Fast and Faithful: Scalable Neuro-Symbolic Learning and Reasoning with Differentiable Fuzzy $\mathcal{EL}^{++}$

Anonymous Author(s)

Abstract

The unification of neural learning and symbolic reasoning remains a foundational challenge in AI, forcing a persistent trade-off between logical rigor and computational scale. The field has largely diverged into two paths: expressive frameworks rooted in first-order logic that are formally sound but computationally intractable, and scalable methods, such as geometric embeddings, that sacrifice formal logical guarantees for efficiency. This paper introduces **DF-$\mathcal{EL}^{++}$**, the first end-to-end differentiable framework that transcends this trade-off by unifying PTIME-complete reasoning with neural learning. Our primary contribution is a complete, theoretically-grounded methodology centered on a unified semantic-loss framework: we unite the tractable structure of the description logic $\mathcal{EL}^{++}$ with a Product-based fuzzy semantics, deriving our learning objective directly from the corresponding Göguen implication to ensure high logical fidelity. This principled semantic core is made robust and practical by two supporting innovations: a normalization strategy that re-architects complex axioms for stable optimization, and a novel domain construction technique that prevents model collapse to ensure non-trivial reasoning. Validated on massive, real-world ontologies like SNOMED CT (377K concepts), DF-$\mathcal{EL}^{++}$ demonstrates a unique synergy of scale and performance: it remains computationally efficient where expressive systems fail, while decisively outperforming dominant scalable baselines in a range of knowledge base completion tasks with up to a 42% relative improvement in Hits@1. This work establishes a new, provably sound, and scalable pathway for a new generation of neuro-symbolic systems that are both empirically powerful and logically reliable.

Our implementation, experimental data, and extended results are available at http://github.com/anonymous-ai-researcher/kdd2026 to support reproducibility and future research.

1 Introduction

Neural-symbolic AI represents a fundamental paradigm in modern AI that seeks to integrate the complementary strengths of neural learning and symbolic reasoning into unified computational frameworks [10, 17]. Neural networks excel at pattern recognition from raw sensory data but lack interpretability and logical consistency, while symbolic systems provide transparent reasoning and formal guarantees but struggle with uncertainty and perceptual grounding [41]. This integration has become increasingly critical as AI systems are deployed in domains requiring both empirical adaptability and logical reliability—from scientific discovery where automated reasoning must extract insights from massive datasets [63], to medical diagnosis where interpretable decision-making is essential for clinical trust [21].

The central technical barrier preventing effective neural-symbolic unification lies in *representational incompatibility*: neural networks operate through continuous, distributed activations in high dimensional vector spaces, while symbolic systems manipulate discrete structures through compositional operations governed by formal semantics [32]. This fundamental mismatch manifests in the symbol grounding problem [33]—establishing meaningful correspondences between abstract symbolic representations and their concrete perceptual instances—which has proven resistant to traditional approaches that maintain strict paradigm boundaries.

Pipeline-Based Approaches maintain strict physical separation between neural and symbolic components, processing information sequentially through standardized interfaces. While preserving individual paradigm strengths, this approach fundamentally maintains the representational divide, preventing bidirectional information flow and joint optimization. Representative methods include KBANN [59], NMN [2], ProbLog [51], DeepProbLog [40], ∂ILP [22], ABL [72], NeurASP [69], and DeepStochLog [66].

Constraint-Based Approaches integrate symbolic knowledge as training-time regularization, incorporating logical constraints as auxiliary loss terms without fundamentally altering the neural representational framework. This asymmetric integration discards logical guidance during inference, relegating reasoning to opaque neural computations and failing to deliver the interpretability that motivates neural-symbolic integration. Representative methods include MLN [53], SL [68], SATNet [64], DL2 [23], SBR [18], NMLN [44], RNM [42], SPL [1], DLM [43], and CCN+[25].

Embedding-Based Approaches project neural and symbolic elements into shared vector spaces, replacing discrete logical operations with differentiable similarity computations. While enabling gradient-based learning, this apparent unification transforms principled logical operations into approximate similarity measures, sacrificing the formal semantic foundations that make symbolic reasoning reliable and interpretable. Representative methods include NTP [54], CTP [45], EL Embeddings [38], EmEL++ [49], ELBE [46], BoxEL [67], Box^{2}EL [36], and Hamilton et al. [31].

Program Synthesis Approaches employ neural networks as auxiliary tools for symbolic program generation, typically using neural components to guide search over discrete program spaces. While maintaining symbolic interpretability, integration remains limited to synthesis tasks rather than continuous reasoning, often requiring expensive search procedures that struggle with joint perception-reasoning optimization. Representative methods include DeepCoder [9], DiffLog [55], ∂4 [11], EC2 [20], and Playgol [16].

Recent work has begun addressing representational unification through frameworks that establish formal correspondences between continuous neural outputs and discrete logical structures, e.g., DLM [43], TensorLog [14], LTN [7], logLTN [8], NeuPSL [50], ULLER [62], FALCON [57], and TAR [58]. These approaches create shared representational spaces where neural learning and symbolic reasoning operate jointly through continuous relaxations of logical operations, often employing fuzzy or probabilistic semantics to render discrete structures differentiable. However, these methods universally adopt first-order logic with extended semantics as their representational foundation, inheriting first-order logic's fundamental undecidability and exponential worst-case complexity, which severely limits scalability for real-world knowledge bases containing millions of axioms and assertions.

This computational intractability forces a difficult compromise in neuro-symbolic AI, creating a chasm between expressive but intractable logics and scalable but semantically weak methods. To break this impasse, our work proposes a paradigm shift, pivoting away from attempts to tame intractable logics and instead building upon a foundation of proven computational efficiency. This leads us to the description logic $\mathcal{EL}^{++}$ [4], a language that uniquely balances practical modeling power with the guarantee of polynomial-time reasoning. Our choice is strategically grounded in real-world success: $\mathcal{EL}^{++}$ is not a theoretical toy but the logical backbone of large-scale, mission-critical biomedical ontologies such as SNOMED CT [56] and Gene Ontology (GO) [3].

This paper introduces **Differentiable Fuzzy $\mathcal{EL}^{++}$ (DF-$\mathcal{EL}^{++}$)**, the first framework to resolve this longstanding trade-off by endowing a PTIME-tractable logic with a fully differentiable, theoretically sound fuzzy semantics. DF-$\mathcal{EL}^{++}$ achieves a genuine unification of scalable neural learning and logical reasoning. It operates as a principled refinement engine: a neural network first grounds symbols by producing an initial, uncertain fuzzy knowledge base from data. Our framework then uses gradient-based optimization to repair and refine these beliefs, ensuring they cohere with the ontology's logical constraints while preserving high-confidence empirical evidence. This creates a virtuous cycle where perception is disciplined by logic, and logic is grounded in data.

Our DF-$\mathcal{EL}^{++}$ framework is realized through a cohesive set of technical contributions that move beyond traditional heuristic designs to establish a complete, theoretically-grounded methodology:

- At the core of our framework is a principled choice of **Product-based fuzzy semantics**, selected for its smooth, non-vanishing gradients that are essential for stable optimization. Crucially, we derive our loss function directly from the corresponding **Göguen R-implication**, ensuring that the learning objective is not an arbitrary heuristic but a direct reflection of the underlying logic. This synergy of a differentiable t-norm and its native implication guarantees high logical fidelity throughout the learning process.
- We introduce a normalization framework that re-architects the optimization landscape itself, serving as a crucial bridge between declarative logic and gradient-based learning. It confronts the inherent challenges of deep logical structures—namely, unstable gradients and intractable credit assignment—by systematically decomposing them into a set of shallow, modular forms. This transformation of the logical syntax is not a mere pre-processing step but a necessary precondition for enabling stable and effective neuro-symbolic learning.
- We introduce a novel **semantically-aware domain construction** strategy that prevents the model collapse endemic to less principled approaches. By strategically instantiating "witness" individuals required by the logical axioms, we ensure that the model cannot trivially satisfy constraints by learning empty, meaningless representations for concepts. This technique anchors the learning process in a semantically rich space, forcing the model to engage with the logic in a meaningful way.
- We provide extensive empirical evidence that these theoretical advantages translate to superior real-world performance. On massive biomedical ontologies, DF-$\mathcal{EL}^{++}$ is not only computationally tractable where expressive FOL-based systems fail to converge but also decisively outperforms state-of-the-art geometric baselines in knowledge base completion tasks.

2 Preliminaries: Description Logic $\mathcal{EL}^{++}$

Description logics (DLs) form a family of knowledge representation languages designed as decidable fragments of first-order logic while maintaining sufficient expressivity for practical knowledge representation tasks [5]. Among the various DL dialects, $\mathcal{EL}^{++}$ occupies a unique position: unlike more expressive DLs such as $\mathcal{ALC}$ that achieve EXPTIME-complete reasoning complexity, $\mathcal{EL}^{++}$ maintains polynomial-time reasoning through carefully designed syntactic restrictions while supporting essential modeling constructs, making it the logical foundation for major biomedical ontologies including SNOMED CT [56], Gene Ontology [3], and ChEBI [34].

2.1 Syntax of $\mathcal{EL}^{++}$

Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be mutually disjoint countably infinite sets of **concept names**, **role names**, and **individual names**, respectively. Additionally, let $\mathsf{N_F}$ be a countably infinite set of **feature names** (functional roles), disjoint from the above sets. A **signature** $\Sigma = \langle \mathsf{N_C}, \mathsf{N_R}, \mathsf{N_I}, \mathsf{N_F} \rangle$ defines the vocabulary available for constructing $\mathcal{EL}^{++}$ expressions. The syntax of $\mathcal{EL}^{++}$ concepts (or simply concepts) is defined inductively by the following grammar:

$$C, D ::= \top \mid A \mid C \sqcap D \mid \exists r.C \mid \exists f.C$$

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, $f \in \mathsf{N_F}$, and C, D denote arbitrary concepts. We distinguish between **atomic** concepts, which are either concept names or $\top$, and **complex** concepts otherwise.

An $\mathcal{EL}^{++}$ **TBox** $\mathcal{T}$ is a finite set of axioms of the forms:

$$\begin{aligned} \text{concept inclusions:} \quad & C \sqsubseteq D \\ \text{role inclusions:} \quad & r_1 \circ \cdots \circ r_k \sqsubseteq s \\ \text{domain restrictions:} \quad & \mathrm{dom}(f) \sqsubseteq C \\ \text{range restrictions:} \quad & \mathrm{ran}(f) \sqsubseteq C \end{aligned}$$

where $r_1, \ldots, r_k, s \in \mathsf{N_R}$, $f \in \mathsf{N_F}$, and C, D are concepts. The composition $r_1 \circ \cdots \circ r_k$ represents role chains, enabling the expression of transitivity and other complex role relationships. For example, $\mathit{hasParent} \circ \mathit{hasParent} \sqsubseteq \mathit{hasGrandparent}$ captures that one's parent's parent is one's grandparent.

An $\mathcal{EL}^{++}$ **ABox** $\mathcal{A}$ is a finite set of assertions of the forms:

$$\begin{aligned} \text{concept assertions:} \quad & a : C \\ \text{role assertions:} \quad & (a, b) : r \\ \text{feature assertions:} \quad & (a, b) : f \end{aligned}$$

where $a, b \in \mathsf{N_I}$, $r \in \mathsf{N_R}$, $f \in \mathsf{N_F}$, and C is a concept. Features are distinguished from roles by their *functionality constraint*: while a role can relate an individual to multiple individuals (e.g., $(\mathit{Alice}, \mathit{Bob}) : \mathit{hasChild}$ and $(\mathit{Alice}, \mathit{Carol}) : \mathit{hasChild}$), a feature must be functional, meaning each individual can be related to *at most one* individual via that feature (e.g., $(\mathit{Alice}, 1990\text{-}01\text{-}15) : \mathit{hasBirthDate}$). This restriction makes features particularly useful for modeling deterministic properties such as unique identifiers, birth dates, or registration numbers.

An $\mathcal{EL}^{++}$ **ontology** $O = \langle \mathcal{T}, \mathcal{A} \rangle$ consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. The **signature** of O is defined as $\text{sig}(O) = \text{sig}_C(O) \cup \text{sig}_R(O) \cup \text{sig}_f(O) \cup \text{sig}_I(O)$, where $\text{sig}_C(O)$, $\text{sig}_R(O)$, $\text{sig}_f(O)$, and $\text{sig}_I(O)$ denote the sets of concept names, role names, feature names, and individual names occurring in O, respectively.

