Interprétation d'images

Apports des ontologies et des logiques de description

Jamal Atif - Isabelle Bloch - **Céline Hudelot**

January 30, 2017

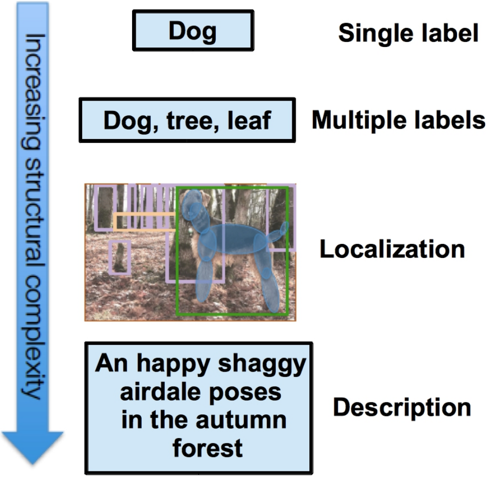
Outline

Semantic image interpretation and annotation

Questions

What is the semantic content of these images? What do they represent?

Semantic image interpretation and annotation



Source : T Berg

Semantic image interpretation and annotation

A hard problem for machines in spite of the increasing performance of sensors and the computing capacities.

Issues [Smeulders 00, Snoek 10]

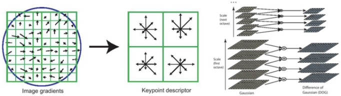
- Sensory gap.
- **Semantic gap**.
- Scaling gap: balance between expressivity/complexity and scaling of models.

Semantic image interpretation and annotation

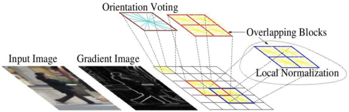
Sensory gap

Image = projection of a reality, often in 3D and continuous, into a discrete and 2D representation.

Numerous advances [Lowe 04, Dalal 05]



SIFT

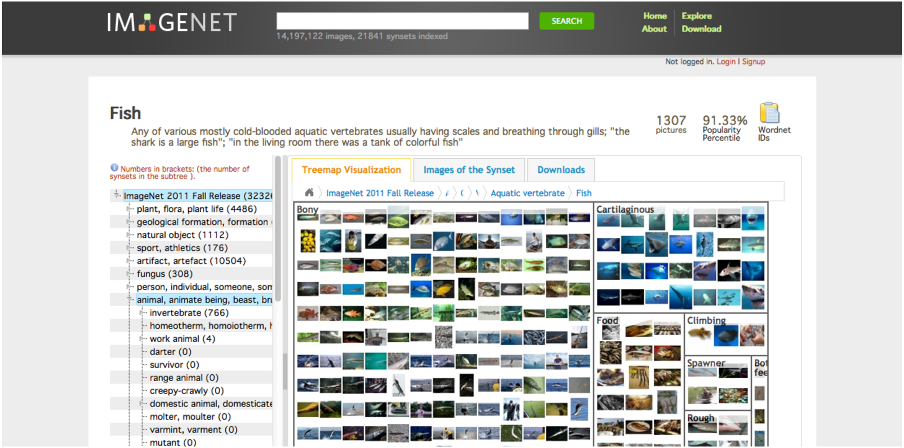


HoG

...

Semantic image interpretation and annotation

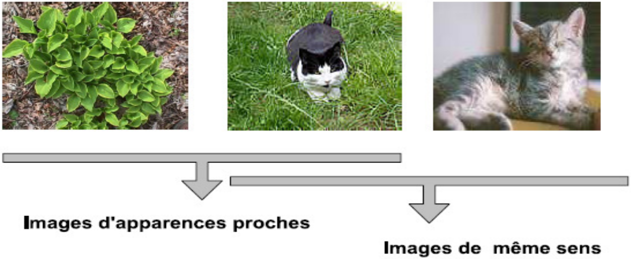
Scale gap



Convolutional Networks (Yann Le Cun) : [Krizhevsky 12, Erhan 14] : challenge ILSVRC : 1000 catégories et 1.461.406 images.

Semantic image interpretation and annotation

Semantic gap



Definition

Lack of coincidence between the information that one can extract from the visual data and the interpretation of these data by a user in a given situation [Smeulders 00].

Known as symbol grounding [Harnad 99] in AI and robotics.

Image and semantics

What is the semantics of this image?

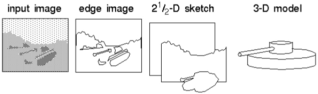
- *A white object on a green background.*
- *An insect.*
- *A white fly on a rose leaf.*

- Image semantics is not inside the image.
- Image interpretation depends on **a priori knowledge.**
- Image interpretation depends on the user objectives.
- Importance of contextual and structural information.

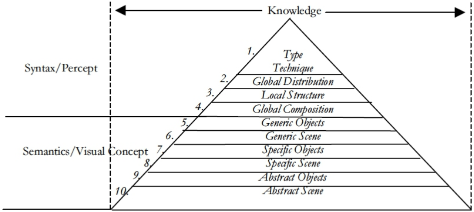
Image and semantics

A multi-level paradigm

Since the early years of CV



D. Marr hierarchy [Marr 82]



Semantic pyramid [Jaimes 00]

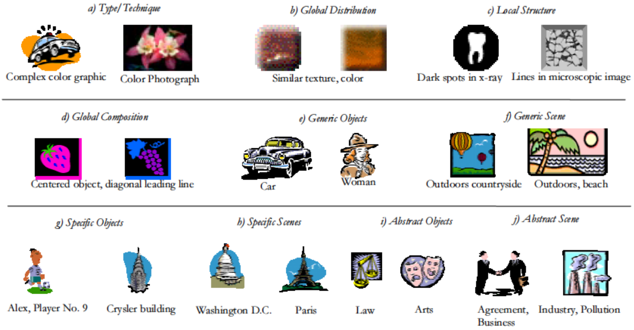
Niveau de la scène

Générique : Paysage de montagne, rallye
Spécifique : Chypre
Abstrait : Sport, Divertissement

Niveau de l'objet

Générique : voiture, voiture de rallye
Spécifique : citroen de Sebastien Loeb

Image and semantics



Jaimes et al.

Image and semantics

A multi-level paradigm

Even in the recent representation learning with deep learning approaches.

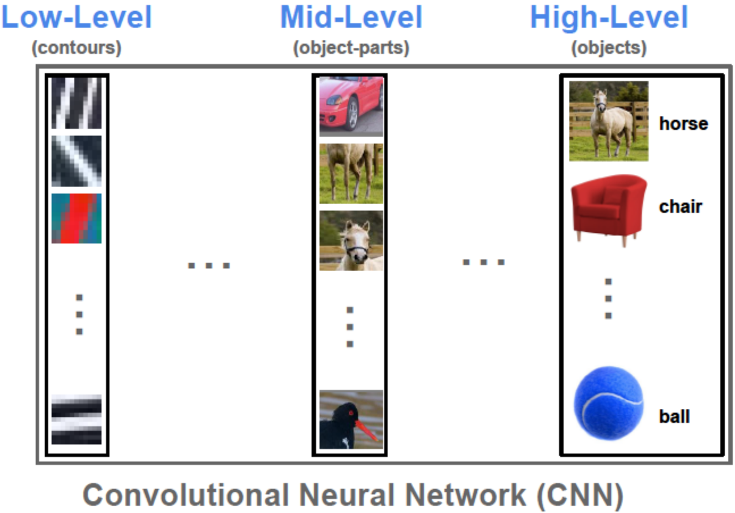


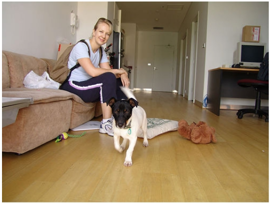
Image and semantics

Several semantics acceptations: from object semantics to structural description semantics.

→

Car: present
Cow: present
Bike: not present
Horse: not present
...

[Duygulu 02, Barnard 03, Lavrenko 03, Djeraba 03, Carneiro 07, Liu 07, Deng 10]



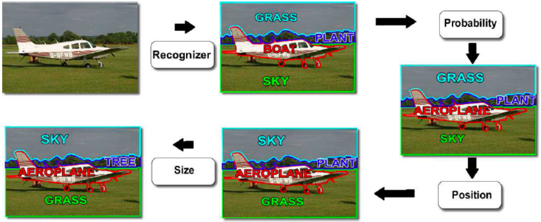
This is a photograph of one person and one brown sofa and one dog. The person is against the brown sofa. And the dog is near the person, and beside the brown sofa."

[Yao 10, Kulkarni 11, Farhadi 10, Farhadi 13, Karpathy 14]

Image and semantics

Importance of contextual and sptatial information

Source : [Parikh 12]



Source : [Galleguillos 10]

Importance of spatial relations in image interpretation

- Spatial reasoning
- Carry an important structural information
- More stable and reliable than object features

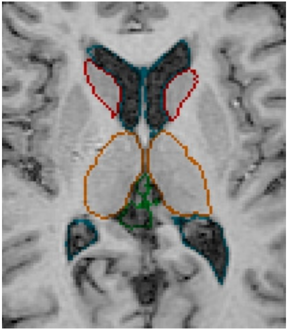


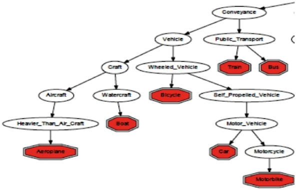
Image and semantics

Importance of prior knowledge

Semantics = a property that emerges from the interaction between data and knowledge [Hanson 78, Santini 01, Hudelot 03].

Connaissances implicites

Connaissances explicites



⇒ Interest of ontologies

Outline

What is an ontology ?

Example from F. Gandon, WIMMICS Team, INRIA

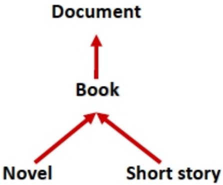
What is the last document that you have read?

Documents

Your answer is based on a **shared** **ontology**

You can reason

I can understand



Ontologies: Definition

Ontology

ethymology: ontos (being, that which is) + logos (science, study, theory)

- Philosophy
  - Study of the nature of being, becoming and reality.
  - Study of the basic categories of being and their relations.
- Computer Science
  - Formal representation of a domain of discourse.
  - Explicit specification of a conceptualization [Gruber 95].

Ref: [Guarino 09]

Ontologies: Definition

ontology

Formal, explicit (and shared) specification of a conceptualization [Gruber 95, Studer 98]

- Formal, explicit specification:
  - a formal language is used to refer to the elements of the conceptualization, e.g. description logics
- Conceptualization:
  - Objects, concepts and other entities and their relationships

Concept

Denoted by:

- a name
- a meaning (intensional definition)
- a set of denoted objects (extensional definition)

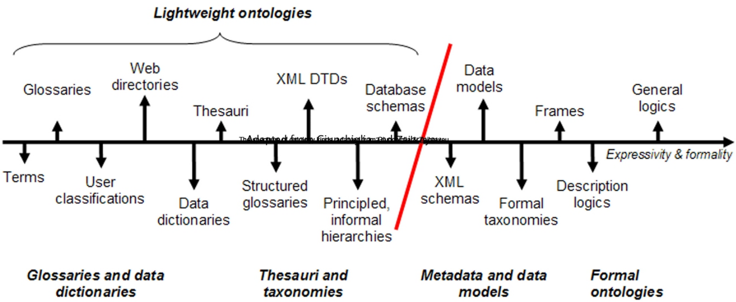
Relation

Denoted by:

- a name
- an intension
- an extension

The different types of ontologies

According to their expressivity



Source : [Uschold 04]

The different types of ontologies

According to their abstraction level

- **Top (or Upper)-level ontology**: very general concepts that are the same across all knowledge domains [Wikipedia] (e.g. DOLCE).
- **Core ontology**: minimal set of concepts and relations used to structure and describe a given domain (e.g. Dublin Core).
- **Domain ontology**: concepts and relations of a specific domain (e.g. FMA).

