Investigating Fuzzy Reasoning in Semantic Image Analysis

S. Dasiopoulou · **I. Kompatsiaris** · **M.G. Strintzis**

Abstract Recent advances in semantic image analysis have brought forth generic methodologies to support concept learning at large scale. The attained performance however is highly variable, reflecting effects related to similarities and variations in the visual manifestations of semantically distinct concepts, much as to the limitations issuing from considering semantics solely in the form of perceptual representations. Aiming to enhance performance and improve robustness, we investigate a fuzzy DLs-based reasoning framework, which enables the integration of scene and object classifications into a semantically consistent interpretation by capturing and utilising the underlying semantic associations. Evaluation with two sets of input classifiers, configured so as to vary with respect to the wealth of concepts' interrelations, outlines the potential of the proposed approach in the presence of semantically rich associations, while delineating the issues and challenges involved.

Keywords fuzzy reasoning, semantic image analysis, semantic integration, fuzzy DLs, inconsistency handling

1 Introduction

Extracting semantic image descriptions is an intricate problem, challenging researchers for decades in the quest for generalisable, yet robust, approaches to alleviate the so called *semantic gap* and support content management services at a level closer to user needs [1–3]. A factor partially accountable for the challenges pertaining to this endeavor lies in the very nature of it, namely in the fact that it involves the handling of significant amount of imprecise and incomplete information. Leaving out issues related to the subjective interpretations that different users may attribute to conveyed meaning, intention, etc., imprecision relates intrinsically to a number of tasks including segmentation, feature extraction, intra- and inter-class variability. Inevitably, managing imprecision is critical not only when dealing with the aforementioned tasks, but also for the subsequent processing that realises the extraction of semantic descriptions.

Informatics and Telematics Institute
Centre for Research and Technology Hellas, Thessaloniki, Greece
E-mail: (dasiop,ikom,strintzis)@iti.gr

Towards this direction, machine learning approaches have gained growing popularity in the past couples of years, as they provide convenient means for discovering and handling knowledge either incomplete or too complex to be explicitly handled. Support Vector Machines (SVMs) and Bayesian networks constitute characteristic examples in this direction, allowing one to learn in a generic fashion classifiers for a significant number of concepts (referring to objects, events, scene descriptions, etc.) [4–7]. Despite the reports of successful applications, the performance remains still highly variable and deteriorates rather dramatically as the number of supported concepts increases. Among the principal causes for the variable performance are effects related to similarities between visual manifestations of semantically distinct concepts and to the variance in the possible manifestations a single concept may have. These effects issue directly to a large extent from the fundamental assumption underlying learning approaches, i.e. that the addressed semantics is captured to a satisfactory degree by features pertaining to their visual manifestations.

Since in many cases semantics goes beyond the capacity of perceptual features, discrepancies result between the learned associations and the intended ones. Consequently, the learnt classifiers often result in the extraction of complementary, overlapping, incomplete, as well as conflicting descriptions. Evidence can be found not only in individual evaluations of research studies in the relevant literature, but also in large scale benchmarks, such as the TRECVID challenge [8], sponsored by the National Institute of Standards and Technology (NIST). Two main challenges confronted repeatedly in the series of annual TRECVID evaluation activities include the deterioration of performance as the number of addressed concepts increases and the variable correspondence between the semantics of the content retrieved using the learned detectors and the semantics alleged by the detectors per se [9, 10]. The efforts during the last edition included the enhancement of robustness, even for a reduced number of concepts, through the utilisation of complementary information beyond visual features, either in the form of taxonomic relations as captured in the LSCOM ontology [11] or through detector combinations [12, 13].

Induced by the aforementioned, we investigate the utilisation of formal semantics in order to interpret the outcome of statistically learned classifiers into a semantically consistent image interpretation. Focusing on approaches deploying perceptual similarity against learned concept models, viz. where the confidence of the extracted descriptions reflects the membership to a concept class, we propose a fuzzy Description Logic (DL) based framework to capture and reason over the extracted descriptions, while handling the underlying vagueness. The input image classifications may be either scene or object level, and their interpretation is realised in three steps, namely designation of scene level characterisation, identification and resolution of inconsistencies with respect to possibly conflicting classifications, and enrichment, where additional inferred descriptions are made explicit.

The rest of the paper is structured as follows. Section 2 explicates the reasons that motivated our investigation into a fuzzy DL based reasoning approach and outlines the contribution with respect to the relevant literature. Section 3 provides a brief introduction into fuzzy DLs and delineates features pertaining to reasoning within the context of image interpretation. Section 4 presents the proposed framework architecture and the details of the individual reasoning tasks involved. Evaluation is given in Section 5, where the proposed framework is assessed against two experimental settings that consider input classifications of loose and rich semantic associations. Additionally, initial results are presented with respect to enriching the background knowledge with co-occurrence information in the form of fuzzy implication. Finally, Section 6 summarises the paper and discusses future directions.

2 Motivation and Contribution

Learning based approaches provide a number of appealing traits, such as generic learning mechanisms and the capability to capture and utilise associations hidden in the examined input data whose explicit handling might otherwise be too complex and strenuous to be effective. Despite their unquestionable role in semantic image analysis tasks, one cannot overlook the limitations that conduce to the variability of the attained performance. As mentioned previously, the problem lies to a large extent in the rather poor utilisation of the semantics underlying the learned concepts and in the imprecision involved in the constituent tasks. Aiming to enhance performance, we focus on a twofold goal, addressing the utilisation of concept semantics while providing the means to cope with imprecision.

The use of explicit knowledge for the purpose of introducing concept semantics is not a new idea. Going back to the 80s and early 90s one finds an extensively rich literature [14, 15] that investigates a wide gamut of knowledge representation schemes, as rigorous as first order logic [16] and as intuitive as the early semantic networks [17]. Among the weaknesses of the early knowledge-directed paradigms were the lack of common representations and reasoning mechanisms that prohibited interoperability and reuse of knowledge. The Semantic Web (SW) initiative[1] changed the scenery, advocating explicit semantics and corresponding representation languages to capture meaning in a formal and interoperable fashion. Since 2004, the Resource Description Framework Schema (RDFS) [18] and the Web Ontology Language (OWL) [19] constitute formal W3C recommendations, while substantial interest has revived in Description Logics (DLs), as it underpins the semantics of the SW languages.

In analogy to the different expressivity features provided by the various knowledge representation formalisms, a critical factor regarding the handling of imprecision concerns the nature of its semantics. Examining the relevant literature, imprecision may manifest in the extracted descriptions either as uncertainty regarding the presence of a specific entity (e.g. the presence of a sky region), or as vagueness regarding the degree to which the statement about the presence of the entity is true. Uncertainty characterises probabilistic approaches such as Bayesian nets, while vagueness is encountered in approaches that deploy similarity (distance) metrics against feature models acquired through training, such as SVMs. The confidence values of the latter reflect the extent of matching against the learned concept models, and as such each model can be taken as a fuzzy set, with distance metrics serving the role of the membership function.

Based on the aforementioned, our initial specifications consisted in the selection of a representation formalism with well-defined semantics and the designation of the targeted imprecision semantics. Perceptual-based similarity approaches constitute a fundamental element in statistical learning for image interpretation, and as such in the current investigation we focused on vague rather than probabilistic information. In this context, the concrete incentives of the proposed fuzzy DLs based reasoning framework issue form the specific traits characterising the application of SVM-based concept classifiers, namely contradictory descriptions that pertain to semantically different interpretations and incomplete descriptions, even at the presence of corresponding classifiers.

To put into perspective in an intuitive manner the implications involved, let us consider two example images and the respective descriptions extracted using the SVM based classifiers of [20], shown in Fig. 1. The extracted descriptions are expressed following the fuzzy DL notation. A statement of the form $(im : \exists contains.Concept_i) \geq n_i$ denotes that the image represented by the individual im contains a region depicting an instance of $Concept_i$ with a

---

$(image : Countryside\_buildings) \geq 0.65$
$(image : Roadside) \geq 0.57$
$(image : Rockyside) \geq 0.44$
$(image : Forest) \geq 0.45$
$(image : Seaside) \geq 0.47$
$(image : \exists contains.Sand) \geq 0.66$
$(image : \exists contains.Sky) \geq 0.95$
$(image : \exists contains.Person) \geq 0.62$
$(image : \exists contains.Foliage) \geq 0.70$

$(image : Rockyside) \geq 0.42$
$(image : Countryside\_buildings) \geq 0.52$
$(image : Seaside) \geq 0.51$
$(image : Forest) \geq 0.52$
$(image : Roadside) \geq 0.71$
$(image : \exists contains.Sky) \geq 0.98$
$(image : \exists contains.Sea) \geq 0.73$
$(image : \exists contains.Person) \geq 0.60$
$(image : \exists contains.Sand) \geq 0.75$

Fig. 1 Example outdoor images and respective descriptions extracted following SVM-based analysis.

degree $\geq n_i$, while a statement $(im : Concept_j) \geq n_j$ denotes that the image at scene level constitutes an instance of $Concept_j$ to a degree $\geq n_j$. As illustrated more than one scene level concepts may be assigned to a single image, and the same may be the case for the concept classifiers performing at object level. Furthermore, we note that since object level classification is performed per image segment, it is possible to have more than one assertions referring to the same concept with varying degrees.