2.2 Semantics of $\mathcal{EL}^{++}$

The semantics of $\mathcal{EL}^{++}$ is defined in terms of an *interpretation* $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, where $\Delta^{\mathcal{I}}$ is a non-empty domain and $\cdot^{\mathcal{I}}$ is the interpretation function that maps each individual name $a \in \mathsf{N_I}$ to an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, each concept name $A \in \mathsf{N_C}$ to a subset $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, each role name $r \in \mathsf{N_R}$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and each feature name $f \in \mathsf{N_F}$ to a partial function $f^{\mathcal{I}} : \Delta^{\mathcal{I}} \rightharpoonup \Delta^{\mathcal{I}}$. The interpretation function $\cdot^{\mathcal{I}}$ is inductively extended to complex concepts and roles as follows:

$$
\begin{aligned}
\top^{\mathcal{I}} &= \Delta^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(\exists r.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} \mid \exists e \in \Delta^{\mathcal{I}} : (d,e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}}\} \\
(\exists f.C)^{\mathcal{I}} &= \{d \in \Delta^{\mathcal{I}} \mid f^{\mathcal{I}}(d) \text{ is defined} \wedge f^{\mathcal{I}}(d) \in C^{\mathcal{I}}\} \\
(r_1 \circ \cdots \circ r_k)^{\mathcal{I}} &= \{(d,e) \mid \exists d_1, \ldots, d_{k-1} : (d,d_1) \in r_1^{\mathcal{I}} \wedge \cdots \\
&\qquad \wedge (d_{k-1}, e) \in r_k^{\mathcal{I}}\}
\end{aligned}
$$

An interpretation $\mathcal{I}$ **satisfies** an axiom/assertion α (or we say α **holds** in $\mathcal{I}$, written $\mathcal{I} \models \alpha$) according to the following conditions:

$$
\begin{aligned}
\mathcal{I} \models C \sqsubseteq D \quad &\text{iff} \quad C^{\mathcal{I}} \subseteq D^{\mathcal{I}} \\
\mathcal{I} \models r_1 \circ \cdots \circ r_k \sqsubseteq s \quad &\text{iff} \quad (r_1 \circ \cdots \circ r_k)^{\mathcal{I}} \subseteq s^{\mathcal{I}} \\
\mathcal{I} \models \text{dom}(f) \sqsubseteq C \quad &\text{iff} \quad \{d \in \Delta^{\mathcal{I}} \mid f^{\mathcal{I}}(d) \text{ is defined}\} \subseteq C^{\mathcal{I}} \\
\mathcal{I} \models \text{ran}(f) \sqsubseteq C \quad &\text{iff} \quad \{f^{\mathcal{I}}(d) \mid d \in \Delta^{\mathcal{I}}, \\
&\qquad f^{\mathcal{I}}(d) \text{ is defined}\} \subseteq C^{\mathcal{I}} \\
\mathcal{I} \models a : C \quad &\text{iff} \quad a^{\mathcal{I}} \in C^{\mathcal{I}} \\
\mathcal{I} \models (a,b) : r \quad &\text{iff} \quad (a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}} \\
\mathcal{I} \models (a,b) : f \quad &\text{iff} \quad f^{\mathcal{I}}(a^{\mathcal{I}}) = b^{\mathcal{I}}
\end{aligned}
$$

An interpretation $\mathcal{I}$ is a **model** of an ontology O (written $\mathcal{I} \models O$) iff every axiom/assertion in O holds in $\mathcal{I}$. A concept C is **satisfiable** w.r.t. O iff there exists a model $\mathcal{I}$ of O and some $d \in \Delta^{\mathcal{I}}$ with $d \in C^{\mathcal{I}}$. An axiom/assertion α is **entailed** by O (written $O \models \alpha$) iff α holds in every model of O. An ontology O is **consistent** iff it has a model.

2.3 Reasoning Tasks in $\mathcal{EL}^{++}$

A distinctive feature of $\mathcal{EL}^{++}$ is that certain reasoning tasks become trivial due to its syntactic restrictions. The absence of negation ensures that every $\mathcal{EL}^{++}$ concept is satisfiable and also every $\mathcal{EL}^{++}$ ontology is consistent, as the language lacks the expressivity to formulate contradiction. This contrasts sharply with more expressive DLs like $\mathcal{ALC}$, where checking concept satisfiability and ontology consistency are fundamental tasks. The fundamental reasoning tasks in $\mathcal{EL}^{++}$ are the following:

Subsumption Checking: A concept C is *subsumed by* a concept D with respect to an $\mathcal{EL}^{++}$ ontology O (written $O \models C \sqsubseteq D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds in every model $\mathcal{I}$ of O.

Instance Checking: An individual a is an *instance* of a concept C with respect to an $\mathcal{EL}^{++}$ ontology O (written $O \models a : C$) if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ holds in every model $\mathcal{I}$ of O.

Theorem 1 (Polynomial-Time Reasoning [4]). *Subsumption and instance checking in $\mathcal{EL}^{++}$ is PTIME-complete.*

3 Differentiable Fuzzy Semantics for $\mathcal{EL}^{++}$

Classical logic (e.g., $\mathcal{EL}^{++}$) assumes crisp semantics, where membership is strictly binary—an element either belongs to a concept or it does not. However, such a rigid interpretation is inadequate for domains characterized by vagueness or uncertainty. To address this, we extend the classical $\mathcal{EL}^{++}$ to a fuzzy setting by replacing crisp set membership with fuzzy membership functions. Importantly, this extension preserves the syntactic structure and polynomial-time reasoning guarantees of classical $\mathcal{EL}^{++}$, ensuring its scalability for large ontologies and knowledge bases.

3.1 Semantics of Fuzzy $\mathcal{EL}^{++}$

The semantics of fuzzy $\mathcal{EL}^{++}$ is defined via a fuzzy interpretation, a concept rooted in Zadeh's fuzzy set theory [70]. It replaces classical binary set membership with a continuous degree of membership, providing a natural way to handle vagueness and uncertainty.

A *fuzzy interpretation* is a pair $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$, where $\Delta^{\mathcal{I}}$ is a non-empty domain and $\cdot^{\mathcal{I}}$ is the fuzzy interpretation function that maps each individual name $a \in \mathsf{N_I}$ to an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, each concept name $A \in \mathsf{N_C}$ to a fuzzy membership function $A^{\mathcal{I}} : \Delta^{\mathcal{I}} \to [0,1]$, each role name $r \in \mathsf{N_R}$ to a fuzzy relation $r^{\mathcal{I}} : \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \to [0,1]$, each feature name $f \in \mathsf{N_F}$ to a partial fuzzy functional relation $f^{\mathcal{I}} : \Delta^{\mathcal{I}} \rightharpoonup \Delta^{\mathcal{I}} \times [0,1]$, where $f^{\mathcal{I}}(d) = (e, \mu)$ indicates that d is functionally related to e with membership degree μ. For example, rather than interpreting the concept `HighRiskPatient` as a crisp set, the fuzzy interpretation assigns a degree of membership to each individual: $A^{\mathcal{I}}(\texttt{Alice}) = 0.95$ indicates that Alice is very likely a high-risk patient, while $A^{\mathcal{I}}(\texttt{Bob}) = 0.40$ reflects a borderline case. This nuanced representation goes beyond binary classification.

The key challenge in fuzzy semantics is extending Boolean operations to the continuous interval $[0,1]$ while preserving logical properties essential for sound reasoning (e.g., must be commutative, associative, monotonic, and satisfy boundary conditions aligned with classical logic) [61]. We need operations that: (1) reduce to standard $\mathcal{EL}^{++}$ when restricted to $\{0,1\}$, (2) provide meaningful semantics for intermediate values, and (3) maintain essential logical properties. **T-norms** [37] ($\otimes$) $T : [0,1]^2 \to [0,1]$ fulfill this role by generalizing conjunction ($\sqcap$) to fuzzy settings.

We evaluate three canonical t-norms for fuzzy conjunction [61]: **Gödel** ($\min(x,y)$), **Product** ($x \cdot y$), and **Łukasiewicz** ($\max(0, x + y - 1)$). Gödel [28] preserves logical idempotency and offers strong interpretability, but it is non-differentiable at $x = y$ and yields sparse gradients (0 or 1), limiting its suitability for gradient-based learning. Łukasiewicz [60] is partially differentiable but suffers from vanishing gradients on a large subdomain ($x + y \leq 1$), introducing instability during training. Product t-norm [30], by contrast, is fully differentiable, strictly monotonic, and preserves all t-norm axioms except idempotency, enabling smooth optimization and robust gradient flow. We therefore adopt **Product** as a principled compromise: it retains core logical properties while offering the differentiability and numerical stability necessary for neural-symbolic learning.

For existential quantification, we adopt the **probabilistic sum** t-conorm (dual to the Product t-norm), interpreting "there exists" as the probability that "at least one witness holds" [61]. Its binary form $S_P(x, y) = x + y - xy = 1 - (1 - x)(1 - y)$ naturally generalizes to multiple witnesses as $1-\prod_e(1-r^{\mathcal{I}}(d,e)\cdot C^{\mathcal{I}}(e))$. This formulation ensures compositional consistency with Product-based conjunction, full differentiability, and smooth gradient flow—overcoming the sparse, unstable gradients induced by supremum-based definitions such as $\sup_e \min(r^{\mathcal{I}}(d, e), C^{\mathcal{I}}(e))$.

Under these choices, the interpretation function extends to complex concepts and roles as follows:

$$\top^{\mathcal{I}}(d) = 1$$
$$(C \sqcap D)^{\mathcal{I}}(d) = C^{\mathcal{I}}(d) \cdot D^{\mathcal{I}}(d)$$
$$(\exists r.C)^{\mathcal{I}}(d) = 1 - \prod_{e\in\Delta^{\mathcal{I}}} (1 - r^{\mathcal{I}}(d, e) \cdot C^{\mathcal{I}}(e))$$
$$(\exists f.C)^{\mathcal{I}}(d) = \begin{cases} \mu \cdot C^{\mathcal{I}}(e) & \text{if } f^{\mathcal{I}}(d) = (e, \mu) \\ 0 & \text{if } f^{\mathcal{I}}(d) \text{ undefined} \end{cases}$$
$$(r_1 \circ \cdots \circ r_k)^{\mathcal{I}}(d, e) = 1 - \prod_{d_1,\ldots,d_{k-1}} (1 - r_1^{\mathcal{I}}(d, d_1)\cdot r_2^{\mathcal{I}}(d_1, d_2) \cdot \ldots \cdot r_k^{\mathcal{I}}(d_{k-1}, e))$$

For neural-symbolic implementation, we note that the direct computation of long product chains can lead to numerical underflow. This is a well-known issue in probabilistic modeling [27], and we address it using a standard technique: performing the computation in log-space [48]. This ensures numerical stability while exactly preserving the Product-based semantics.

A fuzzy interpretation $\mathcal{I}$ *satisfies* an axiom α to degree $n \in [0, 1]$ (written $\mathcal{I} \models_n \alpha$). To define this satisfaction degree in a manner that is theoretically consistent with our choice of the Product T-norm, we must select a corresponding fuzzy implication operator. An inconsistent choice, such as an S-implication like Reichenbach ($x \Rightarrow y = 1 - x + xy$) [52], would create a theoretical schism within our framework, as its construction is based on a fuzzy negation and t-conorm, operators that are not native to the negation-free, conjunction-focused language of $\mathcal{EL}^{++}$ [61].

The only theoretically sound choice that preserves the algebraic structure of our Product-based semantics is its corresponding R-implication—the **Göguen implication** [26]. Derived directly from Product t-norm via the residuation principle [6], it is defined as:

$$\mathcal{J}_G(x, y) = \begin{cases} 1 & \text{if } x \le y \\ y/x & \text{if } x > y \end{cases}$$

This operator provides a robust logical foundation for our framework. For axioms that are universally quantified over the domain (such as concept inclusions),[1] the overall satisfaction degree is the **infimum (inf)** of the implication's truth values across all individuals. This captures the "weakest link" principle [29, 37], ensuring the axiom holds true for the entire domain.

Therefore, we define the satisfaction degrees for all $\mathcal{EL}^{++}$ axiom types as follows. Note that for ABox assertions representing ground facts, the satisfaction degree is simply the membership value itself, as no implication is involved.

$$\mathcal{I} \models_n C \sqsubseteq D \text{ iff } n = \inf_{d\in\Delta^{\mathcal{I}}} \{\mathcal{J}_G(C^{\mathcal{I}}(d), D^{\mathcal{I}}(d))\}$$
$$\mathcal{I} \models_n r_1 \circ \cdots \circ r_k \sqsubseteq s \text{ iff } n = \inf_{d,e\in\Delta^{\mathcal{I}}} \{\mathcal{J}_G((r_1 \circ \cdots \circ r_k)^{\mathcal{I}}(d, e), s^{\mathcal{I}}(d, e))\}$$
$$\mathcal{I} \models_n \mathrm{dom}(f) \sqsubseteq C \text{ iff } n = \inf_{d\in\Delta^{\mathcal{I}}} \{\mathcal{J}_G(\mu_d, C^{\mathcal{I}}(d))\}$$
$$\text{where } \mu_d = \sup\{\mu \mid \exists e. f^{\mathcal{I}}(d) = (e, \mu)\}$$
$$\mathcal{I} \models_n \mathrm{ran}(f) \sqsubseteq C \text{ iff } n = \inf_{e\in\Delta^{\mathcal{I}}} \{\mathcal{J}_G(\nu_e, C^{\mathcal{I}}(e))\}$$
$$\text{where } \nu_e = \sup\{\mu \mid \exists d. f^{\mathcal{I}}(d) = (e, \mu)\}$$
$$\mathcal{I} \models_n a : C \text{ iff } n = C^{\mathcal{I}}(a^{\mathcal{I}})$$
$$\mathcal{I} \models_n (a, b) : r \text{ iff } n = r^{\mathcal{I}}(a^{\mathcal{I}}, b^{\mathcal{I}})$$
$$\mathcal{I} \models_n (a, b) : f \text{ iff } n = \mu \text{ when } f^{\mathcal{I}}(a^{\mathcal{I}}) = (b^{\mathcal{I}}, \mu)$$

A fuzzy interpretation $\mathcal{I}$ is a **fuzzy model** of an ontology $\mathcal{O}$ to degree $n \in [0, 1]$ (written $\mathcal{I} \models_n \mathcal{O}$) iff $n = \inf_{\alpha\in\mathcal{O}}\{\mathcal{I} \models_m \alpha \mid \mathcal{I} \models_m \alpha\}$, i.e., the satisfaction degree of $\mathcal{O}$ is the infimum of satisfaction degrees of all axioms in $\mathcal{O}$. An axiom α is **entailed** by $\mathcal{O}$ to degree n (written $\mathcal{O} \models_n \alpha$) iff $n = \inf_{\mathcal{I}}\{\mathcal{I} \models_m \alpha \mid \mathcal{I} \models_k \mathcal{O} \text{ for some } k > 0\}$, i.e., the entailment degree is the infimum of satisfaction degrees of α across all fuzzy models of $\mathcal{O}$.

