

Video Event Analysis using General Pattern Theory

Fillipe Souza, *University of South Florida, Tampa FL USA*

Sudeep Sarkar, *University of South Florida, Tampa FL USA*

Anuj Srivastava, *Florida State University, Tallahassee FL USA*

Perception = Seeing + Interpretation



What we **perceive** is not the true signal, but an **reconstruction** of what it should be.

We impose structure and pattern on the messy ambiguous signal.

“The analysis of the patterns generated by the world in any modality, with all their naturally occurring **complexity and ambiguity**, with the goal of **reconstructing** the processes, objects and events that produced them” is pattern theory.

- Mumford (1994)

OXFORD MATHEMATICAL MONOGRAPHS

General Pattern Theory

A Mathematical Study of Regular Structures

ULF GRENANDER



OXFORD SCIENCE PUBLICATIONS

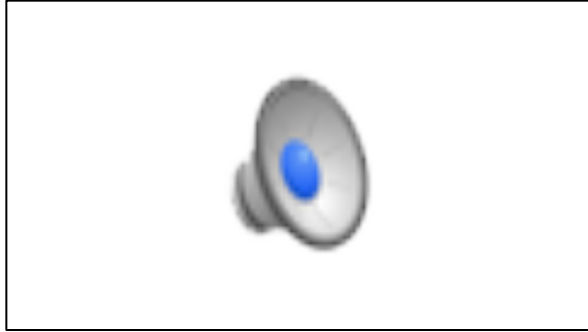


Objectives

- Create algebraic representations with probabilistic superstructures that model patterns in nature.
- Analyze regular structures using algebra, topology, probability, and statistical inference.
- Implement the structures for specific applications using algorithms and code.

Transformations in Video

Interpretation



- Noise/blur
 - object/action classification error
- Domain Warping
 - speed of actions, viewpoint
- Multi-scale superposition
 - (objects, actions) – (activity)
 - (recipe)
- Interruptions
 - many irrelevant objects, and simultaneous events.

Semantic Analysis of Video Events



spoon



stir



pickup



mix



bowl



spatula



stir



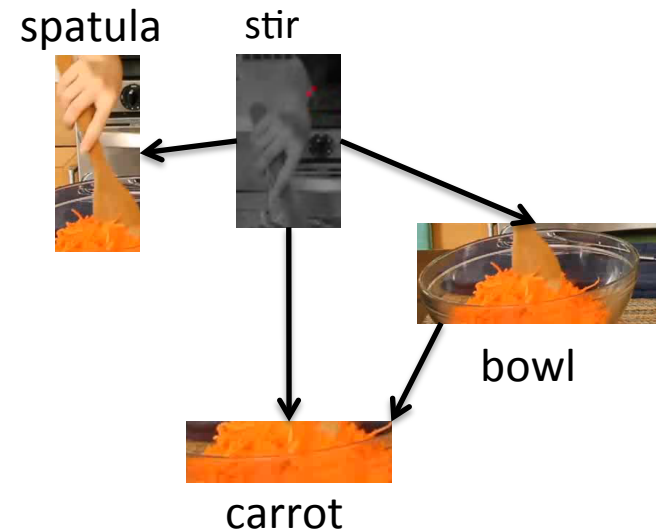
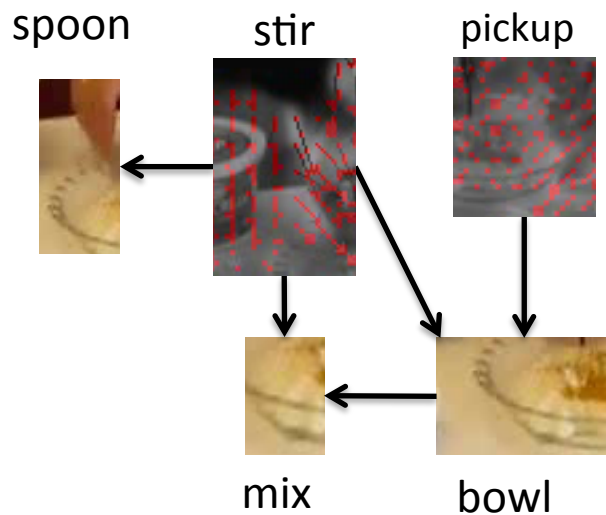
bowl



carrot

primitive concepts

Semantic Analysis of Video Events



Pattern Theory Elements

Generators

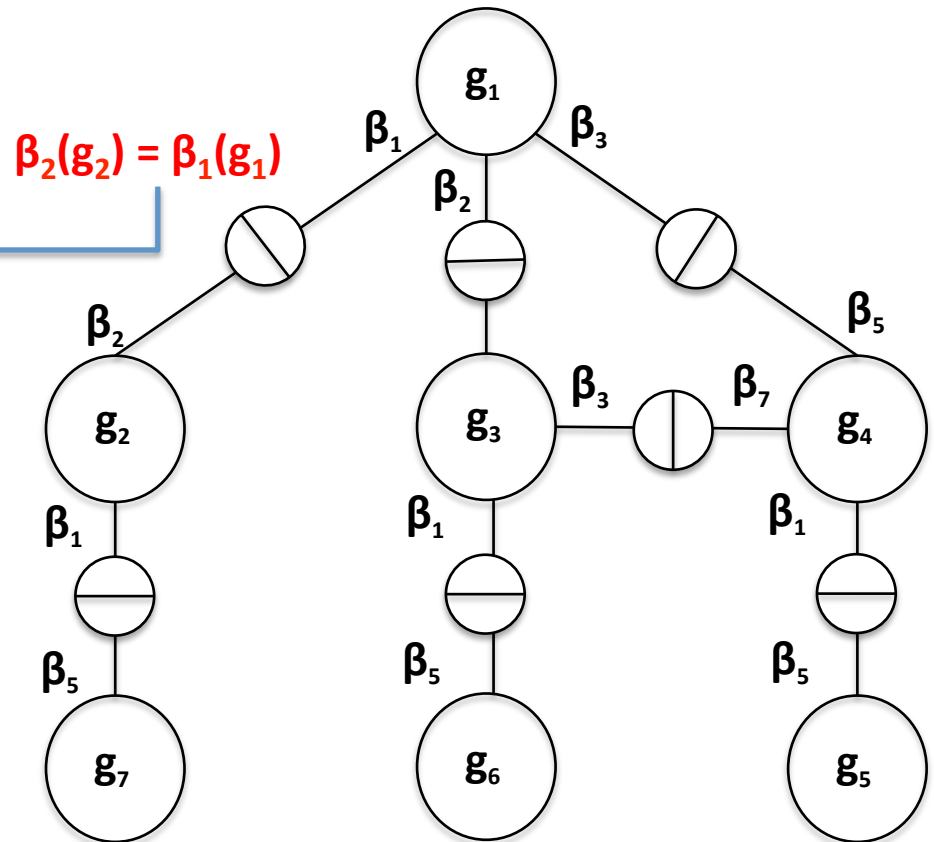
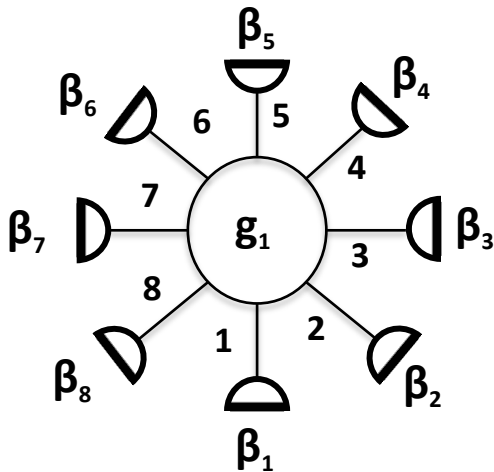
$$g_i \in G : i = 1, \dots, n$$

Bonds

$$\beta_j(g_i) \in B : j = 1, \dots, \omega(g_i)$$

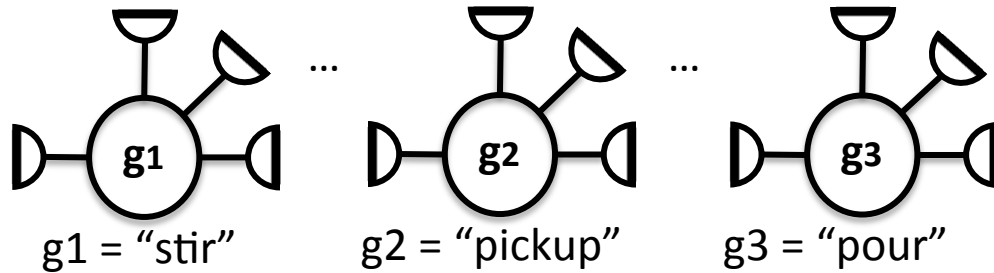
Configurations

$$c = \sigma(g_1, g_2, \dots, g_7) \in C$$

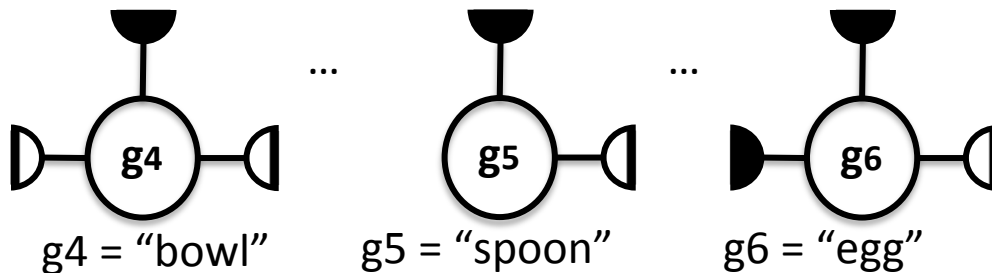


Generator Space for Representing Activities

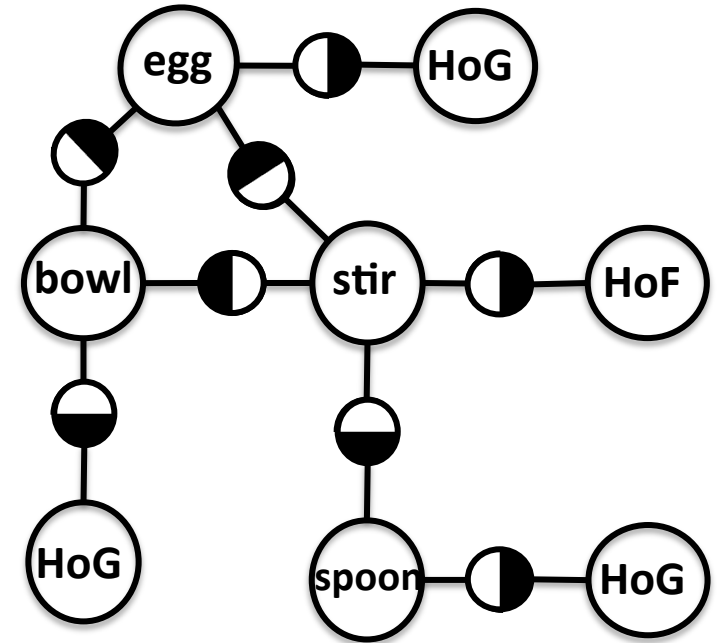
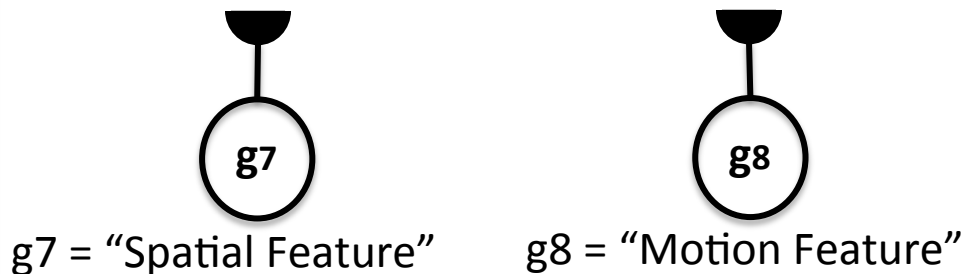
Level 3 = Actions



