

# Robots learning about objects from the Web

June 21, 2017

Barbara Caputo

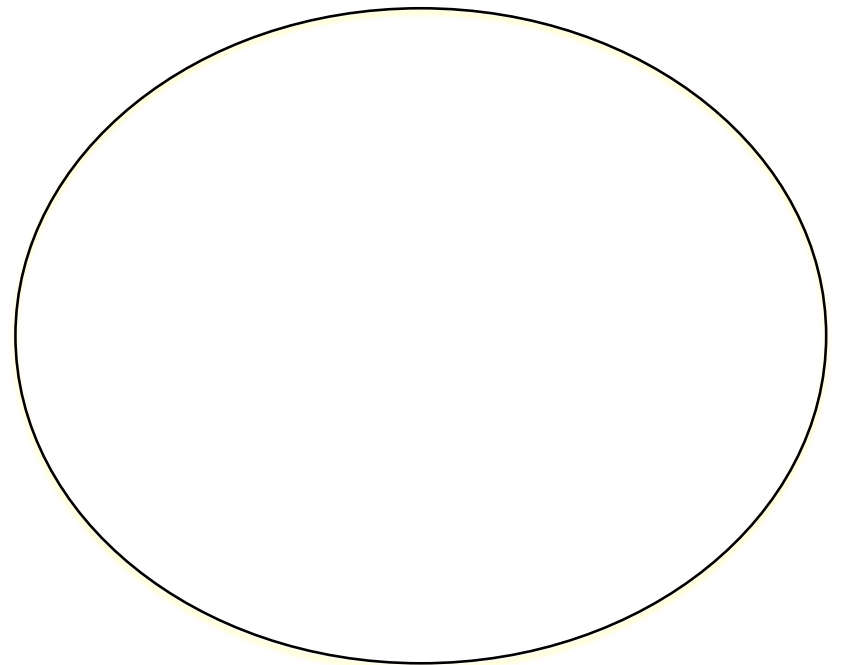
Visual And Multimodal Applied Learning (VANDAL) Laboratory

Department of Computer, Control and Management Engineering



**SAPIENZA**  
UNIVERSITÀ DI ROMA

**Suppose you have a robot at home...**



# Suppose you have a robot at home...

What is this on the table??



## Option A: asking a human

What is this on the table??  
Let me ask Barbara...



# Option A: asking a human

- Robots learn from own experience guided by humans
  - Human-robot interaction [Kjellstrom et al 2011, Billard&Grollman 2013]
  - Learning by demonstration [Argall et al 2009]



## Option B: learning by itself

What is this on the table??  
Let me try to figure it out  
...



## Option B: learning by itself

- Robots learn from own experience alone
  - Developmental robotics [Asada et al 2009, Tsai & Kujpers 2013, 2014..]



# Option C: learning from the Web

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....***





# Option C: learning from the Web



Let's see  
on the Web..



http://www.marialiberati.com/2009/03/07/holy-cannoli/

maria Liberati.c...

## Holy Cannoli !!

March 7th 2009 by Maria Liberati



Holy Cannoli!

Editor: Kate Hollinger

Copyright, 2009, Maria Liberati

The Basic Art of Italian Cooking by Maria Liberati tm

<http://www.marialiberati.com/blog2>

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 20<sup>th</sup> 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.

## Option C: learning from the Web



Let's see  
on the Web..




http://www.marialiberati.com/2009/03/07/holy-cannoli/

Maria Liberati.c...

## Holy Cannoli !!

March 7th 2009 by Maria Liberati



Holy Cannoli!

Editor: Kate Hollinger

Copyright, 2009, Maria Liberati

The Basic Art of Italian Cooking by Maria Liberati tm

http://www.marialiberati.com/blog2

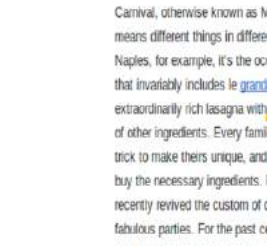
A cannoli, which is Sicilian for "little tube", is an ever-popular pastry de found sprinkled throughout cafes and restaurants in Italy. We thank the region of Italy for these rich desserts that are found today throughout t States as well, thanks to the immigrants who brought the recipe over early 20<sup>th</sup> century. Cannoli originally can early as the first century A.D., served as during the season of Carnevale, a festi before Lent. It is said that cannoli was m pre-Easter celebration similar in some w

An original cannoli (the singular form of

http://www.browneyedbaker.com/2010/01/08/cannoli/

## Cannoli

January 8, 2010 | 49 Comments | Email | Print



...been enjoying cannoli for a long all this holiday season. Sure you I'm talking honest-to-goodness southern's name. I had been hearing but never had the opportunity to try ch them for made and enjoyed. I of the family got together to help seems as much as possible so that photo of the cannoli sitting on the that I took home with me.

... pastry that is made by rolling out form and then deep frying. Once either nuts or dried fruit are added (nuts or chocolate chips, but we

Carnival, otherwise known as Martedì Grasso (Mardi Gras), means different things in different parts of the Peninsula. In Naples, for example, it's the occasion for a lavish banquet that invariably includes le *grande lasagne di Carnevale*, an extraordinarily rich lasagne with ricotta, meatballs, and a host of other ingredients. Every family had a special ingredient or trick to make theirs unique, and many would go into hock to buy the necessary ingredients. In Venice, instead, they have recently revived the custom of donning masks, and going to fabulous parties. For the past century or so they've been holding Carnival parades with satirical floats in the town of Viareggio, on the Tuscan coast. And almost everywhere there are *cenci*, *frittelle*, and other sweets of one kind or another to liven the merrymaking. Among the most glorious are Cannoli. Palermo's fried waters filled with an airy mixture of ricotta, candied fruit, chocolate, and other ingredients, which are so tasty that they have spread not just throughout the island, but to wherever Sicilians have settled, and are now greeted with joy year round.

In the past, when they were "Carnival's Crown," according to Giuseppe Pitre (the doctor and ethnographer who founded Palermo's museum of folk culture in 1906), people gave cannoli to all their friends – by the dozen. One can only imagine the people in the *pastry* shops churning them out and the sweet sugary smell of the *ricotta* filling wafting over the city! They are older, however. Pitre also quotes an anonymous 17th century poet, who said:

## A photograph of five cannoli pastries arranged in a yellow box. The cannoli are filled with a light-colored cream and topped with red cherries and a dusting of powdered sugar. The box is divided into compartments by white paper liners.





# Option C: learning from the Web




Let's see  
on the Web..



Ok, now I  
know what to  
do with it

http://www.marialiberati.com/2009/03/07/holy-cannoli/

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




Holy Cannoli!

Editor: Kate Hollinger  
Copyright, 2009, Maria Liberati  
The Basic Art of Italian Cooking by Maria Liberati tm  
http://www.marialiberati.com/blog2

A cannoli, which is Sicilian for "little tube", is an ever-popular pastry de- found sprinkled throughout cafes and restaurants in Italy. We thank the region of Italy for these rich desserts that are found today throughout the States as well, thanks to the immigrants who brought the recipes over. In the early 20th century, Cannoli originally came from Sicily. It was first found early as the first century A.D., served as during the season of Carnevale, a festi- before Lent. It is said that cannoli was m- pre-Easter celebration similar in some w-

http://www.browneyedbaker.com/2010/01/08/cannoli/

**Cannoli**  
January 8, 2010 | 10:49 Comments | [D] [B] [P]




**About.com Italian Food**

Italian Food Quick/Easy Recipes Le Basi Classic Recipes

See More About: [cannoli](#) [sicilian pastries](#) [carnival recipes](#) [ricotta cream](#)

**Cannoli**

Carnival, otherwise known as Martedì Grasso (Mardigras), means different things in different parts of the Peninsula. In Naples, for example, it's the occasion for a lavish banquet that invariably includes le *grande lasagne di Carnevale*, an extraordinarily rich lasagna with ricotta, meatballs, and a host of other ingredients. Every family had a special ingredient or trick to make theirs unique, and many would go into hock to buy the necessary ingredients. In Venice, instead, they have recently revived the custom of donning masks, and going to fabulous parties. For the past century or so they've been holding Carnival parades with satirical floats in the town of Viareggio, on the Tuscan coast. And almost everywhere there are *cenci*, *frittelle*, and other sweets of one kind or another to liven the merrymaking. Among the most glorious are Cannoli, Palermo's fried waters filled with an airy mixture of ricotta, candied fruit, chocolate, and other ingredients, which are so tasty that they have spread not just throughout the island, but to wherever Sicilians have settled, and are now greeted with joy year round.