Theorem 2 (Complexity Preservation). *Fuzzy subsumption and instance checking in fuzzy $\mathcal{EL}^{++}$ is PTIME-complete.*

4 Neural-Symbolic Learning Framework

This section details the complete framework for learning from logical constraints. We present the three core components that translate our theoretical semantics into a practical, end-to-end optimizable system: a normalization procedure for axioms, a neural parameterization of the fuzzy interpretation, and a unified loss function derived directly from logical principles.

4.1 Normalization

The path from fuzzy to differentiable semantics requires structured axiom representations. Complex nested expressions like $(A \sqcap B) \sqsubseteq \exists r.(C \sqcap D)$ combine multiple logical operations, making it difficult to design targeted loss functions that enforce specific reasoning patterns. We establish a normalization procedure that decomposes complex axioms into elementary forms, each containing at most one logical operator beyond subsumption. This decomposition enables specialized neural optimization strategies while preserving semantic equivalence. A fuzzy $\mathcal{EL}^{++}$ TBox is in **normal form (NF)** if every concept inclusion has one of the following forms:

[1] In DLs, TBox axioms carry an **implicit universal quantifier**; they are terminological statements that must hold for all individuals in any valid model. It is crucial to distinguish this implicit quantification, which is part of the semantics of $\sqsubseteq$, from an explicit universal quantifier constructor (e.g., $\forall r.C$ in $\mathcal{ALC}$), which $\mathcal{EL}^{++}$ omits to ensure computational tractability.

NF1: $A \sqsubseteq B$ NF3: $A \sqsubseteq \exists r.B$ NF5: $A \sqsubseteq \exists f.B$
NF2: $A_1 \sqcap A_2 \sqsubseteq B$ NF4: $\exists r.A \sqsubseteq B$ NF6: $\exists f.A \sqsubseteq B$

where $r \in \mathsf{N_R}$ is a role name, $f \in \mathsf{N_F}$ is a feature name, and $A, A_1, A_2, B \in \mathsf{N_C}$ denote concept names or the top concept.

Note that role inclusions ($r_1 \circ \cdots \circ r_k \sqsubseteq s$) and feature restrictions ($\mathrm{dom}(f) \sqsubseteq C, \mathrm{ran}(f) \sqsubseteq C$) already have simple structures and thus require no normalization. Similarly, ABox assertions maintain their atomic forms ($(a : C, (a, b) : r, (a, b) : f)$) as they directly represent ground facts without complex logical structures.

The normalization process systematically decomposes complex axioms by apply the following transformation rules. Each rule NR1-NR7 identifies exactly one complex concept $\hat{C}$ and replaces it with a fresh concept name A_{new}, while introducing an auxiliary axiom to preserve semantics ($\Longrightarrow$ indicates this transformation). For instance, rule NR2 transforms $\hat{C} \sqcap D \sqsubseteq B$ into two simpler axioms: $\hat{C} \sqsubseteq A_{\text{new}}$ (defining the intermediate concept) and $A_{\text{new}} \sqcap D \sqsubseteq B$ (using the intermediate concept). Rule NR8 uses logical equivalence ($\Leftrightarrow$) to split conjunctive consequents without introducing new symbols. Exhaustive application yields only the six normal forms that can be directly optimized by neural networks.

Apply the following transformation rules exhaustively:

NR1: $\hat{C} \sqsubseteq \hat{D} \Rightarrow \hat{C} \sqsubseteq A_{\text{new}}, A_{\text{new}} \sqsubseteq \hat{D}$

NR2: $\hat{C} \sqcap D \sqsubseteq B \Rightarrow \hat{C} \sqsubseteq A_{\text{new}}, A_{\text{new}} \sqcap D \sqsubseteq B$

NR3: $D \sqcap \hat{C} \sqsubseteq B \Rightarrow \hat{C} \sqsubseteq A_{\text{new}}, D \sqcap A_{\text{new}} \sqsubseteq B$

NR4: $\exists r.\hat{C} \sqsubseteq B \Rightarrow \hat{C} \sqsubseteq A_{\text{new}}, \exists r.A_{\text{new}} \sqsubseteq B$

NR5: $B \sqsubseteq \exists r.\hat{C} \Rightarrow A_{\text{new}} \sqsubseteq \hat{C}, B \sqsubseteq \exists r.A_{\text{new}}$

NR6: $\exists f.\hat{C} \sqsubseteq B \Rightarrow \hat{C} \sqsubseteq A_{\text{new}}, \exists f.A_{\text{new}} \sqsubseteq B$

NR7: $B \sqsubseteq \exists f.\hat{C} \Rightarrow A_{\text{new}} \sqsubseteq \hat{C}, B \sqsubseteq \exists f.A_{\text{new}}$

NR8: $B \sqsubseteq D \sqcap E \Leftrightarrow B \sqsubseteq D, B \sqsubseteq E$

where $r \in \mathsf{N_R}$ is a role name, $f \in \mathsf{N_F}$ is a feature name, $B \in \mathsf{N_C}$ denotes a concept name, D, E denote arbitrary concepts, $\hat{C}, \hat{D}$ denote complex concepts (neither concept names nor $\top$), and $A_{\text{new}} \in \mathsf{N_C}$ denotes a fresh concept name.

Theorem 3 (Semantic Preservation). *For any fuzzy $\mathcal{EL}^{++}$ ontology $\mathcal{O}$, the normalization process produces $\mathcal{O}'$ in polynomial time such that:*

(1) $\mathcal{O}'$ is a conservative extension of $\mathcal{O}$ [39];
(2) For any axiom α using only symbols from $\mathrm{sig}(\mathcal{O})$: $\mathcal{O} \models_n \alpha$ iff $\mathcal{O}' \models_n \alpha$ for all $n \in [0, 1]$;
*(3) The size of $\mathcal{O}'$ is **linear** in the size of $\mathcal{O}$.*

This theorem ensures that normalization preserves all logical consequences while maintaining computational efficiency. The conservative extension property guarantees that new concept names do not affect the semantics of original concepts, enabling faithful neural-symbolic integration.

4.2 Interpretation Domain and Embeddings

4.2.1 Semantically-Aware Domain Construction. A challenge in differentiable fuzzy $\mathcal{EL}^{++}$ is constructing a semantically adequate interpretation domain $\Delta^{\mathcal{I}}$. A naive approach, where the domain consists solely of individuals explicitly named in the ABox ($\Delta_{explicit}$), is semantically inadequate and theoretically flawed. Such a domain risks learning *degenerate solutions*, where the optimization process may trivially satisfy TBox axioms by allowing concepts to collapse into empty sets. For instance, the axiom $\mathtt{A} \sqsubseteq \exists \mathtt{r}.\mathtt{B}$ becomes vacuously true if the model learns that the interpretation of concept $\mathtt{A}$ is empty, a failure mode that prevents meaningful reasoning.

Our **semantically-aware domain construction** strategy is not an ad-hoc heuristic, but a principled and necessary precondition for sound reasoning. By systematically introducing "witness" individuals, we ensure that the domain is sufficiently expressive to model all logical constraints mandated by the TBox. Existential witnesses serve as concrete instantiations required by $\exists$-restrictions, guaranteeing that the model cannot ignore these constraints. More subtly, concept witnesses act as semantic anchors for every subconcept in the ontology, ensuring each concept is grounded in the model and learns a non-trivial, non-empty representation. This principled expansion transforms the domain from a simple set of entities into a structured canvas for reasoning, upon which a robust and logically coherent fuzzy interpretation can be learned.

Explicit Individuals. All individuals explicitly mentioned in the ABox are included:

$$\Delta_{\text{explicit}} = \{a \in \mathsf{N_I} \mid a \in \mathrm{sig}(\mathcal{A})\}$$

Concept Witnesses. For each subconcept[2] C in the normalized TBox, we add a witness e_C:

$$\Delta_{\text{concept}} = \{e_C \mid C \in \mathrm{sub}(\mathcal{T})\}$$

This guarantees that every concept is grounded in the model.

Existential Witnesses. For each axiom $A \sqsubseteq \exists r.B$ and each individual d such that $A^{\mathcal{I}}(d) > 0$, we introduce a fresh witness $w_{d,r,B}$ and enforce[3]:

$$r^{\mathcal{I}}(d, w_{d,r,B}) \approx 1, \quad B^{\mathcal{I}}(w_{d,r,B}) \approx 1$$

Role Chain Expansion. For any role chain axiom $r_1 \circ r_2 \sqsubseteq s$, we construct intermediate nodes w_1, w_2 such that:

$$d \xrightarrow{r_1} w_1 \xrightarrow{r_2} w_2 \quad \Rightarrow \quad d \xrightarrow{s} w_2$$

This ensures chain closure and supports transitive inferences.

Feature Targets. For each feature f in an existential $\exists f.C$, we construct a set of target candidates $\{t_f^{(1)}, \ldots, t_f^{(\lambda)}\}$ and define:

$$f^{\mathcal{I}}(d, e) = \mathrm{softmax}_e \left(\mathbf{F}_{f,d,e} / \tau_f \right), \quad \text{ensuring } \sum_e f^{\mathcal{I}}(d, e) = 1$$

The entire interpretation domain is:

$$\Delta^{\mathcal{I}} = \Delta_{\text{explicit}} \cup \Delta_{\text{concept}} \cup \Delta_{\text{exist}} \cup \Delta_{\text{chain}} \cup \Delta_{\text{feature}}$$

The size of $\Delta^{\mathcal{I}}$ is polynomially bounded:

$$|\Delta^{\mathcal{I}}| = O(|\mathcal{T}| \cdot |\mathrm{sig_I}(\mathcal{O})| + \mathrm{size}(\mathcal{T}) + |\mathcal{T}| \cdot \log |\mathrm{sig_C}(\mathcal{O})|),$$

[2]See Chapter 3 of [5] for standard subconcept definitions of DL concepts.
[3]The '≈ 1' notation signifies an optimization target, not a hard assignment. In our framework, this enforcement is realized by adding specific penalty terms to the overall loss function. For instance, one could add squared error losses $(1 - r^{\mathcal{I}}(d, w_{d,r,B}))^2$ and $(1 - B^{\mathcal{I}}(w_{d,r,B}))^2$, which are minimized during training to drive the corresponding membership degrees towards 1.

where $|\mathcal{T}|$ denotes the total number of axioms in the normalized TBox and size($\mathcal{T}$) the total number of symbols in the TBox.

4.2.2 Concept and Role Embeddings. To facilitate gradient-based learning, we parameterize the fuzzy interpretation. Each concept $C \in \mathsf{N_C}$ is embedded as a vector $\mathbf{c}_C \in \mathbb{R}^{|\Delta^{\mathcal{I}}|}$, and each role $r \in \mathsf{N_R}$ as a matrix $\mathbf{R}_r \in \mathbb{R}^{|\Delta^{\mathcal{I}}| \times |\Delta^{\mathcal{I}}|}$. Their fuzzy membership degrees are then computed via the sigmoid function $\sigma(\cdot)$:

$$C^{\mathcal{I}}(d_i) = \sigma(\mathbf{c}_{C,i})$$
$$r^{\mathcal{I}}(d_i, d_j) = \sigma(\mathbf{R}_{r,ij})$$

Features $f \in \mathsf{N_F}$, parameterized by an embedding $\mathbf{F}_f$, require special treatment to maintain their functionality constraint. We achieve this by applying softmax normalization over potential targets d_j:

$$f^{\mathcal{I}}(d_i, d_j) = \text{softmax}_j \left(\frac{\mathbf{F}_{f,i,j}}{\tau_f} \right)$$

where the temperature $\tau_f > 0$ controls the sharpness of the distribution, ensuring that $\sum_j f^{\mathcal{I}}(d_i, d_j) = 1$ for each individual d_i.

4.2.3 Differentiable and Semantically Sound Evaluation. A key advantage of our chosen Product-based fuzzy semantics is its inherent suitability for gradient-based optimization. The Product t-norm itself is a smooth function providing rich gradient flow. However, a practical challenge arises in the evaluation of operators involving products over large sets, such as the existential quantifier ($\exists r.C$) and role chains. The formula $1 - \prod_e (1 - x_e)$ can be prone to numerical underflow when many terms $(1 - x_e)$ are close to zero.