Outline

Ontologies for image interpretation

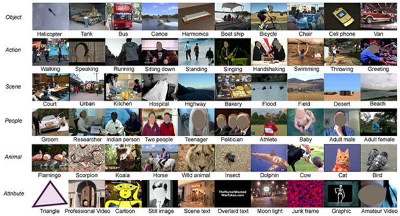
A growing interest since 2001

Various objectives:

- Providing an **unified vocabulary** for the description and annotation of image content.
  - e.g. MPEG-7 ontologies.
- **Structuring** the vocabulary and the database for **large-scale** image problems.
  - e.g. visual ontologies (LabelMe, ImageNet, Visipedia).
- Representing the application domain knowledge for **reasoning** and for **guiding** the interpretation process.
  - e.g. formal ontologies based on description logics.

Ontologies for an unified and standardized description of image content

- MPEG-7 ontologies: Boemie, AceMedia, Rhizomik... (see [Dasiopoulou 10b] for a recent review).
  Main motivation: interoperability between applications.
- LSCOM (Large Scale concept ontology for multimedia) [Naphade 06], MediaMill [Habibian 13].
  Main motivation: common vocabulary for video shot description.

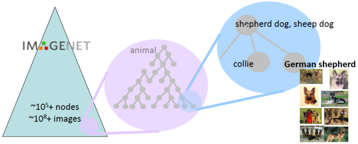


Mainly focused on the descriptive part of ontologies.

Ontologies for structuring the vocabulary and the learning database (1/3)

Main motivation : image classification, annotation and retrieval at large scale [Liu 07, Deng 10].

- Ontologies based on lexical resources (e.g. Wordnet) populated with images:
    - ImageNet [Russakovsky 15], LabelMe [Russell 08], Visipedia [Belongie 16], Visual Genome [Krishna 16]...



ImageNet

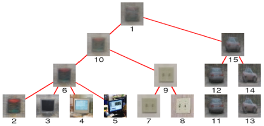
Which concepts are closer ?

- Adequacy of the lexical resources for image interpretation problems ?
- Mainly lightweight ontologies (non-formal, without reasoning capabilities).

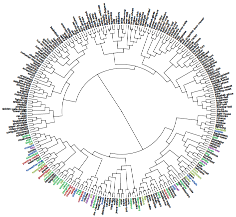
Ontologies for structuring the vocabulary and the learning database (2/3)

Main motivation: hierarchical image classification.

- Visual concept hierarchies inferred from image datasets: [Fei-Fei 05, Marszalek 08, Griffin 08, Sivic 08, Bart 08, Gao 11].



[Sivic 08]



[Griffin 08]

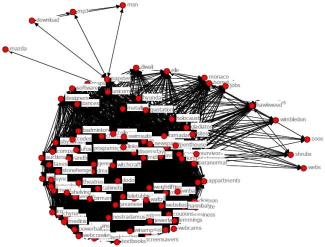
- Mainly hierarchies (no other semantic relations than *is-a*).
- Concepts without semantics (except the leaves).
- Mainly lightweight ontologies (non-formal, without reasoning capabilities).

Ontologies for structuring the vocabulary and the learning database (3/3)

Main motivation: image classification and annotation.

- Ontologies combining text and visual knowledge: [Li 10, Wu 12, Bannour 14, Krishna 16].

Image hierarchy [Li 10]



VCNet [Wu 12]

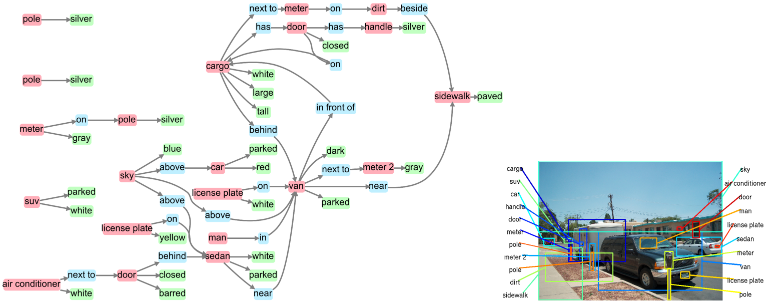
- Dedicated knowledge models.
- Mainly lightweight ontologies (non-formal, without reasoning capabilities).

Ontologies for image captionning

Main motivation: image captionning.

More and more approach, under the dynamics of image captioning to represent objects, attributes of objects and relationships between objects: Scene graphs [**?**], Visual Genome [Krishna 16], Visipedia [Belongie 16]

Scene graphs



Ontologies for image captionning

Main motivation: image captionning.

More and more approach, under the dynamics of image captioning to represent objects, attributes of objects and relationships between objects: Scene graphs [?], Visual Genome [Krishna 16], Visipedia [Belongie 16]

Visual Genome

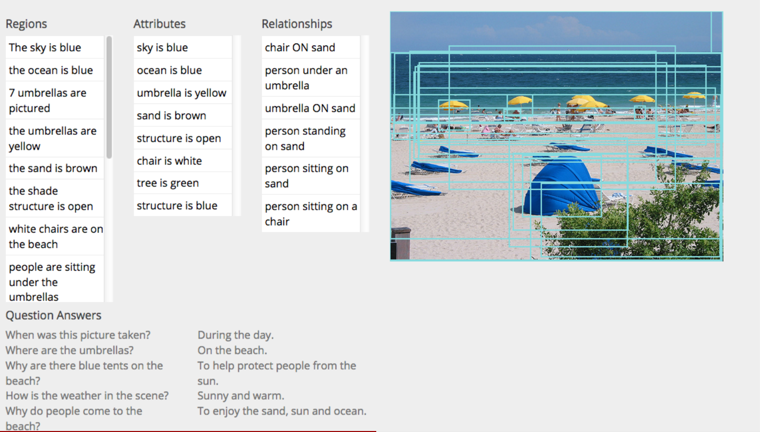


Image interpretation as an ontological driven inference approach

Main motivation : explicit and formal representation of domain and contextual knowledge used to reason and infer the interpretation.

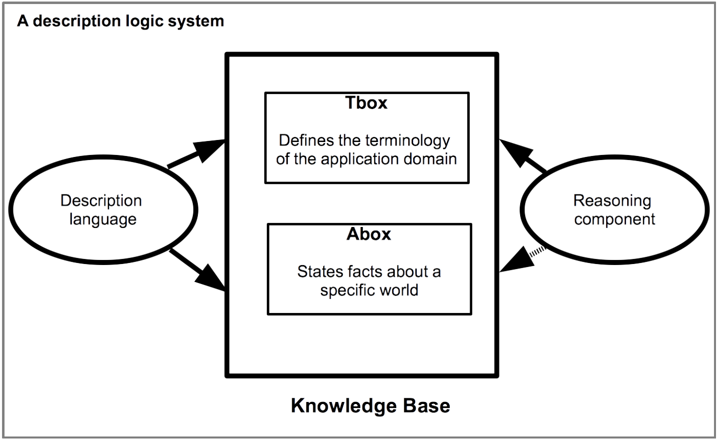
- Annotation and interpretation refinement using basic DLs inference services: [Simou 08, Dasiopoulou 09, Dasiopoulou 10a, Bannour 14].
- *Ontologies* to narrow the semantic gap: [Town 06, Bagdanov 07, Hudelot 08]
- Image interpretation as a non-monotonic reasoning process:
  - Image interpretation as a default reasoning service [Möller 99a, Neumann 08].
  - Abductive reasoning for image interpretation [Peraldi 07, Möller 08, Atif 14, Donadello 14].

Often based on Description Logics (DLs).

Outline

Descriptions logics

- Family of logics for representing structured knowledge.
- Well understood semantics.
- Defined by a set of concepts and role forming operators.
- Compact and expressive and basis of OWL language to represent ontologies.



Description logics : the description language

Syntax of $\mathcal{ALC}$:attributive language with complement

basic language $\mathcal{AL}$ + constructors ($\mathcal{C}$ for the complement $\neg$ operator)

- Signature $\Sigma = (N_C, N_R)$, disjoint sets of concept names and role names respectively.
- Concept descriptions in $\mathcal{ALC}$ are formed according to the following syntax rule:

$$
\begin{aligned}
C, D \longrightarrow & A \mid && \text{(atomic concepts)}\\
& \top \mid && \text{(universal concept)}\\
& \bot \mid && \text{(bottom concept)}\\
& \neg C \mid && \text{(negation)}\\
& C \sqcap D \mid && \text{(conjunction)}\\
& C \sqcup D \mid && \text{(disjunction)}\\
& \forall r.C \mid && \text{(value restriction)}\\
& \exists r.C \mid && \text{(existential restriction).}
\end{aligned}
$$

$A \in N_C$ and $r \in N_R$

Description logics : the description language

Examples of $\mathcal{ALC}$-concept descriptions

- Atomic concepts: $Person, Female, Tutorial, Boring$
- Atomic role: $attends$
- $\mathcal{ALC}$-descriptions:

$$Person \sqcap Female$$

$$Person \sqcap \neg Female$$

$$Person \sqcap \exists attends.Tutorial$$

$$Person \sqcap \forall attends.(Tutorial \sqcap \neg Boring)$$

Description logics : the description language

Semantics of $\mathcal{ALC}$: attributive language with complement

An interpretation $\mathcal{I} = \langle \Delta^{\mathcal{I}}, .^{\mathcal{I}} \rangle$

- $\Delta^{\mathcal{I}}$: a non-empty set, the domain of interpretation
- $.^{\mathcal{I}}$: an interpretation function, which assigns to :
  - every atomic concept $A \in N_C$, a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$,
  - every atomic role $r \in N_R$, a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

Extension to concept descriptions

$$
\begin{aligned}
\top^{\mathcal{I}} &= \Delta^{\mathcal{I}} \\
\bot^{\mathcal{I}} &= \emptyset \\
(\neg C)^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(C \sqcup D)^{\mathcal{I}} &= C^{\mathcal{I}} \cup D^{\mathcal{I}} \\
(\forall r.C)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \forall b.(a,b) \in r^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\} \\
(\exists r.C)^{\mathcal{I}} &= \{a \in \Delta^{\mathcal{I}} \mid \exists b.(a,b) \in r^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}
\end{aligned}
$$

The basic description language $\mathcal{AL}$

Semantics

Equivalence:

$$C \equiv D \text{ if } C^{\mathcal{I}} = D^{\mathcal{I}} \text{ for all interpretations } \mathcal{I}$$

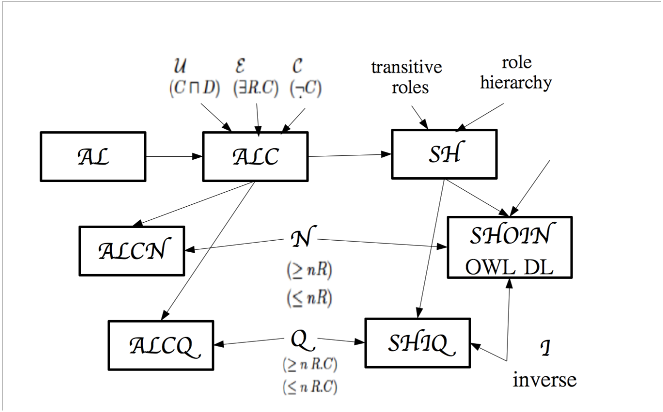
Example

$\forall hasChild.Female \sqcap \forall hasChild.Student$ and $\forall hasChild.(Female \sqcap Student)$ are equivalent.

