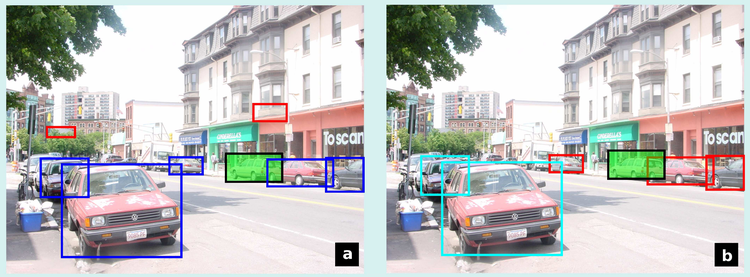
Towards Cautious Collective Inference for Object Verification

José Oramas M. , Luc De Raedt, Tinne Tuytelaars

WACV 14

Abstract

- Use most certain object hypotheses to bootstrap the prediction of the other objects.
- Consider that objects associate with each other based on underlying "relationships" rather than the class they belong to.



Motivation

- Contextual cues have shown to be of great benefit for object detection.
- Top-ranking object hypotheses from state-of-the-art detectors are usually correct --> good seed hypotheses.
- "Relationships" --> group behaviour in nature and imposed or desired rules on man-made objects.

Challenges

- Defining relations between objects.
- How to select the contextual objects.
- How to define the influence of contextual objects.
- How to combine the local and contextual sources of information.

Defining relations between objects

Object:

Format 1
- x_i: X coordinate in the 2D image space.
- y_i: Y coordinate in the image space.
- θ_i: Azimuth Pose angle
- s_i: Detection score

Format 2 (based on Li et al's, CVPR 2012)
- x_i: X coordinate in the 2D image space.
- y_i: Y coordinate in the image space.
- ρ_i: Scale of the of the object bounding box
- a_i: Aspect Ratio
- s_i: Detection score

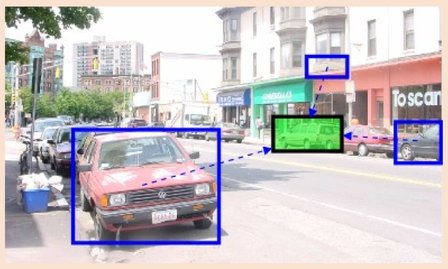
Format 3 (based on Perko et al's, CVIU 2010)
- x_i: X coordinate in the 2D image space.
- y_i: Y coordinate in the image space.
- s_i: Detection score

Relations:

$$r_{ij}^{(RF1)} = (\Delta x_{ij}, \Delta y_{ij}, \Delta\theta_{ij})$$

$$r_{ij}^{(RF2)} = (rx_{ij}, ry_{ij}, r\rho_{ij}, ra_{ij})$$

$$r_{ij}^{(RF3)} = (\Delta x_{ij}, \Delta y_{ij})$$



How to Select the Contextual Objects

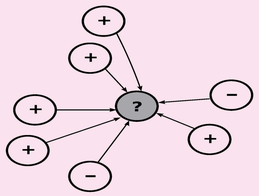
Aggressive Inference

Weighted-Vote Relational Neighbor Classifier [1]

$$wvRN(o_i|N_i) = \frac{1}{z}\sum_{o_j \in N_i} p(o_i|r_{ij}).w_j$$

$$p(o_i|r_{ij}) = \frac{p(r_{ij}|o_i)p(o_i)}{p(r_{ij}|o_i)p(o_i) + p(r_{ij}|\neg o_i)p(\neg o_i)}$$

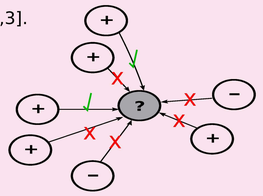
$$w_j = p(o_j|s_j) = \frac{p(s_j|o_j)p(o_j)}{p(s_j|o_j)p(o_j) + p(s_j|\neg o_j)p(\neg o_j)}$$



Cautious Inference

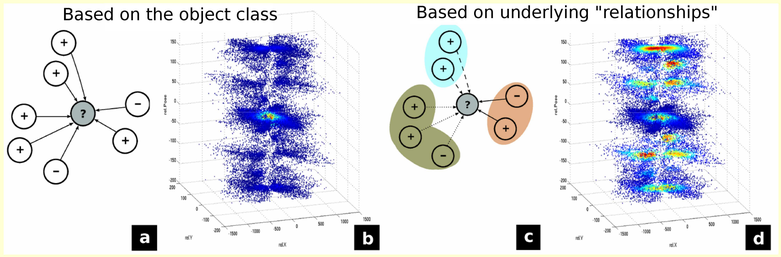
Inspired in the cautious and iterative methods from [2,3].

