Ontology-Based Semantic Image Interpretation

Ivan Donadello[1,2] Luciano Serafini (Advisor)[1]

[1]Fondazione Bruno Kessler, Via Sommarive, 18 I-38123, Trento, Italy

[2]DISI University of Trento, Via Sommarive, 9 I-38123, Trento, Italy

September 23, 2015

Context

- Huge diffusion of digital images in recent years;
- lack of semantic based retrieval systems for images, that is no complex queries: "a person riding a horse on a meadow";
- semantic gap between numerical image features and human semantics;
- need a method that automatically understands the **semantic content of images**.

Relevance:

- semantic content based image retrieval via a query language;
- semantic content enrichment with Semantic Web resource.

Problem Statement

Semantic Image Interpretation (SII) is the task of extracting a graph representing the image content;

Problem Statement

Semantic Image Interpretation (SII) is the task of extracting a graph representing the image content;

- **nodes** represent visible and occluded objects in the image and their properties;

Problem Statement

Semantic Image Interpretation (SII) is the task of extracting a graph representing the image content;

- **nodes** represent visible and occluded objects in the image and their properties;
- **arcs** represent relations between objects;

Problem Statement

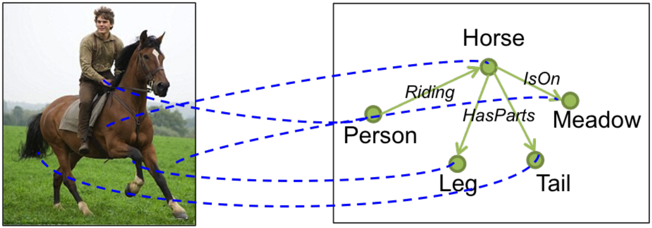
Semantic Image Interpretation (SII) is the task of extracting a graph representing the image content;

- **nodes** represent visible and occluded objects in the image and their properties;
- **arcs** represent relations between objects;
- **alignment** between visible object regions and nodes;

Problem Statement

Semantic Image Interpretation (SII) is the task of extracting a graph representing the image content;

- **nodes** represent visible and occluded objects in the image and their properties;
- **arcs** represent relations between objects;
- **alignment** between visible object regions and nodes;
- an ontology provides the formal semantics and constraints that guide the graph construction;



Aim of the Doctoral Thesis

- Define a theoretical reference framework for SII;
- implementation of a system for SII;
- graph construction guided by mixing:
  - **numeric information** (low-level features of the image);
  - **symbolic information** (high-level constraints available in the ontology);
- perform system evaluation on a ground truth of semantically interpreted images.

State-of-the-art on SII

Logic-Based Works (2014)

- a first description of the image (basic object recognition and their relations) is given;
- model generation (deduction or abduction) by exploiting the ontology.

Neural Networks-based (NN) works (2015)

- Caption generation;

State-of-the-art on SII

Logic-Based Works (2014)

- a first description of the image (basic object recognition and their relations) is given;
- model generation (deduction or abduction) by exploiting the ontology.

Neural Networks-based (NN) works (2015)

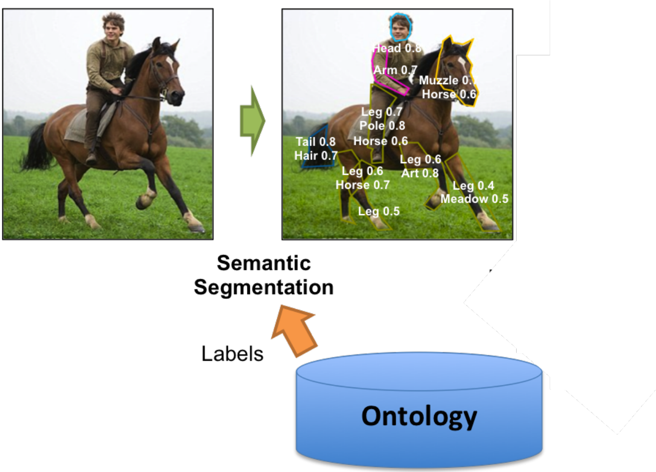
- Caption generation;

