Using Fuzzy DLs to Enhance Semantic Image Analysis

S. Dasiopoulou, I. Kompatsiaris, M.G.Strintzis

Multimedia Knowledge Laboratory
Informatics and Telematics Institute

3rd International Conference on Semantic and Digital Media Technologies
3-5 December 2008, Koblenz, Germany

Outline

- Motivation
- Fuzzy DLs reasoning in semantic image analysis
  - Specifications
  - Tasks
- Evaluation
- Conclusions

Why (explicit) Reasoning in Image Annotation

- Machine learning provides now generic methodologies for supporting more than 100 concepts
  - captures conveniently complex associations between perceptual features and semantics
  - successful application examples, yet versatile general performance

- **Semantics goes beyond perceptual manifestations**
  - *possibly contradictory (Mountain, Sand and Indoor)*
  - *possibly overlapping / complementary (Beach and Sea)*
  - *of restricted abstraction w.r.t. semantic expressiveness (face inside sea vs Swimmer)*

- Learning-based extracted annotations need to be ***semantically interpreted*** into a ***consistent*** description

Semantics goes beyond perceptual manifestations

Discrepancy between semantic expressiveness

Discrepancy between intended and learned semantics

	Best Possible	
Search Topic	*Best Detector*	*AP*
Two visible tennis players on the court	Athlete	0.6501
A goal being made in a soccer match	Stadium	0.3429
Basketball players on the court	Indoor Sports Venue	0.2801
A meeting with a large table and people	Furniture	0.1045
People with banners or signs	People Marching	0.1013
One or more military vehicles	Armored Vehicles	0.0892
Helicopter in flight	Helicopters	0.0791
A road with one or more cars	Car	0.0728
An airplane taking off	Classroom	0.0526
A tall building	Office Building	0.0469
A ship or boat	Cloud	0.0427
George Bush entering or leaving vehicle	Rocket Propelled Grenades	0.0365
Omar Karami	Chair	0.0277
Graphic map of Iraq, Baghdad marked	Graphical Map	0.0269
Condoleeza Rice	US National Flag	0.0237
One or more palm trees	Weapons	0.0225

Snoek et al., "Adding Semantics to Detectors for Video Retrieval", IEEE Multimedia, 2007

Semantics goes beyond perceptual manifestations

$(\text{image}{:}\exists\text{contains.Sand}) \geq 0.75$
$(\text{image}{:}\exists\text{contains.Sky}) \geq 0.87$
$(\text{image}{:}\exists\text{contains.Foliage}) \geq 0.76$
$(\text{image}{:}\exists\text{contains.Conifers}) \geq 0.88$
$(\text{image}{:}\text{Landscape}) \geq 0.92$

- Conifers detector semantics pertain to mountainous scenes
- Sand detector semantics pertains to beach scenes

$(\text{image}{:}\exists\text{contains.Sand}) \geq 0.75$
$(\text{image}{:}\exists\text{contains.Sea}) \geq 0.81$
$(\text{image}{:}\exists\text{contains.Person} \geq 0.67$
$(\text{image}{:}\exists\text{contains.Foliage}) \geq 0.76$
$(\text{image}{:}\exists\text{contains.Grass}) \geq 0.58$
$(\text{image}{:}\text{Beach}) \geq 0.85$
$(\text{image}{:}\text{Beach}) \geq 0.67$

- Sea and Sand detectors entail Beach scene
- Beach scenes entails both Natural and Outdoor scenes

Why Fuzzy Description Logics

- Semantic Web
  - multimedia aware SW
  - interoperablity
  - reuse

- Imperfect information
  - fuzzy (e.g. green region)
  - probabilistic (~ co-occurrence patterns)

Our Approach

- **Goal:** enhance the robustness and completeness of learning-based extracted annotations

- **How:** semantics utilisation
  - to interpret initial annotations
    - semantic integration
    - to detect and resolve inconsistencies
  - to enrich by means of entailment

- **Methodology:** fuzzy DL based reasoning
  - crisp TBox to conceptualise the domain semantics
  - fuzzy ABox to capture the uncertainty of initial annotations