The family of $\mathcal{AL}$ languages

$$\mathcal{AL}[\mathcal{U}][\mathcal{E}][\mathcal{C}][\mathcal{N}][\mathcal{Q}],\cdots$$

Many additional constructors have been introduced.



The family of $\mathcal{AL}$ languages

$\mathcal{ALEN}$ example

$$Person \sqcap (\leq\ 1\ hasChild \sqcup (\geq\ 3\ hasChild \sqcap \exists hasChild.Female))$$

Description logics : terminological knowledge

Terminological axioms

- General Concept Inclusion (GCI)

$$C \sqsubseteq D$$

  C, D are concept descriptions

- Concept definition[a]

$$A \equiv C$$

  A a concept name, C a concept description

[a] abbreviation for $A \sqsubseteq C$ and $C \sqsubseteq A$

TBox

A TBox is a finite set of GCIs

Description logics : terminological knowledge

- An interpretation $\mathcal{I}$ satisfies a GCI $C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$

$$\mathcal{I} \models (C \sqsubseteq D) \Leftrightarrow C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$$

- An interpretation $\mathcal{I}$ satisfies an equality $C \equiv D$ if $C^{\mathcal{I}} \equiv D^{\mathcal{I}}$

$$\mathcal{I} \models (C \equiv D) \Leftrightarrow C^{\mathcal{I}} \equiv D^{\mathcal{I}}$$

- The interpretation $\mathcal{I}$ is a model of a TBox $\mathcal{T}$ iff it satisfies all the GCIs in $\mathcal{T}$
- Two TBoxes are equivalent if they have the same model.

Description logics : terminological knowledge

TBox example

$$
\begin{aligned}
Woman &\equiv Person \sqcap Female \\
Man &\equiv Person \sqcap \neg Woman \\
Mother &\equiv Woman \sqcap \exists hasChild.Person \\
Father &\equiv Man \sqcap \exists hasChild.Person \\
Parent &\equiv Father \sqcup Mother \\
Grandmother &\equiv Mother \sqcap \exists hasChild.Parent \\
MotherWithManyChildren &\equiv Mother \sqcap \geq 3 hasChild
\end{aligned}
$$

Description logics : assertional knowledge

Assertional axioms

- Concept assertion : $C(a)$
- Role assertion : $R(a, b)$
  C a concept description , a, b are individuals names from a set N_I

ABox

An ABox is a finite set of assertions

Interpretation

- Given $\mathcal{I}$, each individual a is mapped to an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
- Unique name assumption: $a^{\mathcal{I}} \neq b^{\mathcal{I}}$
- $\mathcal{I}$ is a model of the ABox $\mathcal{A}$ if it satisfies all its assertions:
  - $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all $C(a) \in \mathcal{A}$
  - $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R$ if for all $R(a, b) \in \mathcal{A}$

Description logics : knowledge base

Knowledge base

A knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$.

The interpretation $\mathcal{I}$ is a model of the knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ iff it is a model of $\mathcal{T}$ and a model of $\mathcal{A}$.

Description logics for knowledge representation

Example in the medical domain

Knowledge in brain imaging

- caudate nucleus: a deep gray nucleus of the telencephalon involved with control of voluntary movement
- the left caudate nucleus is inside the left hemisphere
- it is close to the lateral ventricle
- it is outside (left of) the left lateral ventricle

Excerpt of a corresponding TBox

- $\text{AnatomicalStructure} \sqsubseteq \text{SpatialObject}$
- $\text{LV} \sqsubseteq \text{AnatomicalStructure}$
- $\text{GN} \sqsubseteq \text{AnatomicalStructure}$
- $\text{CN} \sqsubseteq \text{GN}$
- $\text{LV} \equiv \text{RLV} \sqcup \text{LLV}$
- $\text{CN} \equiv \text{RCN} \sqcup \text{LCN}$
- $\text{LCN} \equiv \text{GN} \sqcap \exists\, \text{closeTo.(LLV)} \sqcap \exists\, \text{leftOf.(LLV)}$
- etc.

Description logics: exercice 1

On the blackboard

Students are Persons. Students are following some UEs. Curious students are following the UE Image Interpretation. Axel is a curious student. Persons are descendants of others persons.

1. What are the concepts, the relations, the instances ?
2. Write the corresponding TBox with $\mathcal{ALCN}$.
3. Write the corresponding ABox with $\mathcal{ALCN}$.

Description logics: exercice 2

On the blackboard

We consider the following interpretation

- $\Delta^{\mathcal{I}} = \{t_1, t_2, f_1, f_2, c_1, c_2, j, k, l, m, n\}$
- $Person^{\mathcal{I}} = \{j, k, l, m, n\}$
- $Car^{\mathcal{I}} = \{t_1, t_2, f_1, f_2, c_1, c_2\}$
- $Ferrari^{\mathcal{I}} = \{f_1, f_2\}$
- $Toyota^{\mathcal{I}} = \{t_1, t_2\}$
- $likes^{\mathcal{I}} = \{(j, f_1), (k, f_1), (k, t_2), (l, c_1), (l, c_2), (m, c_1), (m, t_2), (n, f_2), (n, c_2)\}$

Give the interpretations of :

- $\exists likes.Ferrari \sqcap \exists likes.Toyota$
- $\exists likes.Ferrari \sqcap \forall likes.Toyota$
- $\exists likes.Ferrari \sqcap \exists likes.\neg Ferrari$
- $\exists likes.Cars \sqcap \forall likes.\neg(Toyota \sqcup Ferrari)$

Description logics: concrete domains

- A way to integrate *concrete and quantitative qualities* (integers, strings,...) of real world objects with conceptual knowledge [Baader,91].
- A pair (Δ_D, Φ_D) where Δ_D is a set and Φ_D a set of predicates names on Δ_D. Each predicate name P is associated with an arity n and an n-ary predicate $P^{\mathcal{D}} \subseteq \Delta_D^n$.

Examples

- Concrete domain $\mathcal{N}$:
  - Domain: non negative integers.
  - Predicates: $\leq$ (binary predicate) $\leq n$ unary predicate.
  - $\mathsf{Person} \sqcap \exists \mathsf{age}. \leq 20$ denotes a person whose age is less than 20.
- Concrete domain $\mathcal{AL}$, Allen's interval calculus:
  - Domain: intervals.
  - Predicates: built from Allen's basic interval relations.

Description logics: reasoning services

$\Longrightarrow$ Infer implicit knowledge from explicitly one.

- Terminological reasoning.
- Assertional reasoning.

Description logics: reasoning services

Terminological reasoning

Satisfiability

C is satisfiable w.r.t. a TBox $\mathcal{T}$ iff $C^{\mathcal{I}} \neq \emptyset$ for some model $\mathcal{I}$ of $\mathcal{T}$.

Subsumption

C is subsumed by D w.r.t. a TBox $\mathcal{T}$ ($C \sqsubseteq_{\mathcal{T}} D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$.

Equivalence

C is equivalent to D w.r.t. a TBox $\mathcal{T}$ ($C \equiv_{\mathcal{T}} D$) iff $C^{\mathcal{I}} = D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$.

Disjointness

Two concepts C and D are disjoint with respect to $\mathcal{T}$ if $C^{\mathcal{I}} \cap D^{\mathcal{I}} = \emptyset$ for every model $\mathcal{I}$ of $\mathcal{T}$.

Reduction to subsumption

For concepts C, D we have

- C is unsatisfiable $\iff$ C is subsumed by $\bot$;
- C and D are equivalent $\iff$ C is subsumed by D and D is subsumed by C;
- C and D are disjoint $\iff$ $C \sqcup D$ is subsumed by $\bot$.

The statements also hold with respect to a TBox.

Reduction to Unsatisfiability

For concepts C, D we have

- C is subsumed by $D \iff C \sqcap \neg D$ is unsatisfiable;
- C and D are equivalent $\iff$ both $C \sqcap \neg D$ and $\neg C \sqcap D$ are satisfiable;
- C and D are disjoint $\iff C \sqcap D$ is unsatisfiable.

The statements also hold with respect to a TBox.

Reducing Unsatisfiability

Let C be a concept. Then the following are equivalent:

- C is unsatisfiable;
- C is subsumed by $\bot$;
- C and $\bot$ are equivalent;
- C and $\bot$ are disjoint.

The statements also hold with respect to a TBox.

Description logics: reasoning services

Assertional reasoning

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an ontology.

Consistency

$\mathcal{A}$ is consistent with respect to a TBox $\mathcal{T}$, if there is an interpretation that is a model of both $\mathcal{A}$ and $\mathcal{T}$.

Instance checking

a is an instance of C w.r.t. $\mathcal{T}$ iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$. We also write $\mathcal{A} \models C(a)$. The same holds for roles.

Retrieval problem

Given an ABox $\mathcal{A}$ and a concept C, find all individuals a such that $\mathcal{A} \models C(a)$.

Realization problem (dual to the retrieval problem)

Given an individual a and a set of concepts, find *the most specific concepts* (msc) C from the set such that $A \models C(a)$. The mscs are the concepts that are minimal with respect to the subsumption ordering $\sqsubseteq$.

Reduction

- $\mathcal{A} \models C(a)$ iff $\mathcal{A} \cup \{\neg C(a)\}$ is inconsistent;
- C is satisfiable iff $\{C(a)\}$ is consistent.

Subsumption checking

- Structural subsumption
- Semantic tableaux
- etc.

Open world, Closed world

Closed World Assumption

Limitations to what is expressed

- example : ABox : $hasChild(anne, paul)$
- anne has only one child : paul

Open World Assumption: description logics

Open world : no limitations to what is expressed

- example : ABox : $hasChild(anne, paul)$
- anne can have other child than paul
- $(\leq 1 hasChild)(anne)$

Tableau based reasoning

Principle

To prove F : build a tree with :

- The root is labeled with $\neg F$.
- The nodes are labeled by the concepts.
- Node successors are built par some expansion rules.
- A clash at the end of a path if :
  - $C(x) \in \mathcal{A}$ and $\neg C(x) \in \mathcal{A}$
  - $C(x) \in \mathcal{A}$ and $\neg C(y) \in \mathcal{A}$ and $(x = y$ or $y = x)$
  - $\perp(x) \in \mathcal{A}$

Tableau based reasoning

$\sqcap$ rule

Conditions

$\mathcal{A}$ contains $(C_1 \sqcap C_2)(x)$ and does not contain $C_1(x)$ and $C_2(x)$

Action

Prolongation : $\mathcal{A}^{'} = \mathcal{A} \cup \{C_1(x), C_2(x)\}$

Tableau based reasoning

$\sqcup$ rule

Conditions

$\mathcal{A}$ contains $(C_1 \sqcup C_2)(x)$ and does not contain $C_1(x)$ and $C_2(x)$

Action

Branching: $\mathcal{A}^{'} = \mathcal{A} \cup \{C_1(x)\}$ and $\mathcal{A}^{''} = \mathcal{A} \cup \{C_2(x)\}$

Tableau based reasoning

$\exists$ rule

Conditions

$\mathcal{A}$ contains $(\exists R.C)(x)$ and there is no individual z such as $R(x,z)$ and $C(z)$ are also in $\mathcal{A}$