In the case of the first image, the scene level concept *Countryside_buildings* appears as the prevalent one. This is not only because the corresponding scene description is the one with the highest degree among the extracted scene level descriptions, but also because the object level descriptions neither contradict it nor imply preference of any other scene description. To assess the latter, one needs to be aware of the logical associations of the concepts involved. Even in this example, where the utilisation of semantics has the same effect as selecting the description with the highest confidence, the potential for additional enhancement is manifested. Specifically, exploiting the semantics of the *Countryside_buildings* concept, one can infer that the image contains a building depiction, thus that there is a missing assertion that needs to be added. Analogously, one can determine that the extracted *Sand* description is semantically unrelated, and hence should be removed.

Moreover, the exploitation of semantics enables the overall image description to be further enriched, allowing the addition of descriptions referring to concepts such as *Landscape* and *Outdoor*. The latter, though quite trivial for the considered example, can be of great value in the general case as it alleviates the need for training and learning detectors for all concepts of interest, as long as they are semantically connected to other concepts. We note that in the absence of confidence degrees, it would not be possible to acquire ranked estimations regarding the plausibility of the different interpretations. As a result, on one hand all classifications would be rendered equally probable, and preference could be established only on the grounds of the descriptions that would have to be removed (respectively added), risking rather artificial interpretations. On the other hand, depriving all information about perceptual similarity would leave out substantial knowledge regarding the validity of the choices considering the removal or addition of descriptions.

Similar considerations apply for the case of the second image. Taking into account solely the explicit scene level descriptions, the *Roadside* assertion appears to be the more plausible. Going through the object level descriptions and taking into account the semantics of the concepts involved, the simultaneous presence of *Sea* and *Sand* descriptions entails the presence of a *Beach* scene, an implication that implicitly contradicts the dominant *Roadside* assertion. The remaining object level classifications, i.e. the *Person* and *Sky* ones, do not provide any additional knowledge regarding the plausibility of scene level descriptions. Hence, since the implied confidence with respect to *Beach* appears to be larger than the told degree of *Roadside*, it would be desirable to be able to assess *Beach*, and by consequence *Seaside*, as the more plausible image interpretations.

The aforementioned examples, though quite simplistic, outline the potential and motivation for employing explicit semantics while providing the means to handle imprecision. Furthermore, they designate additional specifications regarding the selection of an appropriate knowledge representation formalism, namely the ability to capture expressive semantic associations (including disjointness, subsumption, conjunction) and to handle the vagueness introduced by the accompanying degrees. Based on these considerations, we chose to investigate a reasoning framework based on fuzzy Description Logics (DLs) [21–23] as they present two very appealing traits. First, they are strongly related to OWL (particularly the Lite and DL species); thus they benefit from the Semantic Web initiative towards knowledge sharing, reuse and interoperability, both regarding the background knowledge comprising the domain semantics and the produced image annotations. Second, extending classical DLs, they provide the means to handle vagueness in a formal way, while providing for well-defined reasoning services. The current stage of the relevant literature, presented in the next subsection, elucidates further the motivation underlying the proposed reasoning framework, and outlines its contribution.

2.1 Relevant Work

Going through the relevant literature, one notices that despite the significant amount of imprecision involved, a substantial share of the reported logic-based investigations towards the use of explicit knowledge adopts crisp approaches. In the series of works presented in [24, 25], crisp DLs are proposed for inferring complex descriptions that are modelled in the form of aggregates, i.e. conceptual entities modelled by parts satisfying specific constraints. Interpretation is realised as a deductive process, assuming the availability of primitive descriptions that do not contradict each other. Crisp DLs are also considered in [26], in combination with rules, for video annotation; again, possible conflicts or missed descriptions are not addressed. In [27] DLs are used to interpret perceptual descriptions pertaining to colour, texture and background knowledge into semantic objects. To this end a pseudo fuzzy algorithm is presented to reason over the calculated feature values with respect to the prototypical values constituting the definition of semantic objects. In [28], DLs have been combined with rules in order to perform abductive inference over crisp descriptions and acquire plausible interpretations, which are ranked using cost criteria involving the number of assertions that can be explained and the number of assertions that need to be hypothesised.

There also exist ontology-based approaches that focus more on formalising the transition from low-level descriptors to domain concepts include [29],[30] and [31]. Ontology languages are used to represent both domain specific concepts and visual features such colour, texture and shape, and to link them in a formal fashion. Although such approaches can be very useful for purposes of sharing and reusing knowledge regarding low and high level

relations, the reasoning supported with respect to logical relations among the high-level concepts is not addressed. This is not a matter of the limited datatype support currently provided by ontology languages, but in principle because of the non logical nature of the problem at hand, i.e. the estimation of distance between a given data structure that constitutes a feature model and the measurable feature values.

Indicative works addressing probabilistic information include the following. In [32], a probabilistic approach is suggested as a possible solution to the handling of the ambiguity introduced during the analysis stage; however, no description of relevant experimentation or evaluation is provided. Other works that address probabilistic knowledge include Markov logic networks [33], where first order logic is combined with graphical probabilistic models into a uniform representation; although tested on knowledge about the university domain, the experiences drawn are applicable to information extractions tasks such as image interpretation too. In [34], a Bayesian approach is proposed in order to model and integrate probabilistic dependencies among aggregates connected by hierarchical relations, within a DLs scene interpretation framework; however, no evaluation results are available as implementation is still under way. In [35], a probabilistic framework for robotic applications is presented that tackles uncertainty issuing from noisy sensor data and missing information, employing relational hidden Markov models for the purpose of spatio-temporal reasoning.

We stress out again the discrimination between the probabilistic and the fuzzy perspective, as this is critical for the motivation and grounding of the reasoning approach proposed in this paper. A probability of 0.5 regarding a positive sea classification denotes our ignorance with respect to its presence or not; it does not imply anything about how blue the sea may be. A degree of 0.5 on the other hand, denotes how close the colour of this sea region is with respect to the specific colour attached to the notion of sea in the given context. For further reading on the two perspectives the reader is referred to [36]. Consequently, approaches such as the aforementioned ones that address probabilistic knowledge are in fact complementary to the aspects considered in this paper.

Finally, there are works utilising fuzzy DLs, and in this sense closer to the proposed fuzzy DLs based framework. Specifically, fuzzy DLs have been proposed in [37] for the purpose of semantic multimedia retrieval; the fuzzy annotations however are assumed to be available. In the context of semantic analysis, fuzzy DLs have been only recently explored in [38], where fuzzy DLs are used for inferring semantic concepts based on part-of relations and merging of the respective image regions, and in [39] for document classification; none of the approaches however addresses the possibility of inconsistency in the initially extracted descriptions.

Given the aforementioned, the contribution of the fuzzy DLs based reasoning framework presented in this paper can be summarised in the following.

- The uncertainty, in the form of vagueness, which characterises learning-based extracted descriptions is formally handled and intergraded in the process of image interpretation. Thus, the valuable information encompassed in the extracted degrees can be utilised for computing and ranking the different plausible interpretations.
- The inconsistencies, explicit or implicit, which issued from conflicting descriptions are identified and resolved, contrary to the assumption taken by relevant works regarding consistent initial classifications.

3 Fuzzy DLs

Description Logics (DLs) [40] is a family of knowledge representation formalisms charac-terised by logically grounded semantics and well-defined inference services. Starting from the basic notions of atomic concepts and atomic roles, arbitrary complex concepts can be described through the application of constructors (e.g., $\neg$, $\sqcap$, $\forall$). Terminological axioms (TBox) allow to capture equivalence and subsumption semantics between concepts and re-lations, while real world entities are modelled through concept $(a : C)$ and role $(R(a,b))$ assertions (ABox). The semantics of DLs is formally defined through an interpretation I. An interpretation consists of an non-empty set Δ^I (the domain of interpretation) and an interpretation function $.^I$, which assigns to every atomic concept A a set $A^I \subseteq \Delta^I$ and to every atomic role R a binary relation $R^I \subseteq \Delta^I \times \Delta^I$. The interpretation of complex concepts follows inductively [40].

In addition to the means for representing knowledge about concepts and assertions, DLs come with a powerful set of inference services that make explicit the knowledge implicit in the TBox and ABox. *Satisfiability*, *subsumption*, *equivalence* and *disjointness* constitute the core TBox inferences. Satisfiability allows to check for concepts that correspond to the empty set, subsumption and equivalence check whether a concept is more specific or re-spectively identical to another, while disjointness refers to concepts whose conjunction is the empty set. Regarding the ABox, the main inferences are *consistency checking*, which as-sesses whether there exists a model that satisfies the given knowledge base, and *entailment*, which checks whether an assertion ensues from a given knowledge base.

Fuzzy DLs consider the extension of DL languages with fuzzy set theory [41,42]. More specifically, in the case of a fuzzy DL language, a TBox is defined as a finite set of fuzzy concept inclusion and equality axioms, while the ABox consists a finite set of *fuzzy asser-tions*. A fuzzy assertion [21] is of the form $a : C \bowtie n$ and $(a,b) : R \bowtie n$, where $\bowtie$ stands for $\geq$, $>$, $\leq$, and $<^2$. Assertions defined by and $\geq$ are called positive assertions, while assertions defined by and $\leq$ are called negative.