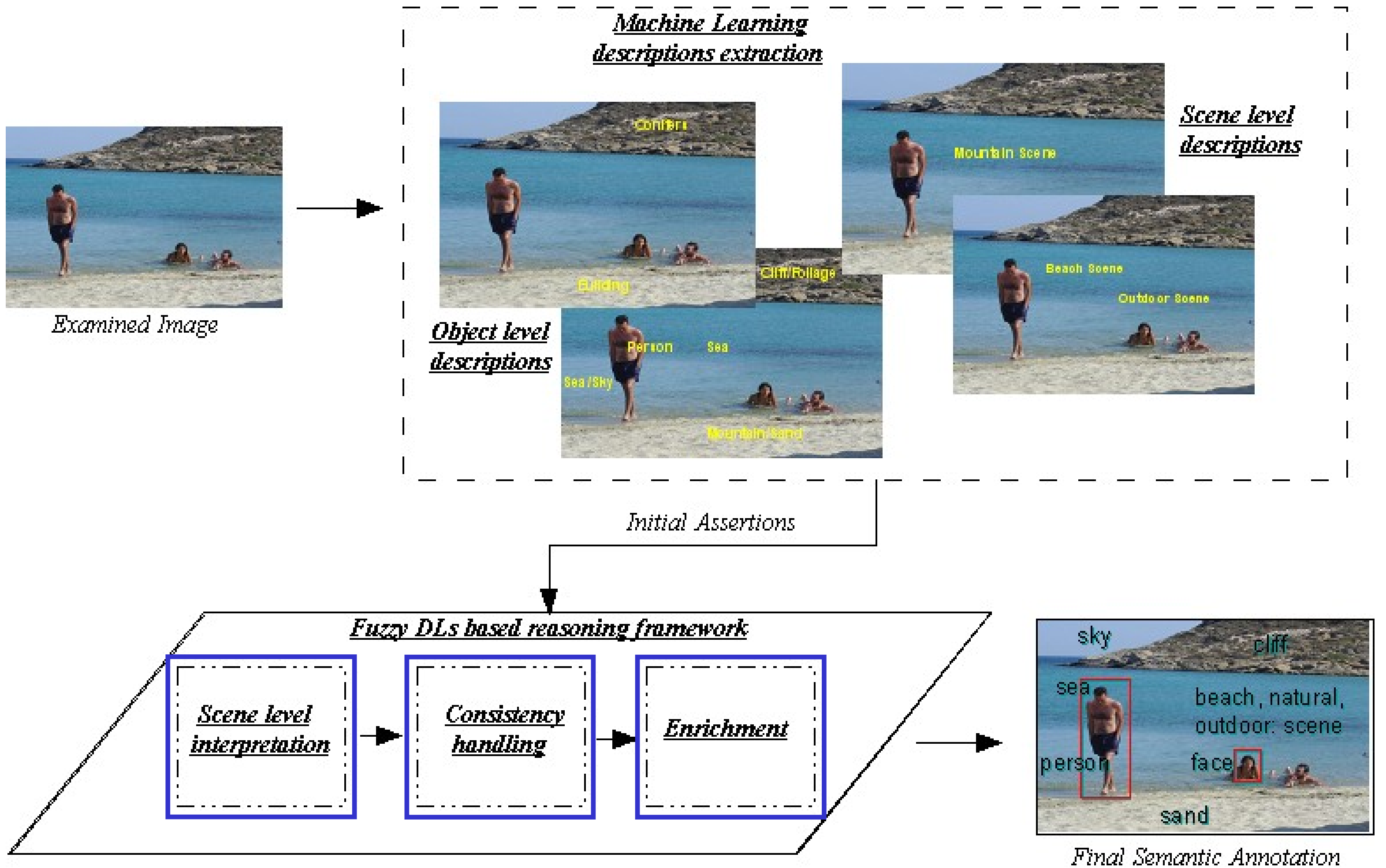
Specifications

- Analysis extracted annotations translate to input assertions
  - descriptions at object / scene level
  - different implementations (black box)

- Annotation degrees express distance from learned feature models
  - concepts as fuzzy sets
  - membership value

