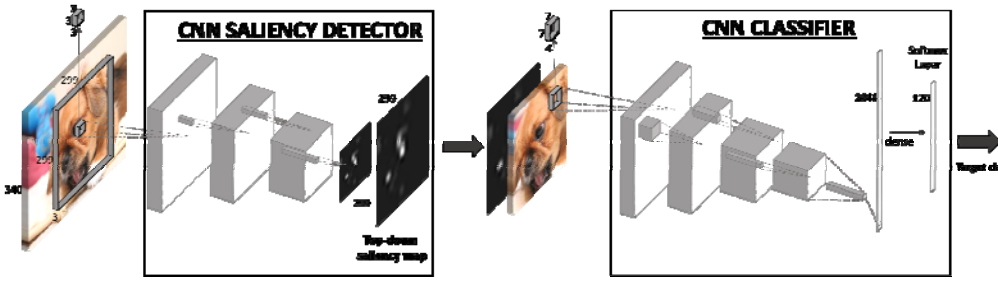
Deeply Incorporating Human Capabilities into Machine Learning Models for Fine-Grained Image Categorization

Ph.D Student: Francesca Murabito – Tutor: Prof. C. Spampinato
PeRCeiVeLab

Dottorato di Ricerca in Ingegneria dei Sistemi, Energetica, Informatica e delle Telecomunicazioni XXXI CICLO

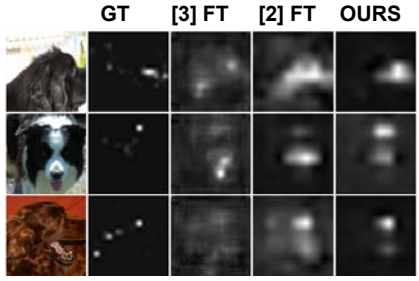
TOP-DOWN SALIENCY DETECTION DRIVEN BY VISUAL CLASSIFICATION

1. A saliency dataset containing about 10 K maps on the 120 Standord Dogs dataset [1]
2. A saliency detection method guided by a visual classification task.
3. A classification network able to exploit such task-based saliency maps in order to improve the fine-grained recognition performance.



Qualitative and Quantitative Performance Analysis

Saliency Dataset	
Number of images	9,861
Number of classes	120
Avg. number of images per class	82.2
Avg. number of fixation points per image	6.2



Dataset	Method	s-AUC	NSS	CC
OUR	SALICON FT[2]	0.837	3.899	0.428
OUR	SalNet FT[3]	0.817	4.174	0.432
OUR	SalClassNet	0.862	4.239	0.461
POET	SALICON FT[2]	0.695	1.669	0.356
POET	SalNet FT[3]	0.660	1.378	0.300
POET	SalClassNet	0.715	1.908	0.387

VISUAL WORLD SEMANTICS FOR FINE-GRAINED IMAGE CLASSIFICATION

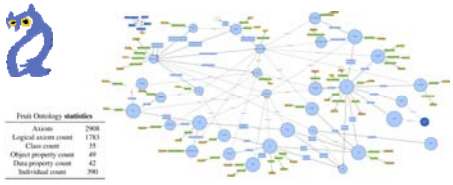
Domain Ontology

Owl classes and properties describing the application domain, i.e. the target class and the context classes

Annotation Ontology

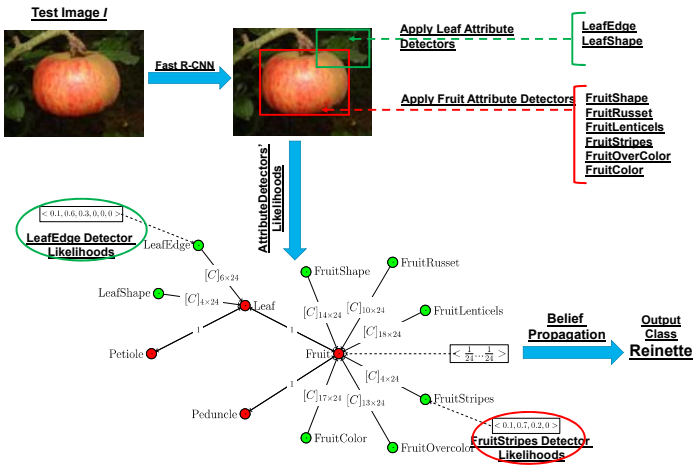
Owl classes and properties necessary to make the ontology readable by the annotation tool

W3C SKOS



An annotation tool which is able to guide and constrain the annotation process with the concepts encoded in the domain ontology

Classification Workflow: from CNN-Detector Likelihoods to Belief Propagation



Performance Analysis

Fruit Dataset

Malus Domestica		Pyrus Communis		Prunus Avium	
Variety	#img	Variety	#img	Variety	#img
Ambrosia	117	Abate	286	Bing	67
Braeburn	83	Anjou	191	Black Tartarian	45
Cameo	70	Conference	181	Burlat	71
Fuji	181	Coscia	166	Ferrovia	77
Golden Delicious	324	Doyenne du Comice	77	Lapins	87
Granny Smith	360	Kaiser	111	Rainer	140
Pink Lady	239	Williams	165	Stella	31
Reinette	242				
Royal Gala	150				
Stark Delicious	411				
Total	2185		1177		518
Total Dataset	3872				

Semantic Classification

Method	Accuracy
GoogleNet[4]	43.5%
ResNet[5]	43.1%
DenseNet[6]	46.2%
Ours	48.6%

Key References

[1] Khosla, A., Jayadevaprakash, N., Yao, B., Fei-Fei, L.. Novel dataset for fine-grained image categorization, in: CVPRW 2011.
[2] Huang, X., Shen, C., Boix, X., Zhao, Q.. Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks, in: ICCV 2015
[3] Pan, J., Sayrol, E., Giro-i Nieto, X., McGuinness, K., O'Connor, N.E., 2016. Shallow and deep convolutional networks for saliency prediction, in: CVPR 2016
[4] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A. Going deeper with convolutions, in CVPR 2015
[5] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition, in CVPR 2016
[6] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. Densely Connected Convolutional Networks, in CVPR 2017