In the past, when they were "Carnival's Crown," according to Giuseppe Pitre (the doctor and ethnographer who founded Palermo's museum of folk culture in 1909), people gave cannoli to all their friends -- by the dozen. One can only imagine the people in the pastry shops churning them out and the sweet sugary smell of the ricotta filling waiting over the city! They are older, however. Pitre also quotes an anonymous 17th century poet, who said:

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

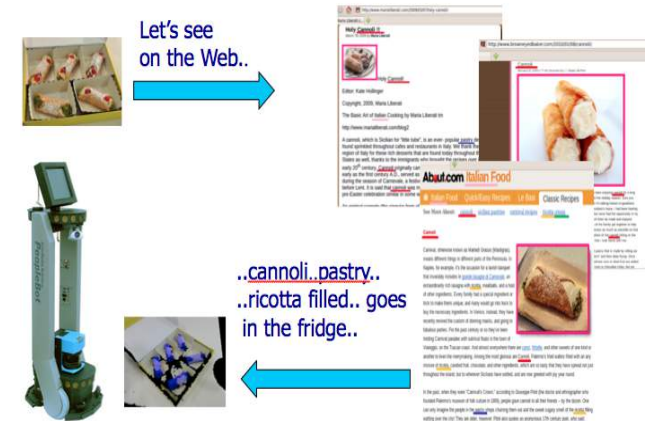


# Option C: learning from the Web

To make this work, we have several challenges, such as:

1. How to extract from the Web information useful for a robot, from a robot query?
2. How to make such information usable by robots?
3. How to update continuously the perceptual and semantic Web-created knowledge bases?
4. ....

***Focus on perceptual knowledge (images, depth); Visual learning tool of choice: deep learning***



**How to find usable perceptual knowledge  
on the Web?**

# 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????**



IM  GENET

# **(Some) Open Issues**

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

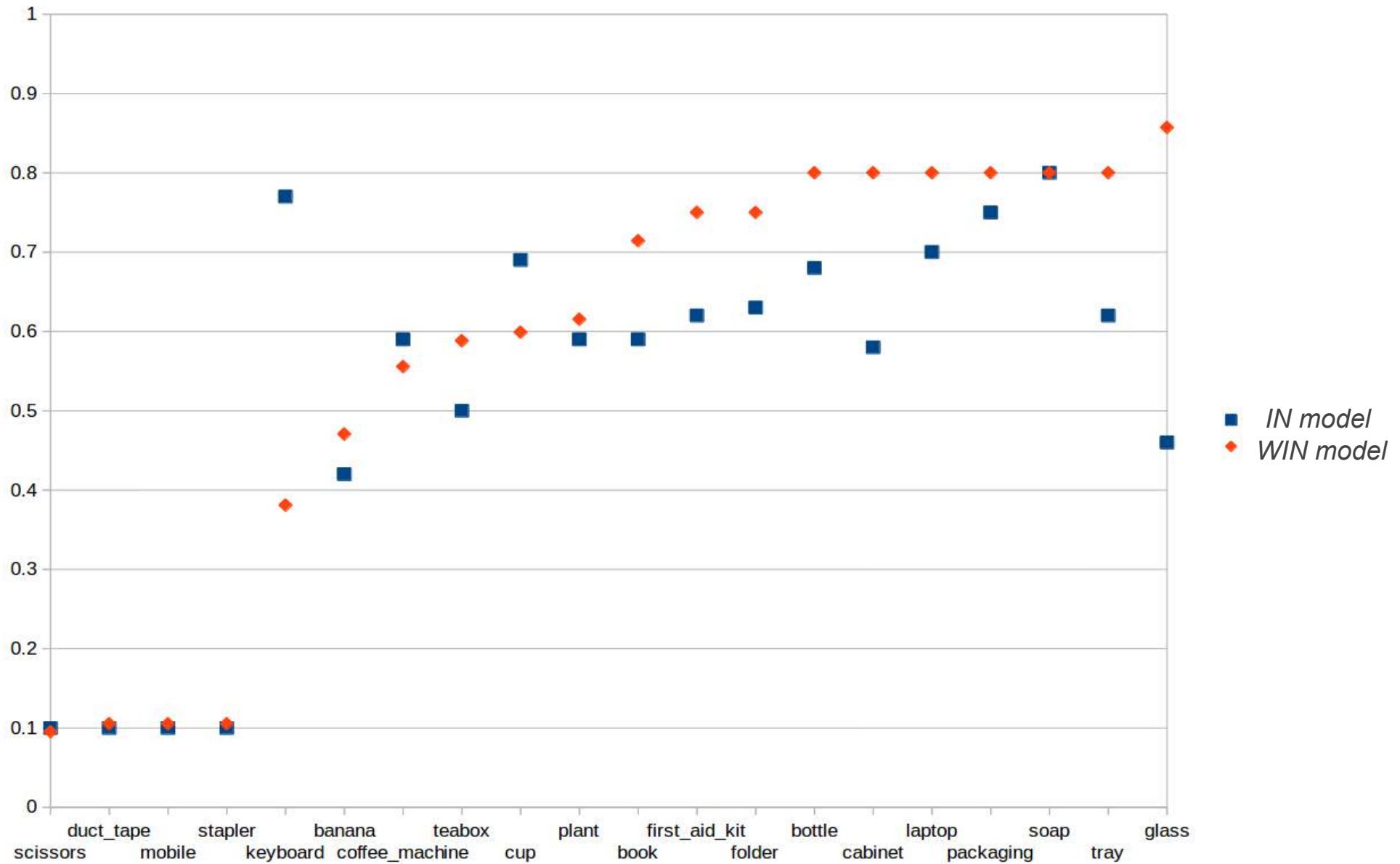
1. What if we could continuously update all its classes?
2. What if we could augment its classes (1k, 3K, ..10K)?
3. What if we could exploit prior knowledge on the task and build on demand task specific databases?

**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**



# 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

# Harvesting the Web –SEO Keywords

- We take advantage of the Search Engine Optimization (SEO) practice of keyword research.
- SEO professionals research keywords, which they use to achieve better rankings in search engines.

## **Tool of choice: MyKeyWorder**

- It references language content from Princeton's Wordnet and Moby Thesaurus.
- It takes a query and returns lists of keywords associated with it, sorting them by relevance and frequency.