To address this while strictly preserving the semantics, we employ a standard technique from probabilistic modeling: performing the computation in **log-space** [48]. Instead of computing a product of terms, we compute the sum of their logarithms. For example, the product term in the existential quantifier is computed as:

$$\prod_{e \in \Delta^{\mathcal{I}}} (1 - x_e) = \exp\left(\sum_{e \in \Delta^{\mathcal{I}}} \log(1 - x_e) \right)$$

This transformation is not an approximation but a mathematically equivalent and numerically stable method of computation. This leads to a crucial theoretical guarantee for our framework:

Proposition 1 (Semantic Equivalence). *Let $[\![\phi]\!]_{fuzzy}$ be the truth value of any $\mathcal{EL}^{++}$ formula ϕ under our Product-based fuzzy semantics, and let $[\![\phi]\!]_{diff}$ be its value computed by our differentiable framework using log-space evaluation. Then, for all formulas ϕ, we have:*

$$[\![\phi]\!]_{diff} = [\![\phi]\!]_{fuzzy}$$

This proposition formally guarantees that our differentiable implementation does not compromise the integrity of the fuzzy logical theory, ensuring that the learning process is guided by a sound and consistent semantics. Finally, to bridge the continuous outputs of our model with discrete logical reasoning for evaluation, we employ an adaptive discretization function:

$$\text{discretize}(v, \alpha) = \begin{cases} \text{asserted} & \text{if } v > \alpha \\ \text{unknown} & \text{if } v \leq \alpha \end{cases}$$

where $\alpha \in (0, 1)$ is a confidence threshold.

4.3 Loss Derived from Fuzzy Semantics

Our learning objective is to find an interpretation $\mathcal{I}$ that satisfies all axioms in the normalized TBox $\mathcal{T}'$. Instead of adopting a generic machine learning loss or designing complex, ad-hoc heuristics, we derive our loss function directly from the Gögen implication that underpins our fuzzy semantics.

The goal is to maximize the satisfaction degree of each axiom, which is equivalent to minimizing its **dissatisfaction**. For any normalized axiom α of the generalized form $LHS \sqsubseteq RHS$ and an individual d, we define the dissatisfaction as $1 - \mathcal{J}_G(LHS^{\mathcal{I}}(d), RHS^{\mathcal{I}}(d))$. Here, $LHS^{\mathcal{I}}(d)$ and $RHS^{\mathcal{I}}(d)$ represent the fuzzy membership degrees of the axiom's premise (Left-Hand Side, the **Head**) and conclusion (Right-Hand Side, the **Tail**) respectively, computed according to the fuzzy semantics defined in Section 3.

Based on the definition of Gögen implication, this dissatisfaction can be expressed as:

$$1 - \mathcal{J}_G(x, y) = \begin{cases} 0 & \text{if } x \leq y \\ 1 - y/x & \text{if } x > y \end{cases}$$

where $x = LHS^{\mathcal{I}}(d)$ and $y = RHS^{\mathcal{I}}(d)$. This piecewise function can be written compactly as $\max(0, 1 - y/x)$. This leads to our unified, theoretically-grounded loss function for any axiom α:

$$\mathcal{L}_{\text{axiom}}(\alpha) = \sum_{d \in \Delta^{\mathcal{I}}} \max\left(0, 1 - \frac{RHS^{\mathcal{I}}(d)}{LHS^{\mathcal{I}}(d) + \epsilon}\right) \tag{1}$$

where ϵ is a small constant (e.g., 10^{-12}) to ensure numerical stability. For instance, to compute the loss for an axiom of the form $A_1 \sqcap A_2 \sqsubseteq B$, the premise degree $LHS^{\mathcal{I}}(d)$ would first be computed as $A_1^{\mathcal{I}}(d) \cdot A_2^{\mathcal{I}}(d)$ before being passed into the loss function.

This loss function offers significant advantages over a standard Hinge loss (e.g., $\max(0, x - y)$) [15]. It captures the *ratio* of violation, providing a more nuanced and semantically meaningful penalty. For instance, a violation from a premise degree of 0.2 to a conclusion degree of 0.1 (a 50% drop in confidence) is penalized more heavily than a violation from 0.9 to 0.8 (an 11% drop), a distinction that a simple difference-based loss would miss.

The total loss for the ontology is then the sum of losses over all normalized axioms, plus a standard regularization term to prevent overfitting. Specifically, we use L2 regularization, where $\mathcal{L}_{\text{reg}}$ is the sum of the squared Frobenius norms of all learnable embedding parameters (concept vectors and role matrices) [27]:

$$\mathcal{L}_{\text{total}} = \sum_{\alpha \in \mathcal{T}'} \mathcal{L}_{\text{axiom}}(\alpha) + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}} \tag{2}$$

This unified loss function is simple, theoretically motivated, and directly enforces the fuzzy semantics of $\mathcal{EL}^{++}$.

Theorem 4 (Soundness of the Learning Objective). *Let O be an $\mathcal{EL}^{++}$ ontology and let $\mathcal{L}_{total}$ be the total loss function defined in Equation (2), assuming a non-negative regularization term. If there exist embeddings for all concepts and roles such that $\mathcal{L}_{total}(O) = 0$, then the resulting fuzzy interpretation $\mathcal{I}$ is a **fuzzy model** of O with a satisfaction degree of 1.*

Table 1: KBC accuracy results on TBox-centric (SNOMED CT, GO) and ABox-centric (Yeast PPI, Human PPI) datasets. Hits@k metrics are reported as percentages. Best results in bold, second-best underlined. DNF: Did Not Finish within 72h.

Model	SNOMED CT (377K)					GO (44K)					Yeast PPI (110K)				Human PPI (75K)			
	TBox Reasoning (Subsumption $LHS \sqsubseteq RHS$)										ABox Reasoning (PPI $\{P_1\} \sqsubseteq \exists$**interacts**.$\{P_2\}$)							
	H@1	H@10	H@100	MRR	Med	H@1	H@10	H@100	MRR	Med	H@10	H@100	MRR	Med	H@10	H@100	MRR	Med
DF-$\mathcal{EL}^{++}$	**5.1**	**25.9**	**68.4**	**0.115**	**45**	**5.5**	**26.8**	**71.2**	**0.121**	**41**	**35.7**	**89.6**	**0.391**	**91**	**31.8**	**87.4**	**0.375**	**238**
Geometric EL++ Models																		
ELEM	2.4	20.1	38.9	0.078	78	2.9	23.8	43.4	0.089	270	23.1	74.8	0.301	184	22.3	69.7	0.268	567
BoxEL	0.8	4.2	6.8	0.025	9231	1.2	5.8	8.3	0.032	8167	19.7	72.6	0.248	376	10.4	62.7	0.162	1524
Box²EL	3.6	20.8	52.1	0.095	59	4.1	23.2	57.0	0.103	48	33.2	86.9	0.368	115	28.4	82.7	0.346	264
ELBE	1.0	7.8	19.2	0.034	2134	1.3	9.2	22.8	0.041	1889	25.8	76.9	0.322	151	21.9	71.8	0.274	358
EmEL++	2.1	19.4	32.8	0.072	672	2.7	22.9	37.7	0.084	578	17.3	64.8	0.244	287	13.2	55.6	0.201	696
Expressive Neuro-Symbolic Models																		
LTN	DNF					0.8	4.2	18.7	0.031	2834	7.1	29.4	0.128	1247	6.8	27.9	0.121	1389
logLTN	DNF					1.1	6.8	24.3	0.047	1689	9.4	38.7	0.167	892	8.9	35.2	0.154	967
FALCON	DNF					1.4	8.7	29.1	0.058	1234	11.2	44.8	0.194	634	10.7	42.1	0.183	708
Faster-LTN	DNF					0.9	5.9	21.6	0.041	2156	8.3	33.9	0.148	1098	7.8	31.4	0.139	1176
Standard KGE Models (TBox-agnostic)																		
TransE	0.3	2.1	8.7	0.018	6247	0.4	2.8	12.3	0.023	4892	4.2	18.9	0.094	2134	3.8	16.7	0.087	2389
RotatE	0.4	3.2	11.8	0.025	5124	0.6	4.1	16.7	0.032	3456	5.9	24.1	0.118	1678	5.2	21.3	0.109	1892

4.4 Complexity Analysis

A key theoretical result underpinning our framework's practical scalability is its computational efficiency, which inherits the PTIME-tractability of the underlying $\mathcal{EL}^{++}$ logic. A formal analysis reveals that the DF-$\mathcal{EL}^{++}$ learning framework has a worst-case time complexity of $O(|\mathcal{T}'| \cdot |N_I|^2)$ per training iteration. This complexity is **polynomial** in the size of the input ontology O, as both the number of axioms after normalization, $|\mathcal{T}'|$, and the number of individuals in our semantically-aware domain, $|N_I|$, are themselves polynomially bounded by the size of O. The overall complexity is derived from the cost of evaluating the loss for each of the $|\mathcal{T}'|$ normalized axioms. The most computationally expensive axioms are those involving existential restrictions or role chains, which require an aggregation over pairs of individuals, costing $O(|N_I|^2)$ per axiom. This polynomial-time guarantee provides the theoretical foundation for our framework's effectiveness on large-scale knowledge bases. A detailed proof of this result is provided in Appendix B.

5 Empirical Evaluation

The primary task to evaluate our framework is Knowledge Base Completion (KBC) [65], a comprehensive inductive reasoning task that assesses a model's ability to generalize from known axioms to predict missing ontological knowledge. KBC provides holistic evaluation across the full range of $\mathcal{EL}^{++}$ constructors, encompassing hierarchical concept relations ($C \sqsubseteq D$) and ABox assertions ($C(a)$, $r(a,b)$). Success requires learning rich representations that capture both hierarchical and relational structure of the ontology.

5.1 Experimental Setup

5.1.1 Datasets and Baselines. We evaluate on four carefully selected real-world datasets representing distinct computational challenges: SNOMED CT (377K concepts) as the definitive scalability benchmark in medical terminology, Gene Ontology (GO, 44K concepts) providing hierarchically rich biological taxonomy, and two protein-protein interaction datasets (Yeast PPI: 110K entities, Human PPI: 75K entities) combining STRING database interactions with GO TBox constraints for ABox reasoning evaluation.

We establish comprehensive comparisons across three methodological categories: (1) **Geometric $\mathcal{EL}^{++}$ models** (ELEM, BoxEL, Box^{2}EL, ELBE, EmEL++) as direct competitors targeting the same logical fragment; (2) **Expressive neuro-symbolic models** (LTN, logLTN, FALCON, Faster-LTN) demonstrating computational challenges of EXPTIME-complete frameworks; (3) **Standard KGE models** (TransE, RotatE) as TBox-agnostic baselines highlighting the importance of structured logical reasoning.

5.1.2 Protocol and Evaluation. Our protocol is designed for rigorous evaluation of inductive reasoning capabilities. For each ontology, we partition the original asserted axioms into training (80%), validation (10%), and test (10%) sets before any normalization to prevent information leakage. We unify diverse prediction tasks through **nominals** (concepts representing single individuals a, denoted as $\{a\}$), transforming concept assertions $a : C$ to $\{a\} \sqsubseteq C$ and role assertions $(a, b) : r$ to $\{a\} \sqsubseteq \exists r.\{b\}$.

For evaluation, we construct ranking tasks by replacing axiom components and ranking the correct completion against candidate entities. We report standard metrics: Hits@K ($K \in \{1, 10, 100\}$), Mean Reciprocal Rank (MRR), and Median Rank (Med).

5.1.3 Implementation Details. All experiments were conducted on servers equipped with NVIDIA RTX 3090 GPUs. Our framework is implemented in PyTorch 1.12 with CUDA 11.6 support. For all models, the Adam optimizer was used with an early stopping strategy (patience of 10 epochs) based on validation MRR.