Level 2 = Objects



Level 1 = Features



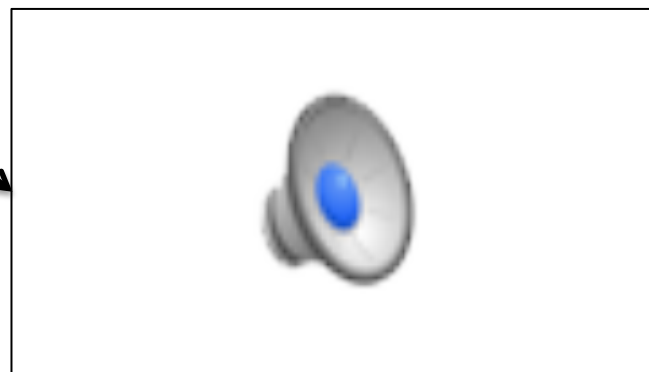
Complete interpretation

Example of Interpretation

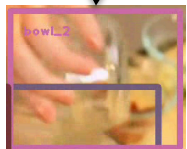


Example of Interpretation

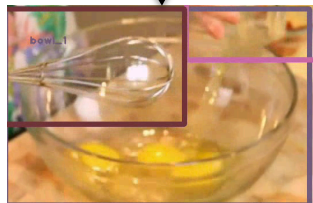
g8
HoF



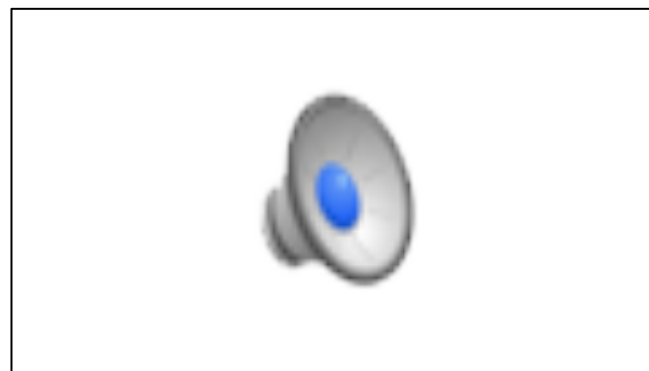
g7
HoG



g6
HoG

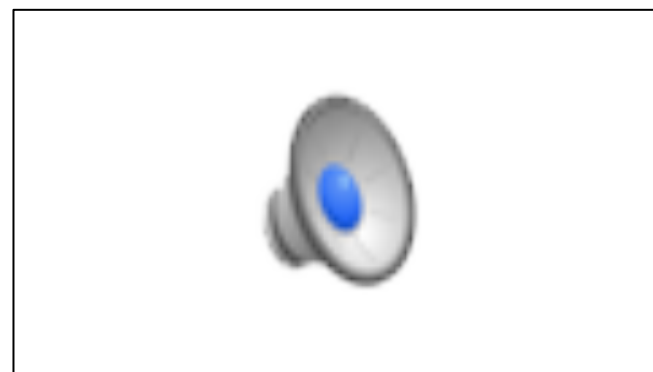
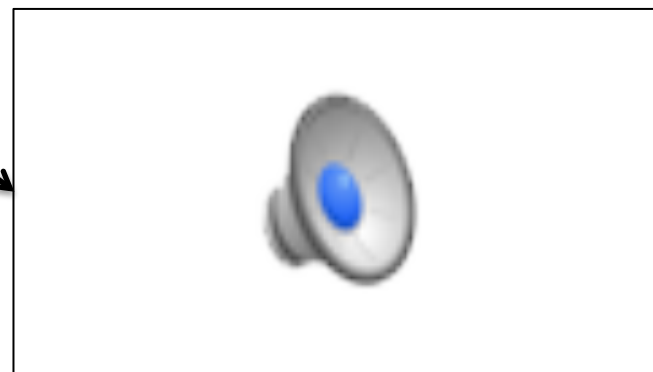
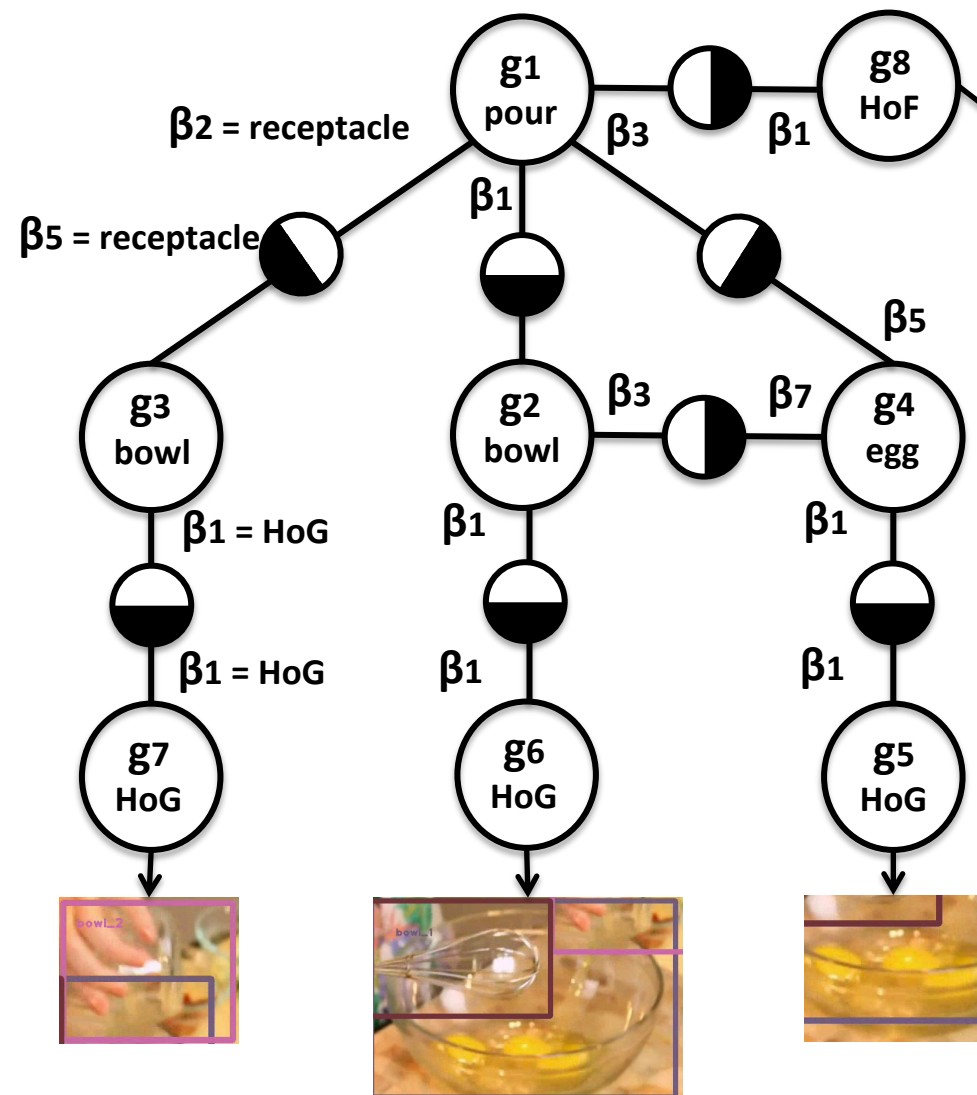


g5
HoG



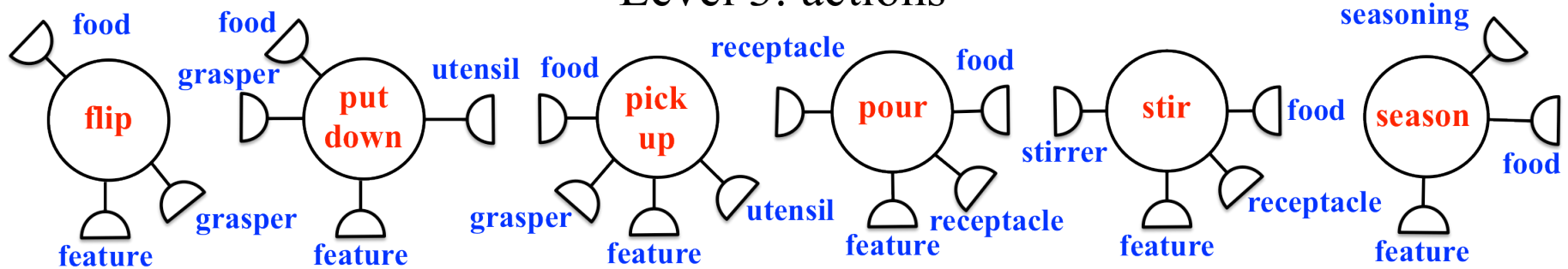
Example of Interpretation

receptacle = {bowl, cup, pan}

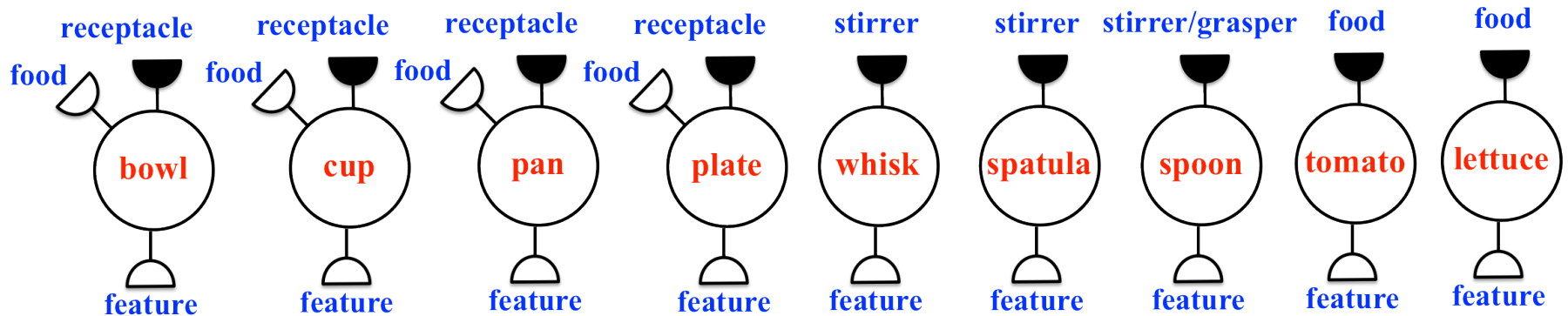


Example of Generator Space to Represent Cuisine Domain Events

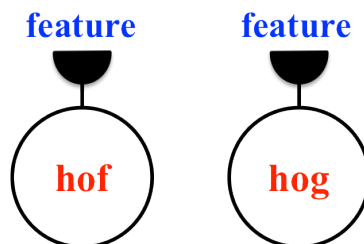
Level 3: actions



Level 2: objects



Level 1: features



Similarity Group

- Define a similarity group S as

$$s : G \rightarrow G \mid s \in S$$

Similarity Group

- Define a similarity group S as

$$s : G \rightarrow G \mid s \in S$$

- Induces a partition of the generator space into equivalence classes (disjoint groups G^α)

$$G = \bigcup_{\alpha \in A} G^\alpha$$

Similarity Group

- Define a similarity group S as

$$s : G \rightarrow G \mid s \in S$$

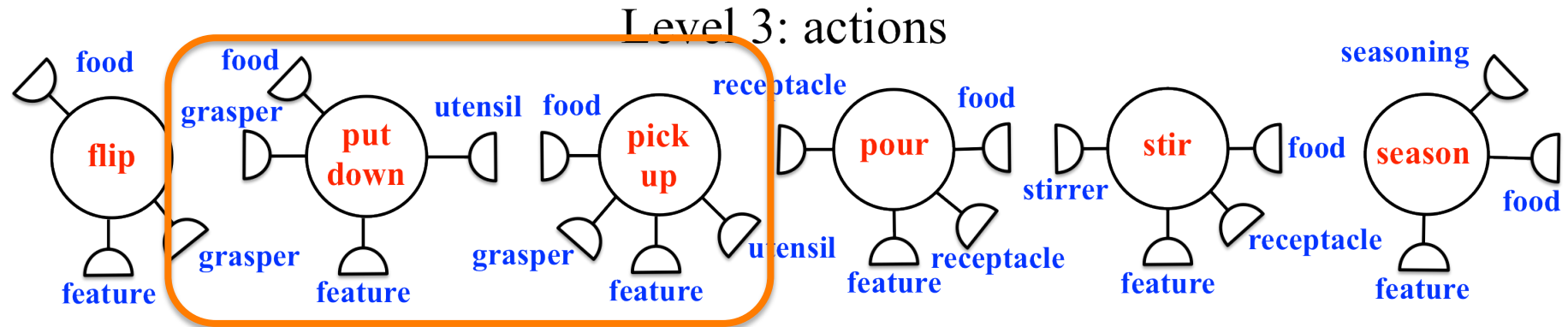
- Induces a partition of the generator space into equivalence classes (disjoint groups G^α)

$$G = \bigcup_{\alpha \in A} G^\alpha$$

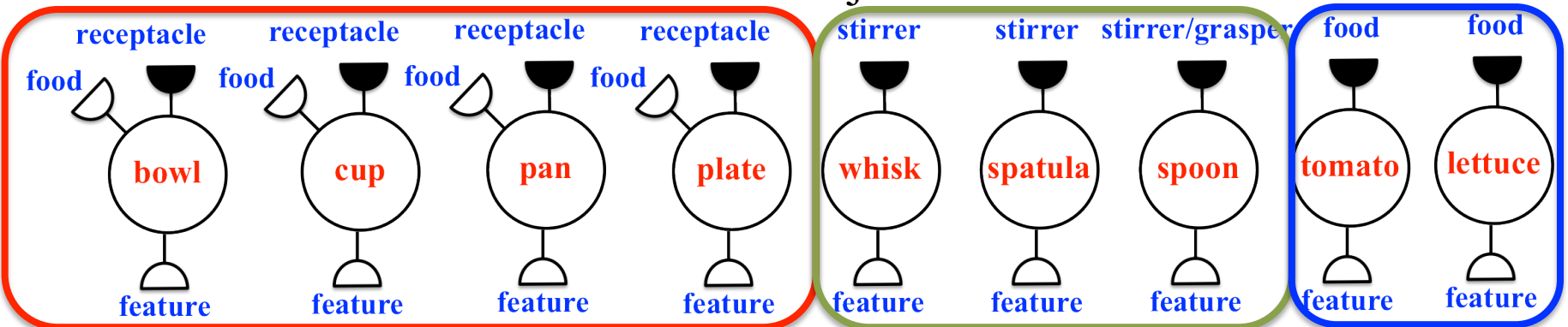
- Such that

$$sg = g' \mid g, g' \in G^\alpha$$

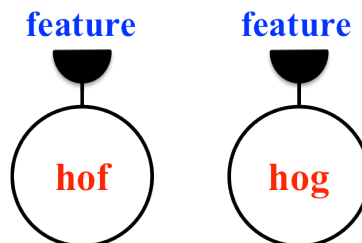
Example of Generator Space to Represent Cuisine Domain Events



Level 2: objects



Level 1: features



Bond Relation

(Local Regularity)

- Bond relation ρ specifies the rules of combination among generators, formally defined as

$$\rho : B \times B \rightarrow \{TRUE, FALSE\}$$

- Determines local regularity of a connected structure of generators

Bond Relation

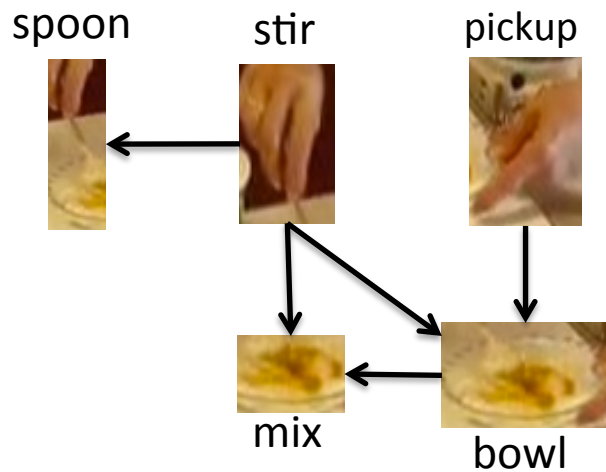
(Local Regularity)