# Harvesting the Web –SEO Keywords

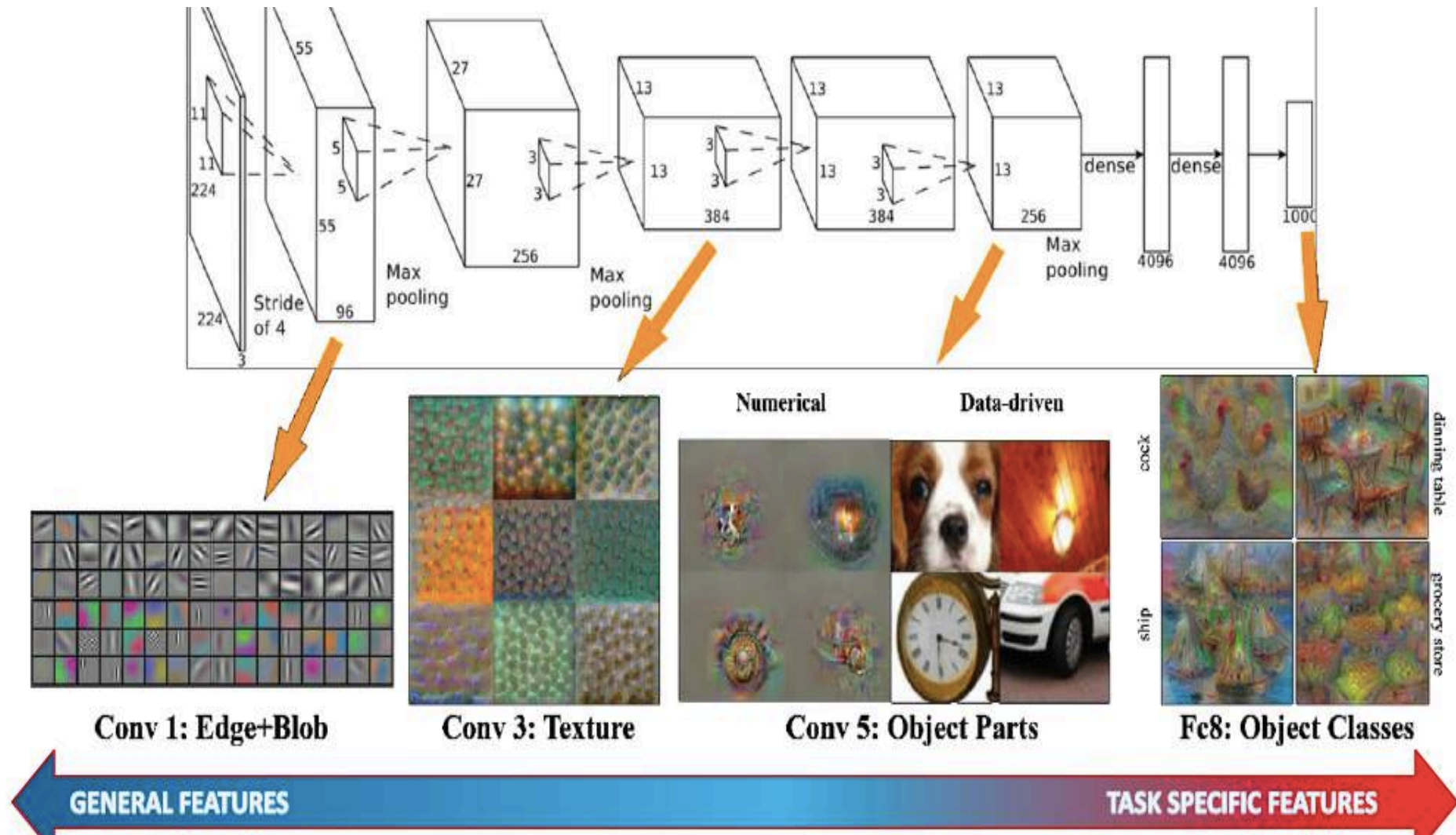


Icecream - WIN



Icecream - WINC

# How does noise affects CNNs?





# How does noise affects CNNs?

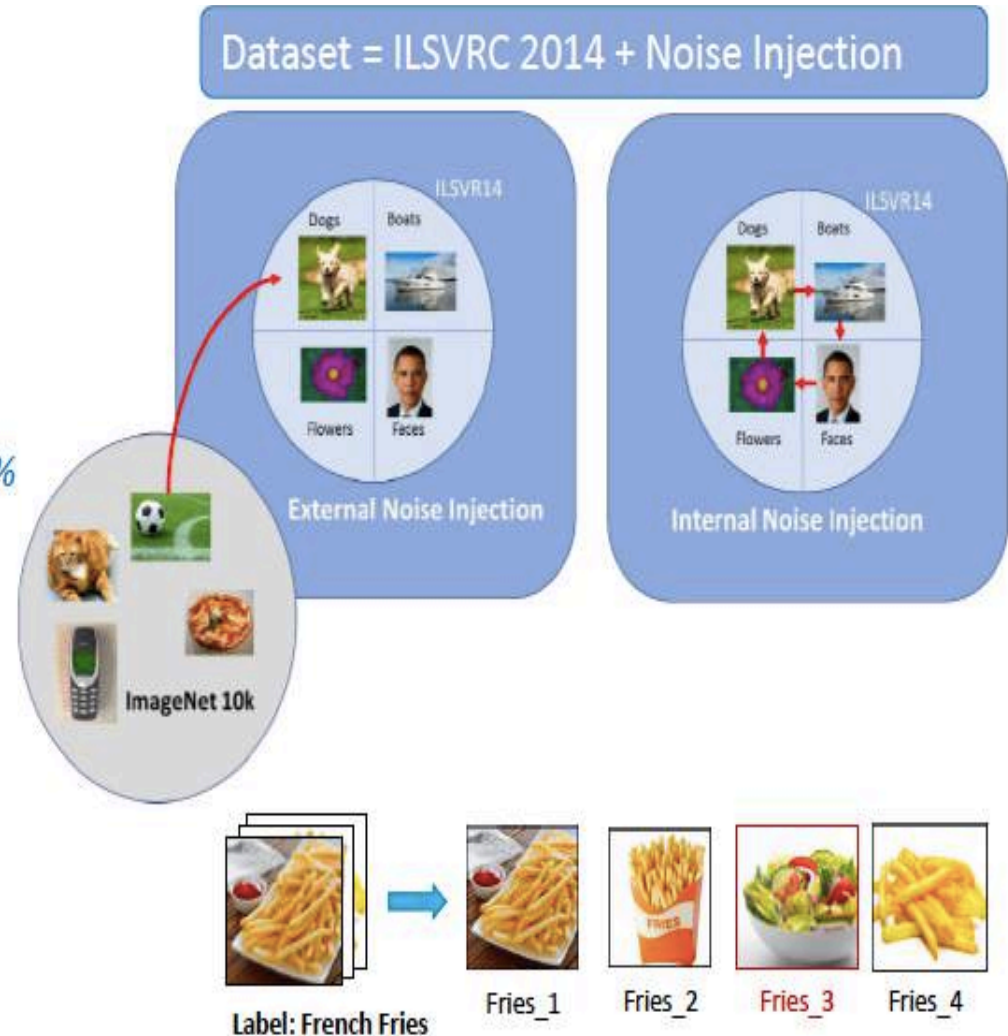
Original dataset	ILSVRC 2014 classification
Object Categories	1000
Number of images	1,281,167 (700-1300 images per cat. )

