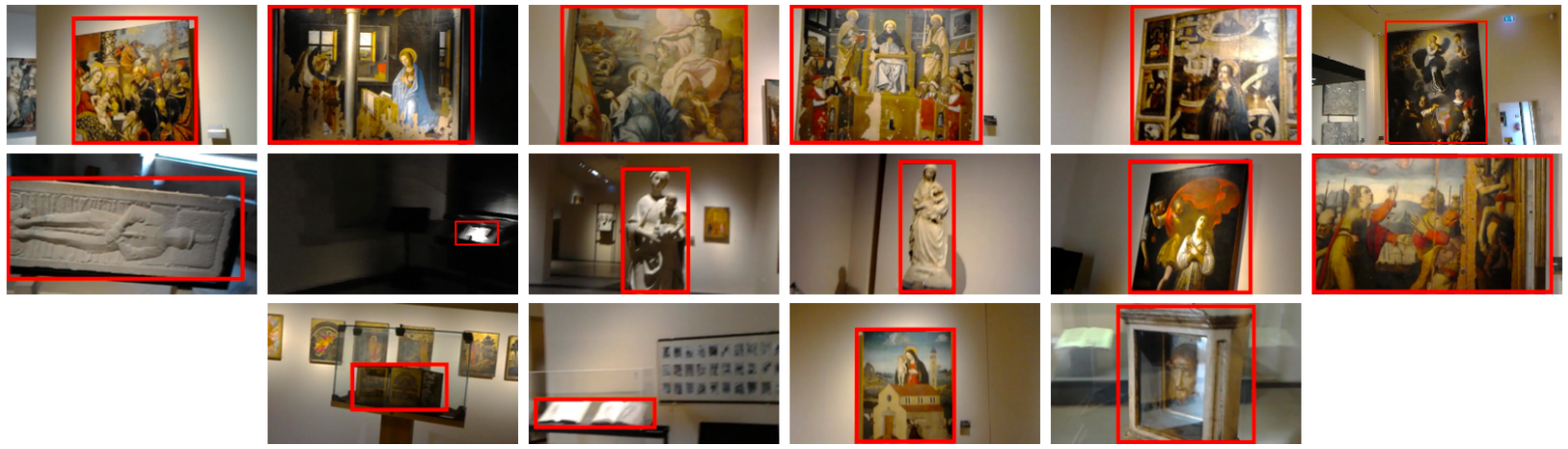


Training and Test with synthetic images

Training with synthetic images and test on real images



Sample synthetic images of the 16 artworks of our dataset.



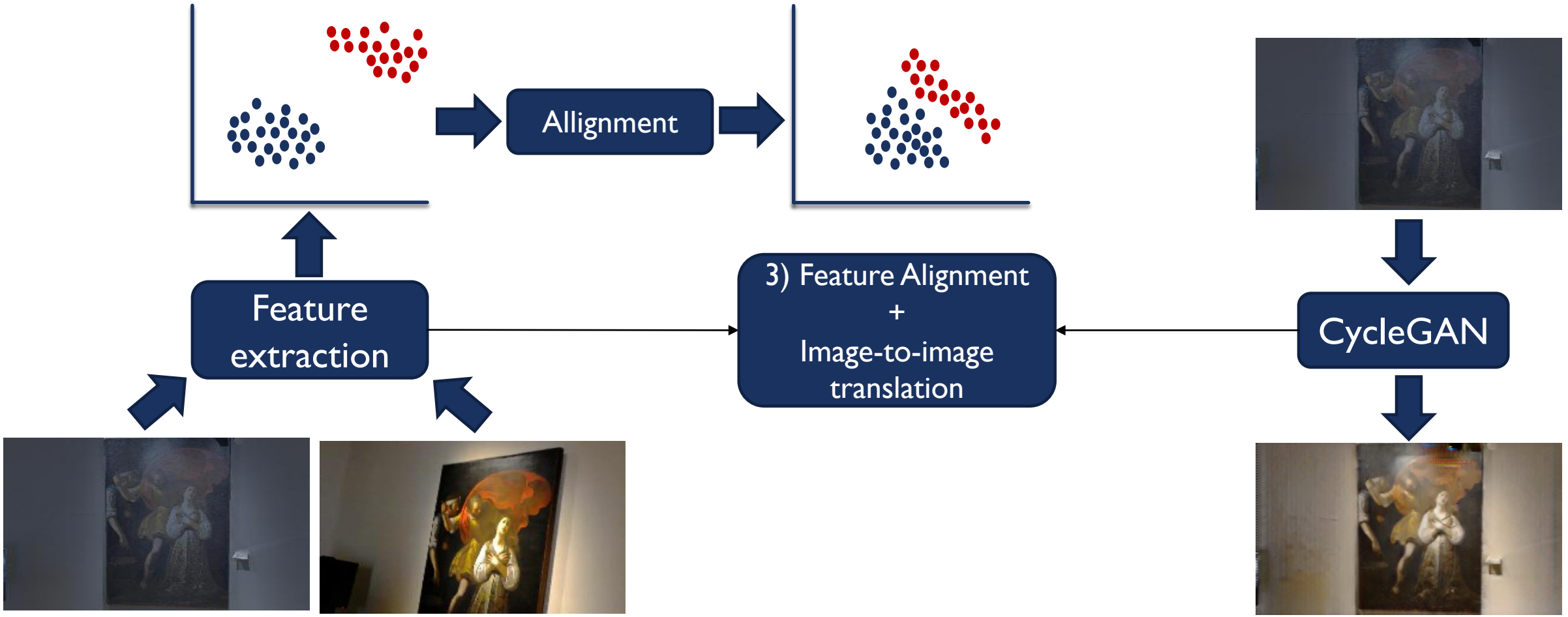
Sample real images of the 16 artworks of our dataset.