- Bond relation ρ of 'INCLUSION' type

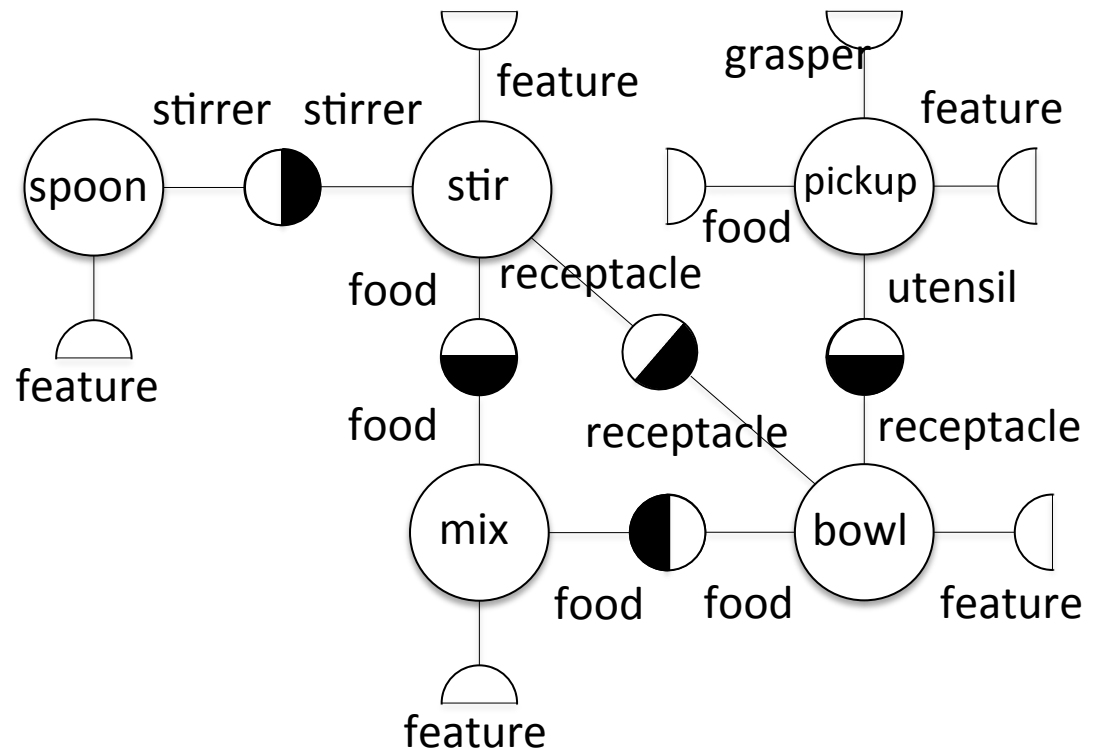
$$B = \{utensil, stirrer, grasper, receptacle, food, feature\}$$

		$\beta_{in}(g_{i'})$					
		utensil	stirrer	grasper	receptacle	food	feature
$\beta_{out}(g_i)$	utensil	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE
	stirrer	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
	grasper	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
	receptacle	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
	food	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
	feature	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE

Configurations: Local Regularity



$$c = \sigma(g_1, g_2, \dots, g_n)$$



This configuration is said to be locally regular since

$$\bigwedge_{\langle \beta', \beta'' \rangle \in c} \rho[\beta_{j'}(g_{i'}), \beta_{j''}(g_{i''})] = TRUE$$

Configuration: Global Regularity

- The topology of configurations are determined by the type of the underlying connector graph

$$\sigma \in \Sigma$$

- Σ =POSET (Partially Ordered Set)
- Links from Actions to Objects to Features
- Can partially order the generators

Regular Configuration

A regular configuration $c = \sigma(g_1, g_1, \dots, g_n)$ is both locally and globally regular, i.e.

$$\text{Local regularity} \rightarrow \bigwedge_{\langle \beta', \beta'' \rangle \in c} \rho[\beta_{j'}(g_{i'}), \beta_{j''}(g_{i''})] = \text{TRUE}$$

$$\text{Global regularity} \rightarrow \sigma \in \Sigma$$

Probability Measure on the Configuration Space

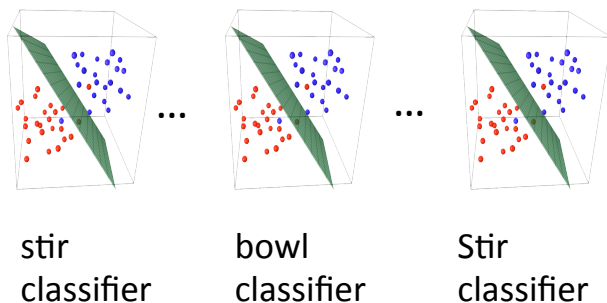
Bond Quantification

Object-Feature
Bonds

Action-Feature
Bonds

Object-Action
Bonds

Concept classifiers

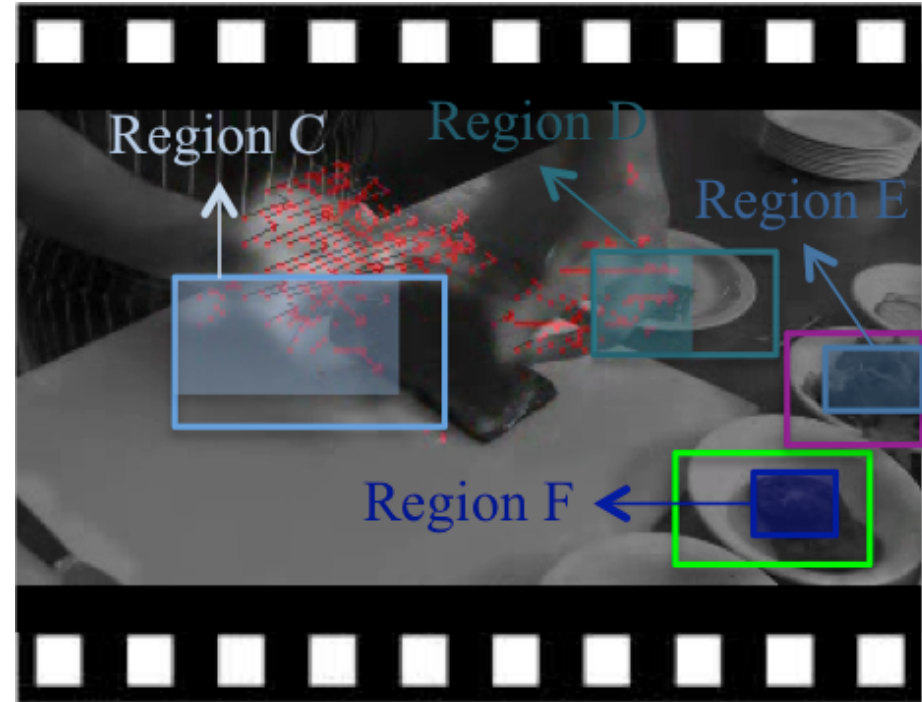
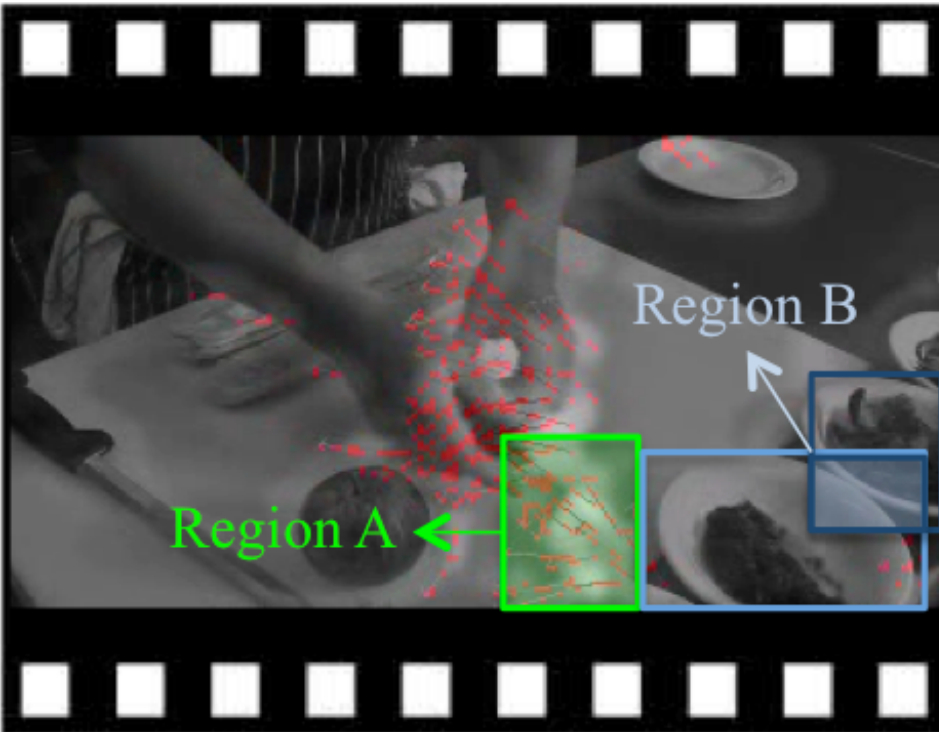


Concept co-occurrence tables

	Pickup	Putdown	Stir	Pour
Bowl	0.25	0.25	0.30	0.20
Spatula	0.15	0.15	0.65	0.05
Carrot	0.01	0.01	0.79	0.19
Plate	0.39	0.58	0.01	0.01

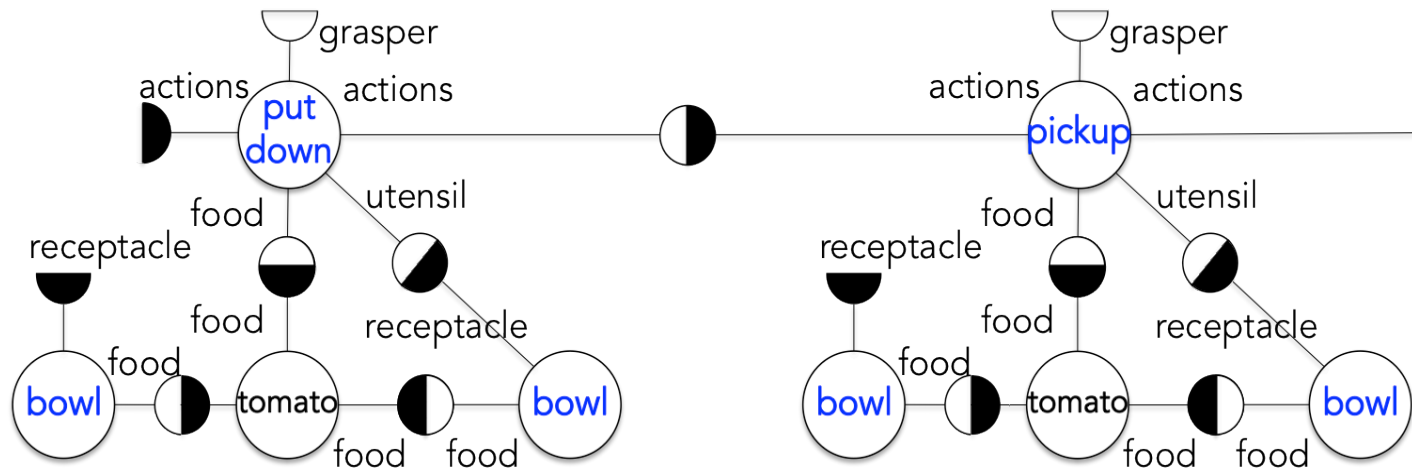
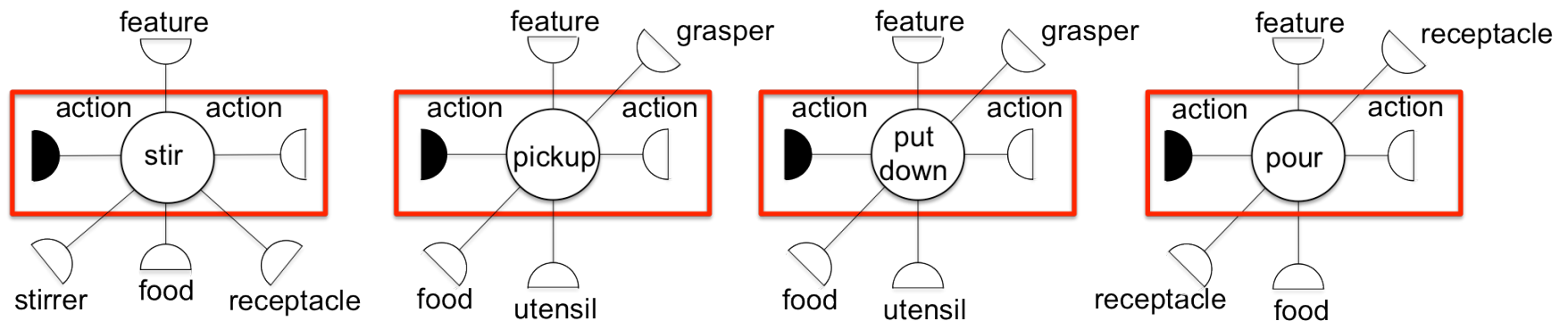
$$A(\beta_{j'}(g_{i'}), \beta_{j''}(g_{i''})) = \exp(\tanh(-w_b f(g_{i'}, g_{i''})))$$

Bond Type: Spatial Coherence



$$a'(\beta', \beta'') = \phi(g_{i'}, g_{i''}) + a(\beta', \beta'')$$

Bond Type: Temporal Coherence

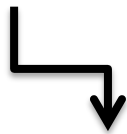


Probability Density Function p on $C(R)$

$$p(c) = \frac{1}{Z} \prod_{(k', k'') \in \sigma} A(\beta_{j'}(g_{i'}), \beta_{j''}(g_{i''})) \prod_i Q(g_i)$$

if the connector graph is allowed to vary, then

$$Z = \sum_{\sigma \in \Sigma} \sum_{c \in C(\sigma)} \prod_{(k', k'') \in \sigma} A(\beta_{j'}(g_{i'}), \beta_{j''}(g_{i''})) \prod_i Q(g_i)$$



consider all possible connectors

Energy of a configuration c

$$E(c) = \log p(c)$$



$$E(c) = \sum_{(k', k'') \in \sigma} a(\beta_{j'}(g_{i'}), \beta_{j''}(g_{i''})) + \sum_i q(g_i) - \log Z$$

Inference Algorithm

Energy Minimization Algorithm

$\sigma(b) \leftarrow \sigma(c)$ $\triangleright$ Let $\sigma(c)$ be the initial best configuration..
configuration..