## DATASET

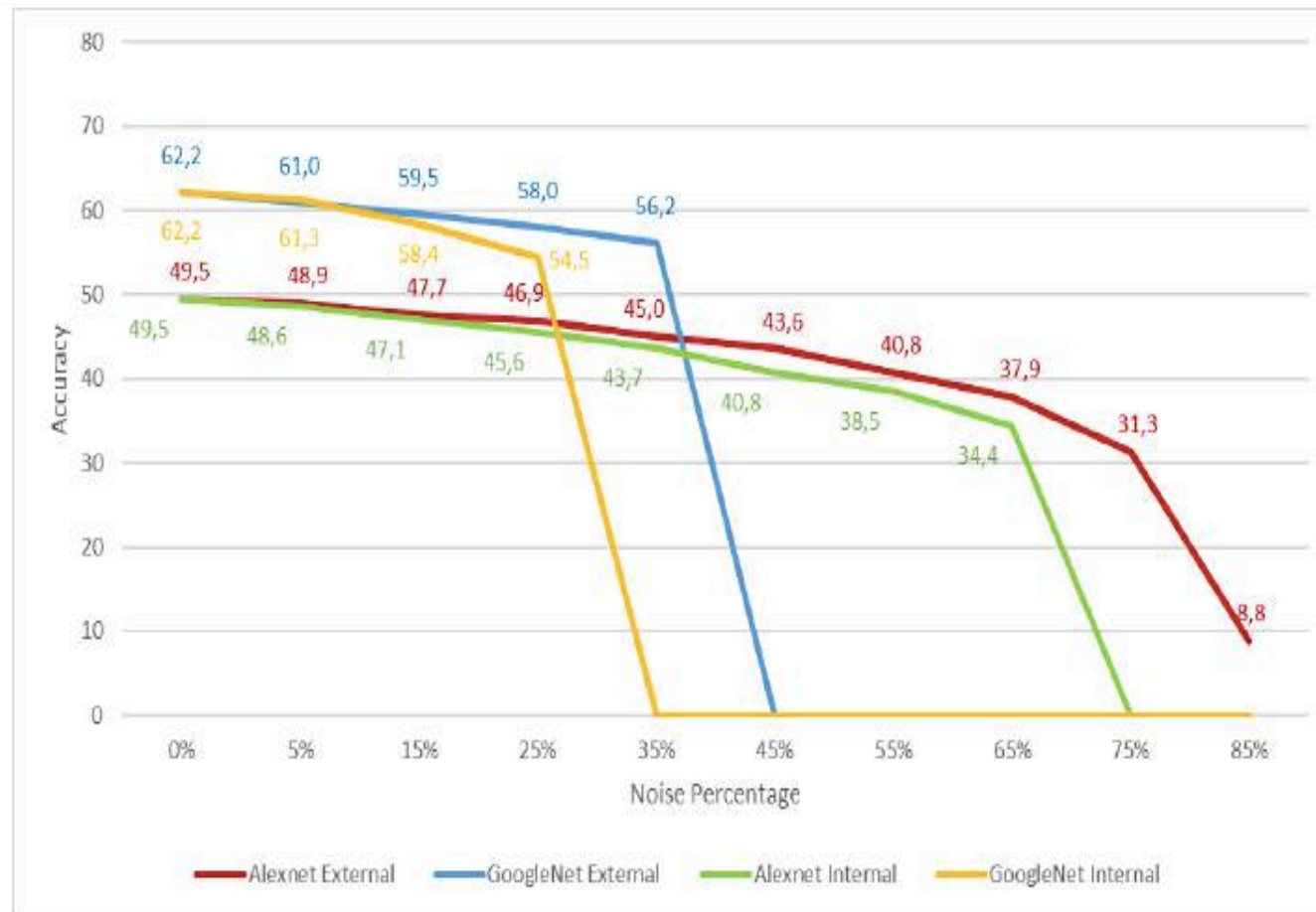
- 2 types of Noise Injection: *Internal and External*
- 9 percentages of Noise Injection: 5% 15% 25% ... 75% 85%

## TRAINING

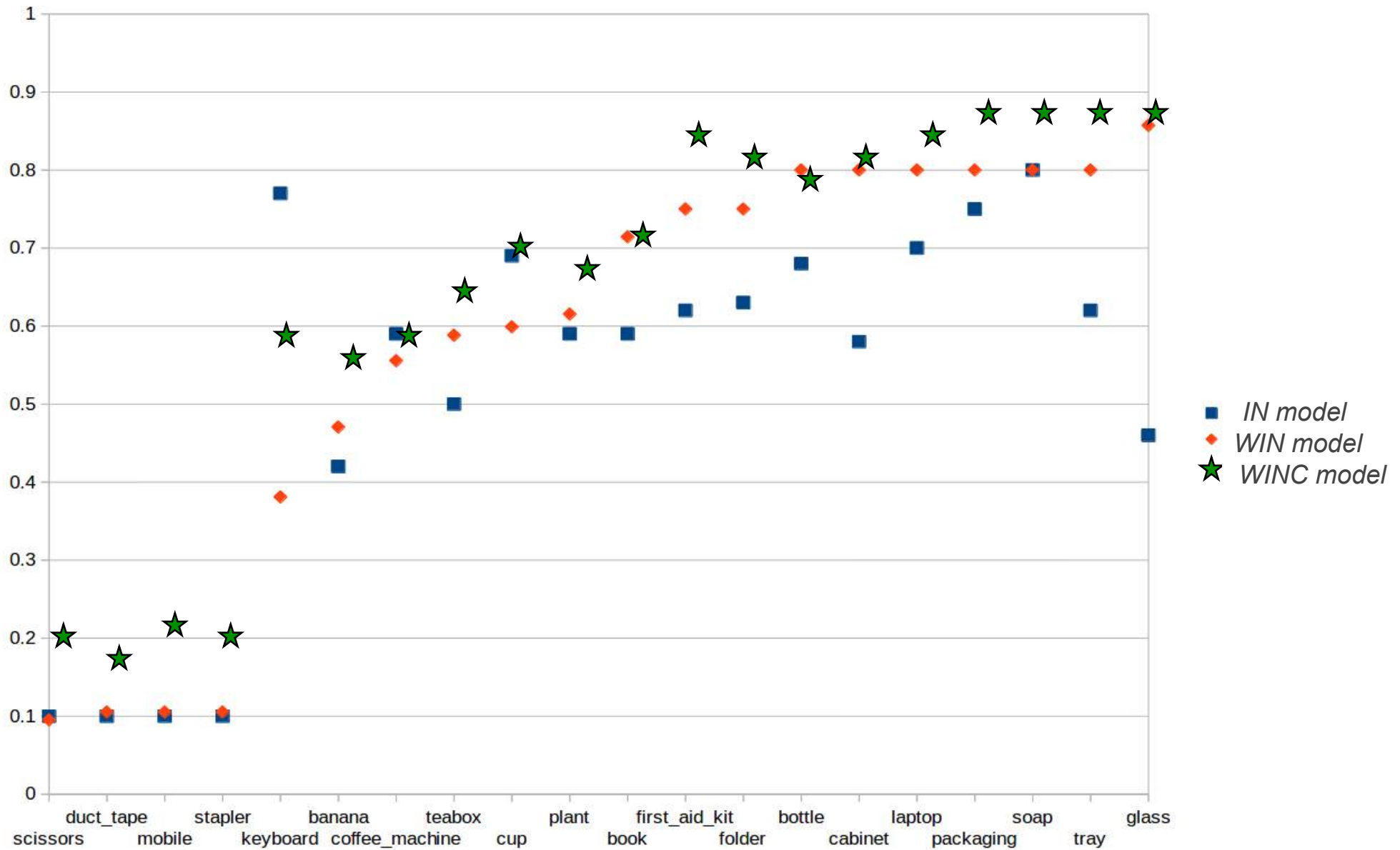
- 2 different architectures: *AlexNet* , *GoogleNet*
- Training details
  - 60 Epochs
  - Stochastic Gradient descent
  - Learning rate 0.01
  - Learning rate update: step
  - Momentum 0.9



# How does noise affects CNNs?



# Results with AlexNet(Web-ImageNet-Cleaned)





# Putting all the pieces together

1. We generate WINC (Web-ImageNet-Cleaned) and train a GoogLeNet on it
2. We use this CNN to make predictions on Bob and Betty
3. We further use semantic context as a heuristic to indicate our preferences from the CNN list of predictions
4. From this preferred subset we rank by the CNN's confidence and pick the highest