Action

$\mathcal{A}' = \mathcal{A} \cup \{R(x,y), C(y)\}$ where y is an individual name which is not in $\mathcal{A}$

Tableau based reasoning

$\forall$ rule

Conditions

$\mathcal{A}$ contains $(\forall R.C)(x)$ and $R(x, y)$ but does not contain $C(y)$

Action

$\mathcal{A}' = \mathcal{A} \cup \{C(y)\}$

Outline

1. Image and semantics
2. What is an ontology ?
3. Ontologies for image understanding: overview
4. Description Logics
5. Description Logics for image understanding
   - Ontologies for interpretation refinement
   - Narrowing the semantic gap
   - Non-monotonic reasoning for image interpretation
     - Default reasoning
     - Abductive reasoning
6. Conclusion

Interpretation refinement using basic DLs inference services

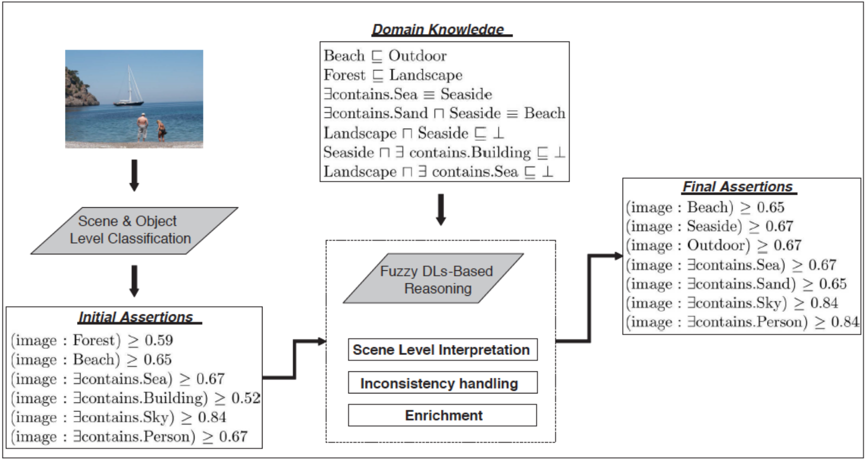
Main principles

- Application domain knowledge is encoded into a TBox.
- A first interpretation of the targeted image is built using computer vision algorithms and translated into ABox assertions.
- Basic reasoning services of DLs such as consistency handling are used to revise the interpretation.
- Fuzzy DLs are used to take into account the imprecision of computer vision algorithms results.

Investigating fuzzy DLs-based reasoning in semantic image analysis [Dasiopoulou 10a].
Building and using fuzzy multimedia ontologies for semantic image annotation [Bannour 14].

Interpretation refinement using basic DLs inference services

Dasiopoulou et al. [Dasiopoulou 10a]



Outline

Narrowing the semantic gap

Main approaches

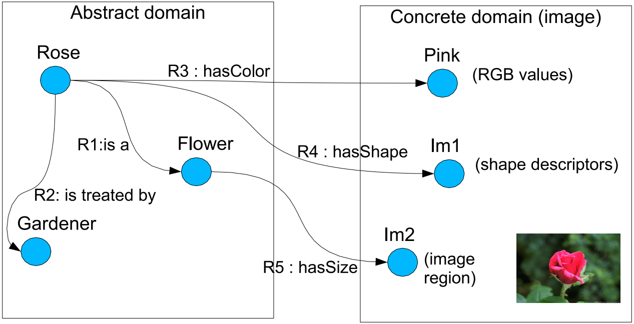
- Building a dedicated visual concept ontology as an intermediate level between image features and application domain concepts:
  [Town 06, Bagdanov 07, Maillot 08, Porello 13, Mezaris 04].
- Using concrete domains to link high level concepts to their specific representations into the image domain:
  [Hudelot 08, Hudelot 14].
  $\Rightarrow$ *operational ontologies for image interpretation.*

A spatial relation ontology for semantic image interpretation

Hudelot et al. [Hudelot 08, Hudelot 14]

Ontologies, concrete domains and semantic gap

Hudelot et al. [Hudelot 08]

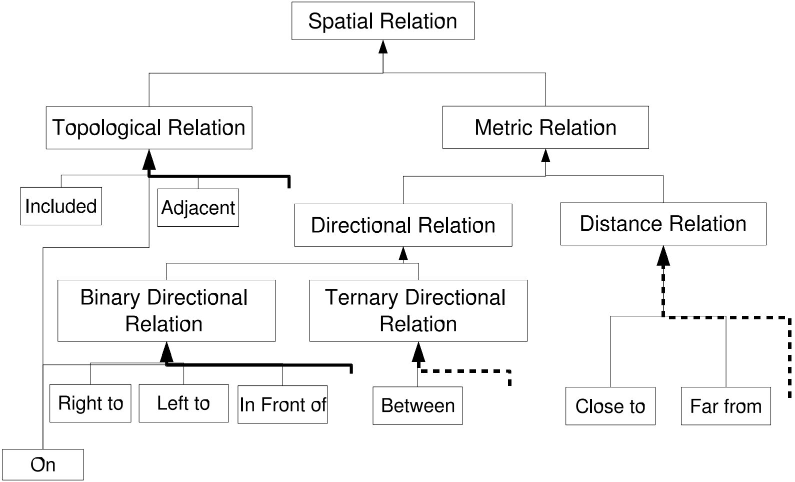


Idea

Each application domain concept is linked to its representation in the image domain: use of concrete domains.

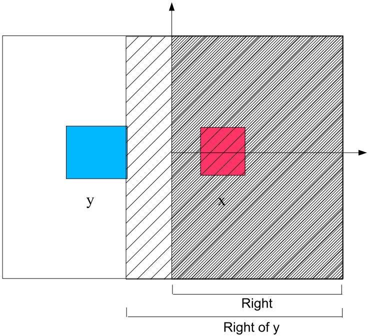
A spatial relation ontology

Hudelot et al. [Hudelot 08]



Formal representation of spatial relations

Hudelot et al. [Hudelot 08]



x is to the right of y: true

Abox:

- y:SpatialObject; x:SpatialObject
- Right_Of_y $\equiv$ Right_Of $\sqcap$ $\ni$hasReferentObject.{y}
- x:SpatialObject $\sqcap$ $\exists$ hasSpatialRelation.Right_Of_y and x:SpatiallyRelatedObject
- C_0 $\equiv$ SpatialRelation $\sqcap$ $\ni$hasReferentObject.{y} $\sqcap$ $\ni$hasTargetObject.{x}

A dedicated logic for spatial reasoning: $\mathcal{ALC}(\mathbf{F})$

Instantiation of the description logic $\mathcal{ALCRP}(D)$ with the concrete domain $\mathbf{F} = (\Delta_{\mathbf{F}}, \Phi_{\mathbf{F}})$.

$\Delta_{\mathbf{F}} = (\mathcal{F}, \leq_{\mathcal{F}}, \wedge, \vee, \emptyset_{\mathcal{F}}, 1_{\mathcal{F}}, t, I)$

A residuated lattice of fuzzy sets defined over the image space $\mathcal{S}$, $\mathcal{S}$ being typically $\mathbb{Z}^2$ or $\mathbb{Z}^3$ for 2D or 3D images, with t a t-norm (fuzzy intersection) and I its residuated implication.

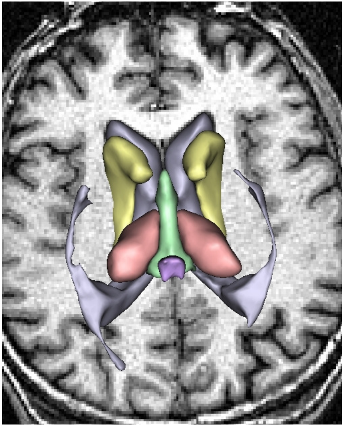
Main predicates of $\Phi_{\mathbf{F}}$:

- μ_X: degree of belonging to the spatial representation of the object X in the spatial domain.
- ν_R: fuzzy structuring element representing the fuzzy relation R in the spatial domain.
- $\delta_{\nu_R}^{\mu_X}$: fuzzy dilation.
- $\varepsilon_{\nu_R}^{\mu_X}$: fuzzy erosion.

Application to brain imaging

Objective:

Progressive recognition of anatomical structures using spatial information.

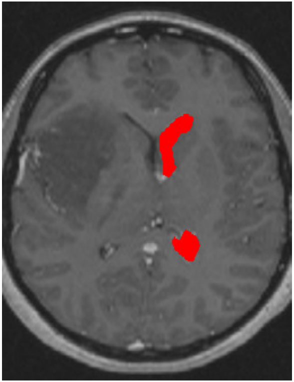


Description of anatomical knowledge

Tbox:

- AnatomicalStructure $\sqsubseteq$ SpatialObject
- GN $\sqsubseteq$ AnatomicalStructure
- RLV $\equiv$ AnatomicalStructure $\sqcap \exists$ hasFR.μ_{RLV}
- LLV $\equiv$ AnatomicalStructure $\sqcap \exists$ hasFR.μ_{LLV}
- LV $\equiv$ RLV $\sqcup$ LLV
- LV $\equiv$ RLV $\sqcup$ LLV
- Right_of $\equiv$ DirectionalRelation $\sqcap \exists$ hasFR.$\nu_{IN\_DIRECTION\_0}$
- Close_to $\equiv$ DistanceRelation $\sqcap \exists$ hasFR.ν_{CLOSE_TO}
- Right_of_RLV $\equiv$ DirectionalRelation $\sqcap \exists$ hasReferentObject.RLV $\sqcap \exists$ hasFR.$\delta^{\mu_{RLV}}_{\nu_{IN\_DIRECTION\_0}}$
- Close_To_RLV $\equiv$ DistanceRelation $\sqcap \exists$ hasReferentObject.RLV $\sqcap \exists$ hasFR.$\delta^{\mu_{RLV}}_{\nu_{CLOSE\_TO}}$
- RCN $\equiv$ GN $\sqcap \exists$ hasSR.(Right_of_RLV $\sqcap$ Close_To_RLV)
- CN $\equiv$ GN $\sqcap \exists$ hasSR.(Close_To_LV)
- CN $\equiv$ RCN $\sqcup$ LCN

Example



Abox:

- c_1: RLV , (c_1,μ_{S_1}): hasFR
- r_1: Right_of, $(r_1,\nu_{IN\_DIRECTION\_0})$: hasFR
- r_2: Close_to, (r_2,ν_{CLOSE_TO}): hasFR

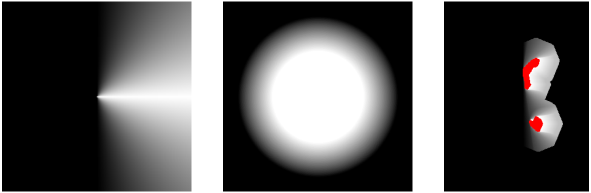
Example

Objective:

- Find some spatial constraints in the image domain on an instance $c2$ of the Left Caudate Nucleus.
- $\Rightarrow$ Find constraints on concrete domains to ensure the satisfiability of the assertions c_2: RCN, (c_2,μ_{S_2}): hasFR

Results using inference and properties

$(\mu_{S_2})^{\mathbf{F}} \leq_{\mathcal{F}} (\delta^{\mu_{S_1}}_{\nu_{IN\_DIRECTION\_0}})^{\mathbf{F}} \wedge (\delta^{\mu_{S_1}}_{\nu_{CLOSE\_TO}})^{\mathbf{F}}$