for $k \leftarrow 1, k_{max}$ **do**

$t \sim U(0, 1)$

if $t < p$ **then**

$\sigma(c') \leftarrow \text{GLOBALPROPOSAL}(\sigma(c''))$

else

$\sigma(c') \leftarrow \text{LOCALPROPOSAL}(G, \sigma(c))$

end if

$T \leftarrow T_0 \times \alpha^k$

if $P(E(\sigma(c')), E(\sigma(c)), T)$ **then**

$\sigma(c) \leftarrow \sigma(c')$

if $E(\sigma(c')) < E(\sigma(b))$ **then**

$\sigma(b) \leftarrow \sigma(c')$

end if

end if

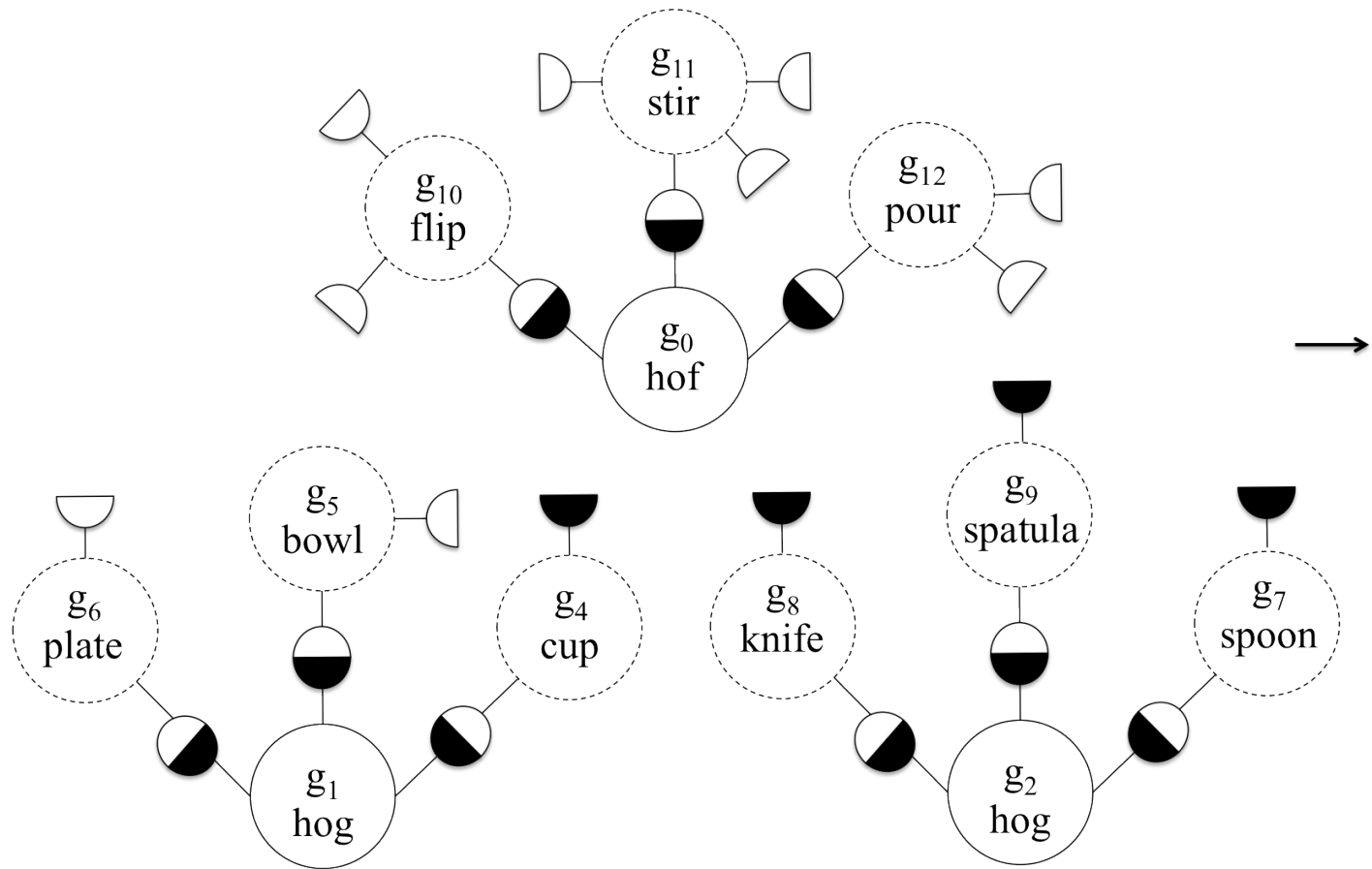
end for

end procedure

MCMC algorithm
with two moves

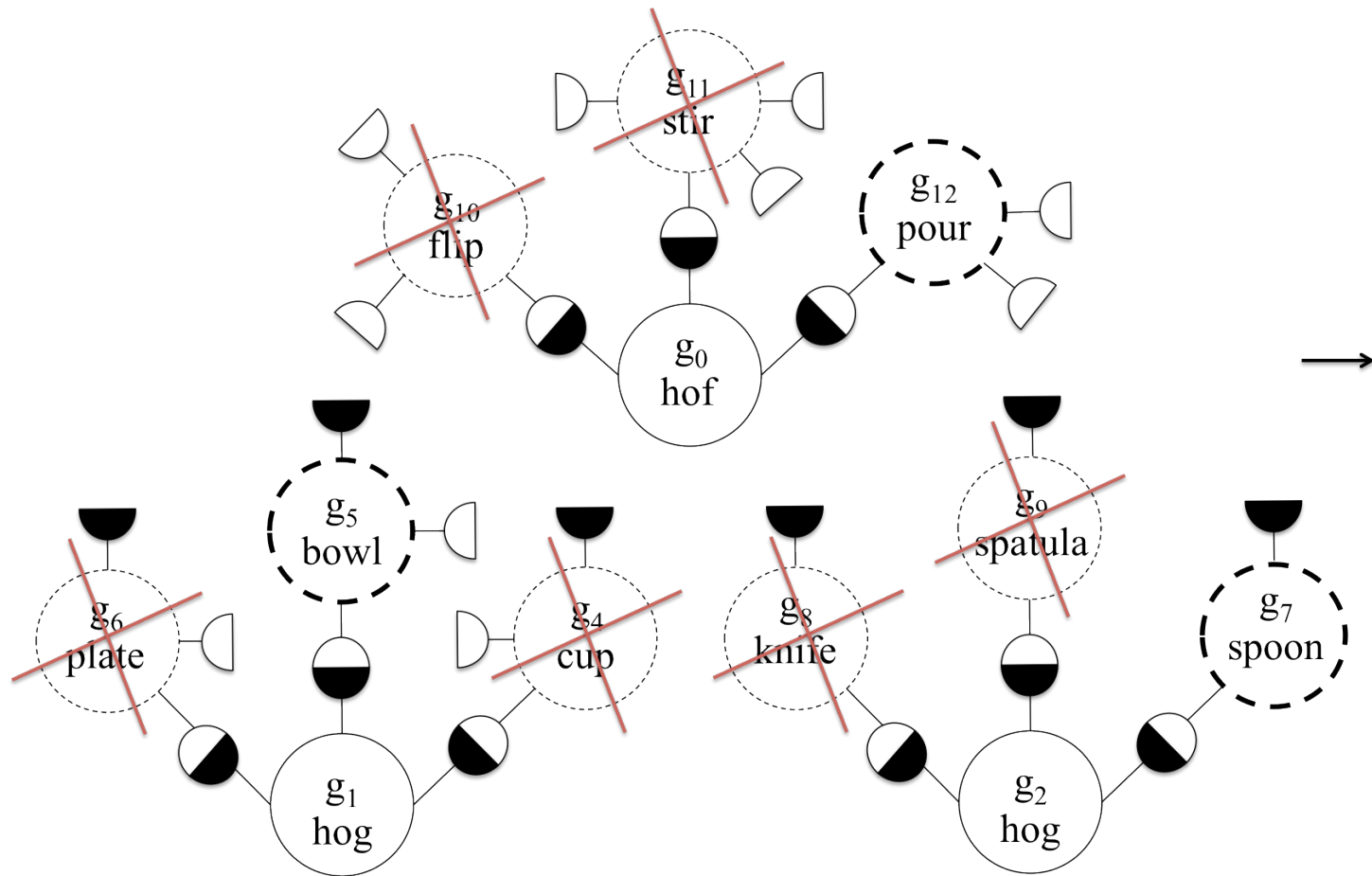
SA's
Temperature
Scheduling

Global Moves in Configuration Space



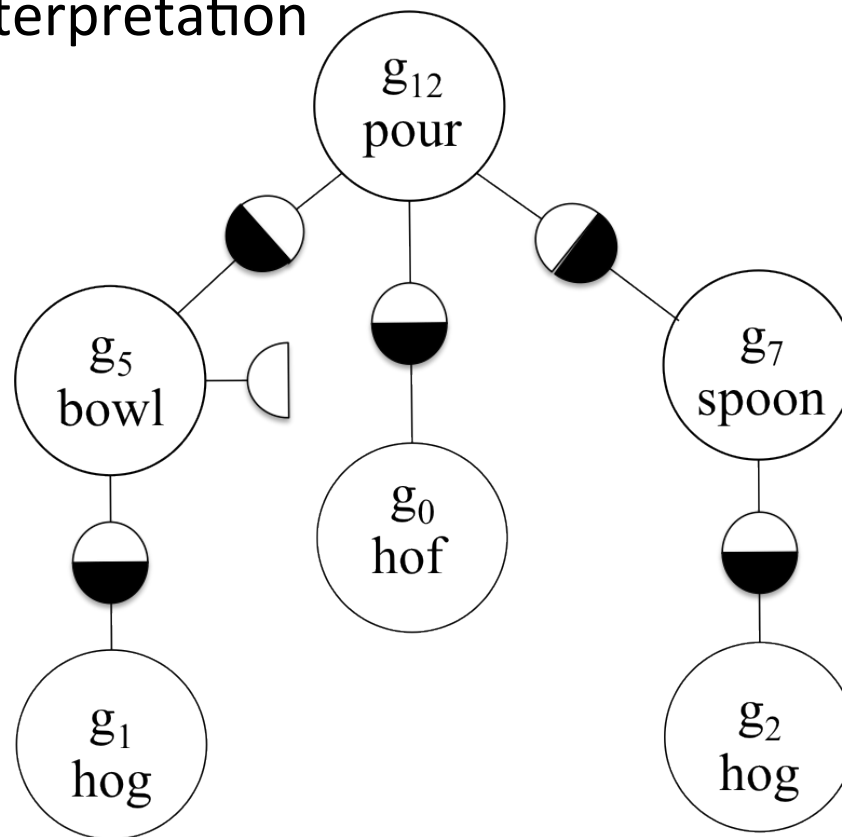
Create a global jump configuration (k best explanations)

Global Moves in Configuration Space



Global Moves in Configuration Space

Add feasible ontological connections to form a semantically meaningful interpretation

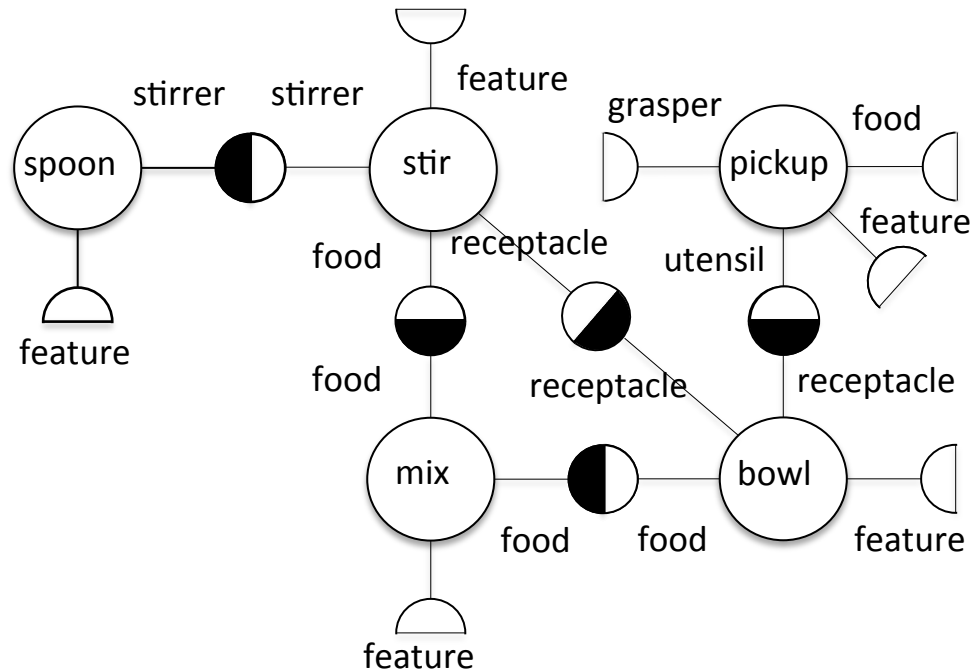


Local Moves in Configuration Space $C(R)$

- Transformation maps

- Generator mapping $s : G \rightarrow G$

$$s\sigma(g_1, g_2, \dots, g_n) = \sigma(sg_1, sg_2, \dots, sg_n) \in C(\sigma)$$

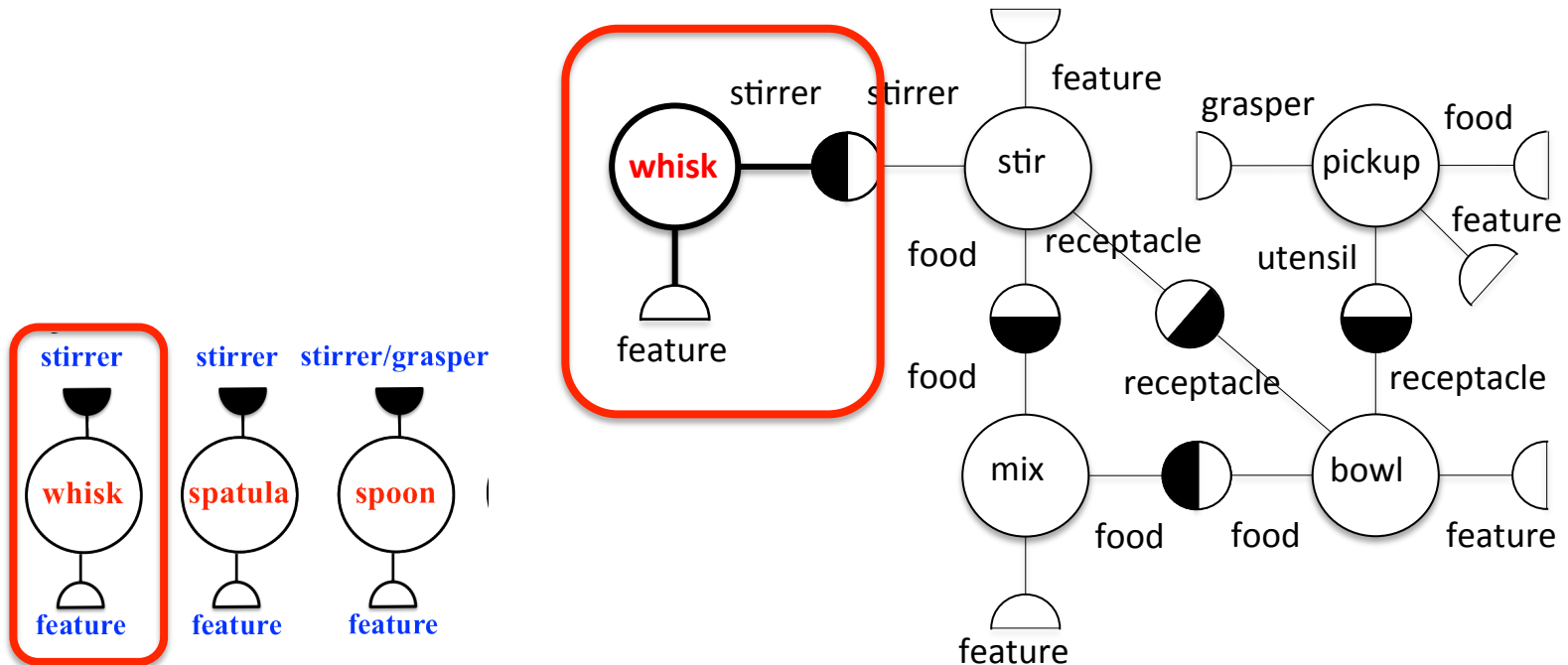


Local Moves in Configuration Space $C(R)$

- Transformation maps

- Generator mapping $s : G \rightarrow G$

$$s\sigma(g_1, g_2, \dots, g_n) = \sigma(sg_1, sg_2, \dots, sg_n) \in C(\sigma)$$