- Ranked list of semantically consistent interpretations

General Framework



Reasoning Task I

- ***Scene level interpretation***
  - involves both asserted and inferred assertions of scene level concepts
  - computes scene level concept hierarchy
  - *Procedure*
    a. remove disjointness axioms
    b. starting from the leaf concepts, maintain between conflicting assertions the one with highest degree
    c. propagates degrees according to fuzzy subsumption semantics to the next level
    d. repeat step b check, if current prevalent assertions contradict the previous level (i.e. have higher plausibility) remove and update accordingly the previous level
    e. ends when reaching the top level concepts

Scene level interpretation demonstration

Domain TBox

$Natural \equiv Outdoors \sqcup \neg ManMade$
$Mountainous \equiv Natural \sqcup \neg Coastal$
$\boxed{Beach \equiv Coastal \sqcap \exists contains.Sand}$
$\boxed{\exists contains.Mountain \sqsubseteq Mountainous}$
$\boxed{\exists contains.Sea \sqsubseteq Coastal}$
$\exists contains.Sand \sqcap Mountainous \sqsubseteq \perp$
$Outdoor \sqcap Indoor \sqsubseteq \perp$

Initial Assertions

(image:Indoor) ≥ 0.67
(image:$\exists$contains.Sea) ≥ 0.73
(image:$\exists$contains.Sand) ≥ 0.58
(image:$\exists$contains.Mountain) ≥ 0.85

Disjointness axioms removed

(image:Indoor) ≥ 0.67
(image:$\exists$contains.Sea) ≥ 0.73
(image:$\exists$contains.Sand) ≥ 0.58
(image:Coastal) ≥ 0.73
(image:Beach) ≥ 0.58
(image:Natural) ≥ 0.73
(image:Outdoor) ≥ 0.73
(image:$\exists$contains.Mountain) ≥ 0.85
(image:Mountainous) ≥ 0.85
(image:Natural) ≥ 0.85
(image:Outdoor) ≥ 0.85

Scene level hierarchy

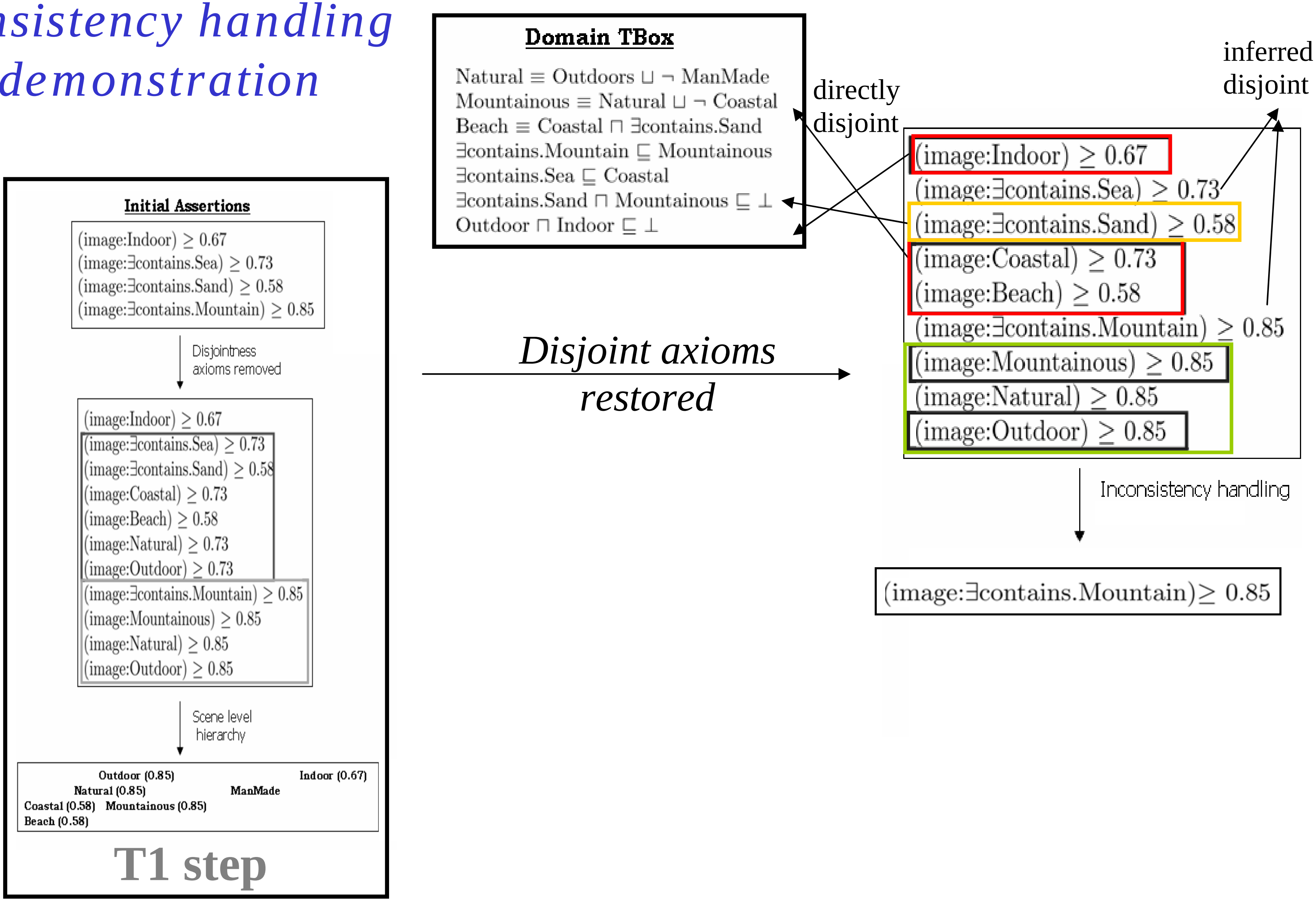
Outdoor (0.85) **Indoor (0.67)**
Natural (0.85) **ManMade**
Coastal (0.58) **Mountainous (0.85)**
Beach (0.58)