Inference details:

$$\mathcal{A} \cup \{c_2 : \mathsf{GN} \sqcap \exists \mathsf{hasSR}.(\mathsf{Right\_of\_RLV} \sqcap \mathsf{Close\_to\_RLV}), (c_2, \mu_{S_2}) : \mathsf{hasFR}\}$$

$$\Big\downarrow \sqcap - rule$$

$$c_2 : \mathsf{GN}, c_2 : \exists \mathsf{hasSR}.(\mathsf{Right\_of\_RLV} \sqcap \mathsf{Close\_to\_RLV})$$

$$\Big\downarrow \exists - rule$$

$$c_3 : \mathsf{Right\_of\_RLV} \sqcap \mathsf{Close\_to\_RLV}, (c_2, c_3) : \mathsf{hasSR}, (c_3, \mu_{S_3}) : \mathsf{hasFR}$$

$$\Big\downarrow \text{Spatial Object Conjunction Rule} \mathcal{R}_\sqcap$$

$$((\mu_{\text{Right\_of\_RLV}}) \sqcap_d (\mu_{\text{Close\_to\_RLV}}))^{\mathbf{F}}$$

$$\Big\downarrow \text{Spatial Object Conjunction Rule} \mathcal{R}_\sqcap$$

$$c_3 : \mathsf{Right\_of\_RLV}, c_3 : \mathsf{Close\_to\_RLV}$$

$$\Big\downarrow \text{Spatial Relation Rule} \mathcal{R}2_{R_X}$$

$$\mu_{S_3} = \delta^{\mu_{S_1}}_{\nu_{IN\_DIRECTION\_0}} \sqcap_d \delta^{\mu_{S_1}}_{\nu_{CLOSE\_TO}}$$

$$\Big\downarrow \text{spatial constraints}$$

$$fit(\mu^{\mathbf{F}}_{S_2}, \mu^{\mathbf{F}}_{S_3}) = fit(\mu^{\mathbf{F}}_{S_2}, (\delta^{\mu_{S_1}}_{\nu_{IN\_DIRECTION\_0}} \sqcap_d \delta^{\mu_{S_1}}_{\nu_{CLOSE\_TO}})^{\mathbf{F}}) = 1$$

$$\Big\downarrow \text{spatial constraints}$$

$$(\mu_{S_2})^{\mathbf{F}} \leq_{\mathcal{F}} (\delta^{\mu_{S_1}}_{\nu_{IN\_DIRECTION\_0}})^{\mathbf{F}} \wedge (\delta^{\mu_{S_1}}_{\nu_{CLOSE\_TO}})^{\mathbf{F}}$$

Outline

1. Image and semantics
2. What is an ontology ?
3. Ontologies for image understanding: overview
4. Description Logics
5. Description Logics for image understanding
   - Ontologies for interpretation refinement
   - Narrowing the semantic gap
   - Non-monotonic reasoning for image interpretation
     - Default reasoning
     - Abductive reasoning
6. Conclusion

Non-monotonic reasoning for image interpretation

Main principles:

Image interpretation is modeled as a non-monotonic reasoning process.

- Default reasoning:
  Non-monotonic logic to formalize reasoning with default assumptions [Reiter 80].
- Abductive reasoning:
  Backward reasoning: from *observations* to *explanations*, Charles Sanders Peirce in the late 19th century.

Image interpretation as a default reasoning service

Default rule

$$\frac{\alpha : \beta_1, \cdots, \beta_n}{\gamma}$$

- α: precondition of the rule.
- β_i: justifications.
- γ: consequent.

Intuitive explanation

Starting with a world description α of what is known to be true, i.e. deducible and it is consistent to assume β_i then conclude γ.

Example

$\forall x, \mathit{plays\_instruments}(x) : \mathit{improvises}(x) / \mathit{jazz\_musician}(x)$[a]

[a] For all x, if x plays an instrument and if the fact that x can improvise is consistent with all other knowledge then we can conclude that x is a jazz musician.

Default reasoning in DL

Reiter's default theory [Reiter 80]

A pair $(\mathcal{W}, \mathcal{D})$ where $\mathcal{W}$ is a set of closed first-order formulae (the world description) and $\mathcal{W}$ a set of default rules.

Terminological default theory [Baader 92]

A pair $(\mathcal{A}, \mathcal{D})$ where:

- $\mathcal{A}$: an ABox.
- $\mathcal{D}$: a finite set of default rules whose preconditions, justifications and consequents are concept terms.

Maintaining decidability

- Default rules have to be closed over the ABox (instanciation with explicitly mentioned ABox individuals).
- Closed default rules: α, β_i, γ are ABox concept axioms (no use of free variables, i.e. TBox concept axioms).

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, Neumann 08]

Use case: topological reasoning for aerial image interpretation

Main idea

- Defaults are used for hypothesis generation regarding the classification of areas in an image.
- Default reasoning generates ABox extensions (hypothesized classifications) consistent with the rest of the knowledge base.

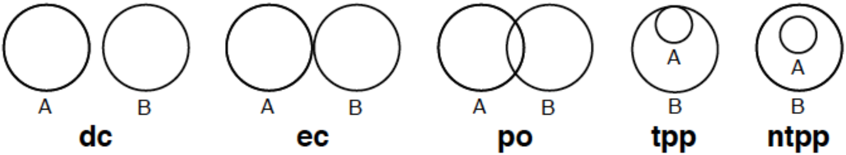
Preliminaries

The description logic $\mathcal{ALCRP}(\mathcal{S}_2)$ for spatial information modeling and reasoning: $\mathcal{ALC}$ with:

- *predicate existence restriction*: $\exists u_1, \ldots, u_n.P$ with P a predicate name from $\mathcal{S}_2$ with arty n and $u_1, \ldots, u_n$ feature chains.
- a concrete domain $\mathcal{S}_2$ defined w.r.t. the topological space $\langle \mathbb{R}^2, 2^{\mathbb{R}^2} \rangle$.

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, Neumann 08]



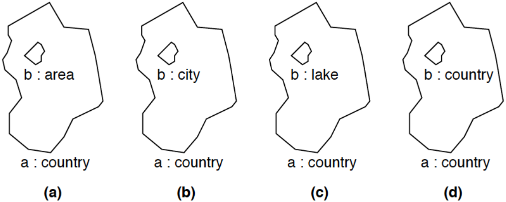
The concrete domain $\mathcal{S}_2$ over the topological space $\langle \mathbb{R}^2, 2^{\mathbb{R}^2} \rangle$

- $\Delta_{\mathcal{S}_2}$: set of non-empty, regular closed subsets of $\mathbb{R}^2$: `regions`
- Set of predicate names:
  - Predicate `is_region` with $\texttt{is\_region}^{\mathcal{S}_2} = \Delta_{\mathcal{S}_2}$ and its negation `is_no_region` with $\texttt{is\_no\_region}^{\mathcal{S}_2} = 0_{\mathcal{S}_2}$
  - 8 basic predicates `dc, ec, po, tpp, ntpp, tppi, eq` (RCC-8 relations)
  - Predicates to name disjunctions of base relations :$p1 - \ldots - p_n$
  - The predicate `dc-ec-po-tpp-ntpp-tppi-ntppi-eq` is called `spatially_related`
  - A binary predicate `inconsistent_relation` with $\texttt{inconsistent\_relation}^{\mathcal{S}_2} = \emptyset$ (negation of `spatially_related`).

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, Neumann 08]

Example



Interpretation problem: generate hypotheses for object b.

$\mathcal{S}_2$ predicates formalization

$$inside \equiv \exists(has\_area)(has\_area).tpp - ntpp$$
$$contains \equiv \exists(has\_area)(has\_area).tppi - ntppi$$
$$overlaps \equiv \exists(has\_area)(has\_area).po$$
$$touches \equiv \exists(has\_area)(has\_area).ec$$
$$disjoint \equiv \exists(has\_area)(has\_area).dc$$

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, Neumann 08]

Example

TBox

$$
\begin{array}{rcl}
area & \equiv & \exists(has\_area).is\_region \\
natural\_region & \equiv & \neg administrative\_region \\
country\_region & \sqsubseteq & administrative\_region \sqcap \\
 & & large\_scale \sqcap area \\
city\_region & \sqsubseteq & administrative\_region \sqcap \\
 & & \neg large - scale \sqcap area \\
lake\_region & \sqsubseteq & natural\_region \sqcap area \\
river\_region & \sqsubseteq & natural\_region \sqcap area
\end{array}
$$

$$
\begin{array}{rcl}
country & \equiv & country\_region \sqcap \\
 & & \forall contains.\neg country\_region \sqcap \\
 & & \forall overlaps.\neg country\_region \sqcap \\
 & & \forall inside.\neg country\_region \\
city & \equiv & city\_region \sqcap \\
 & & \exists inside.country\_region \\
lake & \sqsubseteq & lake\_region \\
river & \sqsubseteq & river\_region \sqcap \\
 & & \forall overlaps.\neg lake\_region \sqcap \\
 & & \forall contains.\bot \sqcap \\
 & & \forall inside.\neg lake\_region
\end{array}
$$

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, Neumann 08]

Example

Abox

$\{a : country, b : area, (a, b) : contains, (b, a) : inside\}$

Spatioterminological default rules

$$d_1 = \frac{area : city}{city} \quad d_2 = \frac{area : lake}{lake} \quad d_3 = \frac{area : city}{city}$$

Closed spatioterminological default rules, $d_i(ind)$

e.g.

$$d_1(a) = \frac{\{a : area\} : \{a : city\}}{\{a : city\}}$$

6 different closed defaults can be obtained $(d_1(a), d_1(b), d_2(a), d_2(b), d_3(a), d_3(b))$

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, Neumann 08]

Example

Default rules reasoning

$$d_1 = \frac{area : city}{city}$$

- $d_1(a)$: cannot be applied.
  Contradiction between $a : city$ and $a : country$ in the Abox. *country_region* and *city_region* are disjoint in the TBox (due to *large_scale* and $\neg$*large_scale*).
- $d_1(b)$: can be applied.
  Abox extension:

$$\{a : country, b : area, b : city, (a, b) : contains, (b, a) : inside\}$$

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, Neumann 08]

Example

Default rules reasoning

$$d_2 = \frac{area : lake}{lake}$$

- $d_2(a)$: cannot be applied.
  Contradiction between $a : lake$ and $a : country$ in the Abox. *administrative_region* and *natural_region* are disjoint.
- $d_2(b)$: can be applied.
  Abox extension:

$$\{a : country, b : area, b : lake, (a, b) : contains, (b, a) : inside\}$$

But if Abox contains $d_1(a)$, $d_2(b)$ cannot be applied $\Longrightarrow$ two possible extensions.

Spatioterminological default reasoning

Moller et al. approach [Möller 99b, **?**]

Example

Default rules reasoning, cont'd

$$d_3 = \frac{area : country}{country}$$

- $d_3(a)$ cannot be applied. Its conclusion is already entailed by the ABox.
- $d_3(b)$ cannot be applied. The consequent $b : country$ makes the Abox inconsistent because a is already known as a country.

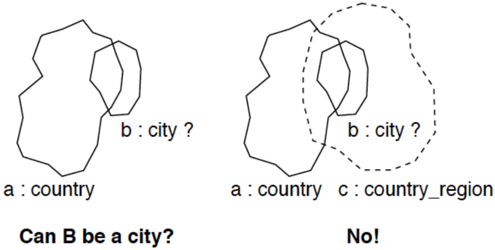
$$\mathcal{A} \models (a : \forall contains.\neg country\_region)$$

$$(a, b) : contains, b : country \implies b : country\_region$$

Spatioterminological default reasoning.