Limitations

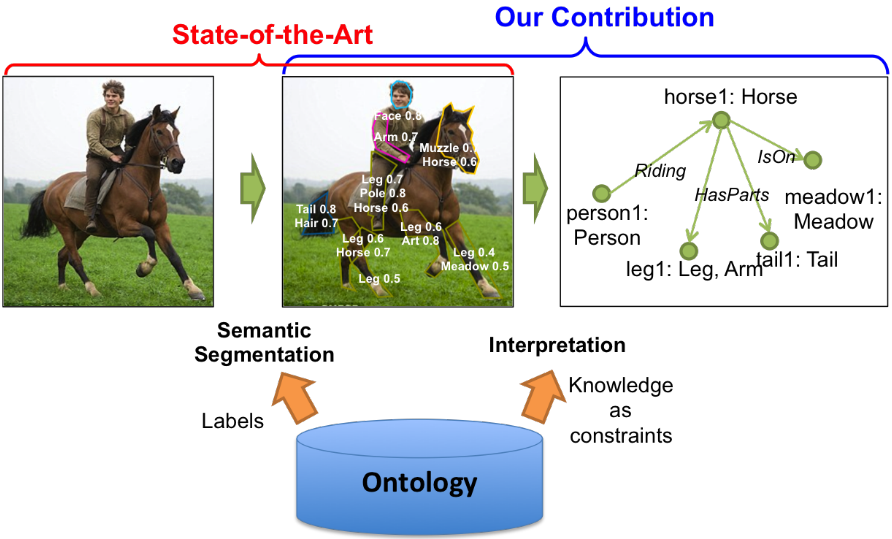
- Logic-based works: no consideration for low-level features;
- NN works: no formal semantics and a priori knowledge.

SII Pipeline

SII Pipeline

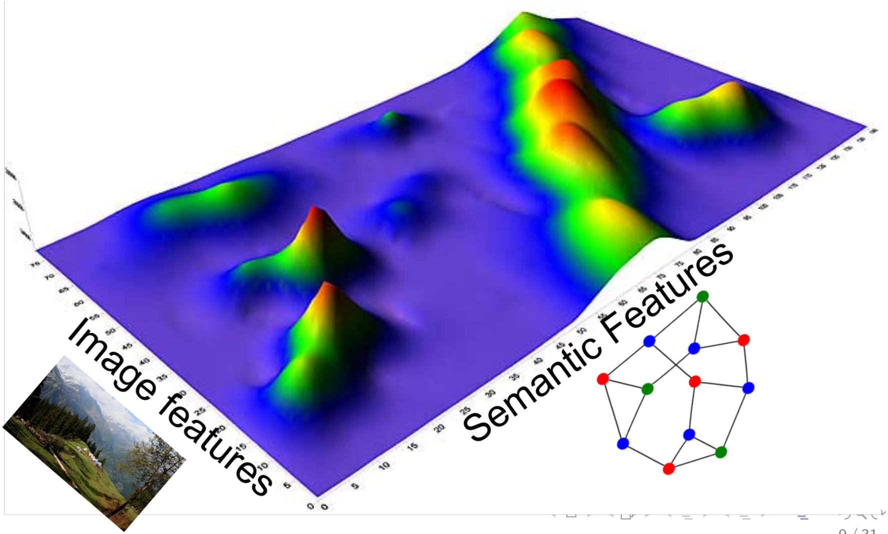


SII Pipeline



Our Vision of SII

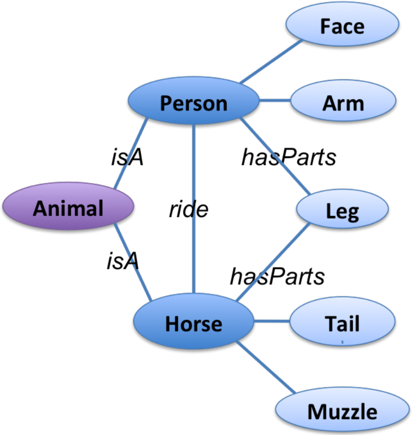
Finding the maximum of a joint search space composed of semantic features and image features.

