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

November 26, 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;
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;
- a **logical theory** (ontology) provides the formal semantics and constraints that guide the graph construction;
Applications

Caption Generation

*A person is riding a horse on a green meadow.*
Applications

Caption Generation

A person is riding a horse on a green meadow.

Story Telling

Your holidays started with a riding on the mountains, then you drunk good wine…
Applications

Caption Generation

A person is riding a horse on a green meadow.

Story Telling

Your holidays started with a riding on the mountains, then you drunk good wine…

Video Description Generation

In this scene Fantozzi tries to play billiard
Applications

Caption Generation

A person is riding a horse on a green meadow.

Video Description Generation

In this scene Fantozzi tries to play billiard

Story Telling

Your holidays started with a riding on the mountains, then you drunk good wine...

Content-Based Image Retrieval

select ?image
where

?image hasContent ?p.
?p type 'person'.
?p isRiding ?h.
?h type 'horse'
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;

Caption:
- bouquet of red flowers
- tablet
- bottle of water
- glass of water with ice and lemon
- cup of coffee
- dining table with breakfast items
- plate of fruit
- banana slices
- fork
- a person sitting at a table
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

State-of-the-Art

Our Contribution

Semantic Segmentation

Interpretation

Ontology

Knowledge as constraints
We use a Description Logic (DL) as logical language of the theory:

- DL is a family of logical languages;
- it is the most prominent formalism for knowledge representation;
- more expressive than propositional logic;
- well defined semantics (as in First Order Logic FOL);
- **decidable** fragment of FOL with efficient inference services (reasoning);
- relevant applications (Semantic Web);
We focus on $SHIQ$;

let $\Sigma$ the alphabet of the logic: $\Sigma = \langle \Sigma_C, \Sigma_R, \Sigma_I \rangle$, we have symbols for concepts, relationships and individuals;

well formed formulas (or concepts) are defined by the grammar:

$$
C, D \ := \ A \mid \neg C \mid C \cap D \mid C \cup D \mid \exists R.C \mid \forall R.C \mid (\geq n)R.C \mid (\leq n)R.C
$$

An axiom is an expression of the form:

<table>
<thead>
<tr>
<th>T-box axioms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$C \sqsubseteq D$</td>
<td>concept inclusion axiom</td>
</tr>
<tr>
<td>$R \sqsubseteq S$</td>
<td>role inclusion axiom</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A-box axioms</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$C(a)$</td>
<td>concept class assertion</td>
</tr>
<tr>
<td>$R(a, b)$</td>
<td>role assertion</td>
</tr>
</tbody>
</table>
Theoretical Framework: DL Syntax

- Examples of axioms:
  - Duck ⊆ ∀Loves.Duck
  - Rich ⊆ ¬Poor
  - BlackCat ⊆ Cat ⊆ ∃hasColour.Black
  - Car ⊆ (≤ 4)hasParts.Wheel

- An ontology (Ω), or logical theory, is a set of axioms.
Theoretical Framework: DL Semantics

A **DL interpretation** is a map between symbols of the logical theory and the reality. Formally it is a pair $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$:

- $\Delta^\mathcal{I}$ is a non-empty set called the **interpretation domain**;
- $\cdot^\mathcal{I}$ is the **interpretation function** that assigns:
  - concepts to subsets of $\Delta^\mathcal{I}$;
  - individuals to elements of $\Delta^\mathcal{I}$;
  - relationships to binary relations in $\Delta^\mathcal{I}$. 
A **DL interpretation** is a map between symbols of the logical theory and the reality. Formally it is a pair $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$:

- $\Delta^\mathcal{I}$ is a non-empty set called the **interpretation domain**;
- $\cdot^\mathcal{I}$ is the **interpretation function** that assigns:
  - concepts to subsets of $\Delta^\mathcal{I}$;
  - individuals to elements of $\Delta^\mathcal{I}$;
  - relationships to binary relations in $\Delta^\mathcal{I}$.