Moller et al. approach [Möller 99b, ?]

Example 2



Subtle inferences due to topological constraints

Abox

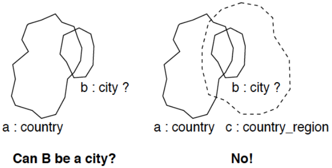
$$\{a : country, b : area, (a, b) : overlaps, (b, a) : overlaps\}$$

$\Longrightarrow$ the default rule $d_1(b)$ cannot be applied to conclude that object b is a city.

Spatioterminological default reasoning.

Moller et al. approach [Möller 99b, ?]

Example 2



$$\mathcal{A} = \{a : country, b : area, (a,b) : overlaps, (b,a) : overlaps\}$$

$(b,a) : overlaps, b : city \implies b : city\_region \sqcap \exists inside.country\_region \implies \not\models (a : country\_region)$ (since $(b,a) : overlaps$).

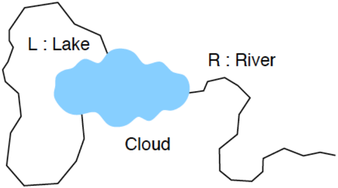
Remark

- Due to $\exists$, there exists an implicit individual c which is a *country_region* such that $(b,c) : inside$ holds.
- Impossible due to topological constraints (b inside c and c not overlap with a or does not contain a).
- No way to conclude that b could possibly be a city.

Spatioterminological default reasoning.

Moller et al. approach [Möller 99b, **?**]

Example 3



Incomplete spatial information

Abox

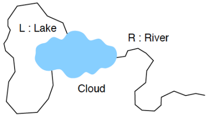
$$\{l : lake, r : river\}$$

We can conclude that the spatial relationship between the river and the lake is either *ec* or *dc*.

Spatioterminological default reasoning.

Moller et al. approach [Möller 99b, ?]

Example 3



Incomplete spatial information

Restricted default theories with ABox patterns

$$d_4 = \frac{\{x : lake, y : river, (x,y) : spatially\_related : country\} : \{(x,y) : disjoint\}}{\{(x,y) : disjoint\}}$$

$$d_4 = \frac{\{x : lake, y : river, (x,y) : spatially\_related : country\} : \{(x,y) : touches\}}{\{(x,y) : touches\}}$$

Closing the patterns yields 8 different closed defaults.

Abductive reasoning

- Abduction using safe rules (Peraldi et al. [Peraldí 09]).
- Concept abduction (Atif et al. [Atif 14]).

Abductive reasoning

Sort of *backward reasoning* from a set of observations to a cause.

Definition

Given a knowledge base $\mathcal{K}$ and a formula $\mathcal{O}$ representing an observation with $\mathcal{K} \not\models \mathcal{O}$, we look for an explanation formula $\mathcal{H}$ such that $\mathcal{H}$ is satisfiable w.r.t. $\mathcal{K}$ and

$$\mathcal{K} \cup \mathcal{H} \models \mathcal{O}$$

holds.

Case of image interpretation

- Scene = observation.
- Interpretation = look for the *best* explanation considering a terminological knowledge part about the scene context.

Abductive reasoning and description logics

Distinct abductive problems [Elsenbroich 06]

Let $\mathcal{L}$ be a DL, $\mathcal{K}$ a knowledge base in $\mathcal{L}$

- Concept abduction
- ABox abduction
- TBox abduction
- Knowledge base abduction

Abduction using safe rules

Multimedia interpretation as abduction

Peraldi et al. [Peraldí 09]

Ontology-based reasoning techniques for multimedia interpretation and retrieval

Möller et al. [Möller 08]

Multimedia interpretation as an abduction problem.

Peraldi et al. [Peraldí 09]

Main idea: Abduction as a non-standard retrieval inference service

Observations are used to constitute queries that have to be answered by acquiring what should be added to the knowledge base in order to positively answer to a query

Use of conjunctive queries

Structure of the form {`head` | `body`}:

$$\{(X_1, \cdots, X_n) \mid atom_1, \cdots, atom_m\}, \text{ with}$$
$$atom = C(X), R(X, Y), (X = Y)$$

- head: list of variables for which we like to compute bindings
- body: query atoms

Example: $\{x \mid \exists y \exists z (ChildOf(x, y) \land ChildOf(x, z) \land Married(y, z))\}$

Query answer: set of bindings for variables in the head

Formalisation

Peraldi et al. [Peraldí 09]

Abduction inference

Given a set of ABox assertions Γ in form of a query and a KB, $\Sigma = (\mathcal{T}, \mathcal{A})$, derive all sets of ABox assertions Δ (explanations) such that Δ is consistent w.r.t the ontology Σ ($\Sigma \cup \Delta$ is satisfiable) and:

- $\Sigma \cup \Delta \models \Gamma$.
- Δ is a minimal explanation for Γ, i.e. there exists no other explanation Δ' in the solution set that is not equivalent to Δ and it holds that $\Sigma \cup \Delta' \models \Delta$.

Formalisation

Peraldi et al.[Peraldí 09]

Multimedia abduction:

- $\Sigma = (\mathcal{T}, \mathcal{A})$, a knowledge base on the application domain with $\mathcal{A}$ assumed empty.
- $\Gamma = \Gamma_1 \cup \Gamma_2$, set of Abox assertions, encoding low level extracted information from images (objects and their spatial relationships):
  - Γ_1: bona fide assertions, assumed to be true by default.
  - Γ_2: assertions requiring fiats (aimed to be explained).
- Abduction process : compute Δ, a set of ABox explanations, such that

$$\Sigma \cup \Gamma_1 \cup \Delta \models \Gamma_2$$

The process is implemented as (boolean) query answering.

Illustration on an example

Peraldi et al. [Peraldí 09]

ABox Γ : low-level image analysis results

$$
\begin{array}{rcl}
pole_1 & : & Pole \\
human_1 & : & Human \\
bar_1 & : & Bar \\
\{bar_1, human_1\} & : & near
\end{array}
$$

Σ, a Tbox and DL-safe rules on the athletics domain

$$
\begin{array}{rcl}
Jumper & \sqsubseteq & Human \\
Pole & \sqsubseteq & Sports\_Equipment \\
Bar & \sqsubseteq & Sports\_Equipment \\
Pole \sqcap Bar & \sqsubseteq & \\
Pole \sqcap Jumper & \sqsubseteq & \\
Jumper \sqcap Bar & \sqsubseteq & \\
Jumping\_Event & \sqsubseteq & \exists_{\leq 1} hasParticipant.Jumper \\
Pole\_Vault & \sqsubseteq & Jumping\_Event \sqcap \exists hasPart.Pole \sqcap \exists hasPart.Bar \\
High\_Jump & \sqsubseteq & Jumping\_Event \sqcap \exists hasPart.Bar \\
near(Y, Z) & \leftarrow & Pole\_Vault(X), hasPart(X, Y), Bar(Y), \\
 & & hasPart(X, W), Pole(W), hasParticipant(X, Z), Jumper(Z) \\
near(Y, Z) & \leftarrow & High\_Jump(X), hasPart(X, Y), Bar(Y), \\
 & & hasParticipant(X, Z), Jumper(Z)
\end{array}
$$

Illustration on an example

Peraldi et al. [Peraldí 09]

ABox Γ : low-level image analysis results

$pole_1$	:	$Pole$
$human_1$	:	$Human$
bar_1	:	Bar
$\{bar_1, human_1\}$	:	$near$

- $\Gamma_1 = \{pole_1 : Pole, human_1 : Human, bar_1 : Bar\}$
- $\Gamma_2 = \{(bar_1, human_1) : near\}$
- Boolean query $Q_1 := \{() \mid near(bar_1, human_1)\}$

Peraldi et al. [Peraldí 09]

Possible explanations:

- $\Delta_1 = \{new\_ind_1 : Pole\_Vault, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, new\_ind_2) : hasPart, new\_ind_2 : Pole, (new\_ind_1, human_1) : hasParticipant, human_1 : Jumper\}$
- $\Delta_2 = \{new\_ind_1 : Pole\_Vault, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, pole_1) : hasPart, (new\_ind_1, human_1) : hasParticipant, human_1 : Jumper\}$
- $\Delta_3 = \{new\_ind_1 : High\_Jump, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, human1) : hasParticipant, human_1 : Jumper\}$

Preference score :

$$S_p(\Delta) := S_i(\Delta) - S_h(\Delta), \text{ with}$$
$$S_i(\Delta) := |\{i \mid i \in inds(\Delta) \text{ and } i \in inds(\Sigma \cup \Gamma_1)\}|$$
$$S_h(\Delta) := |\{i \mid i \in inds(\Delta) \text{ and } i \in new\_inds\}|$$

Peraldi et al. [Peraldí 09]

- Δ_1 incorporates $human_1$ and bar_1 from Γ_1, then $S_i(\Delta_1) = 2$.
- Δ_1 hypothesizes two new individuals: new_ind_1, new_ind_2, then $S_h(\Delta_1) = 2$.

$\Longrightarrow$ $S_p(\Delta_1) = 0$

- $S_p(\Delta_2) = 3 - 1 = 2$.
- $S_p(\Delta_3) = 2 - 1 = 1$.

$\Longrightarrow$ Δ_2 represents the 'preferred' explanation:

$\Delta_2 = \{new\_ind_1 : Pole\_Vault, (new\_ind_1, bar_1) : hasPart, (new\_ind_1, pole_1) : hasPart, (new\_ind_1, human_1) : hasParticipant, human_1 : Jumper\}$

The image should better be interpreted as showing a pole vault and not a high jump.

Multimedia interpretation as concept abduction

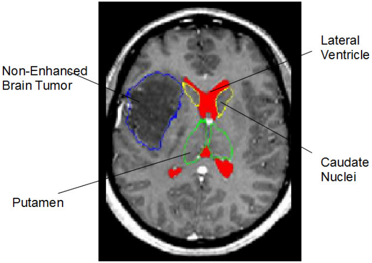
Explanatory reasoning for image understanding using formal concept analysis and description logics.

Atif et al. [Atif 14]

Brain image understanding

Atif et al. [Atif 14]

Image interpretation



Pathological brain with small deforming peripheral tumor

Interpretation as an abduction process

$\mathcal{K} \models (\gamma \rightarrow \varphi)$

Computing of the *best explanation* from observations φ given some a priori expert knowledge $\mathcal{K}$ encoded in description logics.