A crucial component of our training protocol is **negative sampling**. For each positive axiom in a training batch, we generate

Table 2: KBC efficiency analysis across four datasets

Model	SNOMED CT (377K)					GO (44K)					Yeast PPI (110K)					Human PPI (75K)				
	Params (M)	Epochs	T/Ep (min)	Total (h)	Mem (GB)	Params (M)	Epochs	T/Ep (min)	Total (h)	Mem (GB)	Params (M)	Epochs	T/Ep (min)	Total (h)	Mem (GB)	Params (M)	Epochs	T/Ep (min)	Total (h)	Mem (GB)
DF-$\mathcal{EL}^{++}$	**8.2**	**35**	**42.3**	**24.7**	**12.4**	**8.1**	**32**	**4.8**	**2.6**	**6.2**	**8.4**	**41**	**19.7**	**13.5**	**10.3**	**8.3**	**38**	**11.2**	**7.1**	**8.7**
Geometric EL++ Models																				
ELEM	12.4	42	38.1	26.7	11.8	12.2	39	4.2	2.7	5.9	12.6	47	17.8	13.9	9.4	12.5	44	9.8	7.2	8.1
BoxEL	15.7	47	41.2	32.3	15.8	15.5	43	4.7	3.4	7.8	15.9	52	21.1	18.3	12.1	15.8	49	12.6	10.3	10.2
Box²EL	22.1	52	47.8	41.5	22.1	21.8	48	6.1	4.9	11.2	22.5	58	26.4	25.5	17.8	22.3	54	15.9	14.3	14.6
ELBE	16.2	45	44.2	33.2	16.2	16.0	41	5.6	3.8	8.1	16.5	50	24.1	20.1	13.2	16.3	47	14.3	11.2	10.9
EmEL++	16.8	48	45.7	36.6	18.3	16.6	44	5.5	4.0	9.2	17.1	53	23.8	21.0	14.7	16.9	50	13.7	11.4	12.1
Expressive Neuro-Symbolic Models																				
LTN	28.4	>180	>65.0	DNF	>32	28.2	156	8.7	22.6	18.4	28.7	178	37.2	110.4	26.3	28.5	167	24.1	67.1	22.1
logLTN	31.2	>175	>68.0	DNF	>28	31.0	148	8.9	21.9	16.7	31.5	169	39.8	112.1	24.8	31.3	159	25.7	68.1	20.4
FALCON	26.7	>170	>62.0	DNF	>30	26.5	142	9.1	21.5	17.9	27.0	161	35.6	95.6	25.9	26.8	151	22.9	57.6	21.7
Faster-LTN	24.9	>90	>58.0	DNF	>24	24.7	76	6.3	8.0	14.2	25.2	89	28.4	42.1	21.1	25.0	84	18.7	26.2	17.8
Standard KGE Models (TBox-agnostic)																				
TransE	6.8	65	15.2	16.5	8.9	6.7	58	2.1	2.0	4.4	6.9	67	8.1	9.0	6.7	6.8	62	5.4	5.6	5.8
RotatE	8.4	58	21.7	21.0	11.2	8.3	52	3.2	2.8	5.6	8.5	59	12.8	12.6	8.9	8.4	55	7.9	7.2	7.3

a set of $\omega = 50$ negative samples by corrupting either the head or the tail component, chosen uniformly at random. We employ a self-adversarial negative sampling loss with a fixed margin of $\delta = 1.0$. To prevent overfitting, the total loss for all models includes a standard **L2 regularization term** with weight $\lambda_{reg} = 1 \times 10^{-5}$ applied to all learnable embeddings.

To ensure a fair comparison, hyperparameters for all baseline models were carefully tuned via grid search on dedicated validation sets. The search space included learning rates $\{1 \times 10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}\}$, embedding dimensions $\{100, 200, 400\}$, and batch sizes $\{256, 512, 1024\}$. For our own model, we report results using the optimal configuration found via the same search: a learning rate of 2×10^{-4}, an embedding dimension of 200, and a batch size of 512 for SNOMED CT and 1024 for other datasets.

Finally, to guarantee the statistical significance and reproducibility of our results, all reported metrics are the **average of 5 runs with different random seeds**. The complete hyperparameter details for all baselines are provided in Appendix C.3.

5.2 Results and Analysis

Table 5 demonstrates DF-$\mathcal{EL}^{++}$'s consistent superiority across all datasets and metrics. Our framework achieves substantial improvements over the strongest geometric baseline (Box^{2}EL): 42% relative improvement in Hits@1 on SNOMED CT (5.1% vs 3.6%), 34% on GO (5.5% vs 4.1%), and 8–12% on PPI datasets, with similarly strong improvements in MRR and median rank metrics. Complete results with standard deviations are provided in Appendix D.

The primary reason for this success lies in our framework's high logical fidelity. Unlike geometric models that approximate logical constraints through spatial relationships, our loss function derives directly from fuzzy logical semantics, forcing learned embeddings to maintain strict consistency with ontological structure. This is particularly evident in precision-oriented metrics (Hits@1, MRR) where exact ranking matters most.

The results on SNOMED CT offer definitive validation of our polynomial complexity advantages. Crucially, the efficiency gains detailed in Table 2 are achieved **not by sacrificing accuracy, but alongside it**. DF-$\mathcal{EL}^{++}$ significantly exceeds the accuracy of geometric models like Box²EL while requiring significantly less training time (24.7h vs. 41.5h). While expressive neuro-symbolic methods (LTN, FALCON) fail to converge within 72-hour limits, our framework remains highly efficient. This unique combination of SOTA accuracy and scalability is the primary contribution of our work.

The dramatic performance differential with standard KGE models (2–3× improvements in Hits@1) confirms the fundamental importance of incorporating structured logical reasoning. Cross-dataset consistency demonstrates robust generalization capabilities, establishing our framework as a highly scalable and accurate approach for neural-symbolic ontological reasoning.

6 Conclusion and Future Work

This paper addresses a fundamental challenge in neural-symbolic AI: reconciling the logical fidelity of symbolic reasoning with the computational scalability required for real-world applications. We introduced DF-$\mathcal{EL}^{++}$, a framework that demonstrates these two requirements need not be mutually exclusive. Our core contribution is a theoretically-grounded methodology that unifies the PTIME-tractability of $\mathcal{EL}^{++}$ with deep learning optimization through principled fuzzy semantics and a loss function derived directly from logical implication. The empirical results validate our theoretical claims: DF-$\mathcal{EL}^{++}$ achieves SOTA accuracy on large-scale KBC tasks while maintaining polynomial-time complexity, consistently outperforming both geometric approximation methods and expressive but computationally intractable alternatives.

Our work opens several promising research directions. First, we plan to investigate end-to-end training scenarios where gradients from our reasoning module fine-tune upstream neural perception models, creating fully integrated learning-reasoning pipelines. Second, while our framework already scales to hundreds of thousands of concepts, we aim to explore sparse-matrix representations to handle knowledge bases with billions of entities. Finally, extending the framework to learn new logical axioms from data would bridge our deductive approach with Inductive Logic Programming [47], enabling systems that can both apply and discover knowledge.

References

[1] Kareem Ahmed, Stefano Teso, Kai-Wei Chang, Guy Van den Broeck, and Antonio Vergari. 2022. Semantic Probabilistic Layers for Neuro-Symbolic Learning. In *Advances in Neural Information Processing Systems 35 (NeurIPS'22)*.

[2] Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. 2016. Neural Module Networks. In *Proc. CVPR'16*. IEEE Computer Society, 39–48.

[3] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, Midori A Harris, David P Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C Matese, Joel E Richardson, Martin Ringwald, Gerald M Rubin, and Gavin Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25, 1 (2000), 25–29.

[4] Franz Baader, Sebastian Brandt, and Carsten Lutz. 2005. Pushing the EL Envelope. In *Proc. IJCAI'05*. Professional Book Center, 364–369.

[5] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. 2017. *An Introduction to Description Logic*. Cambridge University Press.

[6] Michal Baczynski. 2004. Residual implications revisited. Notes on the Smets-Magrez Theorem. *Fuzzy Sets Syst.* 145, 2 (2004), 267–277.

[7] Samy Badreddine, Artur S. d'Avila Garcez, Luciano Serafini, and Michael Spranger. 2022. Logic Tensor Networks. *Artif. Intell.* 303 (2022), 103649.

[8] Samy Badreddine, Luciano Serafini, and Michael Spranger. 2023. logLTN: Differentiable Fuzzy Logic in the Logarithm Space. *CoRR* abs/2306.14546 (2023).

[9] Matej Balog, Alexander L. Gaunt, Marc Brockschmidt, Sebastian Nowozin, and Daniel Tarlow. 2017. DeepCoder: Learning to Write Programs. In *Proc. ICLR'17*. OpenReview.net.

[10] Tarek R. Besold, Artur S. d'Avila Garcez, Sebastian Bader, Howard Bowman, Pedro M. Domingos, Pascal Hitzler, Kai-Uwe Kühnberger, Luís C. Lamb, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, Hoifung Poon, and Gerson Zaverucha. 2021. Neural-Symbolic Learning and Reasoning: A Survey and Interpretation. In *Neuro-Symbolic Artificial Intelligence: The State of the Art*. Frontiers in Artificial Intelligence and Applications, Vol. 342. IOS Press, 1–51.

[11] Matko Bosnjak, Tim Rocktäschel, Jason Naradowsky, and Sebastian Riedel. 2017. Programming with a Differentiable Forth Interpreter. In *Proc. ICML'17 (Proceedings of Machine Learning Research, Vol. 70)*. PMLR, 547–556.

[12] Stephen P. Boyd and Lieven Vandenberghe. 2014. *Convex Optimization*. Cambridge University Press.

[13] Xianjie Chen, Roozbeh Mottaghi, Xiaobai Liu, Sanja Fidler, Raquel Urtasun, and Alan L. Yuille. 2014. Detect What You Can: Detecting and Representing Objects Using Holistic Models and Body Parts. In *Proc. CVPR'14*. IEEE Computer Society, 1979–1986.

[14] William W. Cohen, Fan Yang, and Kathryn Mazaitis. 2020. TensorLog: A Probabilistic Database Implemented Using Deep-Learning Infrastructure. *J. Artif. Intell. Res.* 67 (2020), 285–325.

[15] Corinna Cortes and Vladimir Vapnik. 1995. Support-Vector Networks. *Mach. Learn.* 20, 3 (1995), 273–297.

[16] Andrew Cropper. 2019. Playgol: Learning Programs Through Play. In *Proc. IJCAI'19*. ijcai.org, 6074–6080.

[17] Artur S. d'Avila Garcez, Krysia Broda, and Dov M. Gabbay. 2002. *Neural-symbolic learning systems - foundations and applications*. Springer.

[18] Michelangelo Diligenti, Marco Gori, and Claudio Saccà. 2017. Semantic-based regularization for learning and inference. *Artif. Intell.* 244 (2017), 143–165.

[19] Ivan Donadello, Luciano Serafini, and Artur S. d'Avila Garcez. 2017. Logic Tensor Networks for Semantic Image Interpretation. In *Proc. IJCAI'17*. ijcai.org, 1596–1602.

[20] Kevin Ellis, Lucas Morales, Mathias Sablé-Meyer, Armando Solar-Lezama, and Josh Tenenbaum. 2018. Learning Libraries of Subroutines for Neurally-Guided Bayesian Program Induction. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*. 7816–7826.

[21] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639 (2017), 115–118.

[22] Richard Evans and Edward Grefenstette. 2018. Learning Explanatory Rules from Noisy Data. *J. Artif. Intell. Res.* 61 (2018), 1–64.

[23] Marc Fischer, Mislav Balunovic, Dana Drachsler-Cohen, Timon Gehr, Ce Zhang, and Martin T. Vechev. 2019. DL2: Training and Querying Neural Networks with Logic. In *Proc. ICML'19 (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 1931–1941.

[24] Ross B. Girshick. 2015. Fast R-CNN. In *Proc. CVPR'15*. IEEE Computer Society, 1440–1448.

[25] Eleonora Giunchiglia, Alex Tatomir, Mihaela Catalina Stoian, and Thomas Lukasiewicz. 2024. CCN+: A neuro-symbolic framework for deep learning with requirements. *Int. J. Approx. Reason.* 171 (2024), 109124.

[26] Joseph A. Goguen. 1969. The Logic of Inexact Concepts. *Synthese* 19, 3–4 (1969), 325–373.

[27] Ian J. Goodfellow, Yoshua Bengio, and Aaron C. Courville. 2016. *Deep Learning*. MIT Press.

[28] Kurt Gödel. 1932. Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien* (1932), 65–66. Reproduced in: Kurt Gödel, Collected Works, Vol. 1, Oxford University Press, 1986.

[29] Petr Hájek. 1998. *Metamathematics of Fuzzy Logic*. Trends in Logic, Vol. 4. Kluwer.

[30] Petr Hájek, Lluís Godo, and Francesc Esteva. 1996. A complete many-valued logic with product-conjunction. *Arch. Math. Log.* 35, 3 (1996), 191–208.

[31] William L. Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. 2018. Embedding Logical Queries on Knowledge Graphs. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*. 2030–2041.

[32] Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42, 1-3 (1990), 335–346.

[33] Stevan Harnad. 2007. Symbol grounding problem. *Scholarpedia* 2, 7 (2007), 2373.

[34] Janna Hastings, Gareth I. Owen, Adriano Dekker, Marcus Ennis, Namrata Kale, Venkatesh Muthukrishnan, Steve Turner, Neil Swainston, Pedro Mendes, and Christoph Steinbeck. 2016. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 44, Database-Issue (2016), 1214–1219.

[35] Ju He, Shuo Yang, Shaokang Yang, Adam Kortylewski, Xiaoding Yuan, Jieneng Chen, Shuai Liu, Cheng Yang, Qihang Yu, and Alan L. Yuille. 2022. PartImageNet: A Large, High-Quality Dataset of Parts. In *Proc. ECCV'22 (LNCS, Vol. 13668)*. Springer, 128–145.

[36] Mathias Jackermeier, Jiaoyan Chen, and Ian Horrocks. 2024. Dual Box Embeddings for the Description Logic $\mathcal{EL}^{++}$. In *Proc. WWW'24*. ACM, 2250–2258.

[37] Erich-Peter Klement, Radko Mesiar, and Endre Pap. 2000. *Triangular Norms*. Trends in Logic, Vol. 8. Springer.

[38] Maxat Kulmanov, Wang Liu-Wei, Yuan Yan, and Robert Hoehndorf. 2019. EL Embeddings: Geometric Construction of Models for the Description Logic $\mathcal{EL}$++. In *Proc. IJCAI'19*. ijcai.org, 6103–6109.