Reasoning Task II

- ***Consistency handling***
  - performs over the initial set of annotations

  - Procedure
    - restore disjointness axioms semantics
    - remove all explicit assertions conflicting T1 interpretation
      - object & scene level
    - removes all inferred (if anymore) assertions conflicting T1 interpretation
      - first object level (order matters in this case)
      - second scene level
    - removal of inferred assertions, i.e. assertions referring to complex concepts is performed w.r.t. to the semantics of the operands involved in the axioms they participate
    - in case of more than one consistent (final) interpretations apply economy criteria
      - number of assertions removed of assertions
      - average plausibility of removed assertions

Consistency handling demonstration



Tasks I & II from a more formal perspective (1)

- Semantic integration of knowledge bases
  - integrated axioms & assertions may introduce conflicts
  - removal of axioms / assertions to reach satisfiable knowledge base
- Various approaches
  - stratified ontology
  - enhanced tableaux-based expansion tracking the axioms involved in an inconsistency
  - removal of whole axioms vs parts of axioms
  - ...

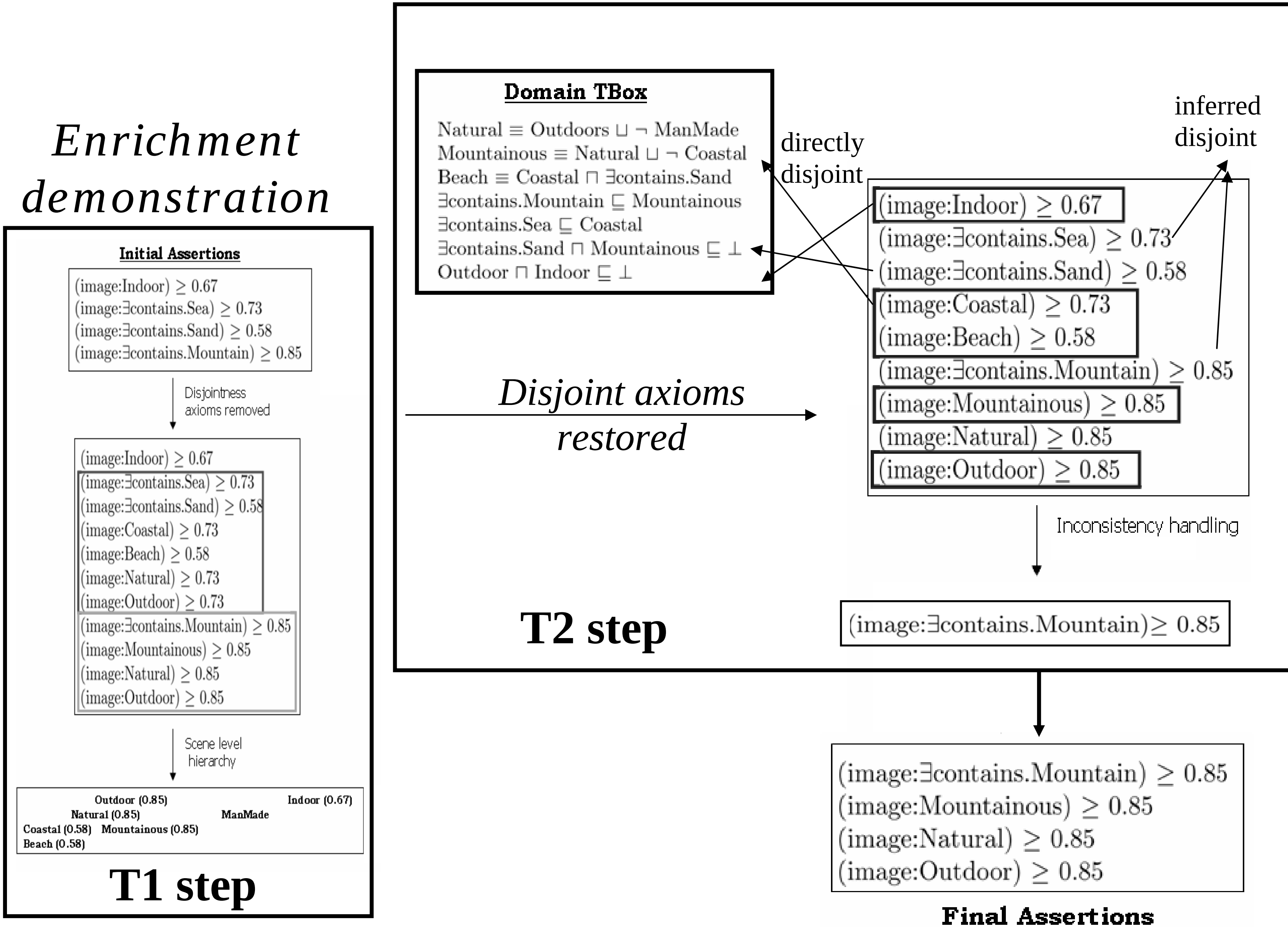
Tasks I & II from a more formal perspective (2)

- Traits
  - only assertions can be removed
    - axioms capture commonsense knowledge
  - consistency at scene level precedes object level consistency
    - first level: scene assertions
    - second level: object assertions
  - fuzzy assertions, i.e. “prioritised” facts
- Implementation
  - extends reverse tableaux-based methodologies with fuzzy information consideration
  - introduces a “stratified” perspective