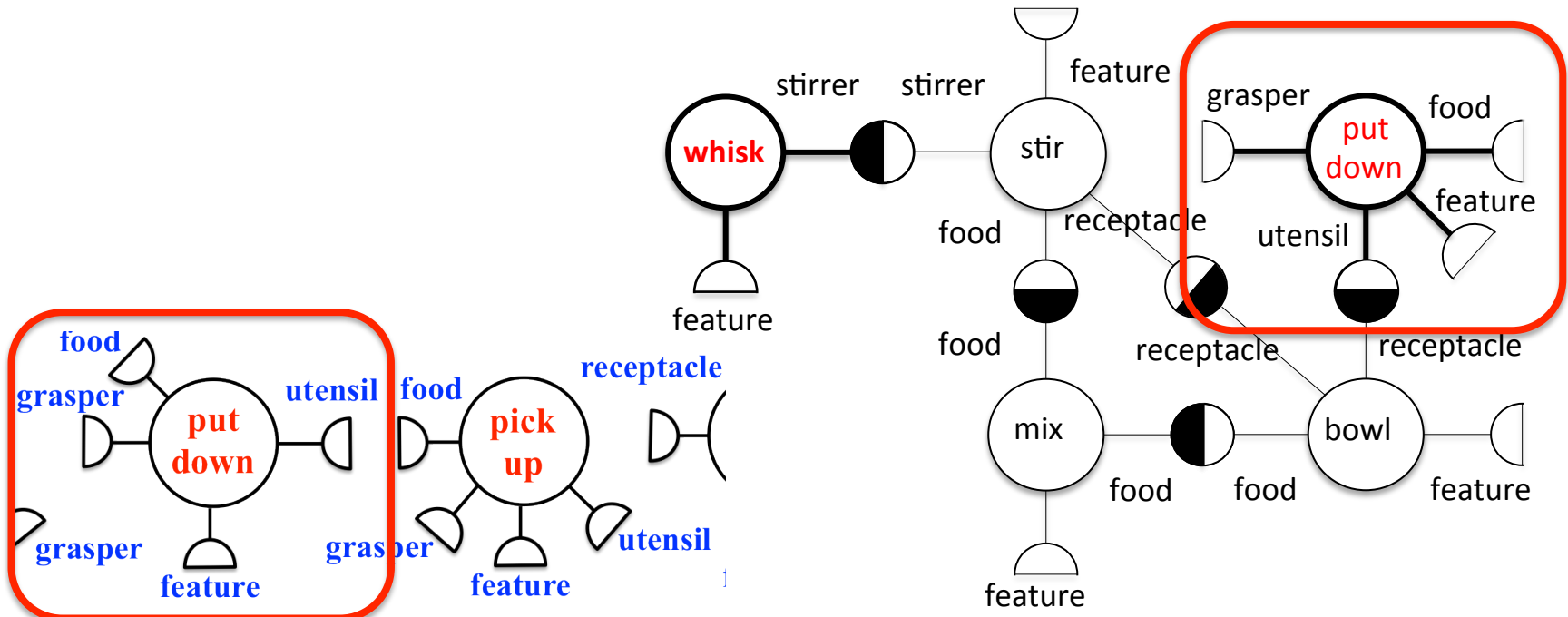


Local Moves in Configuration Space $C(R)$

- Transformation maps

- Generator mapping $s : G \rightarrow G$

$$s\sigma(g_1, g_2, \dots, g_n) = \sigma(sg_1, sg_2, \dots, sg_n) \in C(\sigma)$$

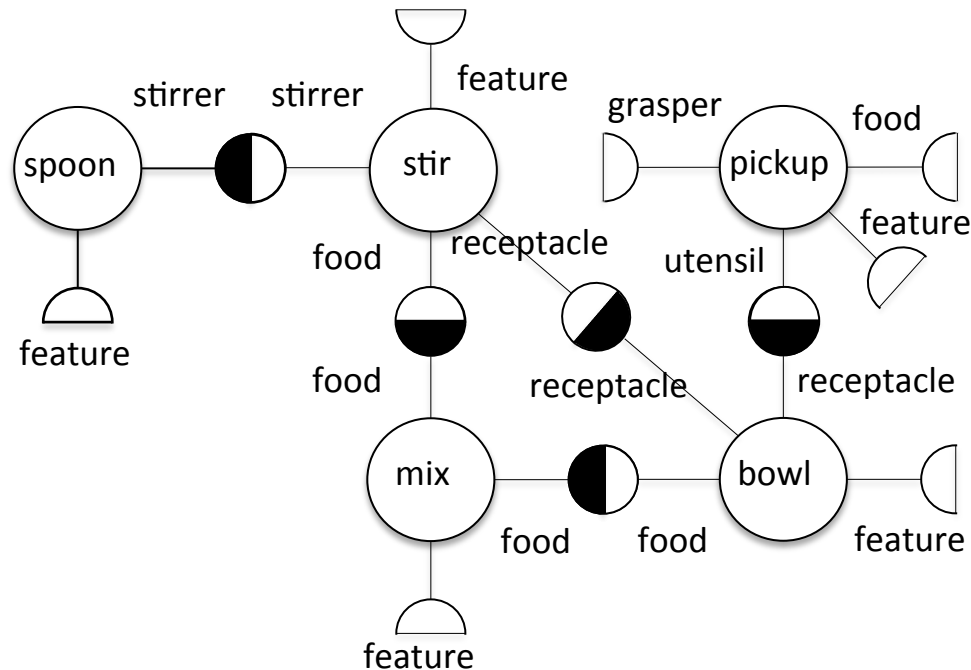


Local Moves in Configuration Space $C(R)$

- Transformation maps

- Simple moves $T : \Sigma \rightarrow \Sigma$

$$T\sigma(g_1, g_2, \dots, g_n) = \sigma'(g_{1'}, g_{2'}, \dots, g_{n'}) \in C(R)$$



Local Moves in Configuration Space $C(R)$

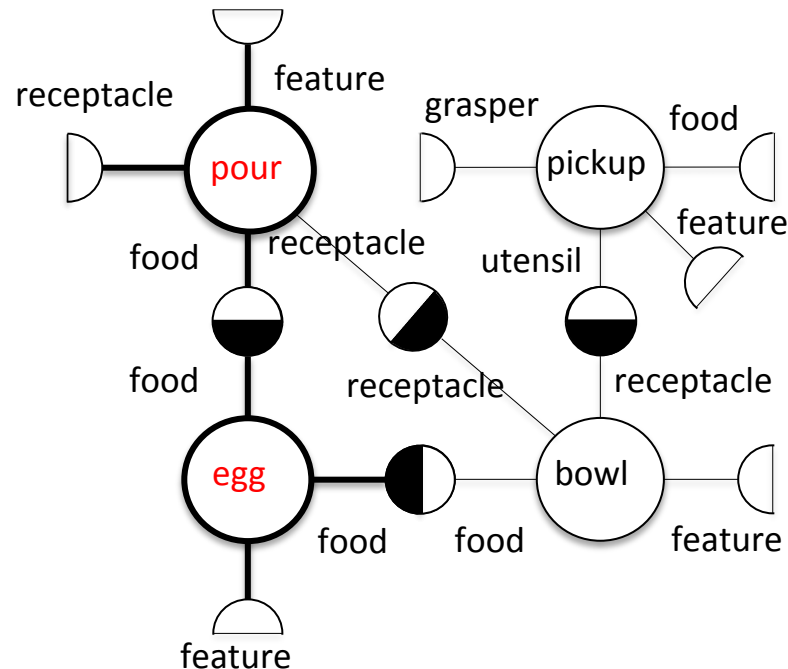
- Transformation maps

- Simple moves $T : \Sigma \rightarrow \Sigma$

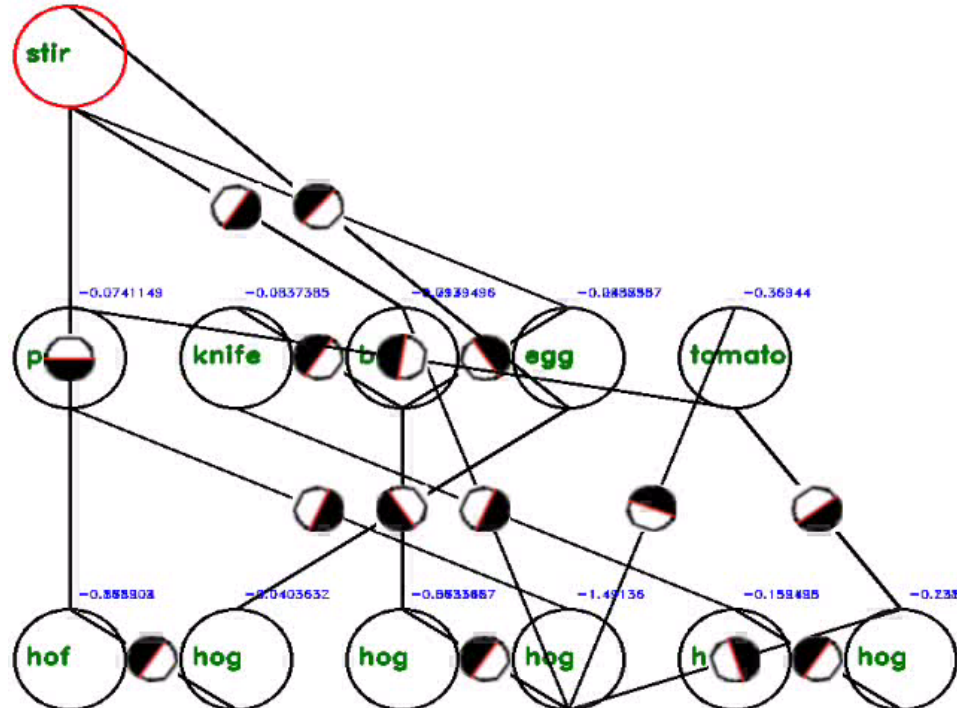
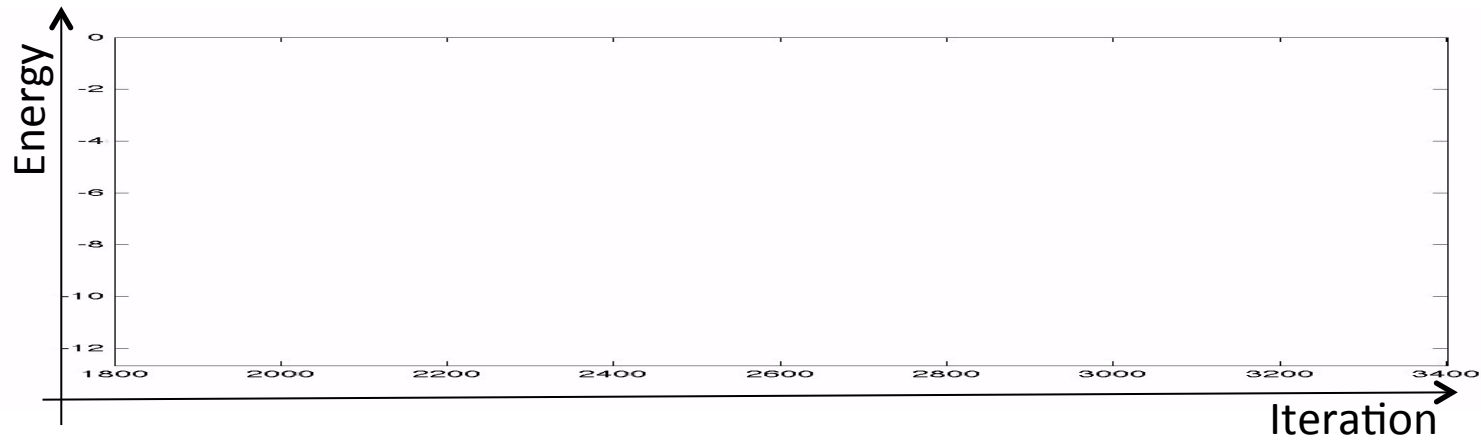
$$T\sigma(g_1, g_2, \dots, g_n) = \sigma'(g_{1'}, g_{2'}, \dots, g_{n'}) \in C(R)$$

- Multiple Simple Transformations

- Remove spoon generator
 - Replace stir generator by pour generator
 - Replace mix generator by egg generator