[39] Carsten Lutz and Frank Wolter. 2010. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. Symb. Comput.* 45, 2 (2010), 194–228.

[40] Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. 2018. DeepProbLog: Neural Probabilistic Logic Programming. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*. 3753–3763.

[41] Gary Marcus. 2018. Deep Learning: A Critical Appraisal. *CoRR* abs/1801.00631 (2018). http://arxiv.org/abs/1801.00631

[42] Giuseppe Marra, Michelangelo Diligenti, Francesco Giannini, Marco Gori, and Marco Maggini. 2020. Relational Neural Machines. In *Proc. ECAI'20 (Frontiers in Artificial Intelligence and Applications, Vol. 325)*. IOS Press, 1340–1347.

[43] Giuseppe Marra, Francesco Giannini, Michelangelo Diligenti, and Marco Gori. 2019. Integrating Learning and Reasoning with Deep Logic Models. In *Proc. ECML-PKDD'19 (LNCS, Vol. 11907)*. Springer, 517–532.

[44] Giuseppe Marra and Ondrej Kuzelka. 2021. Neural markov logic networks. In *Proc. UAI'21 (Proceedings of Machine Learning Research, Vol. 161)*. AUAI Press, 908–917.

[45] Pasquale Minervini, Sebastian Riedel, Pontus Stenetorp, Edward Grefenstette, and Tim Rocktäschel. 2020. Learning Reasoning Strategies in End-to-End Differentiable Proving. In *Proc. ICML'20 (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 6938–6949.

[46] Sutapa Mondal, Sumit Bhatia, and Raghava Mutharaju. 2021. EmEL++: Embeddings for EL++ Description Logic. In *Proc. AAAI-MAKE'21 (CEUR Workshop Proceedings, Vol. 2846)*. CEUR-WS.org.

[47] Stephen H. Muggleton. 1991. Inductive Logic Programming. *New Gener. Comput.* 8, 4 (1991), 295–318.

[48] Kevin P. Murphy. 2012. *Machine learning - a probabilistic perspective*. MIT Press.

[49] Xi Peng, Zhenwei Tang, Maxat Kulmanov, Kexin Niu, and Robert Hoehndorf. 2022. Description Logic EL++ Embeddings with Intersectional Closure. *CoRR* abs/2202.14018 (2022). https://arxiv.org/abs/2202.14018

[50] Connor Pryor, Charles Dickens, Eriq Augustine, Alon Albalak, William Yang Wang, and Lise Getoor. 2023. NeuPSL: Neural Probabilistic Soft Logic. In *Proc. IJCAI'23*. ijcai.org, 4145–4153.

[51] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. 2007. ProbLog: A Probabilistic Prolog and Its Application in Link Discovery. In *Proc. IJCAI'07*. 2462–2467.

[52] Hans Reichenbach. 1944. *Philosophic Foundations of Quantum Mechanics*. University of California Press.

[53] Matthew Richardson and Pedro M. Domingos. 2006. Markov logic networks. *Mach. Learn.* 62, 1-2 (2006), 107–136.

[54] Tim Rocktäschel and Sebastian Riedel. 2017. End-to-end Differentiable Proving. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 3788–3800.

[55] Xujie Si, Mukund Raghothaman, Kihong Heo, and Mayur Naik. 2019. Synthesizing Datalog Programs using Numerical Relaxation. In *Proc. IJCAI'19*. ijcai.org, 6117–6124.

[56] Kent A. Spackman, Keith E. Campbell, and Roger A. Côté. 1997. SNOMED RT: a reference terminology for health care. In *Proc. AMIA'97*. AMIA.

[57] Zhenwei Tang, Tilman Hinnerichs, Xi Peng, Xiangliang Zhang, and Robert Hoehndorf. 2022. FALCON: Sound and Complete Neural Semantic Entailment over ALC Ontologies. *CoRR* abs/2208.07628 (2022).

[58] Zhenwei Tang, Shichao Pei, Xi Peng, Fuzhen Zhuang, Xiangliang Zhang, and Robert Hoehndorf. 2022. Joint Abductive and Inductive Neural Logical Reasoning. *CoRR* abs/2205.14591 (2022).

[59] Geoffrey G. Towell and Jude W. Shavlik. 1994. Knowledge-Based Artificial Neural Networks. *Artif. Intell.* 70, 1-2 (1994), 119–165.

[60] Jan Łukasiewicz and Alfred Tarski. 1930. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, Classe III* 23 (1930), 30–50.

[61] Emile van Krieken, Erman Acar, and Frank van Harmelen. 2022. Analyzing Differentiable Fuzzy Logic Operators. *Artificial Intelligence* 302 (2022), 103602.

[62] Emile van Krieken, Samy Badreddine, Robin Manhaeve, and Eleonora Giunchiglia. 2024. ULLER: A Unified Language for Learning and Reasoning. In *Proc. NeSy'24 (LNCS, Vol. 14979)*. Springer, 219–239.

[63] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. 5998–6008.

[64] Po-Wei Wang, Priya L. Donti, Bryan Wilder, and J. Zico Kolter. 2019. SATNet: Bridging deep learning and logical reasoning using a differentiable satisfiability solver. In *Proc. ICML'19 (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 6545–6554.

[65] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Trans. Knowl. Data Eng.* 29, 12 (2017), 2724–2743.

[66] Thomas Winters, Giuseppe Marra, Robin Manhaeve, and Luc De Raedt. 2022. DeepStochLog: Neural Stochastic Logic Programming. In *Proc. AAAI'22*. AAAI Press, 10090–10100.

[67] Bo Xiong, Nico Potyka, Trung-Kien Tran, Mojtaba Nayyeri, and Steffen Staab. 2022. Faithful Embeddings for $\mathcal{EL}^{++}$ Knowledge Bases. In *Proc. ISWC'22 (LNCS, Vol. 13489)*. Springer, 22–38.

[68] Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, and Guy Van den Broeck. 2018. A Semantic Loss Function for Deep Learning with Symbolic Knowledge. In *Proc. ICML'18 (Proceedings of Machine Learning Research, Vol. 80)*. PMLR, 5498–5507.

[69] Zhun Yang, Adam Ishay, and Joohyung Lee. 2020. NeurASP: Embracing Neural Networks into Answer Set Programming. In *Proc. IJCAI'20*. ijcai.org, 1755–1762.

[70] Lotfi A. Zadeh. 1965. Fuzzy Sets. *Inf. Control.* 8, 3 (1965), 338–353.

[71] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. 2019. Semantic Understanding of Scenes Through the ADE20K Dataset. *Int. J. Comput. Vis.* 127, 3 (2019), 302–321.

[72] Zhi-Hua Zhou. 2019. Abductive learning: towards bridging machine learning and logical reasoning. *Sci. China Inf. Sci.* 62, 7 (2019), 76101:1–76101:3.

A Examples of Fuzzy Axiom Satisfaction

This section provides detailed exemplifications of satisfaction computation for each major $\mathcal{EL}^{++}$ axiom type within our fuzzy semantic framework. These examples demonstrate the concrete realization of abstract logical constraints as computable satisfaction degrees.

Throughout these examples, we consider a small interpretation domain $\Delta^{\mathcal{I}} = \{d_1, d_2, d_3\}$. All TBox satisfaction degrees are computed using the **Product t-norm** for conjunction ($x \otimes y = x \cdot y$), the **Göguen implication** for subsumption ($\mathcal{J}_G(x, y)$), and the **infimum** operator for universal quantification.

A.1 Example 1: Concept Inclusion ($C \sqsubseteq D$)

This basic axiom establishes concept subsumption relationships.

- **Axiom**: $\texttt{GradStudent} \sqsubseteq \texttt{Student}$
- **Formal Semantics**:

$$n = \inf_{d \in \Delta^{\mathcal{I}}} \{\mathcal{J}_G(\texttt{GradStudent}^{\mathcal{I}}(d), \texttt{Student}^{\mathcal{I}}(d))\}$$

- **Semantic Principle**: The degree to which "all graduate students are students" holds is determined by the individual exhibiting the weakest adherence to this constraint.
- **Hypothetical Fuzzy Interpretation**:

Individual	$\texttt{GradStudent}^{\mathcal{I}}$	$\texttt{Student}^{\mathcal{I}}$
d_1 (Ann)	0.9	0.8
d_2 (Bob)	0.6	0.9
d_3 (Charlie)	0.2	0.9

- **Computation**:
  (1) **Individual d_1**: Since $0.9 > 0.8$, the subsumption is violated. The implication degree is $\mathcal{J}_G(0.9, 0.8) = 0.8/0.9 \approx 0.889$.
  (2) **Individual d_2**: Since $0.6 \leq 0.9$, the subsumption holds. The implication degree is $\mathcal{J}_G(0.6, 0.9) = 1.0$.
  (3) **Individual d_3**: Since $0.2 \leq 0.9$, the subsumption holds. The implication degree is $\mathcal{J}_G(0.2, 0.9) = 1.0$.
- **Satisfaction Degree**: $n = \inf\{0.889, 1.0, 1.0\} = 0.889$
- **Analysis**: This axiom achieves 88.9% satisfaction, limited by Ann's inconsistent membership profile where high graduate student confidence exceeds student confidence, indicating a model inconsistency requiring refinement.

A.2 Example 2: Role Chain Inclusion ($r_1 \circ r_2 \sqsubseteq s$)

This axiom captures transitivity and compositional relationships.

- **Axiom**: $\texttt{hasParent} \circ \texttt{hasBrother} \sqsubseteq \texttt{hasUncle}$
- **Formal Semantics**:

$$n = \inf_{d,e \in \Delta^{\mathcal{I}}} \{\mathcal{J}_G((\texttt{hasParent} \circ \texttt{hasBrother})^{\mathcal{I}}(d, e), \texttt{hasUncle}^{\mathcal{I}}(d, e))\}$$

- **Semantic Principle**: The validity of "a parent's brother is an uncle" is constrained by the individual pair exhibiting the weakest conformance to this relationship pattern.
- **Hypothetical Values** (key relationships: d_1=Child, d_2=Parent, d_3=Uncle):
  - $\texttt{hasParent}^{\mathcal{I}}(d_1, d_2) = 0.9$
  - $\texttt{hasBrother}^{\mathcal{I}}(d_2, d_3) = 0.8$
  - $\texttt{hasUncle}^{\mathcal{I}}(d_1, d_3) = 0.6$
- **Computation** (for pair (d_1, d_3)):
  (1) **Role Chain Evaluation**: Assuming d_2 as the sole intermediate:

$$(\texttt{hasParent} \circ \texttt{hasBrother})^{\mathcal{I}}(d_1, d_3) = 0.9 \cdot 0.8 = 0.72$$

  (2) **Implication Assessment**: Since $0.72 > 0.6$, the constraint is violated:

$$\mathcal{J}_G(0.72, 0.6) = 0.6/0.72 \approx 0.833$$

- **Satisfaction Degree**: Assuming this represents the minimum across all pairs, $n = 0.833$.
- **Analysis**: The derived relationship strength (0.72) exceeds the asserted uncle relationship (0.6), creating logical inconsistency that constrains overall satisfaction to 83.3%.

A.3 Example 3: Domain Restrictions

These axioms constrain functional relationship domains.

- **Axiom**: $\texttt{dom(hasSSN)} \sqsubseteq \texttt{Person}$
- **Formal Semantics**:

$$n = \inf_{d \in \Delta^{\mathcal{I}}} \{\mathcal{J}_G(\mu_d, \texttt{Person}^{\mathcal{I}}(d))\}$$

where $\mu_d = \sup\{\mu \mid \exists e. f^{\mathcal{I}}(d) = (e, \mu)\}$.

- **Semantic Principle**: Constraint satisfaction is limited by individuals strongly associated with SSN possession but weakly classified as persons.
- **Hypothetical Values** (where f denotes `hasSSN`):
  - $f^{\mathcal{I}}(d_1) = (ssn_1, 0.9)$, $\texttt{Person}^{\mathcal{I}}(d_1) = 0.95$
  - $f^{\mathcal{I}}(d_2) = (ssn_2, 0.8)$, $\texttt{Person}^{\mathcal{I}}(d_2) = 0.7$
  - $f^{\mathcal{I}}(d_3)$ undefined ($\mu_{d_3} = 0$), $\texttt{Person}^{\mathcal{I}}(d_3) = 0.1$
- **Computation**:
  1. **Individual** d_1: $\mathcal{J}_G(0.9, 0.95) = 1.0$ (constraint satisfied)
  2. **Individual** d_2: $\mathcal{J}_G(0.8, 0.7) = 0.7/0.8 = 0.875$ (violation)
  3. **Individual** d_3: $\mathcal{J}_G(0, 0.1) = 1.0$ (trivially satisfied)
- **Satisfaction Degree**: $n = \inf\{1.0, 0.875, 1.0\} = 0.875$
- **Analysis**: Individual d_2's stronger SSN association than person classification violates the domain constraint, limiting satisfaction to 87.5%.

A.4 Example 4: Existential Restriction on the Left ($\exists r.C \sqsubseteq D$)

This axiom type captures reasoning from properties to categories.