A fuzzy set $C \subseteq D$ is defined by its membership function (μ_C), which given an object of the universe D returns the membership degree of that object with respect to the set C. By using membership functions, the notion of the interpretation function is extended to that of a *fuzzy interpretation function*. In accordance to the crisp DLs case, the fuzzy interpretation function is a pair $I = (\Delta^I, .^I)$ where Δ^I is a non-empty set of objects called the domain of interpretation, and $.^I$ is a *fuzzy*, this time, interpretation function which maps: an individual a to an element $a^I \in \Delta^I$, i.e., as in the crisp case, a concept name A to a membership func-tion $A^I : \Delta^I \to [0,1]$, and a role name R to a membership function $R^I : \Delta^I \times \Delta^I \to [0,1]$ [21, 22]. Several fuzzy operators exist in the literature that implement the functions (t-norms, co-norms, negation and implication) which extend the classical Boolean conjunction, disjunc-tion, negation and implication to the fuzzy case. Table 1 shows the corresponding semantics for the language *f*-SHIN under Zadeh logic. The reasoning services definitions are adapted analogously. For example, concept satisfiability requires the existence of an interpretation under which there will be an individual belonging to this concept with a degree $n \in (0,1]$.

Two main efforts exist currently that address formally both the semantics and the cor-responding reasoning algorithms. In [23,43,44], the DL language *SHIN* has been extended according to fuzzy set theory leading to the so called f-*SHIN*. The fuzzy extensions address the assertion of individuals and the extension of the language semantics. In [45], a fuzzy

---

2 Intuitively a fuzzy assertion of the form $a : C \geq n$ means that the membership degree of the individual a to the concept C is at least equal to n.

Table 1 Fuzzy interpretation of DL constructors under Zadeh semantics [23].

$$\top^I = 1$$
$$\perp^I = 0$$
$$(\neg C)^I = 1\text{-}C^I(d)$$
$$(C \sqcap D)^I = \min\{C^I(d), D^I(d)\}$$
$$(C \sqcup D)^I = \max\{C^I(d), D^I(d)\}$$
$$(\forall R.C)^I = \inf_{d' \in \Delta} max\{1 - R^I(d,d'), C^I(d')\}$$
$$(\exists R.C)^I = \sup_{d' \in \Delta} min\{R^I(d,d'), C^I(d')\}$$

extension of *SHOIN(D)* is presented, which constitutes a continuation of earlier works of the authors on extending *ALC*, *SHIF*, and *SHIF(D)* to fuzzy versions [21, 22]. In addition to extending the *SHOIN(D)* semantics to f-*SHOIN(D)*, the authors present a set of interesting features: concrete domains as fuzzy sets, fuzzy modifiers such as *very* and *slightly*, and fuzziness in entailment and subsumption relations.

Additionally to the theoretic foundations for the fuzzy extensions, respective reasoning algorithms have been presented and implemented, namely the *Fuzzy Reasoning Engine*[3] (FiRE) and the *fuzzyDL*[4] engine. FiRE [38] supports querying an f-*SHIN* knowledge base for satisfiability, consistency, subsumption, and entailment, under Zadeh's semantics; general concept inclusions, roles and datatype support are among the planned future extensions. fuzzyDL [46] supports fuzzy *SHIF* semantics extended by the features aforementioned. It supports the Łukasiewicz's and Gödel's t-norm, t-conorm and fuzzy implication, as well as the Kleene-Dienes implication.

As already implied by the examples presented in Section 2, fuzzy DLs can be straightforwardly used to manage the outcome of classifiers with respect to background domain knowledge. Using the available constructors and corresponding inclusion and equality axioms, one can construct the terminological knowledge (TBox) that captures the semantics of the examined application domain. The assertional knowledge (ABox) that includes the fuzzy assertions corresponds to the analysis extracted descriptions. Typical DL inference services can be used afterwards to check whether the extracted descriptions violate the logical model of the domain leading to an inconsistency, or to compute the concepts to which a certain image instance belongs to.

However, in order to address the issues described in Section 2, namely the determination of the more plausible scene level description, the subsequent removal of incoherent descriptions and further enrichment of the overall interpretation, a reasoning framework coordinating the evoked DL inferences is needed. Towards this end, the handling of inconsistency constitutes a central requirement. As illustrated in Section 2, such inconsistency may refer to assertions at the same level of granularity, that is between scene level descriptions or between object level description, or to assertions pertaining to different levels. Handling inconsistencies in DLs knowledge bases usually refers to approaches targeting revision of the terminological axioms [47, 48]. In the examined problem however, the inconsistencies result from the limitations in associating semantics with visual features. Thus it is the ABox that needs to be appropriately managed. The adopted methodology is described in the following section, where the proposed fuzzy DLs based reasoning framework is detailed.

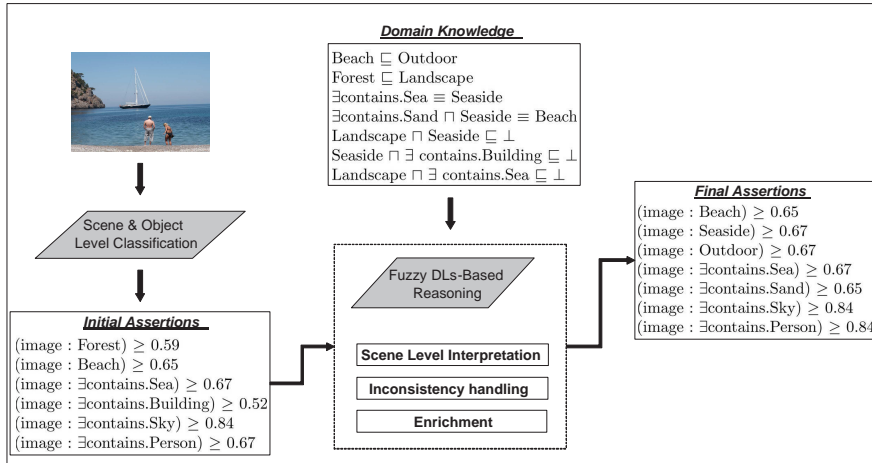
[3] http://www.image.ece.ntua.gr/~nsimou
[4] http://faure.isti.cnr.it/~straccia/software/fuzzyDL/fuzzyDL.html

Fig. 2 Architecture of the proposed fuzzy DLs based reasoning framework.

4 A Fuzzy DL-based Reasoning Framework for Enhancing Semantic Analysis

Figure 2 depicts the architecture of the proposed fuzzy DLs based reasoning framework. The input consists of a set of scene and object level assertions representing the descriptions extracted through the application of learning based analysis techniques. No assumptions are made with respect to the particular implementation followed. The interpretation of the initial assertions consists in their semantic integration into a consistent overall image interpretation, and is realised in three steps. First, the more plausible scene level descriptions are determined, by utilising the subsumption relations among the considered scene level concepts. Next, the inconsistencies in the initial descriptions with respect to the previously computed scene level interpretation are resolved, resulting in a ranked list of plausible interpretations. The highest ranked interpretation is passed to the final step, where by means of logical entailment, inferred assertions are made explicit.

Notably, the first two steps constitute a special case of knowledge integration, as only assertions, but no axioms, are allowed to be removed. The reason for the latter, as also discussed in Section 5, lies in the fact that the axioms address domain-specific yet generic knowledge, rather than conceptualisations customised to specific traits of the examined dataset. As knowledge integration and inconsistency resolving for DLs ontologies have been extensively treated in the literature [49, 50], the approach followed within the proposed reasoning framework, builds on the grounds of existing results by appropriately adjusting them in order to meet the peculiarities introduced within the examined application context. In the sequel, the details of the individual tasks are given.

4.1 Scene level interpretation

The possible logical associations between concepts referring to object level and concepts referring to scene level induce that all assertions need to be taken into account in order to infer the degrees of confidence pertaining to scene level descriptions. To accomplish this,

Table 2 Scene level interpretation.

Scene level interpretation algorithm
Input: scene level concepts hierarchy H_{SC}, input assertions A (scene and object level)
Output: glb for the path of scene level concept hierarchy with highest confidence

  1: populate scene concept hierarchy (compute inferred degrees)
  2: for all hierarchy levels L_i starting from the most specific
  3:   for all scene level concepts $SC_{ij} \in L_i$
  4:     if (disjoint SC_{ik} SC_{il} ...){
  5:       add assertion SC_{ij} with $d_{ij} = \max\{d_{ik}, d_{il}, ...\}$ to *scene interpretation list*
  6:       move remaining SC_{ik}, k≠j, assertions to *inconsistent scene concept list*
  7:     }
  8:     if L_i different than lower hierarchy level{
  9:       if $SC_{ij} \in L_i$ is disjoint to $SC_{(i-1)j} \in L_{i-1}$
10:         move $SC_{(i-1)j}$ and its subsumes to *inconsistent scene concept list*
11:       if $SC_{ij} \in L_i$ subsumes $SC_{(i-1)j} \in L_{i-1}$
12:         if $d_{ij} \ ¡ \ d_{(i-1)j}$
13:           $d_{ij} := d_{(i-1)j}$
14:     }
15: return *scene interpretation list*

all disjointness axioms are removed. Thereby, we allow for the exploitation of all interrelations implicit in the initial descriptions. The next step consists in the bottom-up traversal of the scene level concept hierarchy for the purpose of determining the more plausible scene descriptions.

More specifically, at each level of the hierarchy, we check whether the concepts for which (told or inferred) assertions exist are disjoint. In the presence of disjointness, the assertion with the highest degree of confidence is preserved as more plausible, while the rest are marked as inconsistent. Once the examination at a given level is completed, consistency is examined between the set of currently selected assertions SA_i and those computed at previous hierarchy level SA_{i-1}. If an assertion a of SA_{i-1} contradicts the assertions belonging in SA_i, then a is marked as inconsistent. The process is repeated until the root of the hierarchy is reached and results in discriminating the scene level concept assertions into two sets, namely the set of scene descriptions comprising the more plausible interpretation, and the set of scene descriptions marked as inconsistent. Table 2 summarises the described procedure, where S denotes the list of scene level concepts and A the complete list of assertions extracted by means of analysis.