DATASET

- 16 artworks located in "Galleria Regionale di Palazzo Bellomo"
- 75244 synthetic images:
  - 51284 training set
  - 23960 test set
- 2190 real images:
  - 1502 training set
  - 688 test set

More information about the tool used can be found here: S. Orlando, A. Furnari, G. M. Farinella (2020). Egocentric Visitor Localization and Artwork Detection in Cultural Sites Using Synthetic Data . Pattern Recognition Letters - Special Issue on Pattern Recognition and Artificial Intelligence Techniques for Cultural Heritage.
http://iplab.dmi.unict.it/SimulatedEgocentricNavigations/

DOMAIN ADAPTATION METHODS



RESULTS AFTER DIFFERENT AMOUNTS OF TRAINING ITERATIONS

	Training Iterations						
Model	6K	12K	22K	32K	42K	52K	62K
F. RCNN	2.27%	**9.67%**	5.79%	3.58%	3.33%	3.81%	3.62%
RetinaNet	9.83%	**14.44%**	13.22%	12.31%	12.09%	12.44%	11.97%

Performance of Faster RCNN and RetinaNet on real images after different amounts of training iterations on synthetic images.

		Training epochs for CycleGAN					
Model (iter)	N.A.	10	20	30	40	50	60
F. RCNN (62K)	3.62%	25.16%	25.49%	25.51%	26.68%	27.65%	**28.25%**
RetinaNet (62K)	11.97%	27.30%	32.14%	**34.15%**	32.66%	32.79%	32.82%
F. RCNN (12K)	9.67%	29.93%	32.84%	33.95%	31.45%	**34.19%**	31.58%
RetinaNet (12K)	14.44%	34.51%	35.45%	34.84%	35.34%	**35.76%**	35.74%

Results obtained transforming real images to synthetic at test time. The models have been trained onsynthetic images. N.A. stands for No Adaptation.