Inference Process Demo



Three Best Interpretations

LEVEL 3

2: stir

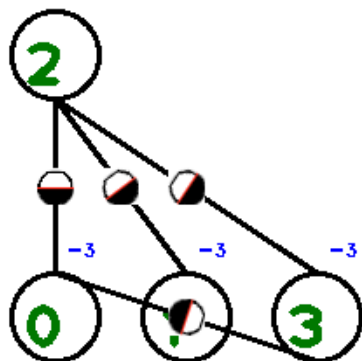
LEVEL 2

0: carrot_1

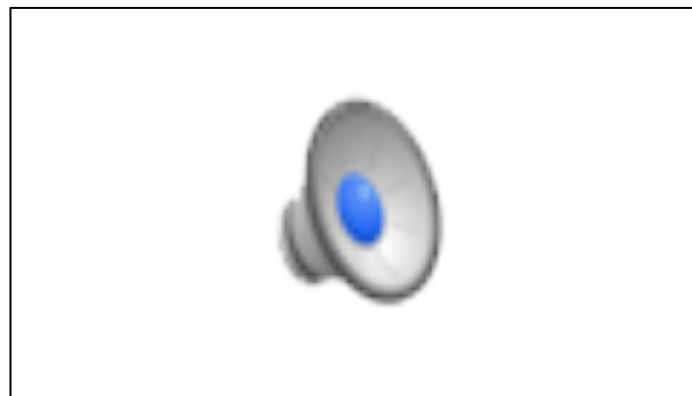
1: spatula_1

3: bowl_2

LEVEL 1



Ground-truth interpretation



Input video

LEVEL 3

11: stir

LEVEL 2

6: spatula

7: bowl

8: carrot

9: pepper

10: bowl

LEVEL 1

0: hof

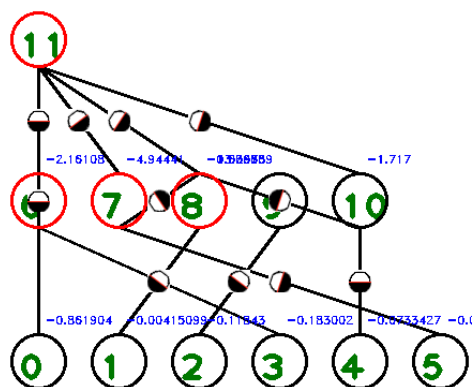
1: hog

2: hog

3: hog

4: hog

5: hog



1st Best

LEVEL 3

10: stir

LEVEL 2

6: spatula

7: bowl

8: carrot

9: bowl

11: oil

LEVEL 1

0: hof

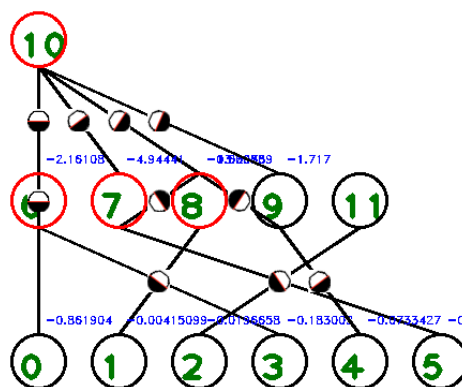
1: hog

2: hog

3: hog

4: hog

5: hog



2nd Best

LEVEL 3

11: putdowr

LEVEL 2

6: tomato

7: bowl

8: spatula

9: spoon

10: plate

LEVEL 1

0: hog

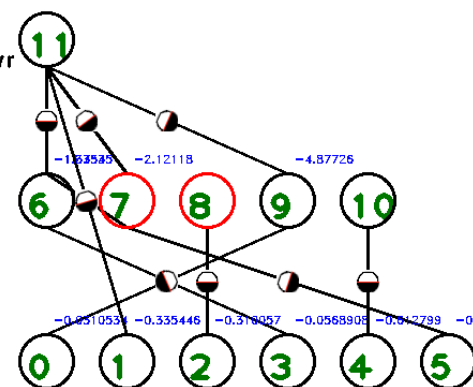
1: hof

2: hog

3: hog

4: hog

5: hog

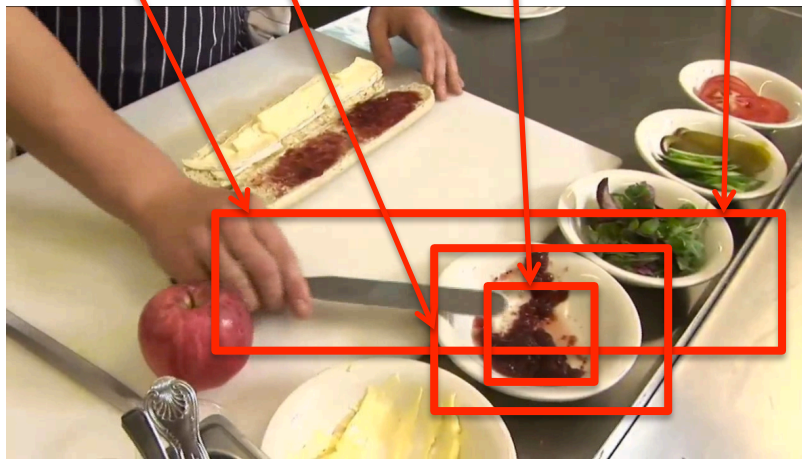
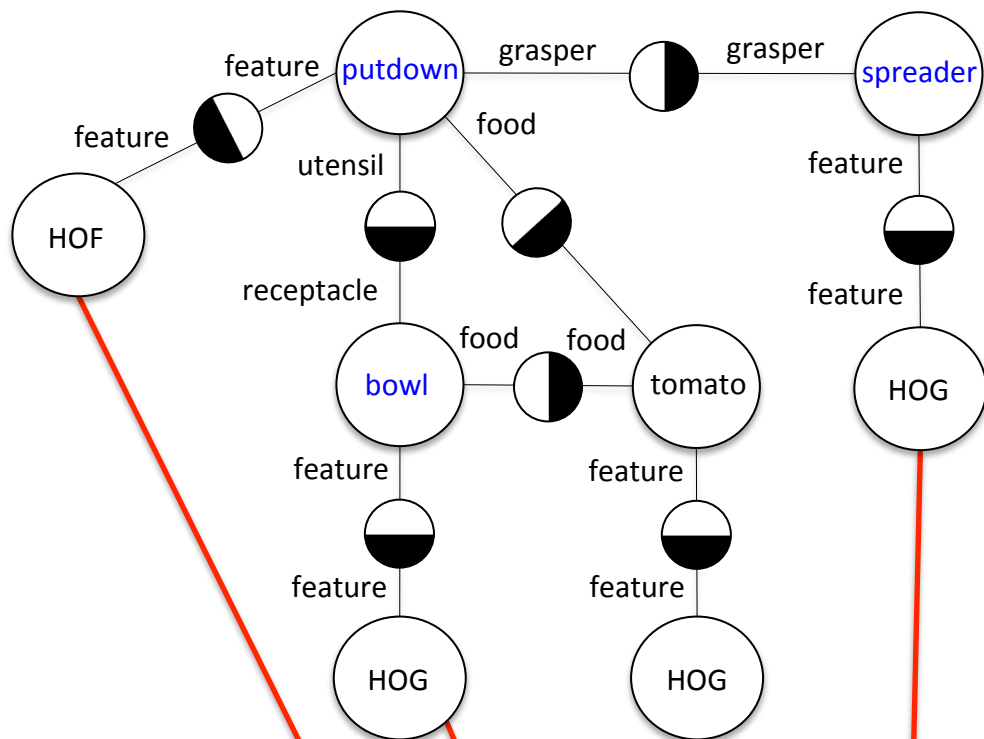


3rd Best

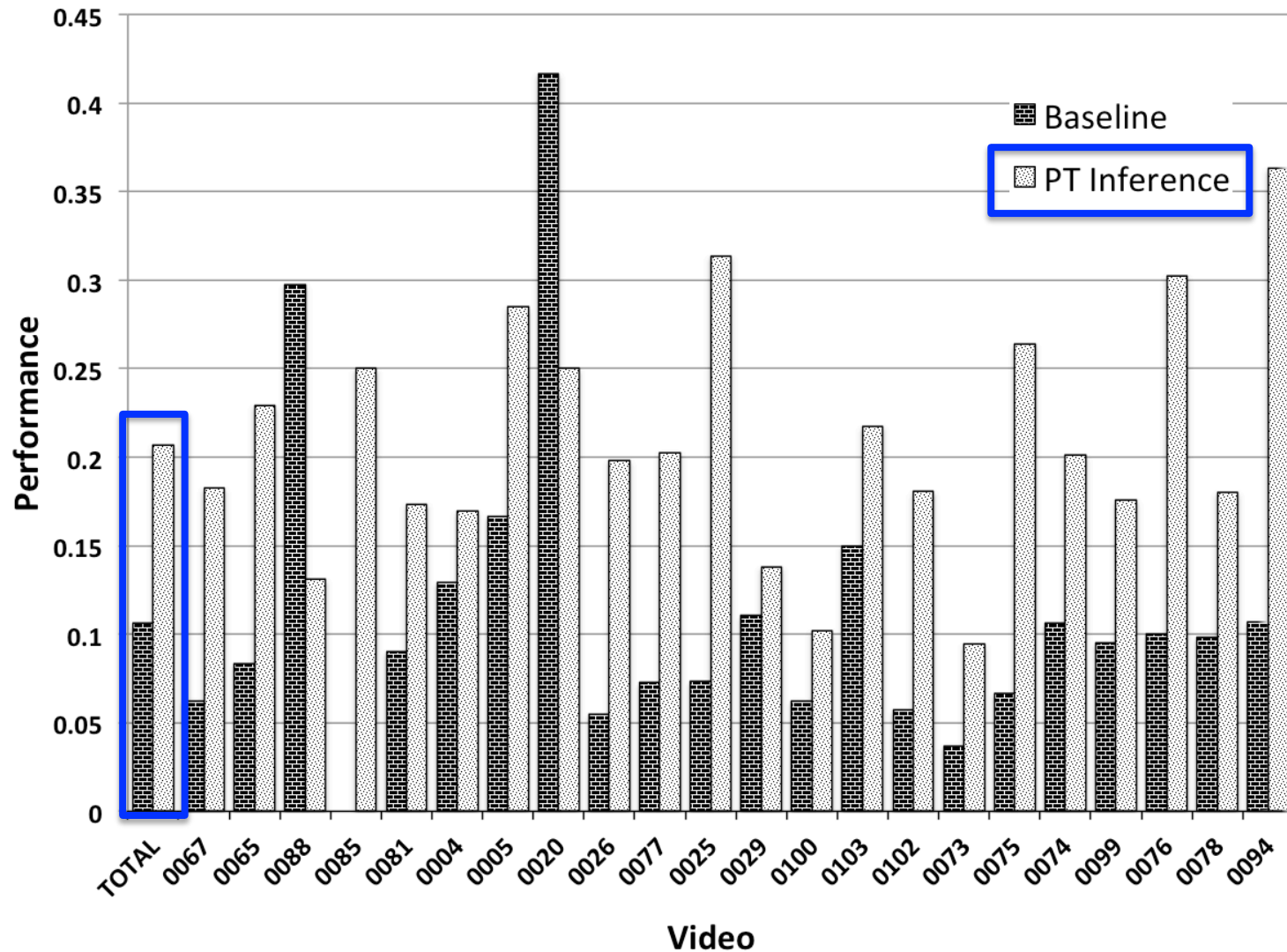
Experimental Results on

YouCook Dataset of Corso et al.

Breakfast Dataset of Serre et al.