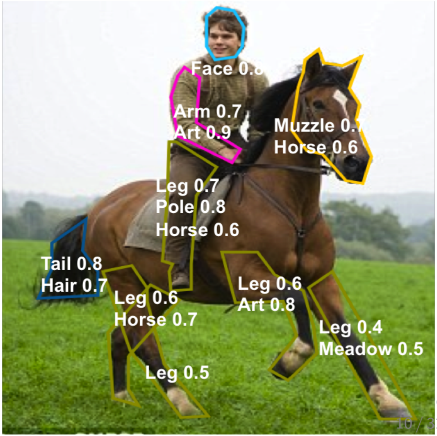


Theoretical Framework

Background Knowledge encoded in a Description Logic ontology $\mathcal{O}$.

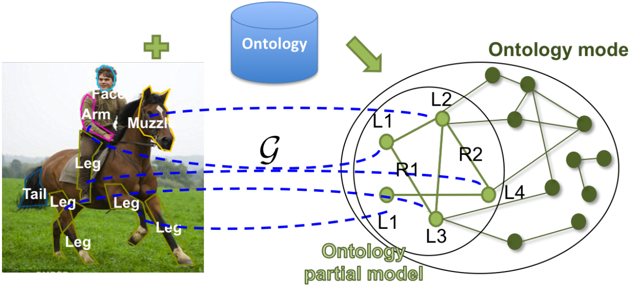
Labelled picture is a pair $\mathcal{P} = \langle S, L \rangle$ where S are segments of the image, L are (weighted) labels from Σ.





The Partial Model

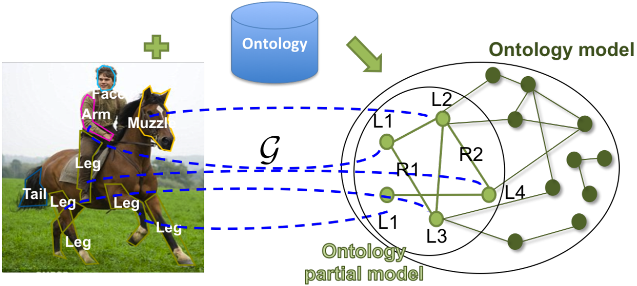
- A picture is a partial view of the real world;



- **A partial model $\mathcal{I}_p$ is a structure that can be extended to a model of $\mathcal{O}$;**

The Partial Model

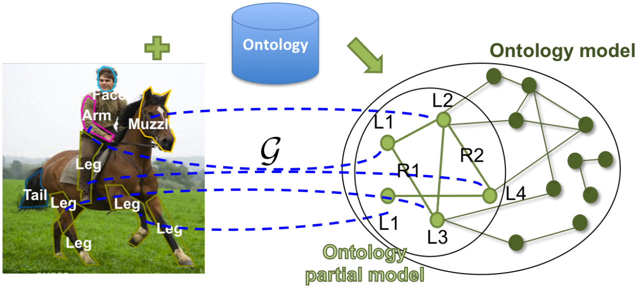
- A picture is a partial view of the real world;



- **A partial model $\mathcal{I}_p$ is a structure that can be extended to a model of $\mathcal{O}$;**
- . A partial model of an ontology $\mathcal{O}$ is an interpretation $\mathcal{I}_p = (\Delta^{\mathcal{I}_p}, \cdot^{\mathcal{I}_p})$ of $\mathcal{O}$: there exists a model $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with $\Delta^{\mathcal{I}_p} \subseteq \Delta^{\mathcal{I}}$ and $\cdot^{\mathcal{I}_p}$ is a restriction of $\cdot^{\mathcal{I}}$ on $\Delta^{\mathcal{I}_p}$.

The Partial Model

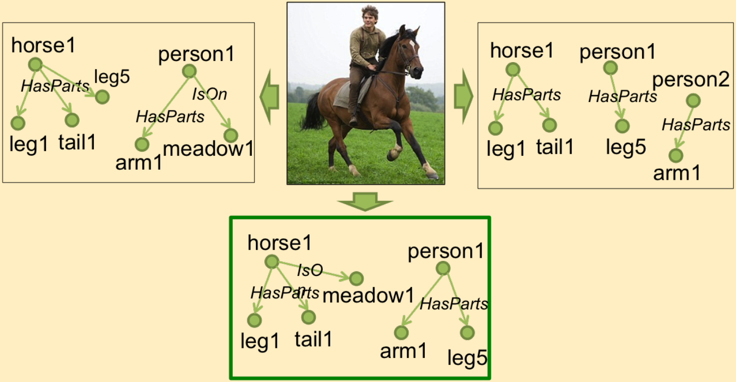
- A picture is a partial view of the real world;