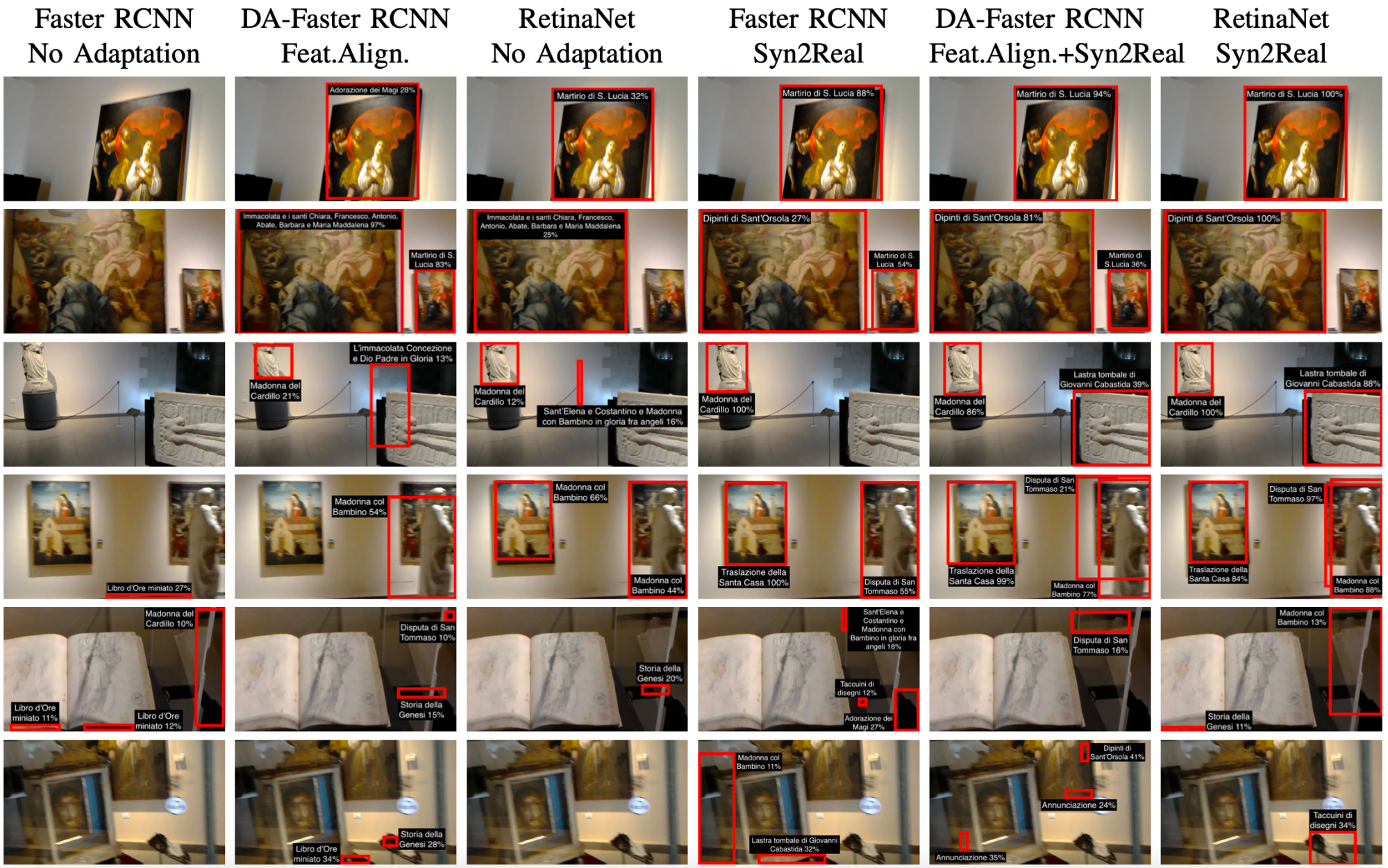
		Training epochs for CycleGAN					
Model	N.A.	10	20	30	40	50	60
F. RCNN	9.67%	18.76%	20.92%	21.22%	23.17%	24.45%	**26.03%**
RetinaNet	14.44%	40.13%	44.29%	46.05%	47.89%	49.96%	**55.54%**

Results obtained training the models on synthetic images transformed to real and tested on real images. N.A. stands for No Adaptation.

RESULTS

Object Detector	Adaptation	mAP
Faster RCNN	None	9.67%
RetinaNet	None	14.44%
Faster RCNN	Real2Syn (Test set)	34.19%
RetinaNet	Real2Syn (Test set)	35.76%
Faster RCNN	Syn2Real (labeled Training set)	26.03%
RetinaNet	Syn2Real (labeled Training set)	**55.54%**
DA-Faster RCNN	Feat.Align.	12.94%
DA-Faster RCNN	Feat.Align.+Real2Syn (Test set and Unlabeled Training set)	19.88%
DA-Faster RCNN	Feat.Align.+Syn2Real (labeled Training set)	33.20%