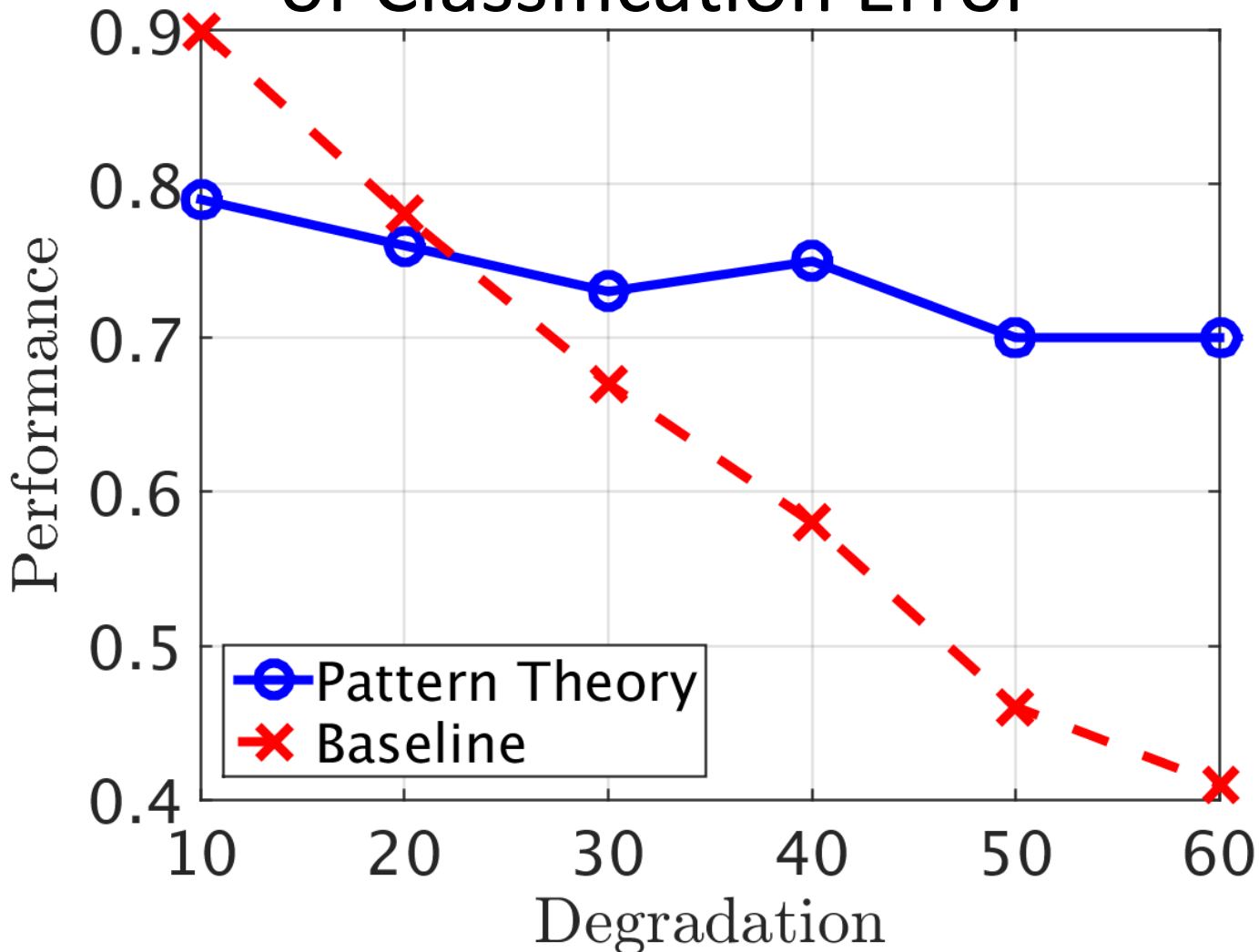


Performance Improvement by 100%

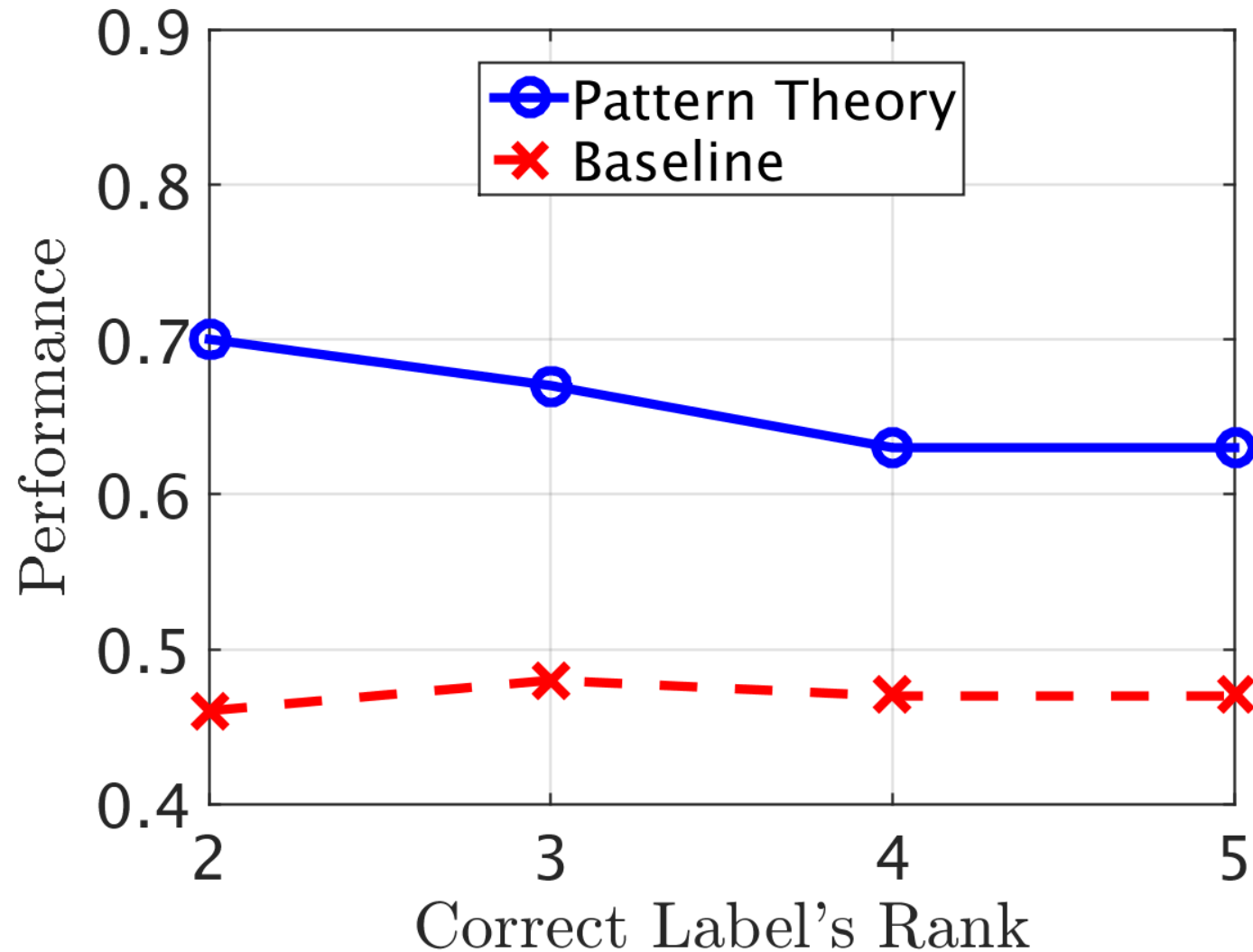


Souza et al., Pattern Theory-based Interpretation of Activities, ICPR 2014, Best Scientific Paper Award.

Robustness under Increasing Rates of Classification Error

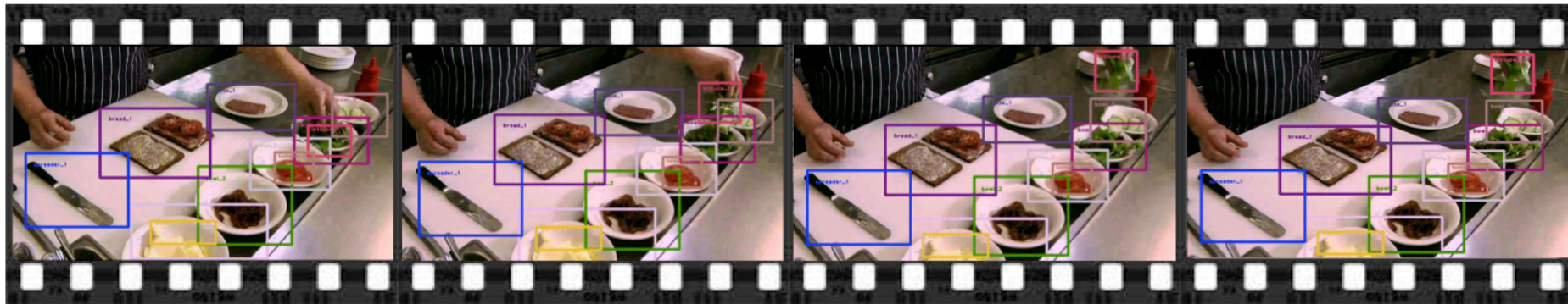


Robustness under Varying Severity Levels of Classification Errors (for 50% miss rate)

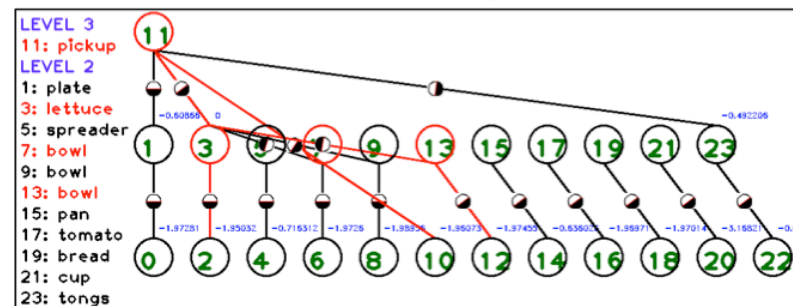


Interruptions: Object Clutter

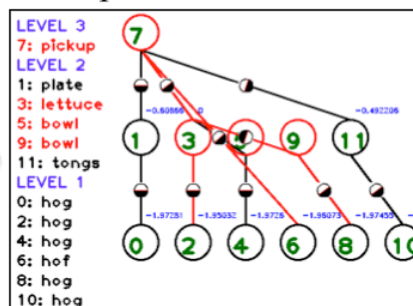
a) *Sequence of frames depicting the contents of a video shot: “pick up lettuce from bowl”*