- **A partial model $\mathcal{I}_p$ is a structure that can be extended to a model of $\mathcal{O}$;**
- . A partial model of an ontology $\mathcal{O}$ is an interpretation $\mathcal{I}_p = (\Delta^{\mathcal{I}_p}, \cdot^{\mathcal{I}_p})$ of $\mathcal{O}$: there exists a model $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with $\Delta^{\mathcal{I}_p} \subseteq \Delta^{\mathcal{I}}$ and $\cdot^{\mathcal{I}_p}$ is a restriction of $\cdot^{\mathcal{I}}$ on $\Delta^{\mathcal{I}_p}$.
- A **semantically interpreted picture** is a triple $(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_{\mathcal{O}}$;

The Most Plausible Partial Model

Many partial models for a picture



Searching for the partial model that best fits the picture content, i.e. the **most plausible partial model**.

The Semantic Image Interpretation Problem

Formalization

- A **cost function** $\mathcal{S}$ assigns a cost to semantically interpreted pictures $(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_{\mathcal{O}}$;
- $\mathcal{S}(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_{\mathcal{O}}$ expresses the gap between low-level features of $\mathcal{P}$ and objects and relations encoded in $\mathcal{I}_p$;
- the **most plausible partial model** $\mathcal{I}_p^*$ minimizes $\mathcal{S}$:

$$\mathcal{I}_p^* = \underset{\substack{\mathcal{I}_p \models_p \mathcal{O} \\ \mathcal{G} \subseteq \Delta^{\mathcal{I}_p} \times S}}{\text{argmin}} \mathcal{S}(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_{\mathcal{O}}$$

- the **semantic image interpretation problem** is the construction of $(\mathcal{P}, \mathcal{I}_p^*, \mathcal{G})_{\mathcal{O}}$ that minimizes $\mathcal{S}$.

Case Study: Clustering-Based Cost Function

- Task: **part-whole recognition**, i.e., discovery complex objects from their parts;
- part-whole recognition can be seen as a **clustering problem**;
  - parts of the same object tend to be grouped together;

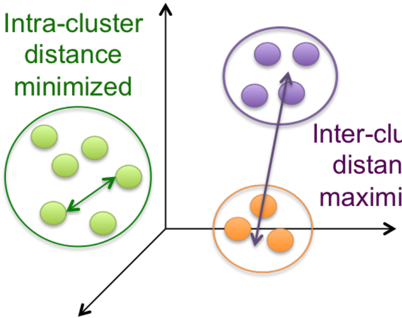
Case Study: Clustering-Based Cost Function

- Task: **part-whole recognition**, i.e., discovery complex objects from their parts;
- part-whole recognition can be seen as a **clustering problem**;
  - parts of the same object tend to be grouped together;

- cost function as a clustering optimisation function.

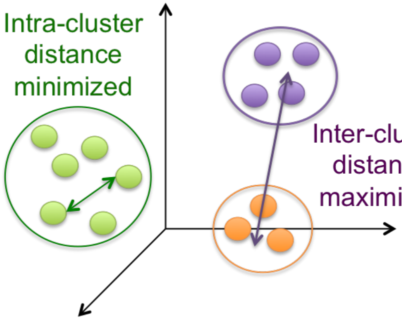
Case Study: Clustering-Based Cost Function

- Clustering: grouping a set of input elements into groups (clusters) such that:



Case Study: Clustering-Based Cost Function

- Clustering: grouping a set of input elements into groups (clusters) such that:

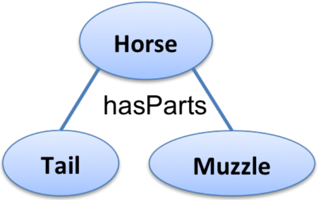


- *clustering solution* of $(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_{\mathcal{O}}$ is $\mathcal{C} = \{C_d \mid d \in \Delta^{\mathcal{I}_p}\}$ where $C_d = \{\mathcal{G}(d') \mid d' \in \Delta^{\mathcal{I}_p}, \langle d, d' \rangle \in \text{hasPart}^{\mathcal{I}_p}\}$;
- d represents the composite object, the **centroid** of the cluster;

Case Study: Clustering-Based Cost Function

Mixing numeric and semantic features:

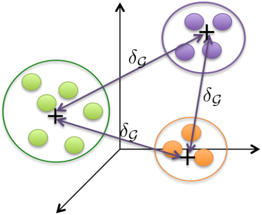
- **grounding distance** $\delta_{\mathcal{G}}(d, d')$: the Euclidean distance between the centroids of $\mathcal{G}(d)$ and $\mathcal{G}(d')$;
- **semantic distance** $\delta_{\mathcal{O}}(d, d')$ is the shortest path in $\mathcal{O}$:



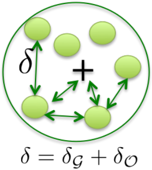
  - if $\text{Muzzle}(d'), \text{Tail}(d'')$ then $\delta_{\mathcal{O}}(d', d'') = 2$;
  - if $\text{Muzzle}(d'), \text{Horse}(d)$ then $\delta_{\mathcal{O}}(d', d) = 1$;

Case Study: Clustering-Based Cost Function

- **Inter-cluster distance Γ**:



- **Intra-cluster distance Λ**:

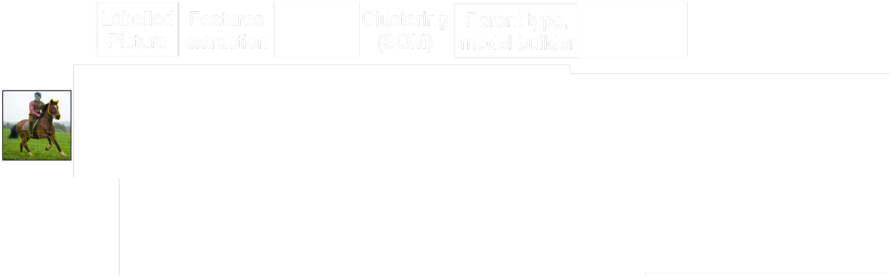


- **Cost function**:

$$\mathcal{S}(\mathcal{P}, \mathcal{I}_{\boldsymbol{p}}, \mathcal{G})_{\mathcal{O}} = \alpha \cdot \Gamma + (1 - \alpha) \cdot \Lambda$$