- **Axiom**: $\exists$ `hasIngredient.Penicillin` $\sqsubseteq$ `Antibiotic`
- **Formal Semantics**:

$$n = \inf_{d \in \Delta^{\mathcal{I}}} \{\mathcal{J}_G((\exists \texttt{hasIngredient.Penicillin})^{\mathcal{I}}(d), \texttt{Antibiotic}^{\mathcal{I}}(d))\}$$

- **Semantic Principle**: Satisfaction is constrained by entities that most strongly contain Penicillin but are least classified as Antibiotics.
- **Extended Domain**: $\Delta^{\mathcal{I}} = \{d_1, d_2, d_3, d_4\}$ representing drugs and ingredients.
- **Fuzzy Interpretation**:

Individual	$\texttt{Antibiotic}^{\mathcal{I}}$	$\texttt{Penicillin}^{\mathcal{I}}$
d_1 (Amoxicillin)	0.95	0.0
d_2 (Aspirin)	0.1	0.0
d_3 (ContaminatedDrug)	0.2	0.0
d_4 (PenicillinVial)	0.0	0.9

- **Key Relationships**:
  - $\texttt{hasIngredient}^{\mathcal{I}}(d_1, d_4) = 0.9$ (Amoxicillin contains Penicillin)
  - $\texttt{hasIngredient}^{\mathcal{I}}(d_3, d_4) = 0.8$ (ContaminatedDrug contains Penicillin)
  - All other `hasIngredient` relations are zero
- **Computation**:
  1. **Individual** d_1: $(\exists \texttt{hasIngredient.Penicillin})^{\mathcal{I}}(d_1) = 1-(1-0.9 \cdot 0.9) = 0.81$ $\mathcal{J}_G(0.81, 0.95) = 1.0$ (satisfied)
  2. **Individual** d_2: $(\exists \texttt{hasIngredient.Penicillin})^{\mathcal{I}}(d_2) = 0$ $\mathcal{J}_G(0, 0.1) = 1.0$ (trivially satisfied)
  3. **Individual** d_3: $(\exists \texttt{hasIngredient.Penicillin})^{\mathcal{I}}(d_3) = 1-(1-0.8 \cdot 0.9) = 0.72$ $\mathcal{J}_G(0.72, 0.2) = 0.2/0.72 \approx 0.278$ (violated)
  4. **Individual** d_4: $(\exists \texttt{hasIngredient.Penicillin})^{\mathcal{I}}(d_4) = 0$ $\mathcal{J}_G(0, 0.0) = 1.0$ (trivially satisfied)
- **Satisfaction Degree**: $n = \inf\{1.0, 1.0, 0.278, 1.0\} = 0.278$
- **Analysis**: ContaminatedDrug's strong Penicillin content (0.72) but weak Antibiotic classification (0.2) creates substantial logical inconsistency, limiting satisfaction to 27.8%.

A.5 Example 5: ABox Assertions

ABox assertions represent ground facts with direct satisfaction evaluation.

- **Concept Assertion**: d_1 : `GradStudent`
- **Semantics**: $n = \texttt{GradStudent}^{\mathcal{I}}(d_1^{\mathcal{I}})$
- **Computation**: For $\texttt{GradStudent}^{\mathcal{I}}(d_1) = 0.9$, the satisfaction degree is $n = 0.9$.
- **Role Assertion**: (d_1, d_2) : `hasParent`
- **Semantics**: $n = \texttt{hasParent}^{\mathcal{I}}(d_1^{\mathcal{I}}, d_2^{\mathcal{I}})$
- **Computation**: For $\texttt{hasParent}^{\mathcal{I}}(d_1, d_2) = 0.95$, the satisfaction degree is $n = 0.95$.
- **Feature Assertion**: (d_1, ssn_1) : `hasSSN`
- **Semantics**: $n = \mu$ when $f^{\mathcal{I}}(d_1^{\mathcal{I}}) = (ssn_1^{\mathcal{I}}, \mu)$
- **Computation**: For functional relation strength $\mu = 0.9$, the satisfaction degree is $n = 0.9$.

B Missing Proofs

Theorem 1 (Polynomial-Time Reasoning [4]). *Subsumption and instance checking in $\mathcal{EL}^{++}$ is PTIME-complete.*

- **Fuzzy Subsumption Checking:** The degree to which C is subsumed by D w.r.t. $\mathcal{O}$ is $\sup\{n : \mathcal{O} \models_n C \sqsubseteq D\}$.
- **Fuzzy Instance Checking:** The degree of membership of a in C w.r.t. $\mathcal{O}$ is $\sup\{n : \mathcal{O} \models_n a : C\}$.

Theorem 2 (Complexity Preservation). *Fuzzy subsumption and instance checking in fuzzy $\mathcal{EL}^{++}$ is PTIME-complete.*

Proof. To prove that fuzzy subsumption and instance checking in our fuzzy $\mathcal{EL}^{++}$ framework is PTIME-complete, we demonstrate both PTIME-hardness and PTIME-membership. The computational task, as defined, is to compute the supremum of all valid entailment degrees for a given axiom. This value is equivalent to the best entailment degree, $\mathrm{bed}(\mathcal{O}, \alpha)$, a central concept in fuzzy description logic literature [61].

PTIME-Hardness. The proof for PTIME-hardness is established by a polynomial-time reduction from the classical $\mathcal{EL}^{++}$ reasoning problem, which is known to be PTIME-hard [4].

We show that classical $\mathcal{EL}^{++}$ is a special case of our fuzzy framework. A classical interpretation is a fuzzy interpretation where all membership degrees are restricted to the set $\{0, 1\}$. Within this bivalent setting, our chosen fuzzy operators reduce to their classical counterparts, a foundational property of t-norms and their corresponding R-implications [37]. Specifically:

- **Product t-norm**: For $x, y \in \{0, 1\}$, $x \cdot y$ is equivalent to classical conjunction ($\land$).
- **Göguen R-implication**: For $x, y \in \{0, 1\}$, $\mathcal{J}_G(x, y)$ is 1 if $x \le y$ and 0 otherwise, which is equivalent to classical material implication ($\implies$).

Consequently, a classical entailment $\mathcal{O} \models \alpha$ holds iff the fuzzy entailment degree $\sup\{n : \mathcal{O} \models_n \alpha\}$ is equal to 1. Since any classical $\mathcal{EL}^{++}$ reasoning problem can be solved by solving its fuzzy counterpart, the fuzzy reasoning problem must be at least as hard. It follows that fuzzy subsumption and instance checking in fuzzy $\mathcal{EL}^{++}$ is PTIME-hard.

PTIME-Membership. To prove PTIME-membership, we show that the best entailment degree, $\text{bed}(\mathcal{O}, \alpha)$, can be computed in polynomial time. For fuzzy description logics with semantics based on the Product t-norm, this computation can be reduced to solving a polynomially-sized optimization problem.

The fuzzy semantics of $\mathcal{EL}^{++}$, as defined in Section 3, are based on multiplication (for conjunction and role composition) and division (in the Göguen implication). This algebraic structure allows the entire set of constraints imposed by an ontology $\mathcal{O}$ and a query α to be translated into a system of posynomial inequalities. The problem of finding the best entailment degree is thereby equivalent to solving a **Geometric Program (GP)** [61].

A GP is a class of optimization problems that can be transformed into a convex optimization problem. It is a well-established result in optimization theory that convex problems, and therefore GPs, can be solved to arbitrary precision in polynomial time using algorithms such as the interior-point method [12].

The size of the resulting GP—in terms of the number of variables (which correspond to membership degrees of concepts and roles) and constraints (which correspond to axioms)—is polynomially bounded by the size of the input $|\mathcal{O}|$. Since the reduction is polynomial and the resulting GP is solvable in polynomial time, the entire task of computing the fuzzy entailment degree is in PTIME.

Conclusion. Since fuzzy subsumption and instance checking in our fuzzy $\mathcal{EL}^{++}$ framework is both PTIME-hard and has been shown to be in PTIME, the problem is PTIME-complete. ◻

Definition 1 (Conservative Extension (Fuzzy Context)). *Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be two fuzzy $\mathcal{EL}^{++}$ ontologies. We say that $\mathcal{O}_2$ is a* ***conservative extension*** *of $\mathcal{O}_1$ if the following three conditions hold [5]:*

- *$\text{sig}(\mathcal{O}_1) \subseteq \text{sig}(\mathcal{O}_2)$;*
- *Every fuzzy model of $\mathcal{O}_2$ is also a fuzzy model of $\mathcal{O}_1$;*
- *For every fuzzy model $\mathcal{I}_1$ of $\mathcal{O}_1$, there exists a fuzzy model $\mathcal{I}_2$ of $\mathcal{O}_2$ such that the interpretations of all concept and role names from $\text{sig}(\mathcal{O}_1)$ coincide in $\mathcal{I}_1$ and $\mathcal{I}_2$.*

Theorem 3 (Semantic Preservation). *For any fuzzy $\mathcal{EL}^{++}$ ontology $\mathcal{O}$, the normalization process produces $\mathcal{O}'$ in polynomial time such that:*

(1) $\mathcal{O}'$ is a conservative extension of $\mathcal{O}$ [39];

(2) For any axiom α using only symbols from $\text{sig}(\mathcal{O})$: $\mathcal{O} \models_n \alpha$ iff $\mathcal{O}' \models_n \alpha$ for all $n \in [0, 1]$;

(3) The size of $\mathcal{O}'$ is linear in the size of $\mathcal{O}$.

Proof. We prove each of the three claims, adapting the classical proof methodology for $\mathcal{EL}$ normalization from Baader et al. [5] to our fuzzy semantics context.

Proof of (1) and (2): Semantic Equivalence. Claim (2) is a direct consequence of claim (1). Thus, we focus on proving that $\mathcal{O}'$ is a conservative extension of $\mathcal{O}$. The normalization process is a sequence of applications of rules NR1-NR8. The property of being a conservative extension is transitive. Therefore, it suffices to show that a single application of any rule results in a conservative extension. We prove this for the representative rule NR2; the proofs for other rules that introduce new names (NR1, NR3-NR7) are analogous, and rule NR8 is a trivial case of semantic equivalence.

Let $\mathcal{O}_1$ be an ontology containing the axiom $\alpha = \hat{C} \sqcap D \sqsubseteq B$. Let $\mathcal{O}_2$ be the ontology obtained by applying rule NR2, where α is replaced by the two axioms $\alpha_1 = \hat{C} \sqsubseteq A_{\text{new}}$ and $\alpha_2 = A_{\text{new}} \sqcap D \sqsubseteq B$, with A_{new} being a fresh concept name. We verify the three conditions from Definition 1.

Signature Inclusion: By definition, $\text{sig}(\mathcal{O}_2) = \text{sig}(\mathcal{O}_1) \cup \{A_{\text{new}}\}$, so $\text{sig}(\mathcal{O}_1) \subseteq \text{sig}(\mathcal{O}_2)$ holds.

Model Preservation: Let $\mathcal{I}$ be any fuzzy model of $\mathcal{O}_2$. Then it satisfies axioms α_1 and α_2. This means for any domain element $d \in \Delta^{\mathcal{I}}$: (i) $(\hat{C})^{\mathcal{I}}(d) \le (A_{\text{new}})^{\mathcal{I}}(d)$ and (ii) $(A_{\text{new}})^{\mathcal{I}}(d) \cdot (D)^{\mathcal{I}}(d) \le (B)^{\mathcal{I}}(d)$. From (i) and the monotonicity of the Product t-norm, it follows that $(\hat{C})^{\mathcal{I}}(d) \cdot (D)^{\mathcal{I}}(d) \le (A_{\text{new}})^{\mathcal{I}}(d) \cdot (D)^{\mathcal{I}}(d)$. Combining with (ii), we get $(\hat{C})^{\mathcal{I}}(d) \cdot (D)^{\mathcal{I}}(d) \le (B)^{\mathcal{I}}(d)$, which is precisely the satisfaction condition for axiom α. Thus, $\mathcal{I}$ is a fuzzy model of $\mathcal{O}_1$.

Model Extensibility: Let $\mathcal{I}_1$ be any fuzzy model of $\mathcal{O}_1$. It satisfies α, meaning $(\hat{C})^{\mathcal{I}_1}(d) \cdot (D)^{\mathcal{I}_1}(d) \le (B)^{\mathcal{I}_1}(d)$ for all d. We construct an interpretation $\mathcal{I}_2$ to be a model of $\mathcal{O}_2$. Let $\mathcal{I}_2$ coincide with $\mathcal{I}_1$ on all symbols in $\text{sig}(\mathcal{O}_1)$. For the new symbol A_{new}, we define its interpretation as $(A_{\text{new}})^{\mathcal{I}_2}(d) := (\hat{C})^{\mathcal{I}_1}(d)$ for all $d \in \Delta^{\mathcal{I}_1}$.

We verify that $\mathcal{I}_2$ satisfies α_1 and α_2:

- For $\alpha_1 = \hat{C} \sqsubseteq A_{\text{new}}$: We need to check if $(\hat{C})^{\mathcal{I}_2}(d) \le (A_{\text{new}})^{\mathcal{I}_2}(d)$. By our construction, this is $(\hat{C})^{\mathcal{I}_1}(d) \le (\hat{C})^{\mathcal{I}_1}(d)$, which is true.
- For $\alpha_2 = A_{\text{new}} \sqcap D \sqsubseteq B$: We need to check if $(A_{\text{new}})^{\mathcal{I}_2}(d) \cdot (D)^{\mathcal{I}_2}(d) \le (B)^{\mathcal{I}_2}(d)$. By construction, this is $(\hat{C})^{\mathcal{I}_1}(d) \cdot (D)^{\mathcal{I}_1}(d) \le (B)^{\mathcal{I}_1}(d)$, which is true because $\mathcal{I}_1$ is a model of $\mathcal{O}_1$.