*[Massouh, Babiloni, Tommasi, Young, Hawes and Caputo, IROS 2017]*

*[Young, Basile, Cabrio, Caputo, Kunze and Hawes, ICRA 2017]*





# Merging visual and semantic cues

## Context

1. Coffee
2. Toaster
3. Oven
4. Plate
5. Pyrex





# Merging visual and semantic cues

## Context

1. Coffee
2. Toaster
3. Oven
4. Plate
5. Pyrex



## Vision

1. Crate
2. Safe
3. Microwave
4. Refrigerator
5. Fire Screen

# Merging visual and semantic cues

## Context

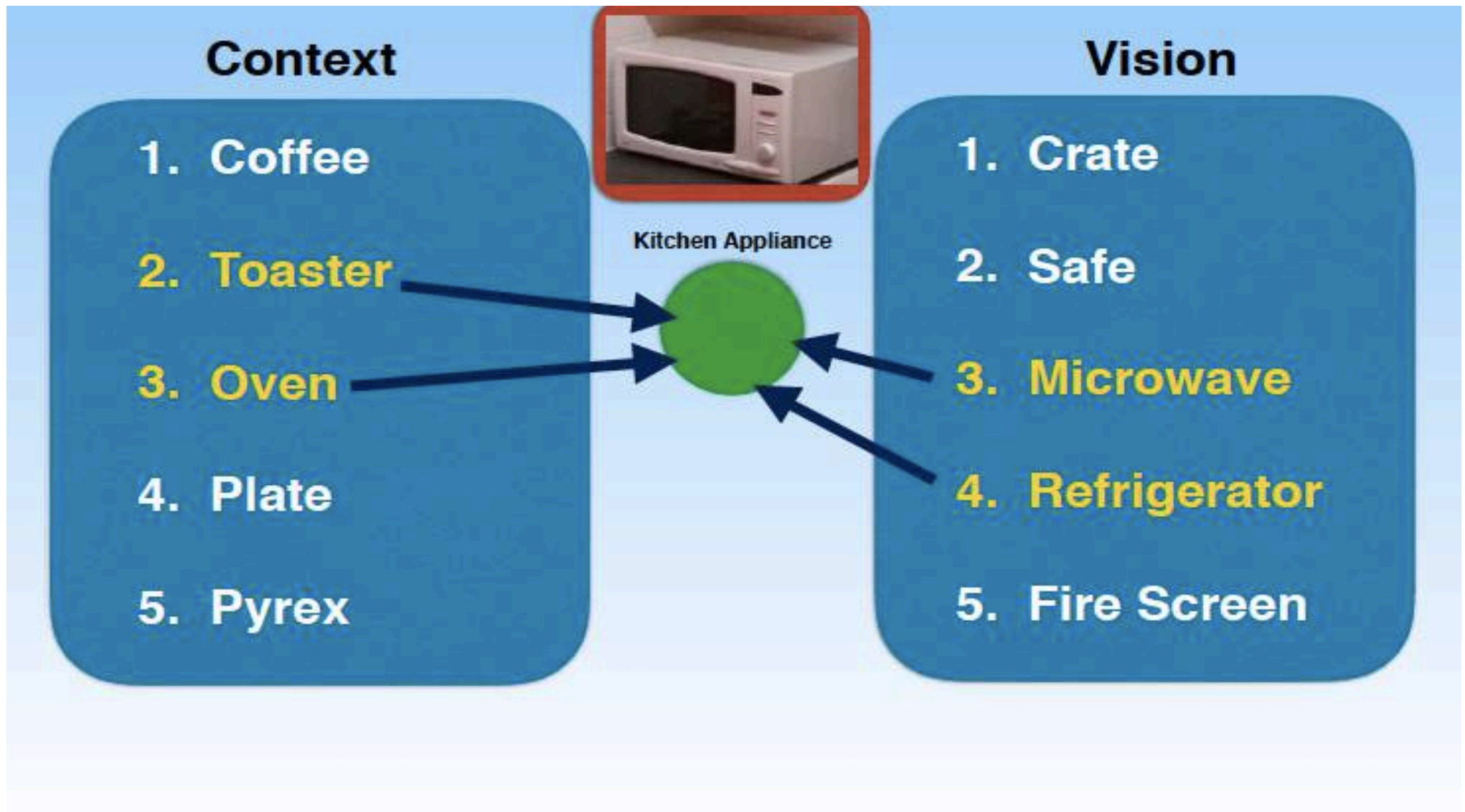
1. Coffee
2. Toaster
3. Oven
4. Plate
5. Pyrex