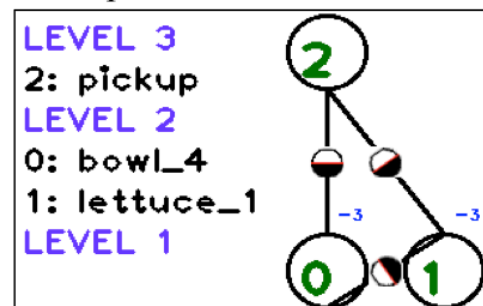
b) Pattern Theory-based Interpretation



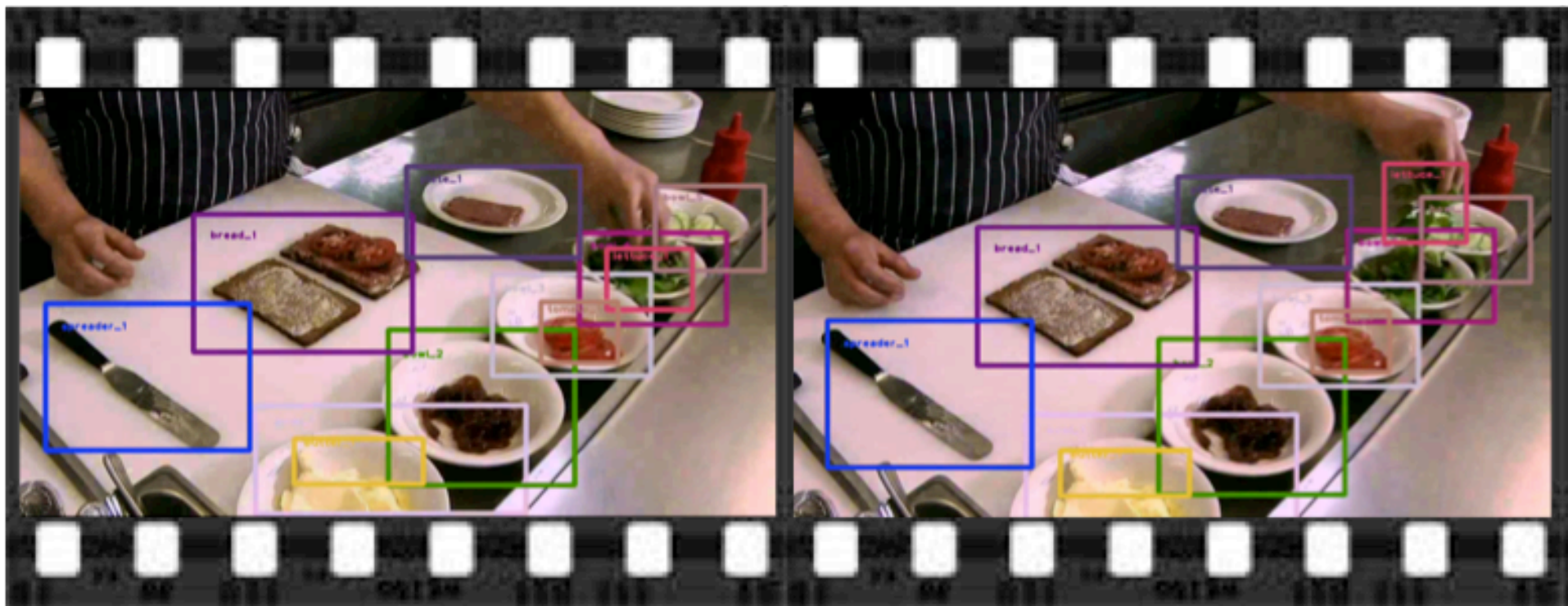
c) Larger Connected Component



d) Ground-truth Interpretation



Interruptions: Object Clutter



Interruptions: Object Clutter

Pattern Theory-based Semantic Interpretation

LEVEL 3

11: pickup

LEVEL 2

1: plate

3: lettuce

5: spreader

7: bowl

9: bowl

13: bowl

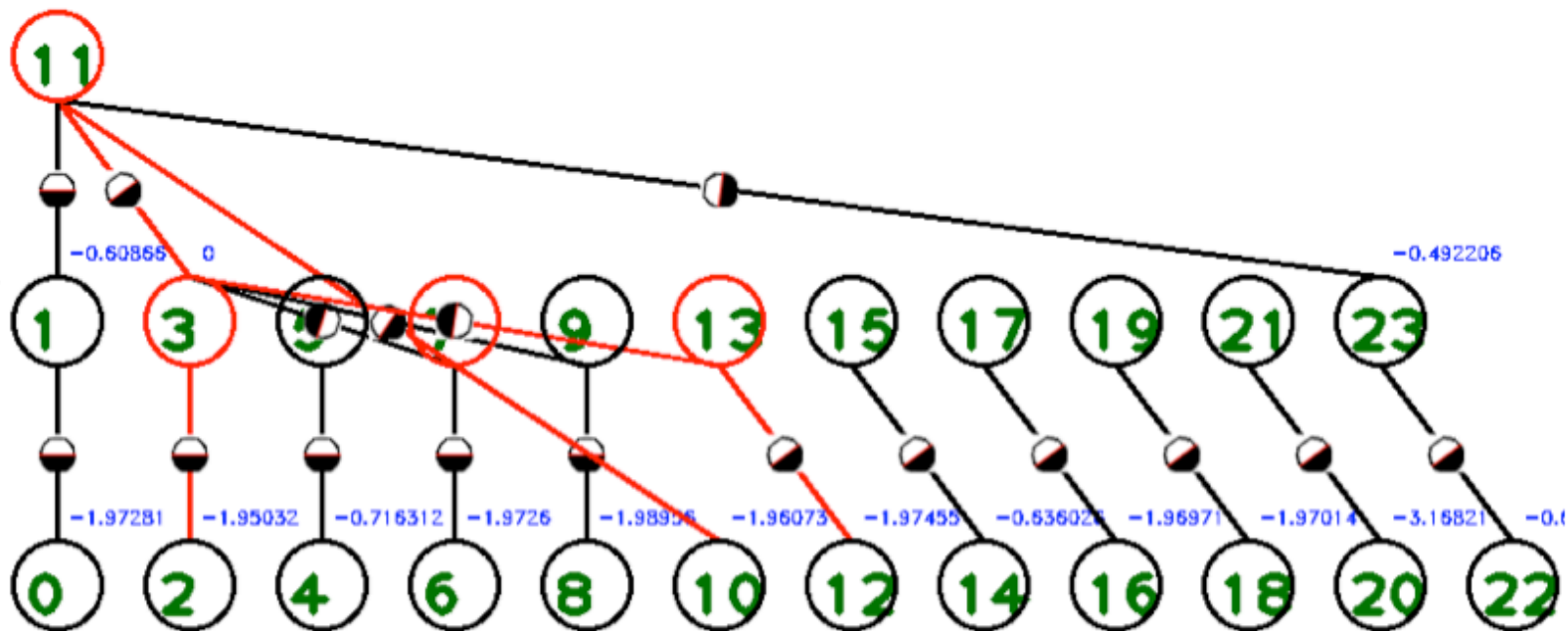
15: pan

17: tomato

19: bread

21: cup

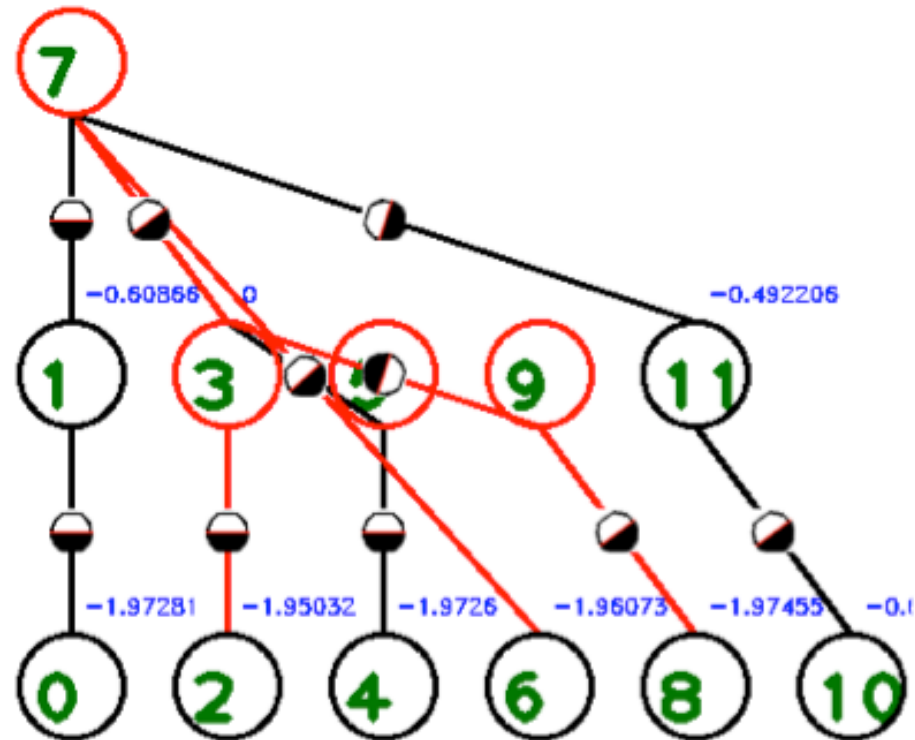
23: tongs



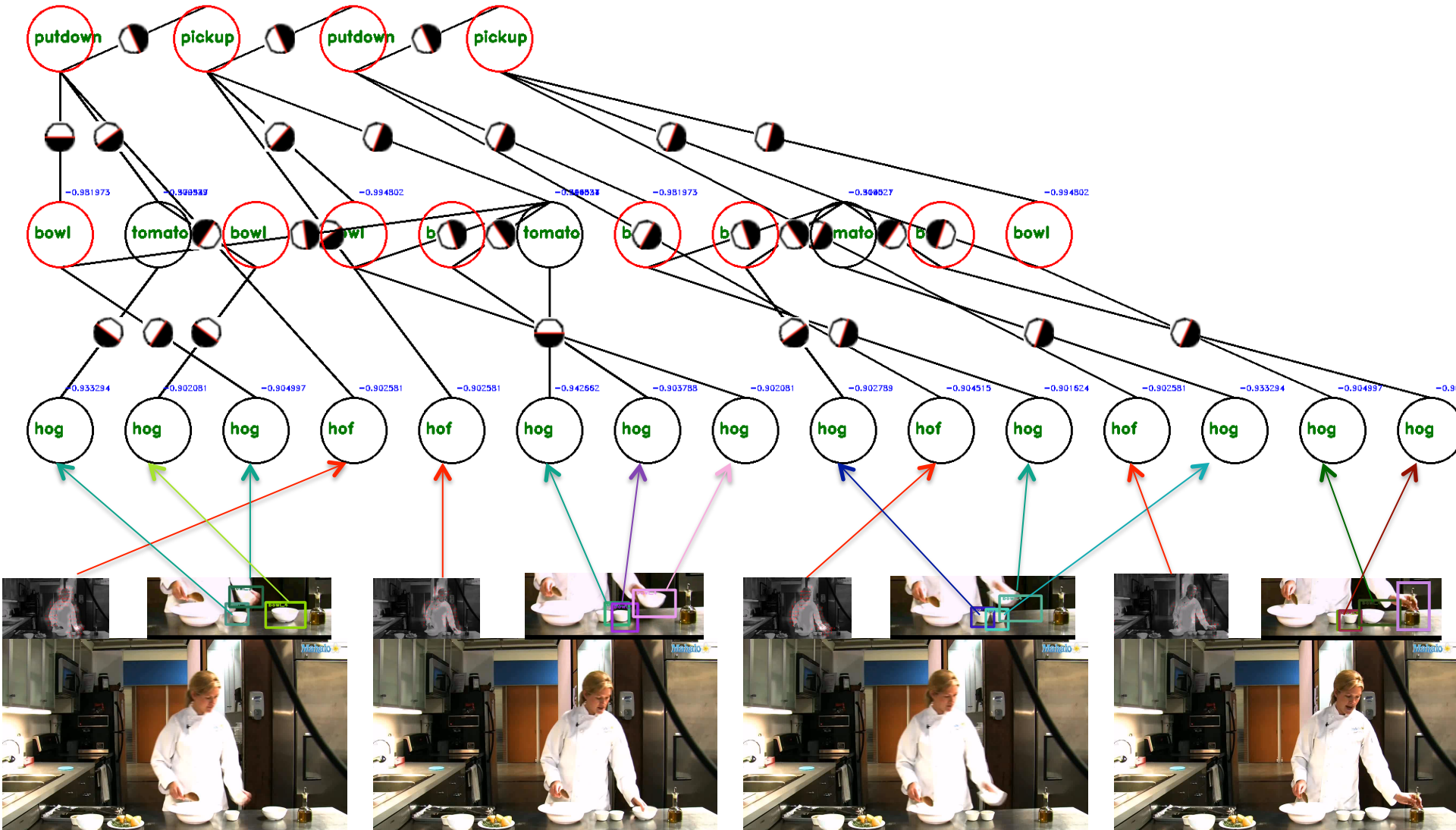
Interruptions: Object Clutter

Largest Connected Component

LEVEL 3
7: pickup
LEVEL 2
1: plate
3: lettuce
5: bowl
9: bowl
11: tongs
LEVEL 1
0: hog
2: hog
4: hog
6: hof
8: hog
10: hog

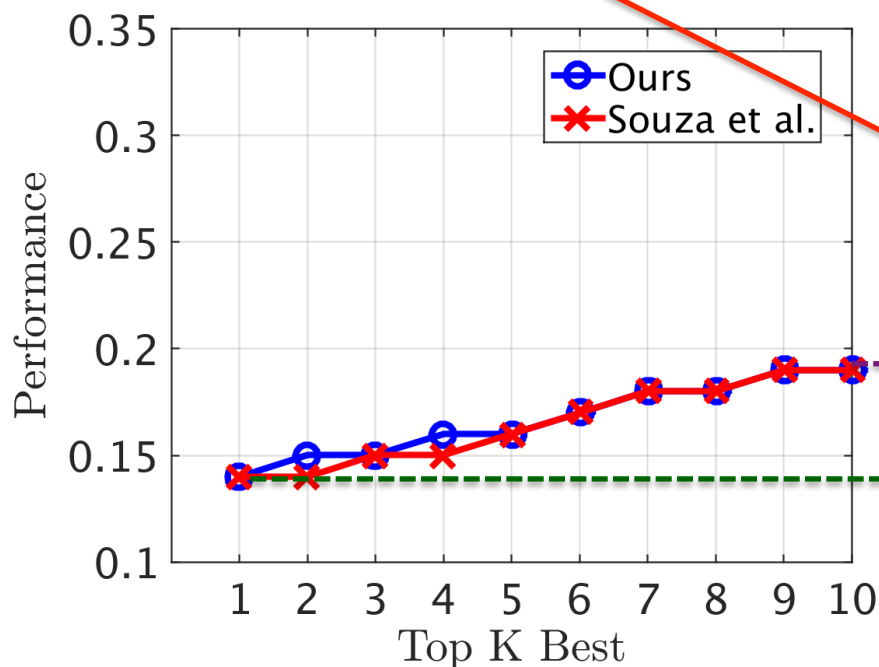


Domain Warping: Sequence of Actions



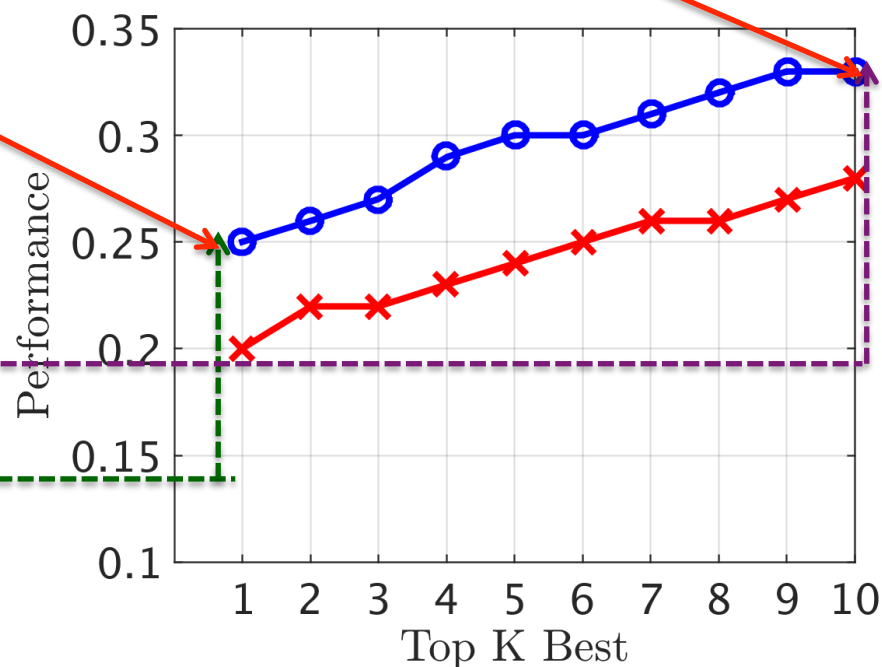
Performance Improvements w/ Temporal Bonds and Long Sequence Analysis

**78% Performance
Improvement
(Best Interpretation)**



Overall Performance
(1 video segment)

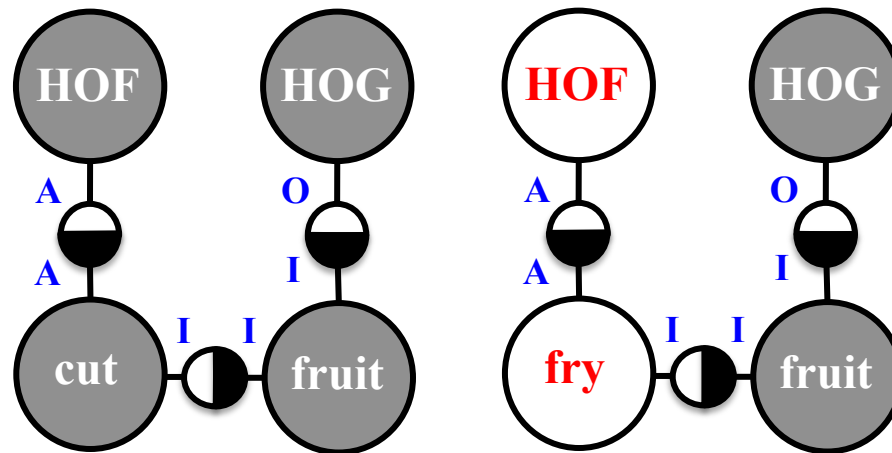
**83% Performance
Improvement
(10 Best Interpretations)**



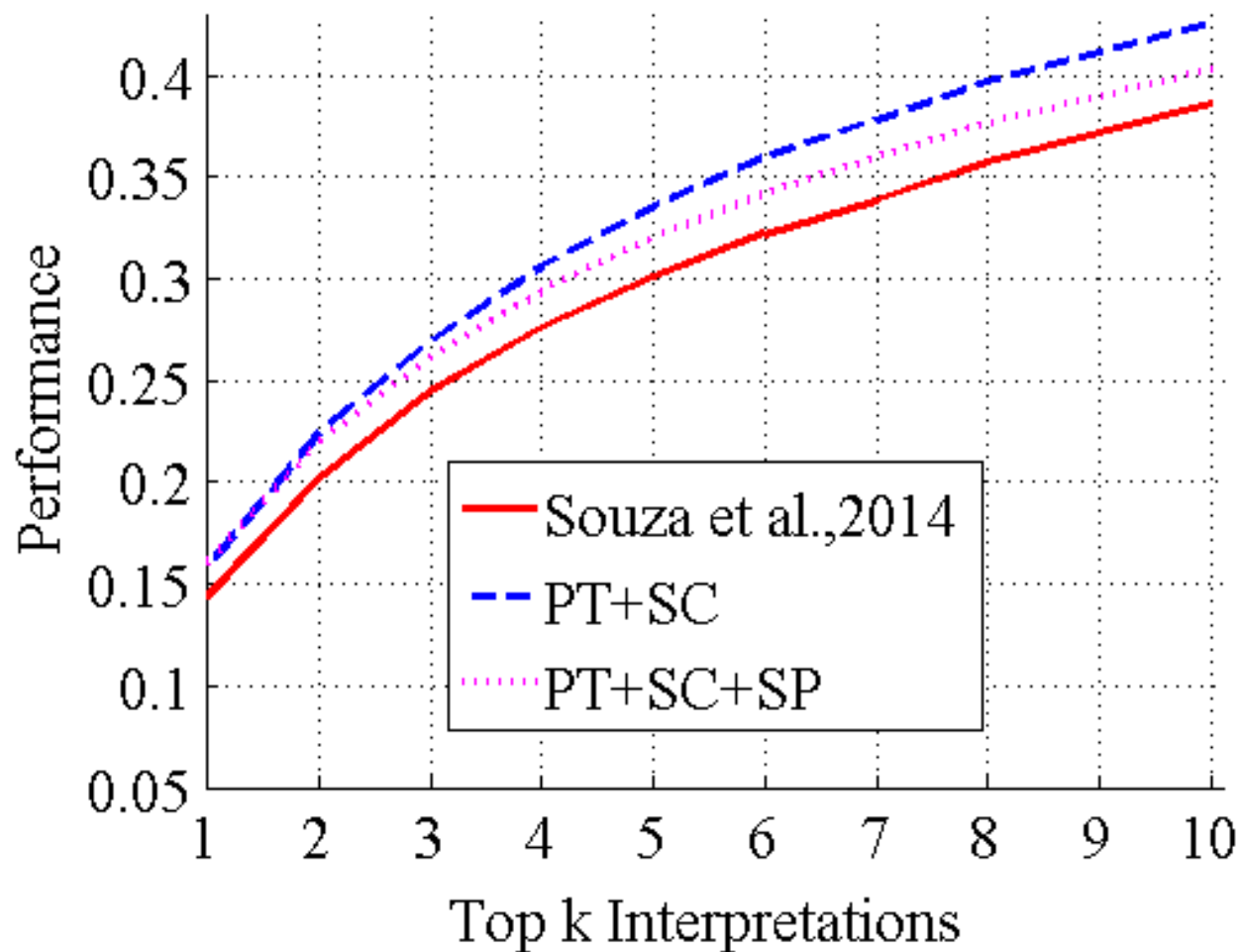
Overall Performance
(4 video segments)

Interruptions: Simultaneous Events

Side by Side
Addition of
Videos of
Two Events



Numeric Performance



Questions?