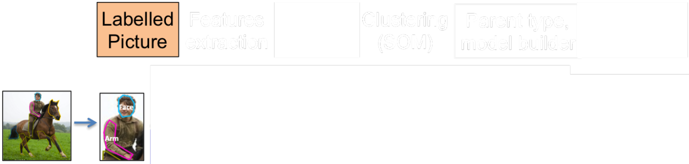
Minimising the Cost Function

The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



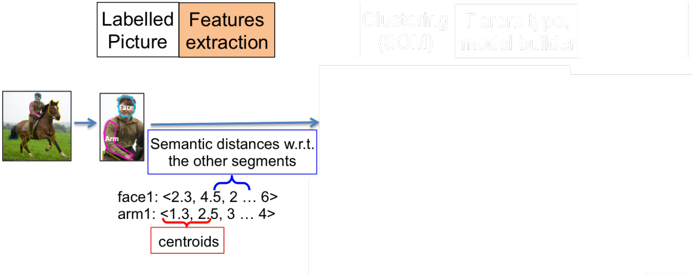
Minimising the Cost Function

The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



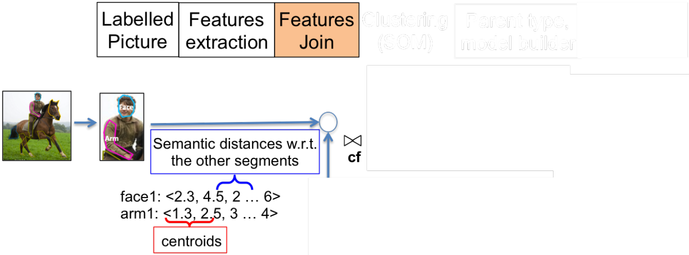
Minimising the Cost Function

The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



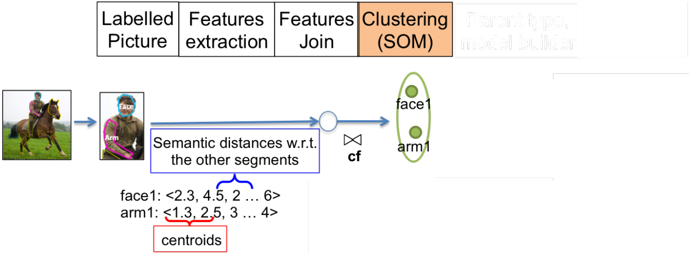
Minimising the Cost Function

The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



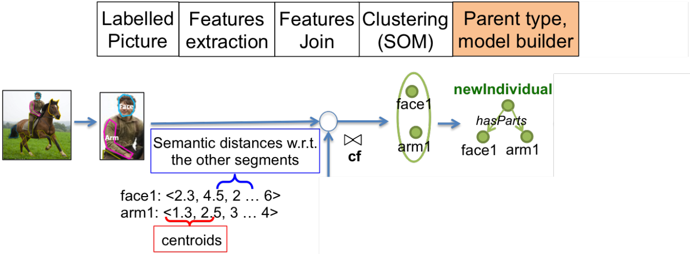
Minimising the Cost Function

The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



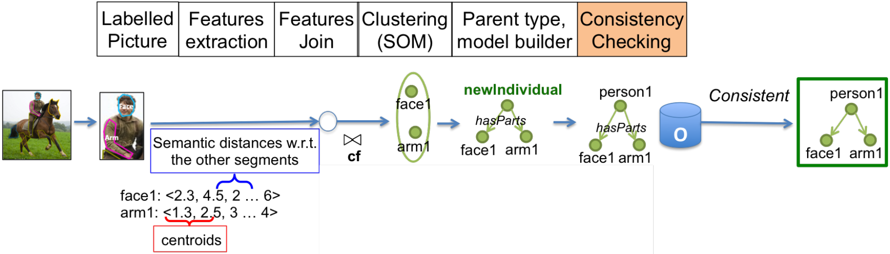
Minimising the Cost Function

The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



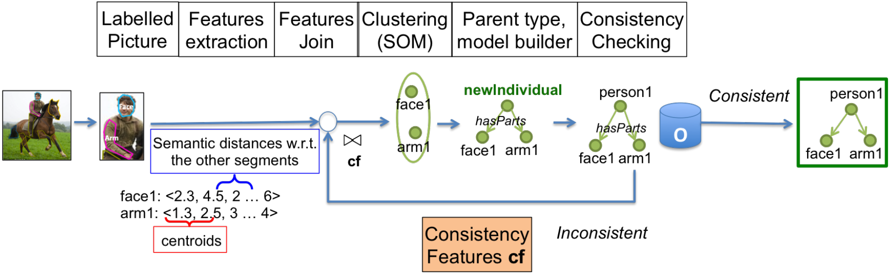
Minimising the Cost Function

The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



Minimising the Cost Function

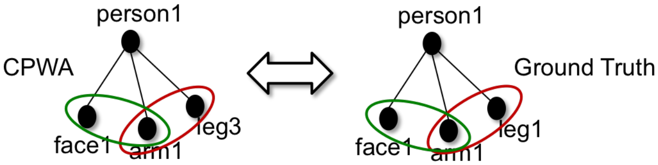
The Clustering Part-Whole Algorithm (CPWA) approximates the minimum of the cost function.



