Barbara Caputo
University of Rome La Sapienza
What is the vision behind ALOOF?
What is the vision behind ALOOF?

What is this on the table??
What is the vision behind ALOOF?

What is this on the table??
Let me take a snapshot
and search on the Web…

..if the robot could learn
from the Web like us....
What is the vision behind ALOOF?

Let’s see on the Web..

**Holy Cannoli!!**
March 7th 2009 by Maria Liberati

A cannoli, which is Sicilian for “little tube”, is an ever-popular pastry dessert found sprinkled throughout cafes and restaurants in Italy. We thank the Sicilian region of Italy for these rich desserts that are found today throughout the United States as well, thanks to the immigrants who brought the recipes over in the early 20th century. Cannoli originally came from the Palermo region of Sicily as early as the first century A.D., served as a treat and given to friends in dozens during the season of Carnevale, a festive European celebration held right before Lent. It is said that cannoli was most likely a symbol of fertility during the pre-Easter celebration similar in some ways to Mardi Gras.
What is the vision behind ALOOF?

Let’s see on the Web..
What is the vision behind ALOOF?

Let’s see on the Web..

..cannoli..pastry..
..ricotta filled.. goes in the fridge..
What is the vision behind ALOOF?

Let’s see on the Web..

Ok, now I know what to do with it.

..cannoli..pastry..
..ricotta filled.. goes in the fridge..
Why this is difficult?

Let’s see on the Web.

Ok, now I know what to do with it.

..cannoli pastry..
..ricotta filled.. goes in the fridge..
Why this is difficult?

How does a robot know it doesn’t know?

Let’s see on the Web...

Ok, now I know what to do with it

cannoli pastry
ricotta filled goes in the fridge...
Why this is difficult?

How does a robot know it doesn’t know?

How to extract from the Web information useful for a robot, from a robot query?

Let’s see on the Web...

Ok, now I know what to do with it...

cannoli, pastry...
ricotta filled... goes in the fridge...
Why this is difficult?

How does a robot know it doesn't know?

How to extract from the Web information useful for a robot, from a robot query?

How to bridge between the Web representations and the own robot representation?
Why this is difficult?

How does a robot know it doesn’t know?

How to extract from the Web information useful for a robot, from a robot query?

Let’s see on the Web...

Ok, now I know what to do with it

...cannoli pastry...
...ricotta filled.. goes in the fridge...

How to use such information to build a semantic object map?

How to bridge between the Web representations and the own robot representation?
Why this is difficult?

**How does a robot know it doesn’t know?**

**How to extract from the Web information useful for a robot, from a robot query?**

**How to bridge between the Web representations and the own robot representation?**

**How to use such information to build a semantic object map?**
The Project Structure

WP5 Integration and Evaluation: BHAM

WP2 Detection and Filling of Knowledge Gaps: BHAM

WP1 Modelling Meta Modal Knowledge: UNIROMA

WP3 Semantic and Perceptual Web Mining: INRIA

WP4 Situated Active Perception: TUW

WP6 Project Management and Dissemination Activities: UNIROMA
Partners

Web Mining
INRIA, UNIROMA

Robot Behaviours
BHAMP, TUW

Robot Perception
TUW, UNIROMA

Integration
BHAMP, INRIA, TUW, UNIROMA

Machine Learning
UNIROMA, INRIA

Robot Systems
BHAMP, TUW

Robot Planning
BHAMP, UNIROMA
Working Scenario

Task1: Where is X?
Task2: What is Y?

Learning from the Web:
object type,
room type,
proximity of furniture,
3D shape models
Key Results @Y2
BHAM, TUW collaboration
How would a robot query the web?

Find a cannoli

What is this?

textual query

perceptual query

Rather than mining the Web on the fly, we might build clean Web-proxies, large enough to cover for most (any?) query

...but isn’t this ImageNet?????
ImageNet as Web proxi
What a CNN is taught mugs look like
What mugs look like to Betty
First Results

1. An AlexNet trained on ImageNet does not generally perform too well on robot data

2. AlexNet(ImageNet) outputs WordNet classes as labels, so a WUP similarity is appropriate

3. WUP score average 0.51 → practically meaningless for most of objects (although microwave, laptops, keyboards have WUPs >0.75)

What are we missing? Can we do better?
(Some) Open Issues

ImageNet is a db created in 2012 of 1K categories — **static, closed world assumption**

1. What if we could continuously update all its classes?
2. What if we could augment its classes (1k, 3K, ..10K)?
3. What if we could exploit prior knowledge on the task and build on demand task specific databases?

**Can we automatically generate larger/task specific databases?**

Proof of concept: generate a copy of ImageNet by downloading images using the class names as queries
What a CNN is taught mugs look like
Results with AlexNet(Web-ImageNet)
(Some) Open Issues

ImageNet is a db created in 2012 of 1K categories – static, closed world assumption

1. What if we could continuously update all its classes?
2. What if we could augment its classes (1k, 3K, ..10K)?
3. What if we could exploit prior knowledge on the task and build on demand task specific databases?

Can we automatically generate larger/task specific databases?

Issues to deal with:
1. How to minimize noisily annotated images when harvesting the Web
2. What is the effect of noise on a given CNN architecture
Results with AlexNet(Web-ImageNet-Cleaned)

ROME, BHAM, INRIA collaboration
More details and more results at the poster session tomorrow!

For the whole story (data, demos, papers) please visit our project web-site:

https://project.inria.fr/aloof/
Thanks