\[
\begin{align*}
\Delta^\mathcal{I} & : \text{people in this room now} \\
\text{student}^\mathcal{I} & \rightarrow \text{SBadaloni} \\
\text{bada}^\mathcal{I} & \rightarrow \text{SBadaloni} \\
\text{amicoDi} & = \{\langle \text{NScattolin, LZenatti}, (\text{LZenatti, ARenier}) \ldots (\text{SGiusto, MRandon}) \rangle \}
\end{align*}
\]
Theoretical Framework: DL Semantics

- We can extend a DL interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ to complex concept expression:

$$
\begin{align*}
(-C)^\mathcal{I} &= \Delta^\mathcal{I} \setminus C^\mathcal{I} \\
(C \cap D)^\mathcal{I} &= C^\mathcal{I} \cap D^\mathcal{I} \\
(C \cup D)^\mathcal{I} &= C^\mathcal{I} \cup D^\mathcal{I} \\
(\exists R.C)^\mathcal{I} &= \{ d \in \Delta^\mathcal{I} \mid \text{for some } (d, d') \in R^\mathcal{I}, \ d' \in C^\mathcal{I} \} \\
(\forall R.C)^\mathcal{I} &= \{ d \in \Delta^\mathcal{I} \mid \text{for all } (d, d') \in R^\mathcal{I}, \ d' \in C^\mathcal{I} \} \\
(\geq n) R.C)^\mathcal{I} &= \{ d \in \Delta^\mathcal{I} \mid \#(\{ d' \in C^\mathcal{I} \mid (d, d') \in R^\mathcal{I} \}) \geq n \} \\
(\leq n) R.C)^\mathcal{I} &= \{ d \in \Delta^\mathcal{I} \mid \#(\{ d' \in C^\mathcal{I} \mid (d, d') \in R^\mathcal{I} \}) \leq n \}
\end{align*}
$$

- An interpretation $\mathcal{I}$ satisfies an axiom $\phi$ if $\phi$ is true under $\mathcal{I}$:

<table>
<thead>
<tr>
<th>T-box axioms</th>
<th>A-box Axioms</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{I} \models C \subseteq D$, iff $C^\mathcal{I} \subseteq D^\mathcal{I}$</td>
<td>$\mathcal{I} \models C(a)$, iff $a^\mathcal{I} \in C^\mathcal{I}$</td>
</tr>
<tr>
<td>$\mathcal{I} \models R \subseteq S$, iff $R^\mathcal{I} \subseteq S^\mathcal{I}$</td>
<td>$\mathcal{I} \models R(a, b)$, iff $\langle a^\mathcal{I}, b^\mathcal{I} \rangle \in R^\mathcal{I}$</td>
</tr>
</tbody>
</table>

- An interpretation $\mathcal{I}$ is a model of $\mathcal{O}$ if it satisfies all its axioms;
- $\mathcal{O}$ imposes constraints of possible states of the world;
- States of the world not consistent with $\mathcal{O}$ are considered impossible.
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 $\Sigma$. 

Diagram:

- **Person**
  - **Animal**: hasParts, isA
  - **Horse**: hasParts, isA
  - **Face**, **Arm**, **Leg**, **Tail**, **Muzzle**

- **Labels**: Face 0.6, Arm 0.7, Art 0.9, Muzzle 0.3, Horse 0.6, Leg 0.7, Pole 0.8, Hair 0.7, Tail 0.8, Leg 0.6, Art 0.8, Leg 0.5, Meadow 0.5
The Partial Model

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

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

![Diagram showing a picture and an ontology model with a partial model overlayed.]
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 $(\mathcal{P}, \mathcal{I}_p, \mathcal{G})$.
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** $S$ assigns a cost to semantically interpreted pictures $(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_O$;
- $S(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_O$ 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 $S$:

$$
\mathcal{I}_p^* = \arg\min_{\mathcal{I}_p \models \mathcal{P}_O} S(\mathcal{P}, \mathcal{I}_p, \mathcal{G})_O \quad \text{subject to} \quad \mathcal{G} \subseteq \Delta \mathcal{I}_p \times S
$$

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