The reason for prioritising scene level assertions over object level ones lies in the different nature of the semantics involved. Scene level concepts, and to be more precise sub-trees in the scene level concept hierarchy, comprise mutually excluding sub-domains. Contrariwise, multiple configurations of assertions at object level may comply with a specific sub-domain. To illustrate this in a more intuitive manner, consider an image for which the extracted descriptions are: $(im : \exists contains.Sea) \geq 0.67$, $(im : \exists contains.Sand) \geq 0.74$ and $(im : Rockyside \geq 0.85)$, and assume a TBox consisting of the axioms $Beach \equiv (\exists contains.Sea) \sqcap (\exists contains.Sand)$ and $Beach \sqcap Rockyside \sqsubseteq \perp$. Both interpretations pertaining to the *Rockyside* sub-domain, i.e. { $(im : \exists contains.Sea) \geq 0.67$, $(im : Rockyside \geq 0.85)$ }, and { $(im : \exists contains.Sand) \geq 0.74$, $(im : Rockyside \geq 0.85)$ } rank higher than the one pertaining to beach sub-domain, i.e. { $(im : \exists contains.Sea) \geq 0.67$, $(im : \exists contains.Sand) \geq 0.74$, $(im : Beach \geq 0.67)$ }, in terms of plausibility as captured in the corresponding degrees.

Table 3 Inconsistency handling.

Inconsistency tracking and resolution algorithm
Input: scene interpretation list SL, input assertions A (scene and object level), domain TBox T
Output: ranked alternative interpretations

1: for all scene level concepts $SC_i \in SL$
2:    if $C_j \sqcap SC_i \sqsubseteq \bot \in T$
3:      add $C_j \sqsubseteq noSC_i$ in T and C_j to *disjoint list*
4: for all $noSC_i \in T$
5:    add $noSC_i \sqsubseteq noConcept$ in T
6: for all C_j for which $C_j \sqcap SC_i \sqsubseteq \bot \in T$
7:    remove told assertions
8: while $(T,A) \models im : NoConcept$ {
9:    apply expansion rules (Table 4)
10:      resolve inconsistencies arising from disjoint object concepts
11:      resolve inconsistencies arising from disjoint scene concepts
12: }
13: compute number of assertions removed and average their degrees
14: return ranked list of alternative consistent interpretations

4.2 Inconsistency handling

Having computed the most plausible scene level interpretation, the next step is to resolve the possible inconsistencies. In the current implementation, the regions depicting an object level concept are represented implicitly through statements of the form $(im : \exists contains.Concept_i) \geq n_i$, and thus inconsistency may arise either among scene level assertions or between scene and object level assertions. The first step towards the identification of inconsistencies is to restore the disjointness axioms that were removed during the scene interpretation task. In order to avoid halting the reasoner, the disjointness axioms are rewritten with respect to the scene level concepts selected at the previous step in the form of subsumption relations to respective *no-concepts*. Practically, given a set of selected scene level concepts $\{S_i, S_j, ..\}$, each axiom of the form $C_k \sqcap S_j \sqsubseteq \bot$ is translated into $C_k \sqsubseteq noS_j$. Furthermore, for all noS_j concepts we introduce an axiom of the form $noS_j \sqsubseteq NoConcept$. Consequently, resolving the possible inconsistencies amounts to tracking the assertions that led to the inference of *NoConcept* instances.

First, we address inconsistencies incurred directly by told descriptions. This translates into checking whether there exist asserted individuals that refer to C_k concepts participating in the introduced $C_k \sqsubseteq noS_j$ axioms. The handling of such assertions is rather straightforward and results in their immediate removal. Clearly, addressing asserted individuals first, prunes the search space during the subsequent tracking of the inferences that lead to an inconsistency.

Next, we consider assertions referring to complex, object and scene level, concepts. Again we start with object level concepts in order to account for cases of conflicts at scene level that are triggered through inference upon object level assertions. Contrary to the previous case, we now need to analyse the involved axioms in order to track the asserted descriptions that cause the inconsistency. Furthermore, these axioms determine which of the descriptions should be removed so as to reach a consistent interpretation. To accomplish this, we build on the grounds of relevant works presented for resolving unsatisfiable DL ontologies [51] and employ a reversed tableaux expansion procedure. We note that the main

⊓-rule	if $(a : C_1 \sqcap C_2) \in L(x)$
	then $L(y)=L(x)\setminus \{(a:C_1)\}$ and
	$L(z)=L(x)\setminus\{(a:C_2)\}$ and
	$L(w)=L(x)\setminus\{(a:C_1 \sqcap C_2)\}$
⊔-rule	if $(a : C_1 \sqcup C_2) \in L(x)$
	then $L(x)=L(x)\setminus\{(a:C_1 \sqcup C_2)\}$
where $C_i \longrightarrow$	A | C ⊓ D | ∃R.D

Table 4 Expansion rules for computing the alternative sets of consistent assertions.

difference with respect to the relevant literature is that in our application framework, we consider solely the removal of assertions, rather than the removal or weakening of terminological axioms. Table 3 summarises the procedure for handling inconsistencies while Table 4 summarises the expansion rules.

The expansion procedure starts having as root node the $(im : NoConcept \geq d_i\}$ assertion, where d_i is the computed inferred degree, and continues until no expansion rule can be applied. As illustrated, in the case of inconsistencies caused by axioms involving the conjunction of concepts, there is more than one way to resolve the inconsistency and reach a consistent interpretation. Specifically, there are many alternative interpretations as the number of all possible disjunctions of size $k = 1,..,N$, where N is the number of conjuncts. In order to choose among them, we rank the set of solutions according to the number of assertions that need to be removed and the average value of the respective degrees.

We note that since all role assertions referring to the role *contains* are assumed to hold with a degree ≥ 1.0, and the regions that depict object level concepts are not explicitly represented, expansion rules are required only for the case of the ⊓ and ⊔ constructors. Consequently, in the case of a domain TBox that utilises additional constructors or when fuzzy role assertions are also considered, the expansion rules would have to be appropriately extended.

4.3 Enrichment

The final step considers the enrichment of the descriptions by means of entailment and is the most straightforward of the three tasks, as it amounts to typical fuzzy DLs reasoning. Once the scene level concepts are selected, and the assertions ensuing inconsistencies either directly or through complex definitions, are resolved, we end up with a semantically consistent set of assertions, whose communication to the fuzzy DLs reasoning engine results in the acquisition of the final semantic image description. To render the inferred descriptions explicit, corresponding queries are formulated and the responses constitute the final outcome of the proposed framework.

Figure 3 displays an example application of the proposed reasoning framework. As illustrated, the applied concept detectors assess the image as both a *Rockyside* and *Seaside* scene. The respective object level descriptions include the concepts *Sky*, *Sea* and *Sand*. Removing the disjointness axioms from the TBox, we obtain the inferred assertions for the concepts *Beach*, *Landscape* and *Outdoor*, where the inferred assertions are grouped with respect to the asserted ones that triggered their inference. The left bottom part of Fig. 3 depicts the scene level concept hierarchy that corresponds to the depicted domain TBox extract, as well as the inferred degrees.
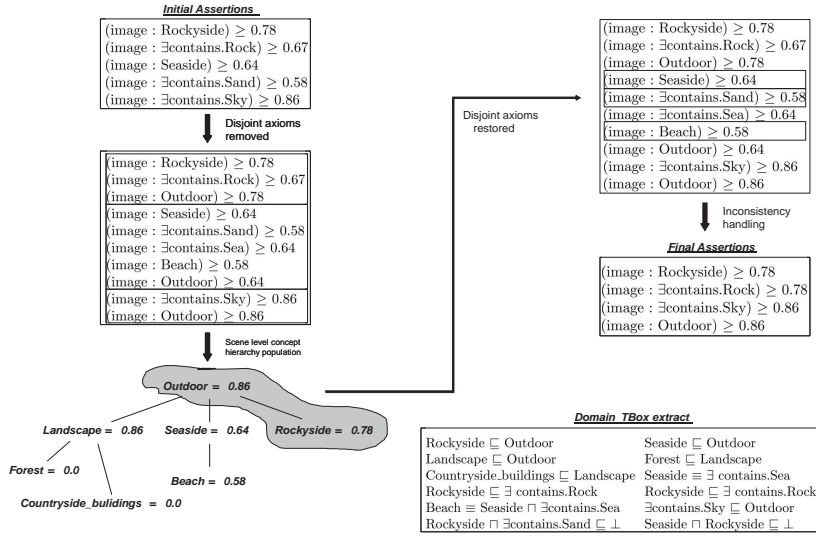
Fig. 3 Example application of the proposed fuzzy DLs based reasoning framework.

Following the afore described procedure, we start from the lowest level of the hierarchy and add the *Beach* concept assertion to the list of plausible scene descriptions. Moving to the next level, there are assertions referring to the *Rockyside* and *Seaside* concepts. Since the two concepts are disjoint, we select the assertion with the highest degree, namely the one referring to the *Rockyside* concept, and add it to the list of plausible scene descriptions. Checking against the previous level, it is the case that *Rockyside* and *Beach* are disjoint. Thus, the *Beach* assertion is removed from the list of plausible descriptions and marked as inconsistent. At the next step, the top level of the hierarchy is reached, which causes the addition of the *Outdoor* concept assertion in the scene interpretation.