Knowledge representation

$$
\begin{aligned}
\mathit{CerebralHemisphere} &\sqsubseteq \mathit{BrainAnatomicalStructure}\\
\mathit{PeripheralCerebralHemisphere} &\sqsubseteq \mathit{CerebralHemisphereArea}\\
\mathit{SubCorticalCerebralHemisphere} &\sqsubseteq \mathit{CerebralHemisphereArea}\\
\mathit{GreyNuclei} &\sqsubseteq \mathit{BrainAnatomicalStructure}\\
\mathit{LateralVentricle} &\sqsubseteq \mathit{BrainAnatomicalStructure}\\
\mathit{BrainTumor} &\sqsubseteq \mathit{Disease}\\
& \quad \sqcap \exists \mathit{hasLocation}.\mathit{Brain}\\
\mathit{SmallDeformingTumor} &\equiv \mathit{BrainTumor}\\
& \quad \sqcap \exists \mathit{hasBehavior}.\mathit{Infiltrating}\\
& \quad \sqcap \exists \mathit{hasEnhancement}.\mathit{NonEnhanced}\\
\mathit{SubCorticalSmallDeformingTumor} &\equiv \mathit{SmallDeformingTumor} \sqcap\\
& \quad \exists \mathit{hasLocation}.\mathit{SubCorticalCerebralHemisphere}\\
& \quad \sqcap \exists \mathit{closeTo}.\mathit{GreyNuclei}\\
\mathit{PeripheralSmallDeformingTumor} &\equiv \mathit{BrainTumor} \sqcap\\
& \quad \exists \mathit{hasLocation}.\mathit{PeripheralCerebralHemisphere}\\
& \quad \sqcap \exists \mathit{farFrom}.\mathit{LateralVentricle}
\end{aligned}
$$

$$
\begin{aligned}
\mathit{LargeDefTumor} &\equiv \mathit{BrainTumor} \sqcap\\
& \quad \exists \mathit{hasLocation}.\mathit{CerebralHem}\\
& \quad \sqcap \exists \mathit{hasComponent}.\mathit{Edema}\\
& \quad \sqcap \exists \mathit{hasComponent}.\mathit{Necrosis}\\
& \quad \sqcap \exists \mathit{hasEnhancement}.\mathit{Enhanced}\\
& \ldots
\end{aligned}
$$

Initial ABox $\mathcal{A}_1$

$\{t_1 : \mathit{BrainTumor};\ e_1 : \mathit{NonEnhanced};\ l_1 : \mathit{LateralVentricle};\ p_1 : \mathit{PeripheralCerebralHemisphere};\ (t_1, e_1) : \mathit{hasEnhancement};\ (t_1, l_1) : \mathit{farFrom};\ (t_1, p_1) : \mathit{hasLocation};\ \ldots\}.$

Interpretation as a concept abduction process

$\mathcal{K} \models \gamma \sqsubseteq O$, with O, main specific concept of t_1, defined as

$$BrainTumor \sqcap \exists hasEnhancement.NonEnhanced \sqcap \exists farFrom.LateralVentricle \sqcap \exists hasLocation.PeripheralCerebralHemisphere$$

A set of possible explanations is :

$$\{DiseasedBrain, SmallDeformingTumoralBrain, PeripheralSmallDeformingTumoralBrain\}$$

The preferred solution according to minimality constraints is:

$$\gamma \equiv PeripheralSmallDeformingTumoralBrain$$

Abduction and logics

Description logics

Where are we ?

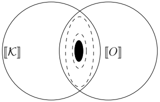
- Only a few works
- Rewriting approach (Modal logics - Description Logics)

Propositional logics (morpho-logics, Bloch et al. [Bloch 02])

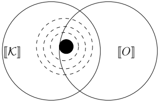
$$[\![\varepsilon(\varphi)]\!] := \varepsilon([\![\varphi]\!]), [\![\delta(\varphi)]\!] := \delta([\![\varphi]\!])$$

Successive erosions of the set of models

- Erosion of the conjunction of the theory with the formula to be explained



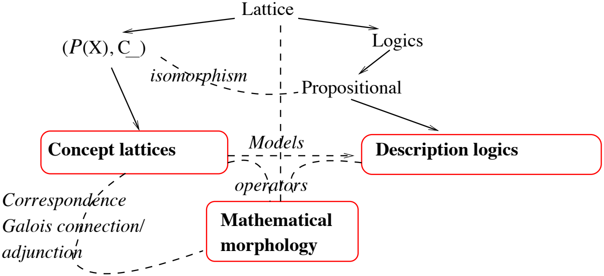
- Erosion of the theory while maintaining the coherence with the formula to be explained



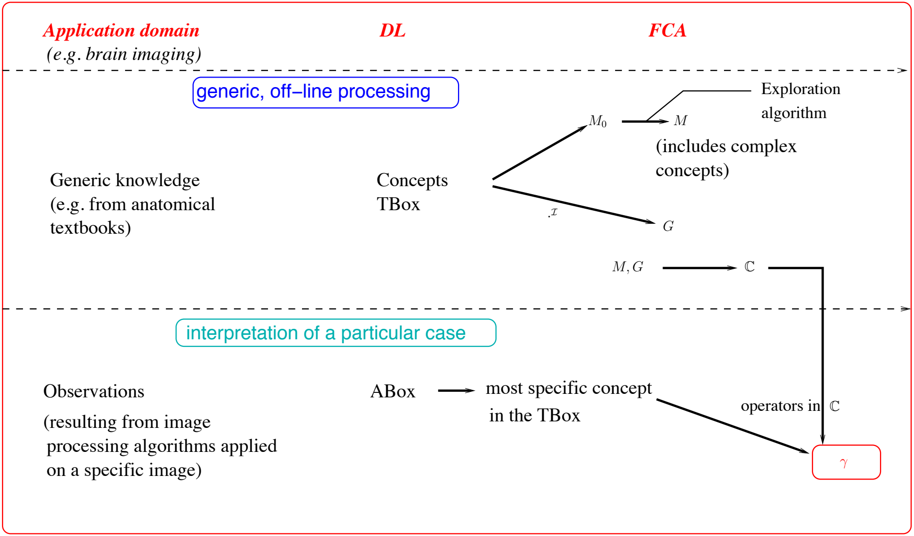
Proposed approach

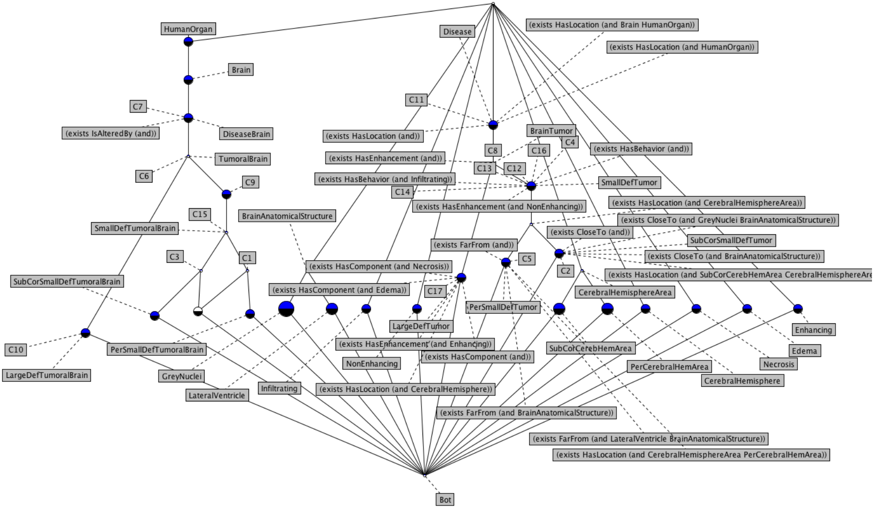
Enrichment of description logics with abductive reasoning services

⇒ Association between three theories :

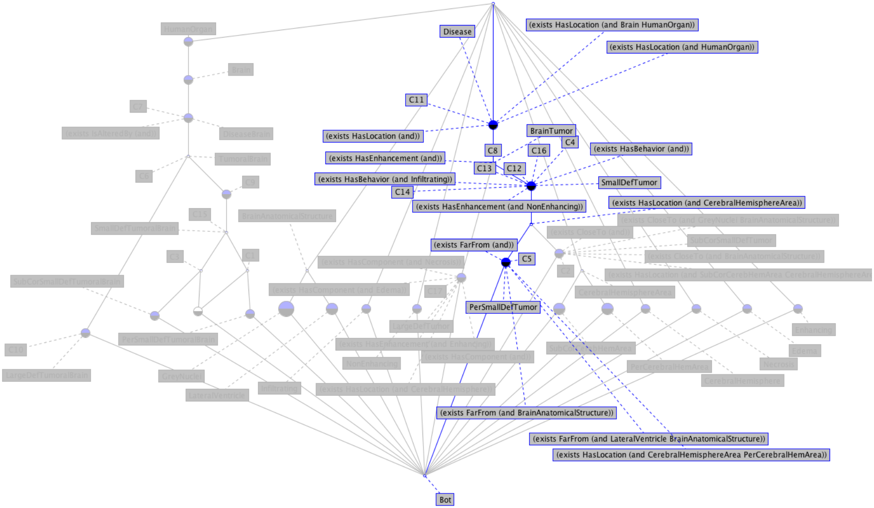


Global scheme





Concept lattice induced from $\mathbb{K}_{brain}$.



Erosion path leading to compute a preferred explanation

Outline

Conclusion

Ontologies and logic-based approaches for image interpretation

- A growing interest in the litterature.
- Main advantages: explicit knowledge encoding for reuse and reasoning processes.
- Need for more convergence between computer vision, machine learning and logics community.

Thanks for your attention

Lourdes Agapito, Michael M. Bronstein & Carsten Rother, editeurs.
Computer vision - ECCV 2014 workshops - zurich, switzerland, september 6-7 and 12, 2014, proceedings, part II, volume 8926 of *Lecture Notes in Computer Science*. Springer, 2015.

J. Atif, C. Hudelot & I. Bloch.
Explanatory reasoning for image understanding using formal concept analysis and description logics.
IEEE Transactions on Systems, Man and Cybernetics: Systems, vol. 44, no. 5, pages 552–570, May 2014.

F. Baader & B. Hollunder.
Embedding Defaults into Terminological Knowledge Representation Formalisms.

In B. Nebel, C. Rich & W. Swartout, editeurs, Principles of Knowledge Representation and Reasoning: Proc. of the Third International Conference (KR'92), pages 306–317. Kaufmann, San Mateo, CA, 1992.

Andrew D. Bagdanov, Marco Bertini, Alberto Del Bimbo, Giuseppe Serra & Carlo Torniai.
Semantic annotation and retrieval of video events using multimedia ontologies.

In International Conference on Semantic Computing (ICSC'07), pages 713 –720, 2007.

Hichem Bannour & Céline Hudelot.
Building and using fuzzy multimedia ontologies for semantic image annotation.
Multimedia Tools Appl., vol. 72, no. 3, pages 2107–2141, 2014.

Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei & Michael I. Jordan.
Matching words and pictures.
Journal of Machine Learning Research, vol. 3, pages 1107–1135, 2003.

Evgeniy Bart, Ian Porteous, Pietro Perona & Max Welling.
Unsupervised learning of visual taxonomies.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), 2008.

Serge J. Belongie & Pietro Perona.
Visipedia circa 2015.
Pattern Recognition Letters, vol. 72, pages 15–24, 2016.

Isabelle Bloch & Jérôme Lang.
Towards mathematical morpho-logics, pages 367–380.

Physica-Verlag HD, Heidelberg, 2002.

Gustavo Carneiro, Antoni B. Chan, Pedro J. Moreno & Nuno Vasconcelos.
Supervised Learning of Semantic Classes for Image Annotation and Retrieval.
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 3, 2007.

Navneet Dalal & Bill Triggs.
Histograms of Oriented Gradients for Human Detection.
In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA, pages 886–893, 2005.

Stamatia Dasiopoulou, Ioannis Kompatsiaris & Michael Strintzis.
Applying Fuzzy DLs in the Extraction of Image Semantics.
In Stefano Spaccapietra & Lois Delcambre, editeurs, Journal on Data Semantics XIV, volume 5880 of *Lecture Notes in Computer Science*, pages 105–132. Springer Berlin / Heidelberg, 2009.

Stamatia Dasiopoulou & Ioannis Kompatsiaris.
Trends and Issues in Description Logics Frameworks for Image Interpretation.

Artificial Intelligence: Theories, Models and Applications, vol. 6040, pages 61–70, 2010.