Reasoning Task III

- ***Enrichment***
  - performs on the set of assertions maintained after step T2
  - Procedure
    - standard fuzzy DLs entailment

Enrichment demonstration



Experimental Results

- Domain of outdoor images (~360 images)
  - developed TBox
- Use of fuzzyDL(*) as inference engine for core fuzzy DLs reasoning services
- Evaluation
  - experiment I: loose semantic connection between scene and object concepts supported by analysis
  - experiment II: stronger semantic interrelations

(*) http://faure.isti.cnr.it/~straccia/software/fuzzyDL/fuzzyDL.html

Outdoor images TBox extract

$\text{Countryside\_buildings} \sqsubseteq \exists\text{contains.Buildings} \sqcap \exists\text{contains.Foliage}$
$\text{Countryside\_buildings} \sqsubseteq \text{Landscape}$
$\exists\text{contains.Forest} \sqcup \exists\text{contains.Grass} \sqcup \exists\text{contains.Tree} \sqsubseteq \exists\text{contains.Foliage}$
$\text{Rockyside} \sqsubseteq \exists\text{contains.Cliff}$
$\text{Rockyside} \sqsubseteq \exists\text{contains.Mountainous}$
$\text{Roadside} \sqsubseteq \exists\text{contains.Road}$
$\text{Roadside} \sqsubseteq \text{Landscape}$
$\exists\text{contains.Sea} \equiv \text{Coastal}$
$\text{Coastal} \sqsubseteq \text{Natural}$
$\exists\text{contains.Forest} \sqsubseteq \text{Landscape}$
$\text{Beach} \equiv \text{Coastal} \sqcap \exists\text{contains.Sand}$
$\text{Beach} \sqsubseteq \text{Natural}$
$\text{Cityscape} \sqsubseteq \text{ManMade}$
$\exists\text{contains.Sky} \sqsubseteq \text{Outdoor}$
$\exists\text{contains.Trunk} \sqsubseteq \exists\text{contains.Tree}$
$\text{Mountainous} \sqcap \text{Coastal} \sqsubseteq \bot$
$\text{Natural} \sqcap \text{ManMade} \sqsubseteq \bot$