$$wvRN(o_i^u|N_i) = \frac{1}{z}\sum_{o_j^k \in (N_i \cap O^k)} p(o_i^u|r_{ij}).w_j$$



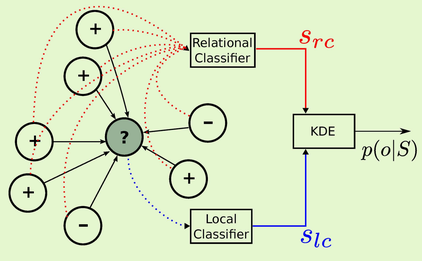
How to define the influence of contextual objects

Relationships are extracted via XMeans clustering [4].



How to combine local and contextual sources of information

Inspired in the score combination method from [2].



$$S = (s_{lc}, s_{rc})$$

$$p(o|S) = \frac{p(S|o)p(o)}{\sum_{(o,\neg o)} p(S|o)p(o)}$$

Experiments

Settings

Dataset :
- KITTI object detection benchmark.
- MIT StreetScenes.

Local Classifiers
- Viewpoint-aware DPM detector from [6] (8 views) .
- DPM detector from [7] (No viewpoint information) .

Using only contextual information

Dataset KITTI benchmark	Relations Representation : RF1				Relations Representation : RF2			
	Class-based Hom.		Relation-based Hom.		Class-based Hom.		Relation-based Hom.	
	Global		Global		Global		Global	
Set all	aggre. 0.29	caut. ***0.38***	aggre. 0.28	caut. ***0.37***	aggre. 0.32	caut. 0.40	aggre. 0.41	caut. ***0.50***
Dataset MIT StreetScenes	**Relations Representation : RF3**				**Relations Representation : RF2**			
	Class-based Hom.		Relation-based Hom.		Class-based Hom.		Relation-based Hom.	
	Global		Global		Global		Global	
Set all	aggre. 0.54	caut. ***0.63***	aggre. 0.51	caut. 0.59	aggre. 0.51	caut. ***0.56***	aggre. 0.49	caut. 0.55

Local + contextual Information [6]

Dataset KITTI benchmark	RF1		RF2	
	Class-based Homophily		Relation-based Homophily	
	Global		Global	
Set all	Detector [6] 0.61 $\pm$0.011	aggre. 0.61 $\pm$0.009	caut. 0.63 $\pm$0.007	aggre. 0.65 $\pm$0.011	caut. ***0.68 $\pm$0.003***
Dataset MIT StreetScenes	**RF3**		**RF2**	
	Class-based Homophily		Class-based Homophily	
	Global		Global	
Set all	Detector [6] 0.69 $\pm$0.006	aggre. 0.77 $\pm$0.001	caut. ***0.80 $\pm$0.028***	aggre. 0.73 $\pm$0.011	caut. 0.76 $\pm$0.014

Local + contextual Information [7]

Dataset KITTI benchmark	RF3		RF2	
	Class-based Homophily		Relation-based Homophily	
	Global		Global	
Set all	Detector [7] 0.65 $\pm$0.005	aggre. 0.68 $\pm$0.007	caut. 0.71 $\pm$0.007	aggre. 0.72 $\pm$0.009	caut. ***0.75 $\pm$0.003***
Dataset MIT StreetScenes	**RF3**		**RF2**	
	Class-based Homophily		Class-based Homophily	
	Global		Global	
Set all	Detector [7] 0.62 $\pm$0.004	aggre. 0.66 $\pm$0.011	caut. ***0.71 $\pm$0.012***	aggre. 0.65 $\pm$0.026	caut. 0.69 $\pm$0.014

Conclusions

- In a Global Neighborhood, Cautious Inference > Aggressive Inference.
- Considering underlying "relationships" is useful for cases where the local information about the unknown object is unavailable.
- Cautious Inference + "relationships" in RF2 format are better for contrained camera settings.
- In most cases, Global Neighborhood > Near Neighborhood.

[1] S. Mackassy and F. Provost, JMLR 2007.
[2] L. McDowell et al., JMLR 2009.
[3] J. Neville et al., SRL WS@AAAI 2000.
[4] A. Pelleg, ICML 2000.
[5] R. Perko and A. Leonardis, CVIU 2010.
[6] R. Lopez et al., WS@ICCV 2011.
[7] P. Felzenszwalb et al., TPAMI 2010.