Stamatia Dasiopoulou, Vassilis Tzouvaras, Ioannis Kompatsiaris & Michael G. Strintzis.
Enquiring MPEG-7 based multimedia ontologies.
Multimedia Tools and Applications (MTAP'10), vol. 46, pages 331–370, 2010.

Jia Deng, Alexander C. Berg, Kai Li & Li Fei-Fei.
What does classifying more than 10,000 image categories tell us?
In Proceedings of the European Conference on Computer Vision (ECCV'10), 2010.

Chabane Djeraba.
Association and Content-Based Retrieval.
IEEE Transaction on Knowledge and Data Engineering (TDDE'03), vol. 15, pages 118–135, January 2003.

Ivan Donadello & Luciano Serafini.
Mixing Low-Level and Semantic Features for Image Interpretation - A Framework and a Simple Case Study.
In Agapito et al. [Agapito 15], pages 283–298.

P. Duygulu, Kobus Barnard, J. F. G. de Freitas & David A. Forsyth.
Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary.
In Proceedings of the European Conference on Computer Vision (ECCV'02), pages 97–112. Springer-Verlag, 2002.

Corinna Elsenbroich, Oliver Kutz & Ulrike Sattler.
A Case for Abductive Reasoning over Ontologies.
In Bernardo Cuenca Grau, Pascal Hitzler, Conor Shankey & Evan Wallace, editeurs, Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions, Athens, Georgia, USA, November 10-11, 2006, volume 216 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.

Dumitru Erhan, Christian Szegedy, Alexander Toshev & Dragomir Anguelov.
Scalable Object Detection Using Deep Neural Networks.
In 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pages 2155–2162, 2014.

Ali Farhadi, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier & David Forsyth.

Every Picture Tells a Story: Generating Sentences from Images.
In Proceedings of the 11th European Conference on Computer Vision: Part IV, ECCV'10, pages 15–29, Berlin, Heidelberg, 2010. Springer-Verlag.

Ali Farhadi & Mohammad Amin Sadeghi.
Phrasal Recognition.
IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 12, pages 2854–2865, December 2013.

L. Fei-Fei & P. Perona.
A Bayesian hierarchical model for learning natural scene categories.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2005), volume 2, pages 524 – 531, 2005.

Carolina Galleguillos & Serge J. Belongie.
Context based object categorization: A critical survey.
Computer Vision and Image Understanding, vol. 114, no. 6, pages 712–722, 2010.

Tianshi Gao & Daphne Koller.
Discriminative learning of relaxed hierarchy for large-scale visual recognition.
In Proceedings of the International Conference on Computer Vision (ICCV'11), 2011.

Gregory Griffin & Pietro Perona.
Learning and Using Taxonomies for Fast Visual Categorization.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), 2008.

Thomas R. Gruber.
Toward principles for the design of ontologies used for knowledge sharing.
International Journal of Human-Computer Studies, vol. 43, no. 5-6, pages 907–928, 1995.

Nicola Guarino, Daniel Oberle & Steffen Staab.
What Is an Ontology?
In Steffen Staab & Rudi Studer, editeurs, Handbook on Ontologies, International Handbooks on Information Systems, pages 1–17. Springer, 2009.

Amirhossein Habibian, Koen E.A. van de Sande & Cees G.M. Snoek.
Recommendations for Video Event Recognition Using Concept Vocabularies.
In Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval, ICMR '13, pages 89–96, New York, NY, USA, 2013. ACM.

Allen R Hanson & Edward M Riseman.
VISIONS: A computer system for interpreting scenes.
Computer vision systems, vol. 78, 1978.

Stevan Harnad.
The Symbol Grounding Problem.
CoRR, vol. cs.AI/9906002, 1999.

C. Hudelot & M. Thonnat.
A cognitive vision platform for automatic recognition of natural complex objects.
15th IEEE International Conference on Tools with Artificial Intelligence, 2003. Proceedings., pages 398–405, 2003.

C. Hudelot, J. Atif & I. Bloch.
Fuzzy Spatial Relation Ontology for Image Interpretation.
Fuzzy Sets and Systems, vol. 159, no. 15, pages 1929–1951, 2008.

Céline Hudelot, Jamal Atif & Isabelle Bloch.
ALC(𝔽): A New Description Logic for Spatial Reasoning in Images.
In Agapito et al. [Agapito 15], pages 370–384.

Alejandro Jaimes & Shih fu Chang.

A Conceptual Framework for Indexing Visual Information at Multiple Levels.
In Storage and Retrieval for Image and Video Databases (SPIE'00), pages 2–15, 2000.

Andrej Karpathy, Armand Joulin & Li Fei-Fei.
Deep Fragment Embeddings for Bidirectional Image Sentence Mapping.
CoRR, vol. abs/1406.5679, 2014.

Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, Michael Bernstein & Li Fei-Fei.
Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations.
2016.

Alex Krizhevsky, Ilya Sutskever & Geoffrey E. Hinton.
ImageNet Classification with Deep Convolutional Neural Networks.
In Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States., pages 1106–1114, 2012.

G. Kulkarni, V. Premraj, S. Dhar, Siming Li, Yejin Choi, A.C. Berg & T.L. Berg.
Baby talk: Understanding and generating simple image descriptions.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1601 –1608, 2011.

V. Lavrenko, R. Manmatha & J. Jeon.
A model for learning the semantics of pictures.
In Neural Information Processing Systems (NIPS'03). MIT Press, 2003.

Li-Jia Li, Chong Wang, Yongwhan Lim, David M. Blei & Fei-Fei Li.
Building and using a semantivisual image hierarchy.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10), pages 3336 –3343, 2010.

Ying Liu, Dengsheng Zhang, Guojun Lu & Wei-Ying Ma.
A survey of content-based image retrieval with high-level semantics.
Pattern Recognition, vol. 40, no. 1, pages 262 – 282, 2007.

David G. Lowe.
Distinctive Image Features from Scale-Invariant Keypoints.

International Journal of Computer Vision, vol. 60, no. 2, pages 91–110, 2004.

Nicolas Maillot & Monique Thonnat.
Ontology based complex object recognition.
Image Vision Comput., vol. 26, no. 1, pages 102–113, 2008.

David Marr.
Vision: A computational investigation into the human representation and processing of visual information.
W. H. Freeman, 1982.

Marcin Marszalek & Cordelia Schmid.
Constructing Category Hierarchies for Visual Recognition.
In Proceedings of the European Conference on Computer Vision (ECCV), pages 479–491, 2008.

Vasileios Mezaris, Ioannis Kompatsiaris & Michael G. Strintzis.
Region-Based Image Retrieval Using an Object Ontology and Relevance Feedback.
EURASIP J. Adv. Sig. Proc., vol. 2004, no. 6, pages 886–901, 2004.

Ralf Möller & Bernd Neumann.

Ontology-based reasoning techniques for multimedia interpretation and retrieval, pages 55–98.
Springer London, London, 2008.

Ralf Möller, Bernd Neumann & Michael Wessel.
Towards Computer Vision with Description Logics: Some Recent Progress.
Integration of Speech and Image Understanding, ICCV Workshop on, vol. 0, page 101, 1999.

Ralf Möller, Bernd Neumann & Michael Wessel.
Towards Computer Vision with Description Logics: Some Recent Progress.
Integration of Speech and Image Understanding, ICCV Workshop on, vol. 0, page 101, 1999.

Ralf Möller & Bernd Neumann.
Ontology-Based Reasoning Techniques for Multimedia Interpretation and Retrieval.
In Yiannis Kompatsiaris & Paola Hobson, editeurs, Semantic Multimedia and Ontologies, pages 55–98. Springer London, 2008.

Milind Naphade, John R. Smith, Jelena Tesic, Shih-Fu Chang, Winston Hsu, Lyndon Kennedy, Alexander Hauptmann & Jon Curtis.

Large-Scale Concept Ontology for Multimedia.
IEEE MultiMedia, vol. 13, pages 86–91, July 2006.

Bernd Neumann & Ralf Möller.
On scene interpretation with description logics.
Image Vision Computing, vol. 26, no. 1, pages 82–101, 2008.

Devi Parikh, C. Lawrence Zitnick & Tsuhan Chen.
Exploring Tiny Images: The Roles of Appearance and Contextual Information for Machine and Human Object Recognition.
IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 10, pages 1978–1991, October 2012.

S. Espinosa Peraldi, A. Kaya, S. Melzer, R. Möller & M. Wessel.
Multimedia Interpretation as Abduction.
In International Workshop on Description Logics (DL'07), 2007.

Irma Sofía Espinosa Peraldí, Atila Kaya & Ralf Möller.
Formalizing Multimedia Interpretation based on Abduction over Description Logic Aboxes.
In Bernardo Cuenca Grau, Ian Horrocks, Boris Motik & Ulrike Sattler, editeurs, Proceedings of the 22nd International Workshop on Description

Logics (DL 2009), Oxford, UK, July 27-30, 2009, volume 477 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.

Daniele Porello, Marco Cristani & Roberta Ferrario.
Integrating Ontologies and Computer Vision for Classification of Objects in Images.
Proceedings of the Workshop on Neural-Cognitive Integration in German Conference on Artificial Intelligence, pages 1–15, 2013.

Raymond Reiter.
A Logic for Default Reasoning.
Artif. Intell., vol. 13, no. 1-2, pages 81–132, 1980.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg & Fei-Fei Li.
ImageNet Large Scale Visual Recognition Challenge.
International Journal of Computer Vision, vol. 115, no. 3, pages 211–252, 2015.

Bryan C. Russell, Antonio Torralba, Kevin P. Murphy & William T. Freeman.
LabelMe: A Database and Web-Based Tool for Image Annotation.

International Journal of Computer Vision, vol. 77, no. 1-3, pages 157–173, 2008.

Simone Santini, Amarnath Gupta & Ramesh Jain.
Emergent Semantics through Interaction in Image Databases.
IEEE Transactions on Knowledge and Data Engineering, vol. 13, no. 3, pages 337–351, 2001.

N. Simou, Th Athanasiadis, G. Stoilos & S. Kollias.
Image indexing and retrieval using expressive fuzzy description logics.
Signal, Image and Video Processing, vol. 2, no. 4, pages 321–335, 2008.

J. Sivic, B. C. Russell, A. Zisserman, W. T. Freeman & A. A. Efros.
Unsupervised Discovery of Visual Object Class Hierarchies.
In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), 2008.

Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta & Ramesh Jain.
Content-based image retrieval at the end of the early years.
IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pages 1349–1380, 2000.

Cees G. M. Snoek & Arnold W. M. Smeulders.
Visual-Concept Search Solved?
IEEE Computer, vol. 43, no. 6, pages 76–78, June 2010.

Rudi Studer, V. Richard Benjamins & Dieter Fensel.
Knowledge Engineering: Principles and Methods.
Data Knowl. Eng., vol. 25, no. 1-2, pages 161–197, 1998.

Christopher Town.
Ontological inference for image and video analysis.
Machine Vision and Applications, vol. 17, no. 2, pages 94–115, 2006.

Michael Uschold & Michael Gruninger.
Ontologies and Semantics for Seamless Connectivity.
SIGMOD Rec., vol. 33, no. 4, pages 58–64, December 2004.

Lei Wu, Xian-Sheng Hua, Nenghai Yu, Wei-Ying Ma & Shipeng Li.
Flickr Distance: A Relationship Measure for Visual Concepts.
Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 5, pages 863 –875, 2012.

B.Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee & Song-Chun Zhu.

I2T: Image Parsing to Text Description.
Proceedings of the IEEE, vol. 98, no. 8, pages 1485 –1508, 2010.