Experiment I – Scene level concepts

	Analysis			Reasoning		
Concept	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**
Indoor	0.00	NaN	NaN	1.00	0.75	0.85
Outdoor	0.99	0.99	0.99	0.99	0.99	0.99
Natural	0.97	0.96	0.97	0.98	0.96	0.97
ManMade	0.18	0.40	0.25	0.18	0.40	0.25
Cityscape	0.18	0.40	0.25	0.18	0.40	0.25
Landscape	0.75	0.63	0.68	0.76	0.68	0.71
Mountainous	0.64	0.28	0.39	0.48	0.30	0.37
Coastal	0.00	NaN	NaN	0.86	0.49	0.63
Beach	0.89	0.30	0.45	0.90	0.31	0.47

Analysis extracted descriptions are ‘semantically treated’, i.e. detection of Beach is considered as positive detection of Outdoor also. Not much impact because of low semantic association between object level and scene level concepts.

Experiment I – Object level concepts

	Analysis			Reasoning		
Concept	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**
Building	1.00	0.17	0.29	0.09	0.83	0.17
Grass	0.06	0.40	0.10	0.01	1.00	0.03
Foliage	0.99	0.70	**0.82**	0.90	0.80	**0.85**
Sky	0.93	0.87	0.89	0.93	0.87	0.89
Cliff	0.98	0.21	**0.35**	0.54	0.42	**0.47**
Tree	0.22	0.65	0.33	0.18	0.58	0.27
Trunk	0.38	0.65	0.48	0.38	0.65	0.48
Sand	0.49	0.37	**0.42**	0.92	0.41	**0.56**
Sea	0.72	0.46	**0.56**	0.88	0.49	**0.63**
Conifers	1.00	0.01	**0.02**	0.50	0.02	**0.03**
Mountain	0.14	0.01	**0.01**	0.43	0.04	**0.06**
Boat	0.10	0.40	**0.16**	0.10	0.50	**0.17**
Road	0.15	0.50	0.23	0.02	0.25	0.03
Ground	0.06	0.57	0.19	0.11	0.57	0.19
Person	0.49	0.54	0.52	0.49	0.54	0.52

Concepts semantically related to scene level concepts are affected the most, e.g. the Sand concept. In general, precision is improved due to the utilisation of disjoint semantics.

Experiment II – Scene level concepts

	Analysis			Reasoning		
Concept	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**
Countryside_buildings	0.30	1.0	**0.46**	0.60	0.86	**0.71**
Rockyside	0.68	0.70	**0.69**	0.68	0.79	**0.74**
Roadside	0.68	0.69	**0.69**	0.68	0.72	**0.70**
Forest	0.75	0.63	**0.69**	0.74	0.68	**0.71**
Coastal	0.85	0.67	**0.75**	0.86	0.72	**0.78**
Outdoor	-	-	-	0.00	1.00	0.99
Indoor	-	-	-	NaN	NaN	NaN
Natural	-	-	-	0.97	1.00	0.98
ManMade	-	-	-	NaN	NaN	NaN
Cityscape	-	-	-	NaN	NaN	NaN
Mountainous	-	-	-	0.67	0.80	0.74
Beach	-	-	-	0.45	0.76	0.57

Higher impact as the analysis supported concepts are characterised are more strongly related to each other.