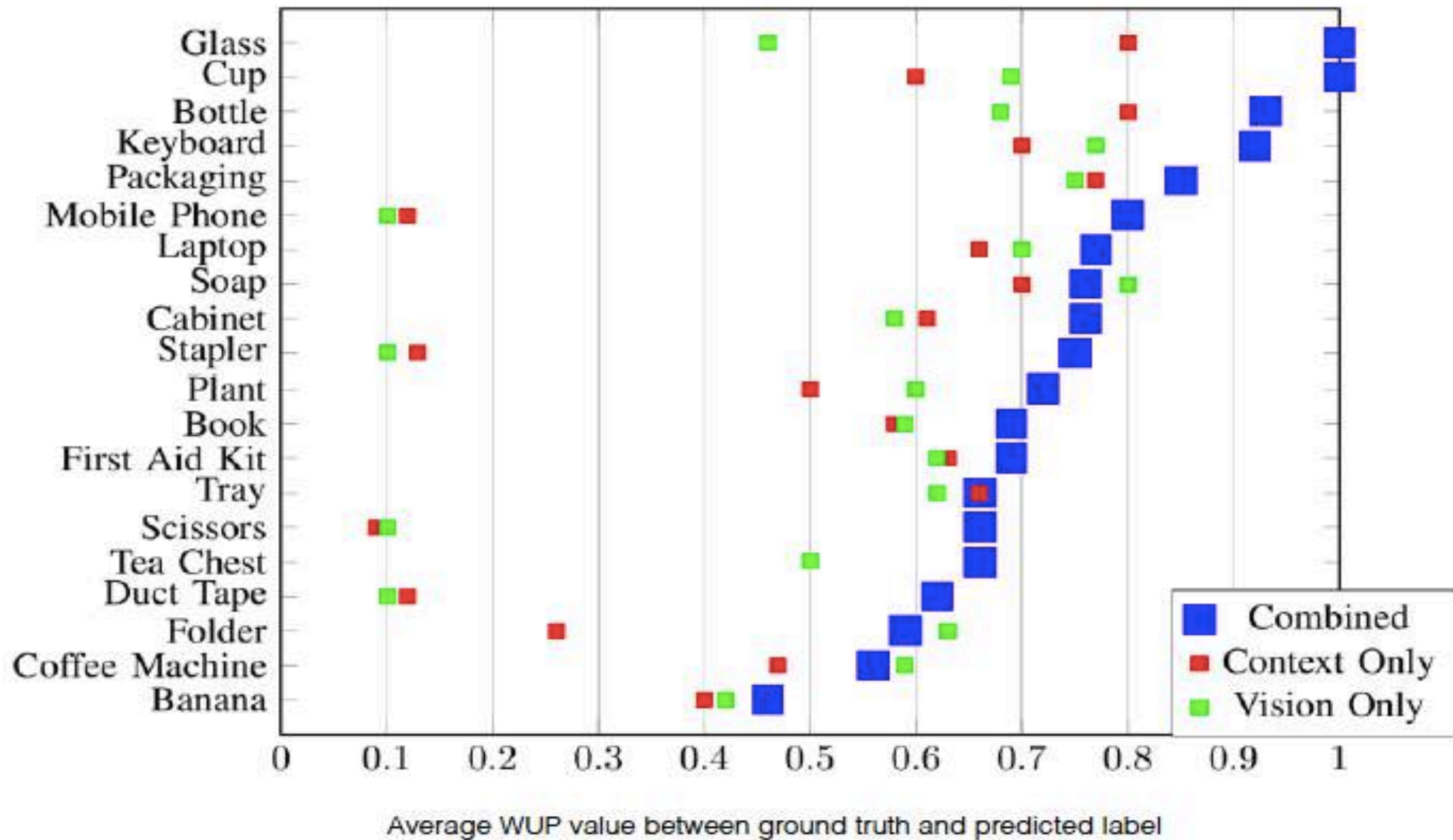
## Vision

1. Crate
2. Safe
3. Microwave
4. Refrigerator
5. Fire Screen

# Merging visual and semantic cues



# Merging visual and semantic cues





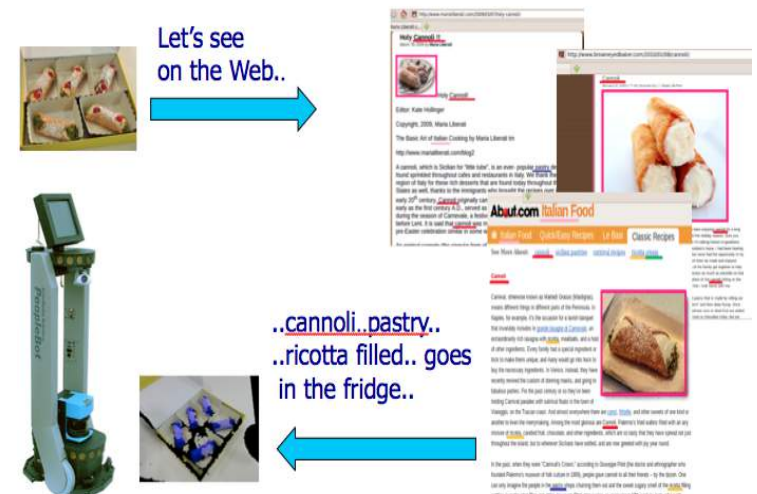
# Learning about objects from the Web

Q:

1. How to find usable perceptual knowledge from the Web?

A:

1. Build automatically Web-based perceptual knowledge bases, generic and task specific



**Moving forward**

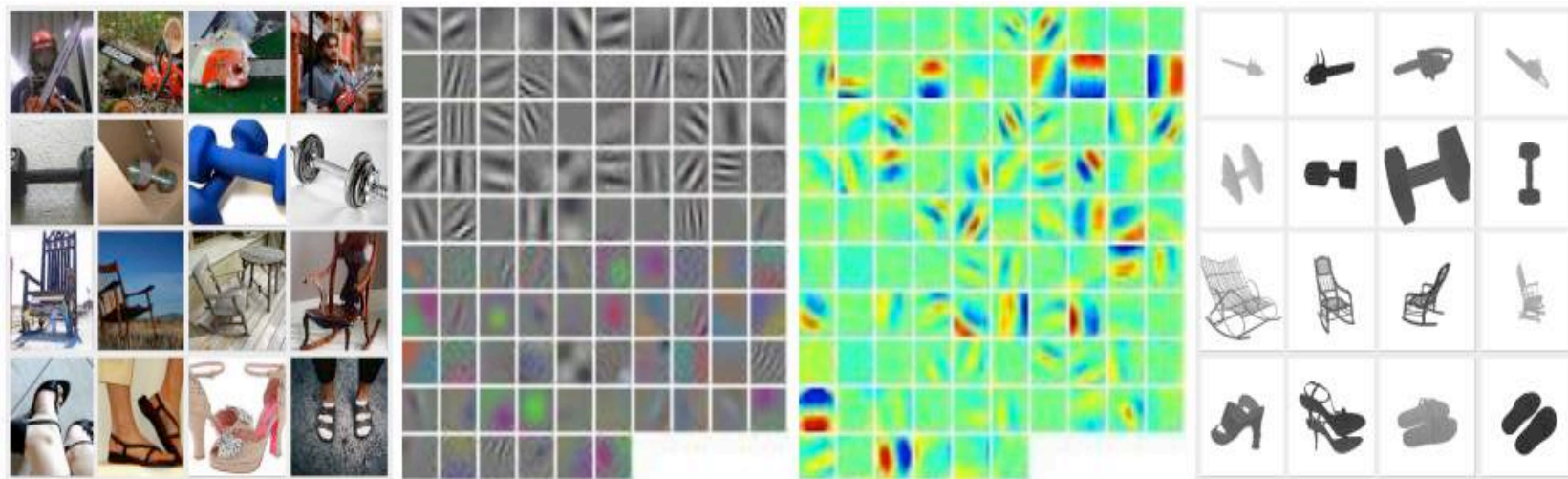
# Harvesting 3D percepts from the Web

*From 3D CAD models to synthetic depth → the VANDAL database*

- ❏ 4.1 millions depth views
- ❏ 9.383 different CAD models (counting morphs)
- ❏ 319 object classes (ILSVRC 2014 subset)