Hence, upon the completion of the scene level interpretation step, the scene level concept hierarchy has been populated, and the sub-tree with the highest degrees is identified, which in our example amounts to the concepts *Rockyside* and *Outdoor*. The inconsistency handling step that follows results in removing the told descriptions referring to the *Sand* and *Seaside* concepts, and thereby the *Sea* and *Beach* descriptions are also removed. For instance, in the case of the *Beach* ≡ *Seaside* ⊓ ∃*contains.Sand* axiom presented in the example of Figure 3, there are three alternatives to a consistent interpretation, obtained by removing either the Seaside assertion, or the ∃*contains.Sand* assertion, or both. In the example of Figure 3, enrichment amounts to the addition of the (image:Outdoor)≥0.86 assertion, and to the update of the (image:∃.contains.Rock) assertion degree to 0.78 from the initial value of 0.67.

5 Experimental Results and Evaluation

To evaluate the feasibility and potential of the proposed fuzzy DLs based reasoning framework, we carried out an experimental implementation, using fuzzyDL as the core fuzzy DLs

Countryside_buildings ⊑ Landscape
∃contains.Buildings ⊑ Countryside_buildings
Countryside_buildings ⊑ ∃contains.Buildings
Grass ⊔ Tree ⊑ Vegetation
Rockyside ⊑ Mountainous
Rockyside ⊑ ∃contains.Rock
Forest ⊑ ∃contains.Tree
∃contains.Vegetation ⊑ Landscape
Roadside ⊑ ∃contains.Road
Roadside ⊑ Landscape
Forest ⊑ Landscape
∃contains.Sea ≡ Seaside
Beach ≡ Seaside ⊓ ∃contains.Sand
∃contains.Sky ⊑ Outdoor
Landscape ⊑ Outdoor
Trunk ⊑ Tree
Forest ⊓ (Roadside ⊔ Countryside_buildings) ⊑ ⊥
Roadside ⊓ Countryside_buildings ⊑ ⊥
Rockyside ⊓ Seaside ⊑ ⊥
Forest ⊓ (Wave ⊔ Sea ⊔ Sand ⊔ Road) ⊑ ⊥
Rockyside ⊓ (Wave ⊔ Sea ⊔ Sand ⊔ Boat) ⊑ ⊥

Table 5 Example extract of the outdoor image domain TBox constructed for evaluation purposes.

Natural	:	outdoor images of natural scenes, including landscape,
	:	mountainous, and seaside ones
Landscape	:	roadside, rural buildings and forest scenes
Mountainous	:	mountain and rockyside scenes
Seaside	:	coastal and beach scenes
ManMade	:	outdoor images of manmade scenes
Cityscape	:	urban scenes

Table 6 Description of scene level concepts for the considered outdoor images dataset.

inference engine. The choice over the available reasoners has been has been the provided expressivity which meets the representation requirements.

Opting for a generic performance assessment, we experimented in the domain of outdoor images, as it allows for wealthy semantics and it is sufficiently broad to avoid hard to generalise observations featuring close domains such as medical imaging. Addressing four categories of outdoor images, namely landscape, mountainous, seaside, and cityscape, a TBox was constructed to capture relations among the involved scene and the object level concepts. Table 6 provides the description of the related scene definitions, while Table 5 illustrates an extract of the developed background knowledge. The considered image dataset consists of 700 images, of which two non overlapping sets of 350 images have been assembled for training and testing purposes respectively. Ground truth has manually generated for the complete dataset at scene and object level.

Utilising the developed TBox, performance is estimated by comparing the reliability of the image descriptions extracted through machine learning to that of the descriptions resulting after the application of the proposed reasoning framework. To this end, we adopted the precision, recall, and F-measure metrics, using the following definitions.

– recall (r): the number of correct assertions extracted/inferred per concept divided by the number of the given concept assertions present in the ground truth image descriptions.
– precision (p): the number of correct assertions extracted/inferred per concept divided by the number of assertions that were extracted/inferred for the given concept.
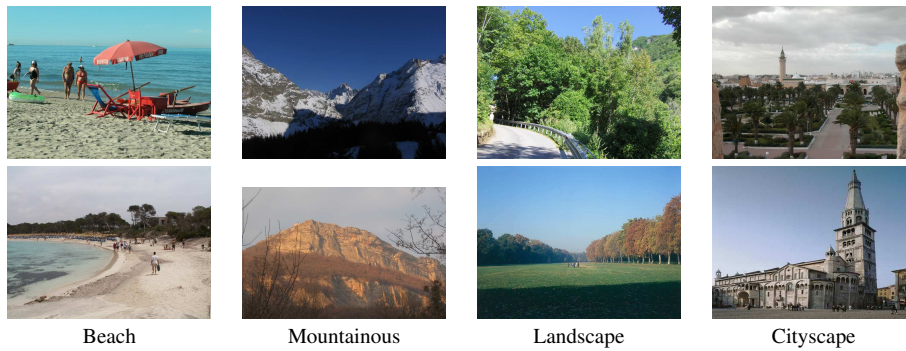
Beach   Mountainous   Landscape   Cityscape

Fig. 4 Experiment I - Example images of the concepts addressed by the learned scene level classifiers.

– F-measure: $2 * p * r/(p + r)$.

To investigate the contribution with respect to the wealth of the semantic associations pertaining to the concept classifiers involved, two experimental settings have been configured. As described in the sequel, in the first experiment the employed concept classifiers are rather poorly related, in terms of both scene and object level concepts; in the second experiment, the employed classifiers address concepts whose interrelations go beyond simple subsumption axioms, hence allowing for more complex inferences. Furthermore, we present initial results with respect the use of fuzzy implication as the meas to model co-occurrence information.

5.1 Experiment I

In the first experiment, the sets of scene C_{scene} and object C_{object} level concepts addressed by the classifiers consist in C_{scene}={Outdoor, Natural, ManMade, Landscape, Mountainous, Beach} and C_{object}={Building, Grass, Vegetation, Rock, Tree, Sea, Sand, Conifers, Boat, Road, Ground, Sky, Trunk, Person} respectively.

Figure 6 illustrates example images of the addressed scene level concepts. As described in Table 6, the concept *Natural* encompasses images belonging to the *Landscape*, *Beach* and *Mountainous* images, while the concept *ManMade* encompasses cityscape images. The only additional concept supported by reasoning is the *Seaside* concept, which represents coastal images that do not necessarily correspond to beach images. Statistical information regarding the frequency distribution of each concept in the train and test datasets is given in Tables 7 and 8. Since the employed object concept classifiers perform on segment level, the distribution of object level concepts is given with respect to the overall number of segments.

Different implementations have been used for the employed classifiers, allowing for overlapping descriptions. Specifically, for scene level classification, the SVM-based approach of [52] and the randomised clustering trees approach of [53] have been used. Similarly, for segment level (object) classification, the distance based feature matching approach based on prototypical values of [31], and the clustering trees approach of [53] have been followed.

Tables 7 and 8 present the attained performance with respect to scene and object level concepts, when applying solely the learned classifiers and when utilising the proposed reasoning framework.

Table 7 Experiment I - Evaluation of analysis and reasoning performance for scene level concepts, and concept distribution in the dataset.

Concept	Ground Truth (%)		Analysis			Reasoning		
	Train	*Test*	*Recall*	*Precision*	*F-M*	*Recall*	*Precision*	*F-M*
Outdoor	100.00	100.00	0.99	0.99	0.99	0.99	0.99	0.99
Natural	91.51	93.70	0.97	0.96	0.97	0.98	0.96	0.97
ManMade	8.49	6.30	0.18	0.40	0.25	0.18	0.40	0.25
Cityscape	8.49	6.30	0.18	0.40	0.25	0.18	0.40	0.25
Landscape	55.12	55.84	0.75	0.63	0.68	0.76	0.68	0.71
Mountainous	18.15	11.88	0.64	0.28	0.39	0.48	0.30	0.37
Beach	10.26	8.52	0.89	0.26	0.40	0.90	0.31	0.47
Seaside	18.24	18.98	-	-	-	0.86	0.49	0.63

Table 8 Experiment I - Evaluation of analysis and reasoning performance for object level concepts, and concept distribution in the dataset.

Concept	Ground Truth (%)		Analysis			Reasoning		
	Train	*Test*	*Recall*	*Precision*	*F-M*	*Recall*	*Precision*	*F-M*
Building	4.12	4.39	0.35	0.17	0.22	0.09	0.83	0.17
Grass	6.73	9.85	1.54	0.40	0.10	0.01	0.94	0.05
Vegetation	9.09	16.67	0.99	0.70	0.82	0.90	0.80	0.85
Rock	4.54	4.47	0.98	0.21	0.35	0.54	0.42	0.47
Tree	3.78	10.08	0.22	0.65	0.33	0.18	0.58	0.27
Sand	2.27	2.19	0.49	0.37	0.42	0.92	0.41	0.56
Sea	4.29	4.32	0.72	0.46	0.56	0.88	0.49	0.63
Conifers	1.01	0.90	1.00	0.01	0.02	0.50	0.02	0.03
Mountain	0.59	0.53	0.14	0.01	0.01	0.43	0.04	0.06
Boat	0.50	1.21	0.10	0.40	0.16	0.10	0.50	0.17
Road	4.04	4.32	0.15	0.50	0.23	0.02	0.25	0.03
Ground	27.27	4.39	0.06	0.57	0.19	0.11	0.57	0.19
Sky	19.02	18.04	0.93	0.87	0.89	0.93	0.87	0.89
Trunk	1.76	3.56	0.38	0.65	0.48	0.38	0.65	0.48
Person	8.67	11.97	0.49	0.54	0.52	0.49	0.54	0.52

As illustrated, in the case of scene level concepts, the utilisation of reasoning matches the performance of the classifiers, while in some cases it introduces a slight improvement. This behaviour is a direct result of the loose semantic relations among the employed scene and object level classifiers. Going through the corresponding TBox (Table 5), ones notices that the scene level concepts are in their majority atomic concepts, participating in subsumption axioms with one another. Hence, additional inferences can issue only from relations involving object level concepts. For the considered set of concept classifiers though, the only type of association between scene level concepts and object level concepts is disjointness. As a result, object level assertions cannot have an effect on the plausibility of scene level assertions.