Experiment II – Object level concepts

	Analysis			Reasoning		
Concept	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**
Building	0.54	0.69	**0.60**	0.62	0.86	**0.72**
Roof	0.33	0.54	**0.41**	0.33	0.75	**0.46**
Grass	0.49	0.42	0.45	0.30	0.52	0.38
Foliage	0.48	0.84	**0.61**	0.86	0.86	**0.86**
Dried-Plant	0.07	0.11	**0.08**	0.07	0.13	**0.10**
Ground	0.26	0.33	0.29	0.26	0.33	0.29
Person	0.75	0.51	0.61	0.75	0.51	0.61
Sky	0.95	0.93	0.94	0.95	0.93	0.94
Cliff	0.65	0.45	**0.53**	0.69	0.70	**0.69**
Tree	0.49	0.52	0.51	0.56	0.47	0.51
Trunk	0.26	0.28	0.27	0.26	0.28	0.27
Sand	0.02	0.10	**0.03**	0.57	0.45	**0.50**
Sea	0.69	0.60	**0.64**	0.85	0.69	**0.76**
Wave	0.25	0.5	0.33	0.25	0.5	0.33
Boat	0.41	0.71	0.52	0.33	0.66	0.44
Road	0.50	0.69	**0.58**	0.69	0.71	**0.70**

Again, higher impact as the analysis supported concepts bear stronger semantic relatedness.

Interesting to note the lower performance for Boat, which is due to analysis mistaken degrees estimation of the scene level concepts

Some Observations

- The application of reasoning in general maintains or enhances performance w.r.t. analysis

- Diversity in classifiers performance
  - e.g. cliff detector is more effective than the rockyside one
    - trade-off: “classifier-customised” TBox vs generic applicable “commonsense” Tbox (*Rockyside* $\sqsubseteq \exists$*contains.Cliff instead of* $\exists$*contains.Cliff* $\sqsubseteq$ *Rockyside)*

- Discrepancies in initial confidence degrees
  - e.g. false high positives for rockyside scenes over coastal ones: may lead to unnecessary object assertions (e.g. the Boat concept)
    - hard to overcome without additional knowledge

Conclusions

- The proposed Fuzzy DLs reasoning enables
  - formal handling of annotations uncertainty semantics
  - utilisation of domain semantics
  - consistent interpretations / descriptions
- The use of explicit semantics is integral in multimedia semantics extractions; yet not the only necessary component
- Largely misestimated degrees can mislead the interpretation

Future Directions

- Investigation of additional knowledge
  - probabilistic information in the form of co-occurrence patterns
  - spatial relations among object level concepts (aligning different segmentation masks)
- Investigation of intermediate representation level
  - link domain definitions with qualitative visual features
    - inconsistent at domain level interpretations are not simply rejected
- Experimentation with descriptions coming from other than image analysis sources
  - text, tags (expressed in ontological terms)
    - provenance-based weights

Thank you for your attention!

Questions?

DLs in brief

- Family of knowledge representation languages characterised by **formal** semantics and **sound & complete** inference algorithms

- **Terminological Box** (TBox): vocabulary (concepts & roles) and interrelations describing the application domain
  - equivalence $\mathrm{Mother} \equiv \mathrm{Woman} \sqcap \exists\ \mathrm{hasChild.Person}$
  - subsumption $\mathrm{Tree} \sqsubseteq \exists\ \mathrm{hasPart.Leaf} \sqcap \exists\ \mathrm{hasPart.Trunk}$
  - complex descriptions inductively build with constructors

- **Assertional ABox** (ABox): facts describing a specific state of the application domain
  - concept assertions $\mathrm{Athlete(John)},\ \mathrm{Woman(Myriam)}$
  - role assertions $\mathrm{hasChild(Myriam{,}John)}$

DLs in brief (cont'd)

- Semantics
  - Interpretation I consists of a non-empty set Δ^I
  - Interpretation function maps each C to $C^I \subseteq \Delta^I$each role to
  $R^I \subseteq \Delta^I \times \Delta^I$d each individual to an object $a^I \in \Delta^I$

- Inference services for TBoxes
  - Satisfiability (is$C^I \neq \oslash$e.g. $\mathrm{Mother} \sqcap \neg \mathrm{Mother}$tisfiable)
  - Subsumption (is$C^I \subseteq D^I$, e.g. $\exists hasChild.Male \sqsubseteq \exists.hasChildPerson$
  - Equivalence (if$C^I \equiv D^I$
  - Disjointness

- Inference services for ABoxes
  - Consistency
  - Entailment (instance checking)