Evaluation

Comparing the predicted partial model with the ground truth, two measures:

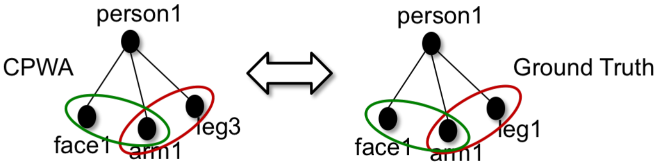
- **grouping (GRP)**:



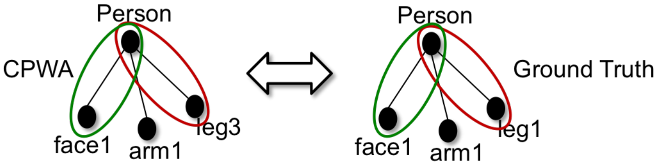
Evaluation

Comparing the predicted partial model with the ground truth, two measures:

- **grouping (GRP)**:



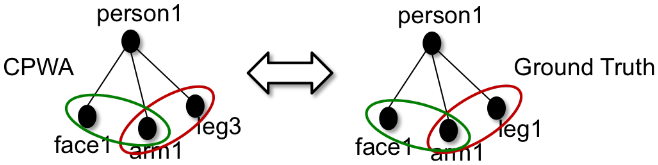
- **complex-object type prediction (COP)**:



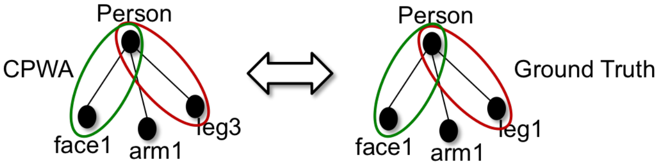
Evaluation

Comparing the predicted partial model with the ground truth, two measures:

- **grouping (GRP)**:



- **complex-object type prediction (COP)**:



- precision, the fraction of predicted pairs that are correct;
- recall, the fraction of correct pairs that are predicted.

Experiments Setting

- **Ground truth** of 203 manually obtained labelled pictures on the urban scene domain;
- manually built **ontology** with basic formalism of meronymy of the domain;
- **task**: discovering complex objects from their parts in pictures.

Results

	$prec_{\mathrm{GRP}}$	rec_{GRP}	$F1_{\mathrm{GRP}}$	$prec_{\mathrm{COP}}$	rec_{COP}	$F1_{\mathrm{COP}}$
CPWA	0.61	0.89	0.67	0.73	0.75	0.74

Experiments Setting

- **Ground truth** of 203 manually obtained labelled pictures on the urban scene domain;
- manually built **ontology** with basic formalism of meronymy of the domain;
- **task**: discovering complex objects from their parts in pictures.

Results

	$prec_{\mathrm{GRP}}$	rec_{GRP}	$F1_{\mathrm{GRP}}$	$prec_{\mathrm{COP}}$	rec_{COP}	$F1_{\mathrm{COP}}$
CPWA	**0.61**	**0.89**	**0.67**	**0.73**	**0.75**	**0.74**
Baseline	0.45	0.71	0.48	0.66	0.69	0.66

- **Baseline**: clustering without semantics;

Experiments and Results

Experiments Setting

- **Ground truth** of 203 manually obtained labelled pictures on the urban scene domain;
- manually built **ontology** with basic formalism of meronymy of the domain;
- **task**: discovering complex objects from their parts in pictures.

Results

	$prec_{\mathrm{GRP}}$	rec_{GRP}	$F1_{\mathrm{GRP}}$	$prec_{\mathrm{COP}}$	rec_{COP}	$F1_{\mathrm{COP}}$
CPWA++	**0.67**	0.81	**0.71**	**0.71**	**0.82**	**0.86**
CPWA	0.61	**0.89**	0.67	0.73	0.75	0.74
Baseline	0.45	0.71	0.48	0.66	0.69	0.66

- **Baseline**: clustering without semantics;
- $\mathrm{CPWA}++$: improved version of CPWA;

Conclusions and Future Work

- Theoretical framework for SII: partial model that minimizes a cost function;
- cost function as a clustering optimization function;
- clustering algorithm that approximates the cost function;
- explicitly using semantics improves the results;
- future work:

Conclusions and Future Work

- Theoretical framework for SII: partial model that minimizes a cost function;
- cost function as a clustering optimization function;
- clustering algorithm that approximates the cost function;
- explicitly using semantics improves the results;
- future work:
  - integrating of semantic segmentation algorithms;
  - generalizing to other relations;
  - extending the evaluation to a standard dataset;
  - using general purposes ontologies;

Thanks for listening

Questions?