To acquire a quantified measure over the effect of reasoning in such a case, we interpreted the response of the scene classifiers according to the respective scene concept semantics. For example, if an image was classified as *Landscape*, it was interpreted as well as positive classification for the concepts *Natural* and *Outdoor*. Thus, as long as a correct classification is attained at some level of the scene concept hierarchy, it is propagated towards the more generic concepts. Consequently, following this scheme, allows us to observe how the application of reasoning compares to this "optimal" scene classification interpretation,

$$(image : Countryside\_buildings) \geq 0.58$$
$$(image : Rockyside) \geq 0.45$$
$$(image : Roadside) \geq 0.59$$
$$(image : Forest) \geq 0.61$$
$$(image : Seaside) \geq 0.65$$
$$(image : \exists contains.Vegetation) \geq 0.72$$
$$(image : \exists contains.Vegetation) \geq 0.62$$
$$(image : \exists contains.Sky) \geq 0.91$$
$$(image : \exists contains.Person) \geq 0.68$$

Fig. 5 Example image and corresponding extracted descriptions where additional knowledge is needed to perform inference.

which includes subsumption relations. We note though that this optimistic perspective taken with respect to the acquired scene classifications, does not reflect the actual behaviour of the classifiers, nor the general case, as it is not uncommon to acquire positive classification results for semantically disjoint concepts or to get negative responses for concepts that are subsumed by concepts for which the classification was positive.

Examining the performance with respect to the object level concepts, the effect induced through reasoning is more interesting. As illustrated in Table 8, four patterns can be observed: i) cases where only the precision is improved, ii) cases where only the recall is improved, iii) cases where both precision and recall improve, and iv) cases where the performance remains invariable to the application of reasoning. As explained in the following, each pattern derives from the type of axioms in which the respective concept participates.

Improvement on precision, such in the case of the *Boat* concept, relates strongly to the utilisation of disjointness, as through the presented inconsistency handling approach, it is ensured that the final assertions comply with the scene level interpretation. Naturally, this harbours the risk of ending up with significantly distorted final descriptions in the case of falsely scene level interpretation. Such an example constitutes the performance deterioration observed with respect to the concept *Road*, in which case the majority of images depicting road were attributed with greater confidence to the *Seaside* or . However, when considering such cases one needs to bear in mind not the interpretation desired based on visual inspection of the examined image, but rather the interpretation appearing more plausible on the grounds of the initial descriptions provided by the classifiers.

Figure 5 illustrates such an example. The two most plausible scene level descriptions are misleading, while the detected object level concepts, though accurate, are not adequate to drive the inference of corresponding scene concept. Additional knowledge, such as co-occurrence of concepts could be exploited to assist and either re-adjust the degrees or trigger an inference. This limitation constitutes one of the reasons motivating as future direction the exploration of reasoning framework that combines fuzzy and probabilistic knowledge. Some very preliminary investigations towards this directions are described in the following, where the modelling of co-occurrence information in the form of fuzzy implication is discussed.

Object level concepts exhibiting improved recall rates, such as *Building*, *Sand* and *Grass*, correspond to concepts that appear on the righthand side of axioms. In practise such concepts may be entailed either by scene level concepts, such is the case for the *Building* and *Sand* concepts, or by other object level concepts, such in the case of the concept *Tree*. Respectively, concepts whose recall is reduced reflect cases of wrongly inferred scene level concepts that resulted in their removal due to disjoint axioms.

Table 9 Overall evaluation for Experiment I.

Concept	Analysis			Reasoning		
	Recall	*Precision*	*F-M*	*Recall*	*Precision*	*F-M*
All concepts	0.68	0.49	0.57	0.68	0.63	0.65
Concepts participating in axioms	0.70	0.49	0.58	0.70	0.64	0.67

| Seaside | Rockyside | Roadside | Countryside buildings | Forest |
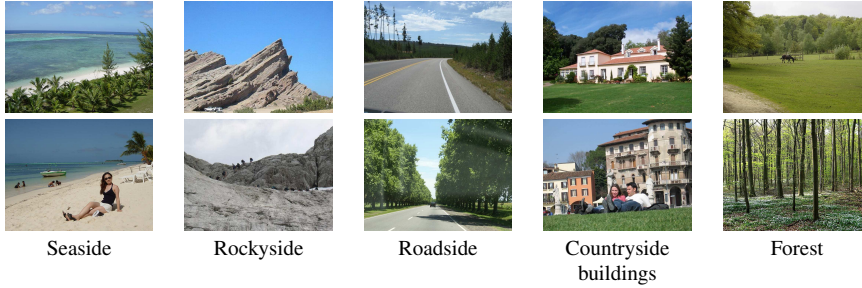
Fig. 6 Experiment II - Example images of the concepts addressed by the learned scene level classifiers.

Cases where both precision and recall are improved correspond to concepts appearing on the righthand side of general inclusion axioms, when the correct scene level concepts have been inferred. Such concepts are characterised by rich semantics and apparently benefit the most from the application of reasoning. Contrariwise, invariable performance indicates atomic concepts participating solely on the left hand side of axioms (e.g. the Trunk and Boat concepts).

Table 9 summarises the average performance of the classifiers and of the proposed reasoning framework for the experiment. The first line presents the performance when all concepts are taken into account, while the second line presents the respective performance when only concepts that participate in axioms are considered. As illustrated, the application of reasoning bears a positive impact in terms of the precision, which is increased from 0.49 to 0.63. Recall remains unchanged for the reasons explained above.

5.2 Experiment II

In the second experiment, the sets of scene C_{scene} and object C_{object} level concepts addressed by the learned classifiers are respectively C_{scene}={Countryside_Buildings, Roadside, Rockyside, Seaside, Forest} and C_{object}={Building, Roof, Grass, Vegetation, Dried-Plant, Sky, Rock, Tree, Sea, Sand, Boat, Road, Ground, Person, Trunk, Wave}. The additional concepts supported by reasoning are *Landscape*, *Mountainous*, *Beach*, and *Outdoor*. *Landscape* is subsumed by *Forest*, *Roadside* and *Countryside_buildings*, while *Mountainous* is subsumed by *Rockyside*. Statistical information regarding the frequency distribution of each concept in the train and test datasets is given in Tables 10 and 11. This time, a single classifier has been employed per concept, following the SVM based approach of [20].

Table 10 compares the performance of the classifiers and of the proposed reasoning framework with respect to scene level concepts. Contrariwise to the previous experiment, only explicitly generated classification are considered. As illustrated, apart from the ex-

Table 10 Experiment II - Evaluation of analysis and reasoning performance for scene level concepts, and concept distribution in the dataset.

	Ground Truth (%)		Analysis			Reasoning		
Concept	Train	Test	Recall	Precision	F-M	Recall	Precision	F-M
Countryside_buildings	16.42	16.05	0.30	1.0	0.46	0.60	0.86	0.71
Rockyside	17.15	17.88	0.68	0.70	0.69	0.68	0.79	0.74
Roadside	18.97	19.34	0.68	0.69	0.69	0.68	0.72	0.70
Forest	19.73	20.45	0.75	0.63	0.69	0.74	0.68	0.71
Seaside	18.24	18.98	0.85	0.67	0.75	0.86	0.72	0.78
Beach	10.26	8.52	-	-	-	0.45	0.76	0.57
Landscape	55.12	55.84	-	-	-	0.97	1.00	0.98
Mountainous	18.15	18.88	-	-	-	0.67	0.80	0.74
Outdoor	100.00	100.00	-	-	-	0.99	1.00	0.99

pected improvement in terms of scene level concepts that are acquired due to subsumption relations, the application of reasoning incurs a more noticeable effect compared to the first experiment. The reasons for this behaviour lie again in the semantics of the concepts involved. Specifically, in this experimental setting, the object level concepts addressed by the classifiers are semantically related with scene level ones not only through disjointness axioms but also by general concept inclusions, thus incurring the inference of scene level concepts from object level ones. For example, the presence of a *Building* assertion triggers the inference of a corresponding *Countryside_buildings* assertion, with an equal or greater degree of confidence. Hence, in combination with the axioms that relate the concepts *Vegetation* and *Grass* with *Landscape* scenes, the application of reasoning allows to improve the recall for *Countryside_buildings*.

Table 11 compares the performance for descriptions at object level. With the exception of the *Boat* and *Grass* concepts, the application of reasoning improves significantly on the performance obtained by the sole application of the classifiers. Again, this is a direct consequence of the fact that the considered object level concepts are characterised by richer semantics with respect to the scene level concepts that constitute their context of appearance. The deterioration in the recall rates of the *Boat* and *Grass* concepts, is again indicative of the risks entailed by a false scene level interpretation. Going through the images for which *Boat* assertions where falsely removed, we observed that the prevailing scene level classifications were not in compliance with the depicted scene. Similar observations hold for the concept *Grass*. A possible way to alleviate such phenomena, apart from the investigation of additional types of knowledge, could be the re-assessment of the terminological axioms describing the domain. However, such approach lurks the risk of ending up with solutions customisable to a specific learning implementations or to specific application domains and datasets.