*Sample renderings with data augmentation*



*Imagenet samples*

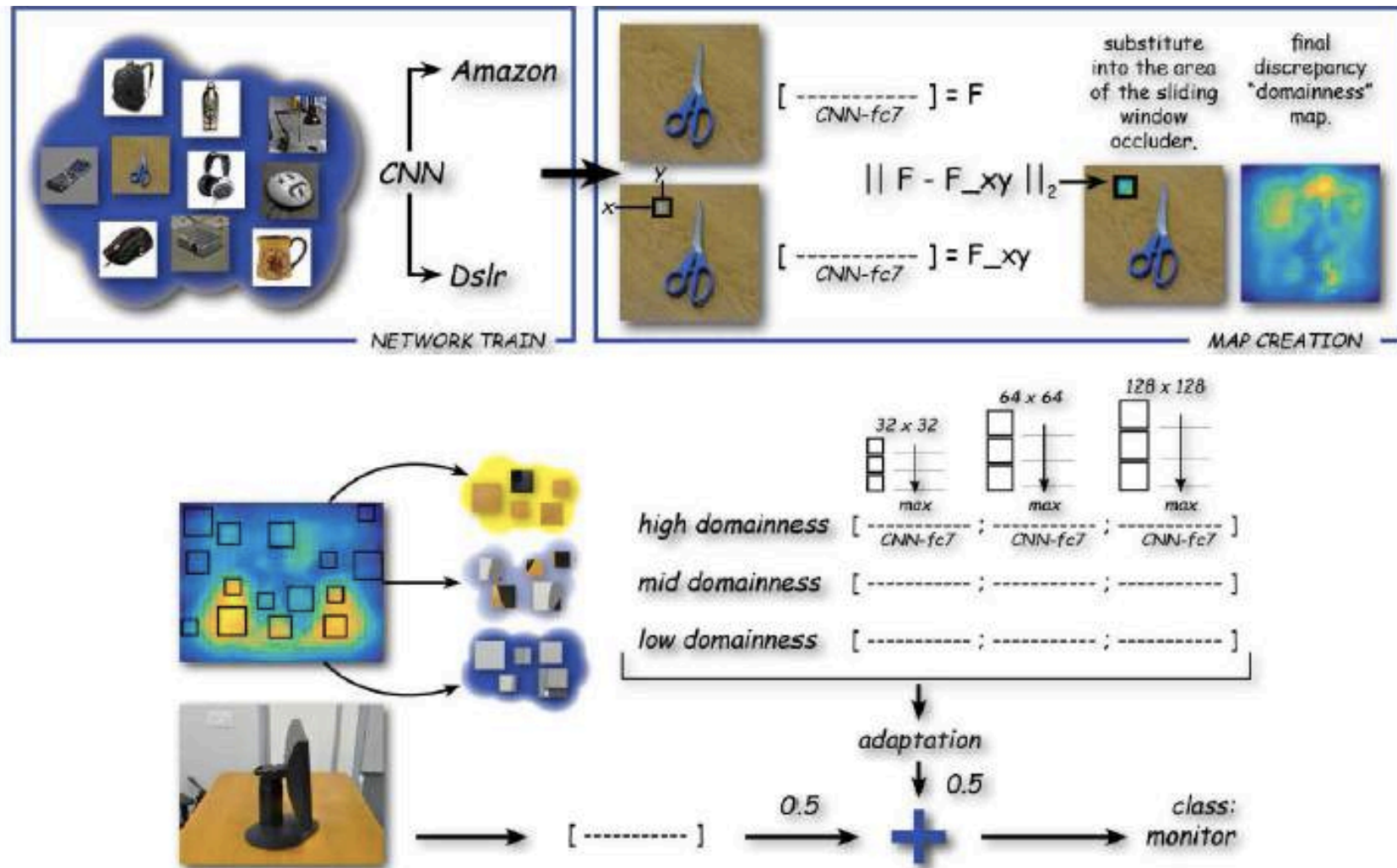
*Imagenet filters*

*DepthNet samples*

*Vandal samples*

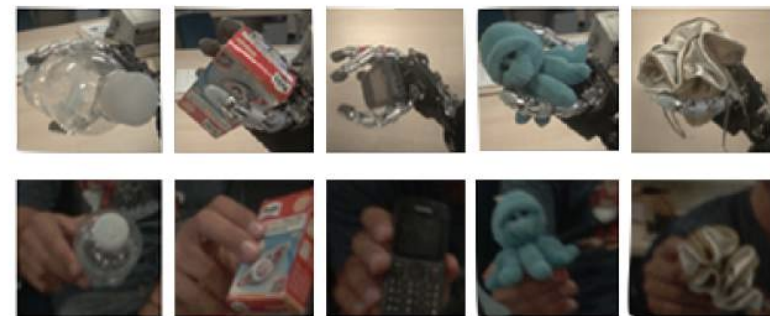
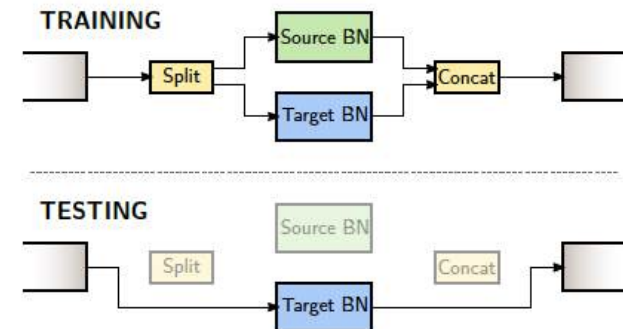
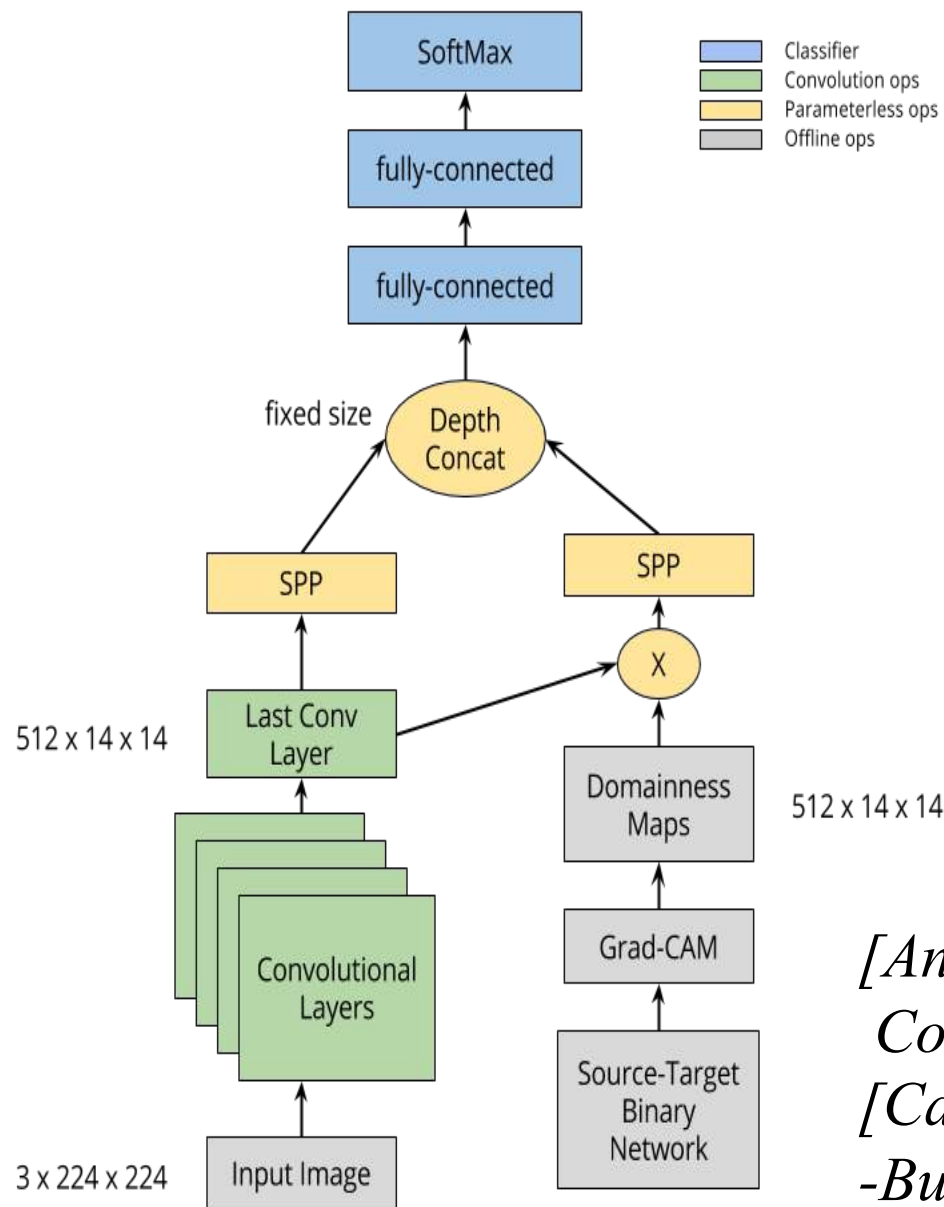
*[Carlucci, Russo and Caputo, ICRA 2017]*

# Can we localize the domain shift?



[Tommasi, Russo, Lanzi and Caputo, ECCV-W 2016]

# Can we embed domainness maps in a CNN?



*[Angeletti, Tommasi and Caputo, ComingSoon 2018]  
 [Carlucci, Porzi, Ricci, Caputo and Rota -Bulo', ICCV17 sub]*

**Learning about objects from the Web:  
where does manipulation fit in?**



# An object is far more than its name...

What is this...?

Identified honey jar. I must check the lid is closed.