Qualitative Results



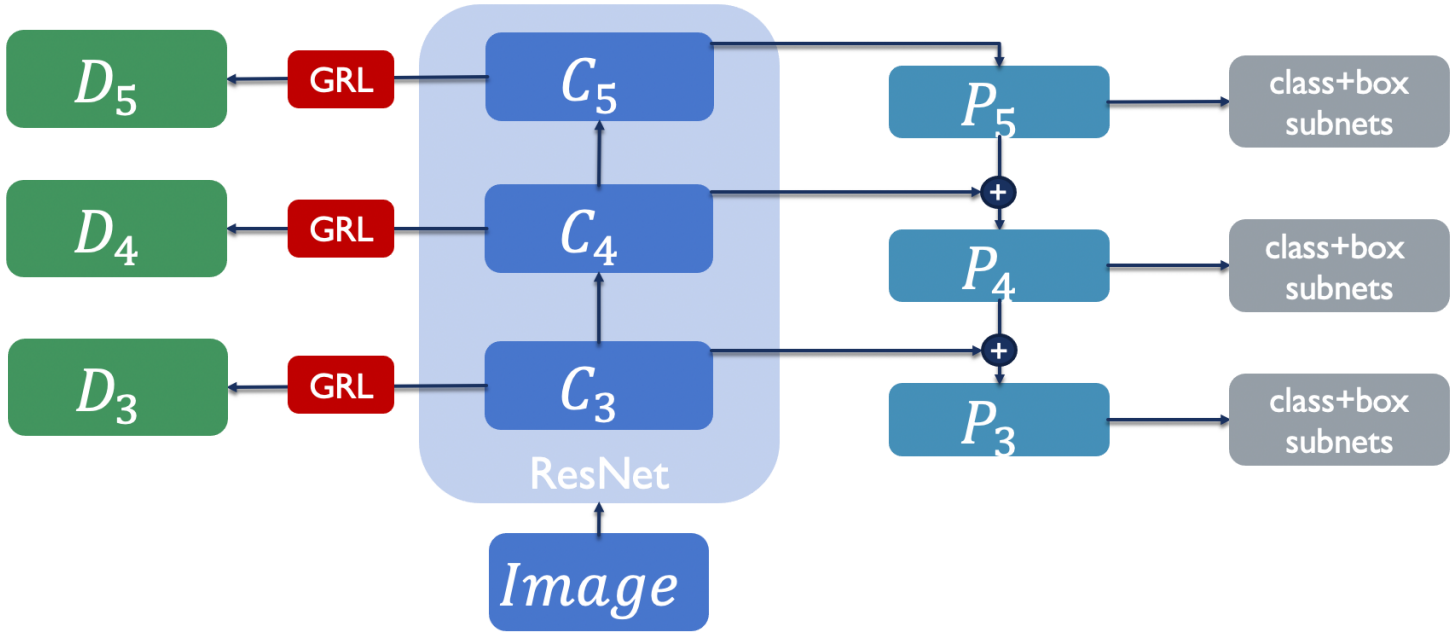
COMPUTATIONAL RESOURCES ANALYSIS

Model	Hours (Days)
RetinaNet (12K iterations)	~ 10 (~ 0.5)
RetinaNet (62K iterations)	~ 65 (~ 3)
Faster RCNN (62K iterations)	~ 131 (~ 5.5)
DA-Faster RCNN	~ 142 (~ 6)
CycleGAN	~ 1470 (~ 61)
CycleGAN + RetinaNet	~ 1535 (~ 64)
CycleGAN + Faster RCNN	~ 1601 (~ 66)
CycleGAN + DA-Faster RCNN	~ 1612 (~ 67)

CONCLUSIONS

- RetinaNet combined with CycleGAN significantly reduce the domain gap between synthetic and real images
- RetinaNet in this case performs better than Faster-RCNN and DA-Faster-RCNN combined with CycleGAN
- CyleGAN significantly increases the performance of each object detector but requires many computational resources
- Combining feature alignment and image-to-image translation techniques allows to achieve better results

EXTENSION: DA-RETINANET



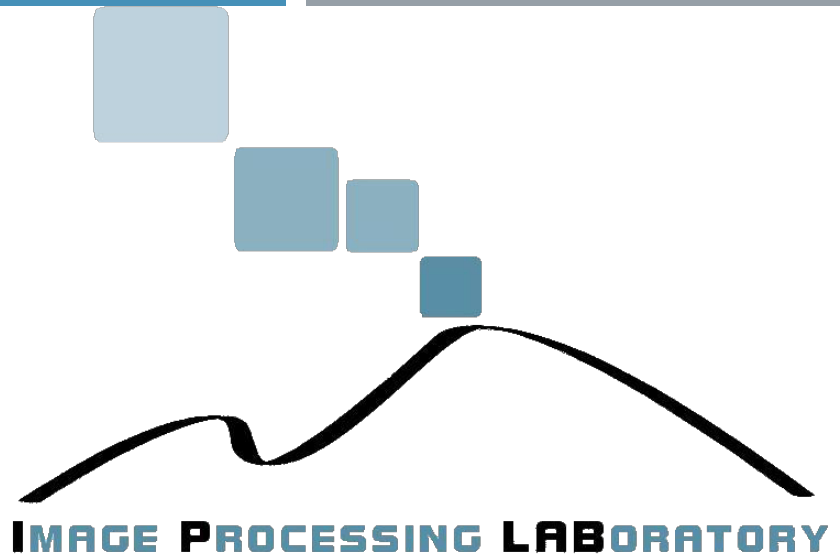
RESULTS

	image-to-image translation		
Model	None	Real2Syn	Syn2Real
DA-Faster RCNN	12.94%	19.88%	33.20%
Strong-Weak	25.12%	33.33%	47.70%
DA-RetinaNet	**31.04%**	**37.49%**	**58.01%**

Paper: https://arxiv.org/pdf/2008.01882v2.pdf

Code: https://github.com/fpv-iplab/DA-RetinaNet

UNIVERSITÀ
degli STUDI
di CATANIA



Unsupervised Domain Adaptation for Object Detection in Cultural Sites

Thanks for your Attention!

G. Pasqualino, A. Furnari, G. M. Farinella

https://iplab.dmi.unict.it/EGO-CH-OBJ-ADAPT/