Similar considerations emerge when examining the not so noticeable effect of reasoning in the recall of scene level concepts such as the *Rockyside* one. Going through the images depicting rocky side scenes, yet failed to recognised as such, we observed that in all cases the classifiers had falsely detected another scene level concept instead; and this, despite the fact that the instantiations of the *Rock* concept were successfully detected for their majority. Adding the axiom $\exists contains.Rock \sqsubseteq Rockyside$, would seem a reasonable approach to improve recall for the *Rockyside* concept, especially since in the examined dataset, the *Seaside* classifier tends to produce higher confidence values than the *Rockyside* one for seaside images, where rock instance appear too. However, it is easy to see that such an amendment

Table 11 Experiment II - Evaluation of analysis and reasoning performance for object level concepts, and concept distribution in the dataset.

Concept	Ground Truth (%)		Analysis			Reasoning		
	Train	Test	Recall	Precision	F-M	Recall	Precision	F-M
Building	4.12	4.39	0.54	0.69	0.60	0.62	0.86	0.72
Roof	1.54	1.89	0.33	0.54	0.41	0.33	0.75	0.46
Grass	6.73	9.85	0.49	0.42	0.45	0.30	0.52	0.38
Vegetation	9.09	16.67	0.48	0.84	0.61	0.86	0.86	0.86
Dried-Plant	0.01	0.04	0.07	0.11	0.08	0.07	0.13	0.10
Rock	4.54	4.47	0.65	0.45	0.53	0.69	0.70	0.69
Tree	3.78	10.08	0.49	0.52	0.51	0.56	0.47	0.51
Sand	2.27	2.19	0.02	0.10	0.03	0.57	0.45	0.50
Sea	4.29	4.32	69	0.60	0.64	0.85	0.69	0.76
Boat	0.50	1.21	0.41	0.71	0.52	0.33	0.66	0.44
Road	4.04	4.32	0.50	0.69	0.58	0.69	0.71	0.70
Sky	19.02	18.04	0.95	0.93	0.94	0.95	0.93	0.94
Ground	27.27	4.39	0.26	0.33	0.29	0.26	0.33	0.29
Person	8.67	11.97	0.75	0.51	0.61	0.75	0.51	0.61
Trunk	1.76	3.56	0.26	0.28	0.27	0.26	0.28	0.27
Wave	0.75	1.14	0.0.25	0.5	0.33	0.25	0.5	0.33

Table 12 Collective evaluation for Experiment II.

Concept	Analysis			Reasoning		
	Recall	Precision	F-M	Recall	Precision	**F-M**
All concepts	0.37	0.65	**0.47**	0.77	0.81	**0.79**
Concepts partici- pating in axioms	0.29	0.61	**0.39**	0.79	0.82	**0.81**

would imbalance the trade off between what constitutes domain semantics and what is mere tuning to the peculiarities of a given dataset.

Finally, Table 12 summarises the average performance. As illustrated, the application of reasoning has a stronger influence compared to the first experiment, reflected on both recall and precision, which are significantly improved.

5.3 Investigating Reasoning with Additional Knowledge

The previously described experimentation configurations allowed to assess the performance of the proposed reasoning framework, and provided useful experiences with respect to the issues and challenges involved. The observations drawn could be summarised in the two following. On one hand, the proposed reasoning framework has the potential to enhance semantic image analysis towards more consistent and complete descriptions, provided that the concepts involved bear an adequate wealth of semantic associations. On the other hand, the inference of the overall description is liable to false interpretations in the presence of heavily distorted classification results, both in terms of the concepts and of their computed degrees. As already mentioned above, a possible way to assist inference in such cases would be the use of additional knowledge. Such knowledge could consider for example the percentage of an image assigned to a given description. Thereby, in cases like that of Figure 5, where normally no scene level concepts can be inferred based on the initial assertions, a

complementary mechanism could favour the *Landscape* concept and its subclasses. Another approach would be be the joint utilisation of probabilistic knowledge, regarding the concept co-occurrence patterns.

Since our focus is currently on the benefits and weaknesses pertaining to the utilisation of fuzzy reasoning, we performed some additional experiments using fuzzy implication to capture co-occurrence information. We note though, that due to the early stage of this effort, no conclusive observations can be drawn. In the following, we briefly describe the semantics of fuzzy implication and present the outcomes of this preliminary investigation.

Fuzzy implication is defined by a function of the form $J : [0,1] \times [0,1] \rightarrow [0,1]$ [42]. In practise this means that the degree of the right hand side expression does not depend only on the degree of the left hand side expression, but is influenced as well by the degree attributed to the implication itself. Thereby, fuzzy implication can serve as means to introduce dataset specific knowledge without impairing the impact of the inferences drawn by the crisp axioms modelling generic semantics.

To get an initial estimation of whether such an approach could be beneficial, we experimented using the following scheme. First, we performed statistical analysis on the manually constructed ground truth to obtain co-occurrence patterns between scene level concepts and object level concepts. The latter share a significant characteristic: although semantically related to specific scene level concepts, they do not necessarily entail information regarding the presence of the respective scene level concepts. Then, the observed concept patterns frequency was used in order to define the degrees of the corresponding fuzzy implication axioms.

Two different schemes have been used for the calculation of the fuzzy implications degrees. In the first case, concepts co-occurrence was measured with respect to the ground truth annotations; thus it reflects in a way an ideal modeling of the dataset specific knowledge. In the second case, co-occurrence was measured taking into account the object level descriptions as obtained by the classifiers and the scene level ground truth annotations. Consequently, in the second case the degrees encompass additional information accounting to an extent for the effectiveness of the classifiers, and thus for the errors introduced by them. Table 13 illustrates the effect on the determination of image descriptions at scene level, using the Kleene-Dienes implication [42], with respect to the two aforementioned schemes for determining concept co-occurrence.

Although both schemes used for the determination of degrees are very simplistic, the introduction of fuzzy implication appears to have the potential to incur a further improvement when compared with the respective values in Table 10. As previously mentioned, these results are only preliminary, and such do not allow for any conclusions to be drawn yet. Further investigation is required with respect to the methodology used to calculate the degrees and the process of selecting which implications to consider. The associations between the concepts aggravate further the difficulty in finding a balance between compensating for classification errors (by opting for higher implication degrees) and avoiding figurative increase in recall values (i.e. improved recall accompanied by significant deterioration in precision).

6 Conclusions and Future Directions

In this paper, we presented a fuzzy DLs-based reasoning framework for the purpose of enhancing the extraction of image semantics through the utilisation of formal knowledge. The deployment of fuzzy DLs allows to formally handle the vagueness encountered in classifications acquired through statistical learning, while the formal semantics allow the integration

Table 13 Evaluation for scene level concepts when using fuzzy implication to model co-occurrence information. The fuzzy implication degrees are acquired from ground truth data and from the classifications respectively.

Concept	Ground Truth Co-Occurrence			Classification Co-Occurrence		
	Recall	*Precision*	*F-M*	*Recall*	*Precision*	*F-M*
Countryside_buildings	0.72	0.86	**0.78**	0.84	0.87	**0.85**
Rockyside	0.74	0.69	**0.71**	0.73	0.80	**0.76**
Roadside	0.68	0.68	**0.68**	0.69	0.76	**0.75**
Forest	0.66	0.75	**0.70**	0.73	0.73	**0.73**
Seaside	0.79	0.80	**0.79**	0.86	0.85	**0.85**

of the initial description into a semantically coherent interpretation. The proposed reasoning framework exploits reasoning not only in order to enrich the image descriptions by means of entailment, but also in order to address and resolve inconsistencies in the initial classifications. Thereby, and free of assumptions regarding the individual classifiers' implementation, it provides the means to reach a consistent final image description and to alleviate limitations often encountered in learning based approaches due to poor semantics utilisation. The experiments, though not conclusive, show promising results, while outlining a number of issues and challenges affecting the attained performance.

Future directions include the extension of the framework in order to handle the representation of individual image regions and spatial relations so as to allow the utilisation of spatial reasoning. Furthermore, as delineated in the conducted evaluation, experimentation towards the possibilities of introducing additional knowledge, either in the form of fuzzy implication or in a probabilistic manner, constitutes another direction towards a more complete framework. The latter is of particular interest, not only because both types of imprecision are met in semantic image analysis, but mostly because they serve complementary purposes.