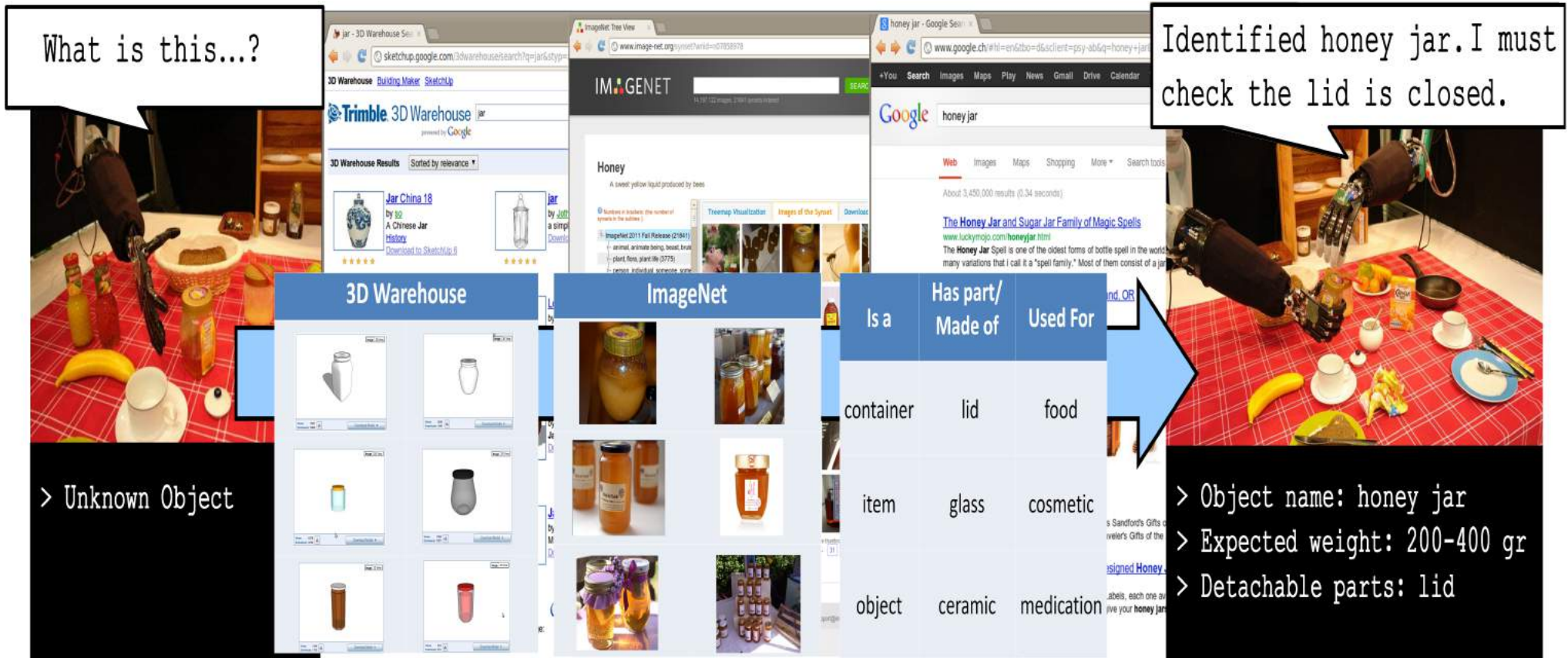
> Unknown Object

3D Warehouse

ImageNet

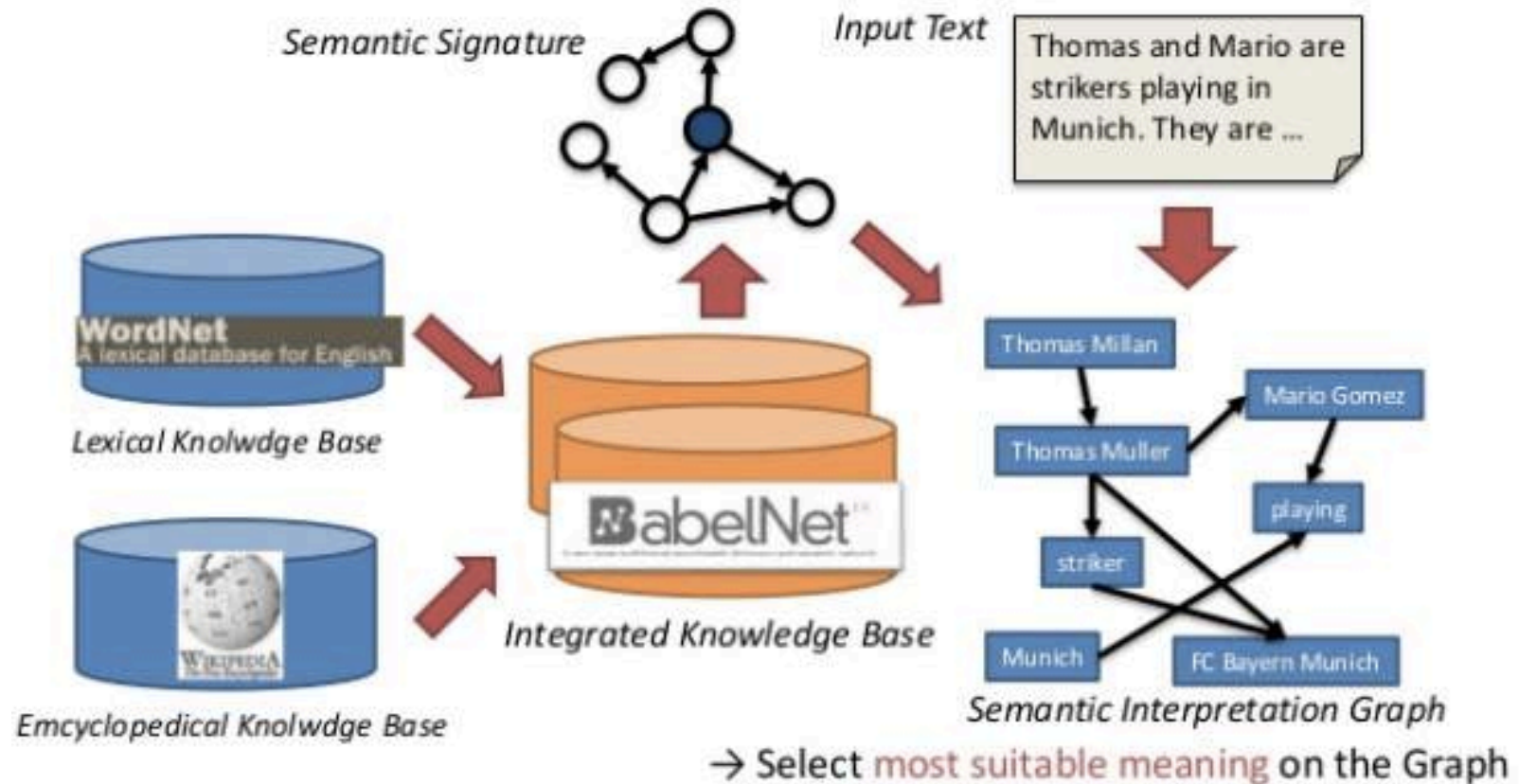
Is a	Has part/ Made of	Used For
container	lid	food
item	glass	cosmetic
object	ceramic	medication

> Object name: honey jar  
> Expected weight: 200-400 gr  
> Detachable parts: lid



*Leveraging over 3D and 2D perceptual data freely available from the Web, linking it to structured semantic knowledge about objects*

# An object is far more than its name...



*Leveraging over 3D and 2D perceptual data freely available from the Web, linking it to structured semantic knowledge about objects*



# A final though



# A final thought



- **Embodied cognition** has been the leading principle in artificial cognitive system research over the last 10 years
- ***BUT*** we humans learn also from externalized knowledge –like books, the Web, virtual reality....

# Thanks for your attention