7 Acknowledgements

References

1. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
2. Naphade, M., Huang, T.: Extracting semantics from audio-visual content: the final frontier in multimedia retrieval. IEEE Transactions on Neural Networks **13**(4), 793–810 (2002)
3. Hanjalic, A., Lienhart, R., Ma, W., Smith, J.: The holy grail of multimedia information retrieval: So close or yet so far away. IEEE Proceedings, Special Issue on Multimedia Information Retrieval **96**(4), 541–547 (2008)
4. Naphade, M., Huang, T.: A probabilistic framework for semantic video indexing, filtering, and retrieval. IEEE Transactions on Multimedia **3**(1), 141–151 (2001)
5. Assfalg, J., Bertini, M., Colombo, C., Bimbo, A.D.: Semantic annotation of sports videos. IEEE Multi-Media **9**(2), 52–60 (2002)
6. Town, C., Sinclair, D.: A self-referential perceptual inference framework for video interpretation. In: International Confernce on Computer Vision Systems (ICVS), Graz, Austria, pp. 54–67 (2003)
7. Snoek, C., Worring, M., van Gemert, J., Geusebroek, J., Smeulders, A.: The challenge problem for automated detection of 101 semantic concepts in multimedia. In: Proc. 14th ACM International Conference on Multimedia, Santa Barbara, CA, USA, pp. 421–430 (2006)

8. Smeaton, A., Over, P., Kraaij, W.: Evaluation campaigns and trecvid. In: MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pp. 321–330. ACM Press, New York, NY, USA (2006)

9. Snoek, C., Huurnink, B., Hollink, L., de Rijke, M., Schreiber, G., Worring, M.: Adding semantics to detectors for video retrieval. IEEE Transactions on Multimedia **9**(5), 975–986 (2007)

10. Hauptmann, A., Yan, R., Lin, W.: How many high-level concepts will fill the semantic gap in news video retrieval? In: in Proc. 6th ACM International Conference on Image and Video Retrieval (CIVR), Amsterdam, The Netherlands, pp. 627–634 (2007)

11. Naphade, M., Kennedy, L., Kender, J., Chang, S.F., Smith, J., Over, P., Hauptmann, A.: A light scale concept ontology for multimedia understanding for trecvid 2005. In: RC23612 W0505-104, Computer Science, IBM Research Report (2005)

12. Snoek, C., van de Sande, K., deRooij, O., Huurnink, B., van Gemert, J., Uijlings, J., He, J., Li, X., Everts, I., Nedovic, V., van Liempt, M., van Balen, R., Yan, F., Tahir, M., Mikolajczyk, K., Kittler, J., de Rijke, M., Geusebroek, J., Gevers, T., Worring, M., Smeulders, A., Koelma, D.: The mediamill trecvid 2008 semantic video search engine. In: TRECvid 2008 Working Notes. National Institute of Standards and Technology (NIST), Washington (2008)

13. Natsev, A., Jiang, W., Merler, M., Smith, J., Tesic, J., Xie, L., Yan, R.: Ibm research trecvid-2008 video retrieval system. In: TRECvid 2008 Working Notes. National Institute of Standards and Technology (NIST), Washington (2008)

14. Rao, A., Jain, R.: Knowledge representation and control in computer vision systems. IEEE Expert pp. 64–79 (1988)

15. Crevier, D., Lepage, R.: Knowledge-based image understanding systems: A survey. Computer Vision and Image Understanding **67**, 161–185 (1997)

16. Reiter, R., Mackworth, A.: A logical framework for depiction and image interpretation. Artificial Intelligence **41**, 125–155 (1990)

17. Niemann, H., Sagerer, G., Schröder, S., Kummert, F.: Ernest: A semantic network system for pattern understanding. IEEE Trans. Pattern Anal. Mach. Intell. **12**(9), 883–905 (1990)

18. Brickley, D., Guha, R.V.: RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February 2004. http://www.w3.org/TR/rdf-schema/

19. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D., Patel-Schneider, P., Stein, L.: OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004. http://www.w3.org/TR/owl-ref/

20. Papadopoulos, G., Mylonas, P., Mezaris, V., Avrithis, Y., Kompatsiaris, I.: Knowledge-assisted image analysis based on context and spatial optimization. Int. J. Semantic Web Inf. Syst. **2**(3), 17–36 (2006)

21. Straccia, U.: Reasoning within fuzzy description logics. J. Artif. Intell. Res. (JAIR) **14**, 137–166 (2001)

22. Straccia, U.: Transforming fuzzy description logics into classical description logics. In: Proc. European Conference on Logics in Artificial Intelligence (JELIA), Lisbon, Portugal, pp. 385–399 (2004)

23. Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J., Horrocks, I.: The fuzzy description logic f-SHIN. In: International Workshop on Uncertainty Reasoning For the Semantic Web (URSW), Galway, Ireland (2005)

24. Moller, R., Neumann, B., Wessel, M.: Towards computer vision with description logics: Some recent progress. In: Proceedings Integration of Speech and Image Understanding, Corfu, Greece, pp. 101–115 (1999)

25. Neumann, B., Moller, R.: On scene interpretation with description logics (FBI-B-257/04) (2004)

26. Bagdanov, A., Bertini, M., DelBimbo, A., Serra, G., Torniai, C.: Semantic annotation and retrieval of video events using multimedia ontologies. In: Proc. IEEE International Conference on Semantic Computing (ICSC), Irvine, CA, USA (2007)

27. Schober, J.P., Hermes, T., Herzog, O.: Content-based image retrieval by ontology-based object recognition. In: Proc. KI-2004 Workshop on Applications of Description Logics (ADL), Ulm Germany (2004)

28. Espinosa, S., Kaya, A., Melzer, S., Möller, R., Wessel, M.: Multimedia interpretation as abduction. In: Proc. International Workshop on Description Logics (DL), Brixen-Bressanone, Italy (2007)

29. J. Hunter J. Drennan, S.L.: Realizing the hydrogen economy through semantic web technologies. IEEE Intelligent Systems Journal - Special Issue on eScience **19**, 40–47 (2004)

30. Little, S., Hunter, J.: Rules-by-example - a novel approach to semantic indexing and querying of images. In: International Semantic Web Conference (ISWC), Hiroshima, Japan, pp. 534–548 (2004)

31. Petridis, K., Bloehdorn, S., Saathoff, C., Simou, N., Dasiopoulou, S., Tzouvaras, V., Handschuh, S., Avrithis, Y., Kompatsiaris, I., Staab, S.: Knowledge representation and semantic annotation of multimedia content. IEE Proceedings on Vision Image and Signal Processing, Special issue on Knowledge-Based Digital Media Processing **153** (2006)

32. Neumann, B., Möller, R.: On scene interpretation with description logics. Image and Vision Computing, Special Issue on Cognitive Vision pp. 82–101 (2007)

33. Richardson, M., Domingos, P.: Markov logic networks. Machine Learning **62**(1-2), 107–136 (2006)
34. Bayesian compositional hierarchies - a probabilistic structure for scene interpretation. Dagstuhl Seminar Proceedings (2008)
35. Qualitative abstraction and inherent uncertainty in scene recognition. Dagstuhl Seminar Proceedings (2008)
36. Dubois, D., Prade, H.: Possibility theory, probability theory and multiple-valued logics: A clarification. Annals of Mathematics and Artificail Intelligence **32**(1-4), 35–66 (2001)
37. Umberto, S., Giulio, V.: Dlmedia: an ontology mediated multimedia information retrieval system. In: Proc. International Workshop on Description Logics (DL), Brixen-Bressanone, Italy (2007)
38. Simou, N., Athanasiadis, T., Tzouvaras, V., Kollias, S.: Multimedia reasoning with f-shin. 2nd International Workshop on Semantic Media Adaptation and Personalization, London, UK (2007)
39. Mylonas, P., Simou, N., Tzouvaras, V., Avrithis, Y.: Towards semantic multimedia indexing by classification and reasoning on textual metadata. Knowledge Acquisition from Multimedia Content Workshop, Genova, Italy (2007)
40. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., P. F. Patel-Schneider, e.: The description logic handbook: Theory, implementation, and applications. In: Description Logic Handbook. Cambridge University Press (2003)
41. Zadeh, L.: Fuzzy sets. Information and Control **8**(32), 338–353 (1965)
42. Klir, G., Yuan, B.: Fuzzy sets and fuzzy logic: Theory and applications. Prentice-Hall (1995)
43. Stoilos, G., Stamou, G., Pan, J.: Handling imprecise knowledge with fuzzy description logic. In: Proc. International Workshop on Description Logics (DL), Lake District, UK (2006)
44. Stoilos, G., Stamou, G., Pan, J., Tzouvaras, V., Horrocks, I.: Reasoning with very expressive fuzzy description logics. J. Artif. Intell. Res. (JAIR) **30**, 273–320 (2007)
45. Straccia, U.: A fuzzy description logic for the semantic web. In: E. Sanchez (ed.) Fuzzy Logic and the Semantic Web, Capturing Intelligence, pp. 73–90. Elsevier (2006)
46. Bobillo, F., Straccia, U.: fuzzydl: An expressive fuzzy description logic reasoner. In: in Proc. International Conference on Fuzzy Systems (FUZZ), Hong Kong, pp. 923–930. IEEE Computer Society (2008)
47. Haase, P., van Harmelen, F., Huang, Z., Stuckenschmidt, H., Sure, Y.: A framework for handling inconsistency in changing ontologies. In: Proc. of International Semantic Web Conference (ISWC), Galway, Ireland, pp. 353–367 (2005)
48. Kalyanpur, A., Parsia, B., Sirin, E., Grau, B.C.: Repairing unsatisfiable concepts in owl ontologies. In: Proc. of European Semantic Web Conference (ESWC), Budva, Montenegro, pp. 170–184 (2006)
49. Bell, D., Qi, G., Liu, W.: Approaches to inconsistency handling in description-logic based ontologies. In: in Proc. OTM Workshops, Vilamoura, Portugal, pp. 1303–1311 (2007)
50. Haase, P., Qi, G.: An analysis of approaches to resolving inconsistencies in dl-based ontologies. In: in Proc. International Workshop on Ontology Dynamics (IWOD), Innsbruck, Austria, pp. 97–109 (2007)
51. Lam, J., Sleeman, D., Pan, J., Vasconcelos, W.: A fine-grained approach to resolving unsatisfiable ontologies. J. Data Semantics **10**, 62–95 (2008)
52. LeBorgne, H., Guérin-Dugué, A., O'Connor, N.: Learning midlevel image features for natural scene and texture classification. IEEE Trans. Circuits Syst. Video Techn. **17**(3), 286–297 (2007)
53. Moosmann, F., Triggs, B., Jurie, F.: Randomized clustering forests for building fast and discriminative visual vocabularies. In: Neural Information Processing Systems (NIPS) (2006)