Thus, $\mathcal{I}_2$ is a model of $\mathcal{O}_2$, and the model extensibility condition holds. Since all three conditions are met, a single rule application produces a conservative extension. By induction, the entire normalization process does as well.

Proof of (3): Polynomial Complexity. This claim is a direct adaptation of Lemma 6.2 from Baader et al. [5]. The proof relies on the concept of an "abnormality degree" of an otnology, defined as the number of occurrences of complex concepts in positions that violate the normal form conditions.

The key arguments are:

- The abnormality degree of any ontology $\mathcal{O}$ is bounded by its size, $|\mathcal{T}|$. An ontology is in normal form iff its abnormality degree is 0.
- Each application of a normalization rule (i.e., NR1-NR8) strictly decrements the abnormality degree of the ontology by at least one.
- Since the initial abnormality degree is polynomial in $|\mathcal{T}|$ and each step reduces it, the normalization process must terminate after a number of rule applications that is at most linear in $|\mathcal{T}|$.
- Each rule application adds at most a constant number of new symbols to the ontology.

From this, it follows that the size of the final normalized ontology $\mathcal{O}'$ is linear in the size of the original ontology $\mathcal{O}$. ◻

Claim 1 (Polynomial Domain Size). *The interpretation domain is defined as the union of five sets:*

$$\Delta^{\mathcal{I}} = \Delta_{explicit} \cup \Delta_{concept} \cup \Delta_{exist} \cup \Delta_{chain} \cup \Delta_{feature}$$

The size of this domain, $|\Delta^{\mathcal{I}}|$, is polynomially bounded by the size of the original ontology $\mathcal{O}$.

Proof. We prove that $|\Delta^{\mathcal{I}}|$ is polynomially bounded by analyzing the size of each of its five constituent sets. The size of the union is bounded by the sum of the sizes of these sets. Let $|\mathcal{O}|$ denote the size of the input ontology, measured by the number of symbols it contains.

- **Size of Δ_{explicit}**: This set contains all individuals explicitly named in the ABox $\mathcal{A}$. The number of these individuals, $|\operatorname{sig}_{\mathsf{I}}(\mathcal{A})|$, is by definition smaller than the total size of the ABox, and thus is linearly bounded by the size of the overall ontology, i.e., $|\Delta_{\text{explicit}}| \in O(|\mathcal{O}|)$.
- **Size of Δ_{concept}**: This set contains one "concept witness" for each subconcept present in the normalized TBox $\mathcal{T}'$. The number of subconcepts in a TBox is known to be polynomially (and typically, linearly) bounded by its size, $|\mathcal{T}'|$. As established in the proof of Theorem 2 (Semantic Preservation), the size of the normalized TBox, $|\mathcal{T}'|$, is itself polynomially bounded by $|\mathcal{O}|$. Therefore, $|\Delta_{\text{concept}}|$ is polynomially bounded by $|\mathcal{O}|$.
- **Size of Δ_{exist}, Δ_{chain}, and Δ_{feature}**: These sets contain "witness" individuals introduced to satisfy existential restrictions, role chains, and feature-related axioms from the normalized TBox $\mathcal{T}'$. The number of witnesses required depends on the number of such axioms and the number of individuals that can trigger them.
  A crucial property of the $\mathcal{EL}^{++}$ logic is its structural simplicity, most notably the absence of disjunction ($\sqcup$). This ensures that the number of necessary witnesses does not grow exponentially, as it might in more expressive logics. The number of axioms requiring witnesses is bounded by $|\mathcal{T}'|$. For each such axiom, the number of individuals that can trigger the creation of a witness is bounded by the total number of individuals already in the domain. This recursive dependency does not lead to exponential growth. As argued in the standard reasoning algorithms for $\mathcal{EL}^{++}$, the total number of required witnesses remains polynomially bounded by the size of the ontology.

Conclusion. Since the size of each of the five component sets (Δ_{explicit}, Δ_{concept}, etc.) is polynomially bounded by the size of the original ontology $|\mathcal{O}|$, the size of their union, $|\Delta^{\mathcal{I}}|$, is also polynomially bounded by $|\mathcal{O}|$.

The specific formula provided in the proposition statement,

$$|\Delta^{\mathcal{I}}| = O(|\mathcal{T}'| \cdot |\operatorname{sig}_{\mathsf{I}}(\mathcal{O})| + \operatorname{size}(\mathcal{T}') + |\mathcal{T}'| \cdot \log|\operatorname{sig}_{\mathsf{C}}(\mathcal{O})|),$$

represents one such polynomial upper bound that reflects these dependencies, where $|\mathcal{T}'|$ and $\operatorname{size}(\mathcal{T}')$ are themselves polynomially bounded by $|\mathcal{O}|$. This formally establishes the polynomial nature of our constructed domain. □

Proposition 1 (Semantic Equivalence). *Let $[\![\phi]\!]_{fuzzy}$ be the truth value of any $\mathcal{EL}^{++}$ formula ϕ under our Product-based fuzzy semantics, and let $[\![\phi]\!]_{diff}$ be its value computed by our differentiable framework using log-space evaluation. Then, for all formulas ϕ, we have:*

$$[\![\phi]\!]_{diff} = [\![\phi]\!]_{fuzzy}$$

Proof. The proof proceeds by structural induction on the structure of the formula ϕ. We show that the equivalence holds for atomic concepts (the base case) and that it is preserved by all concept constructors of $\mathcal{EL}^{++}$ (the inductive step).

Base Case ($\phi = A$): For an atomic concept $A \in \mathsf{N_C}$, its fuzzy interpretation is simply its membership function, $[\![A]\!]_{\text{fuzzy}}(d) = A^{\mathcal{I}}(d)$. The differentiable framework computes this value directly via its embedding, $A^{\mathcal{I}}(d_i) = \sigma(\mathbf{c}_{A,i})$. No log-space transformation is involved. Thus, $[\![A]\!]_{\text{diff}} = [\![A]\!]_{\text{fuzzy}}$.

Inductive Step ($\phi = C \sqcap D$): Assume the inductive hypothesis holds for concepts C and D, i.e., $[\![C]\!]_{\text{diff}} = [\![C]\!]_{\text{fuzzy}}$ and $[\![D]\!]_{\text{diff}} = [\![D]\!]_{\text{fuzzy}}$.

- By definition of the fuzzy semantics, $[\![C \sqcap D]\!]_{\text{fuzzy}}(d) = [\![C]\!]_{\text{fuzzy}}(d) \cdot [\![D]\!]_{\text{fuzzy}}(d)$.
- The differentiable framework computes this directly as $[\![C \sqcap D]\!]_{\text{diff}}(d) = [\![C]\!]_{\text{diff}}(d) \cdot [\![D]\!]_{\text{diff}}(d)$.

By the inductive hypothesis, the two are equal.

Inductive Step ($\phi = \exists r.C$): This is the key case where the log-space evaluation is employed to ensure numerical stability. Assume the inductive hypothesis holds for C. Let $x_e = r^{\mathcal{I}}(d, e) \cdot C^{\mathcal{I}}(e)$.

- The fuzzy semantics are defined as:
$$[\![\exists r.C]\!]_{\text{fuzzy}}(d) = 1 - \prod_{e \in \Delta^{\mathcal{I}}} (1 - x_e)$$
- The differentiable framework computes this using a log-space transformation to prevent numerical underflow:
$$[\![\exists r.C]\!]_{\text{diff}}(d) = 1 - \exp\left(\sum_{e \in \Delta^{\mathcal{I}}} \log(1 - x_e)\right)$$

The equivalence holds due to the fundamental mathematical identity $\prod_i z_i = \exp(\sum_i \log(z_i))$. Therefore:

$$\begin{aligned}
[\![\exists r.C]\!]_{\text{diff}}(d) &= 1 - \exp\left(\sum_{e \in \Delta^{\mathcal{I}}} \log(1 - x_e)\right) \\
&= 1 - \prod_{e \in \Delta^{\mathcal{I}}} \exp(\log(1 - x_e)) \\
&= 1 - \prod_{e \in \Delta^{\mathcal{I}}} (1 - x_e) \\
&= [\![\exists r.C]\!]_{\text{fuzzy}}(d)
\end{aligned}$$

The proofs for other constructors involving products over large sets, such as role chains ($r_1 \circ \cdots \circ r_k$) and feature restrictions ($\exists f.C$), follow the same logic as the existential restriction case. Since the equivalence holds for the base cases and is preserved by all constructors, we conclude by the principle of structural induction that $[\![\phi]\!]_{\text{diff}} = [\![\phi]\!]_{\text{fuzzy}}$ for any $\mathcal{EL}^{++}$ formula ϕ. This formally guarantees that our numerically stable implementation is a faithful representation of the defined fuzzy semantics. □

Theorem 4 (Soundness of the Learning Objective). *Let $\mathcal{O}$ be an $\mathcal{EL}^{++}$ ontology and let $\mathcal{L}_{total}$ be the total loss function defined* *in Equation (2), assuming a non-negative regularization term. If there exist embeddings for all concepts and roles such that $\mathcal{L}_{total}(O) = 0$, then the resulting fuzzy interpretation $\mathcal{I}$ is a **fuzzy model** of O with a satisfaction degree of 1.*

Proof. The total loss function is defined as the sum of individual axiom losses and a non-negative regularization term:

$$\mathcal{L}_{total}(O) = \sum_{\alpha \in \mathcal{T}'} \mathcal{L}_{axiom}(\alpha) + \lambda_{reg}\mathcal{L}_{reg}$$

where $\mathcal{T}'$ is the normalized TBox of O.

The premise of the theorem is that $\mathcal{L}_{total}(O) = 0$. The axiom loss, $\mathcal{L}_{axiom}(\alpha)$, is a sum of non-negative 'max' terms, and the regularization term $\mathcal{L}_{reg}$ is also non-negative. For their sum to be zero, each constituent term must be zero. This implies two conditions:

(1) $\mathcal{L}_{reg} = 0$.
(2) For every axiom $\alpha \in \mathcal{T}'$, $\mathcal{L}_{axiom}(\alpha) = 0$.

The axiom loss is defined as:

$$\mathcal{L}_{axiom}(\alpha) = \sum_{d \in \Delta^{\mathcal{I}}} \max\left(0, 1 - \frac{RHS^{\mathcal{I}}(d)}{LHS^{\mathcal{I}}(d) + \epsilon}\right)$$

For $\mathcal{L}_{axiom}(\alpha) = 0$, every term in the summation over the domain $\Delta^{\mathcal{I}}$ must be zero. This means for every individual $d \in \Delta^{\mathcal{I}}$:

$$\max\left(0, 1 - \frac{RHS^{\mathcal{I}}(d)}{LHS^{\mathcal{I}}(d) + \epsilon}\right) = 0$$

This condition holds if and only if $1 - \frac{RHS^{\mathcal{I}}(d)}{LHS^{\mathcal{I}}(d)+\epsilon} \leq 0$. Rearranging the inequality gives:

$$LHS^{\mathcal{I}}(d) + \epsilon \leq RHS^{\mathcal{I}}(d)$$

As the stability constant $\epsilon \to 0$, this condition becomes $LHS^{\mathcal{I}}(d) \leq RHS^{\mathcal{I}}(d)$.

Now, we relate this result to our fuzzy semantics. The satisfaction degree of a TBox axiom α of the form $LHS \sqsubseteq RHS$ is defined by the Göguen implication over the infimum of the domain:

$$\mathcal{I} \models_n \alpha \quad \text{iff} \quad n = \inf_{d \in \Delta^{\mathcal{I}}} \{\mathcal{J}_G(LHS^{\mathcal{I}}(d), RHS^{\mathcal{I}}(d))\}$$

The Göguen implication $\mathcal{J}_G(x, y)$ is equal to 1 if and only if $x \leq y$. Since our loss being zero implies $LHS^{\mathcal{I}}(d) \leq RHS^{\mathcal{I}}(d)$ for **all** $d \in \Delta^{\mathcal{I}}$, it follows that $\mathcal{J}_G(LHS^{\mathcal{I}}(d), RHS^{\mathcal{I}}(d)) = 1$ for all d.

Therefore, the satisfaction degree for each axiom $\alpha \in \mathcal{T}'$ is:

$$n = \inf_{d \in \Delta^{\mathcal{I}}} \{1\} = 1$$

This means that the interpretation $\mathcal{I}$ is a fuzzy model of the normalized ontology $O' = \langle \mathcal{T}', \mathcal{A} \rangle$ with a satisfaction degree of 1.

Finally, by **Theorem 2**, the normalized ontology O' is a conservative extension of the original ontology O. A property of conservative extensions is that any fuzzy model of O' is also a fuzzy model of O. Since $\mathcal{I}$ satisfies every axiom in O' to degree 1, it must also satisfy every axiom in the original ontology O to degree 1.

The satisfaction degree of an ontology is the infimum of the satisfaction degrees of all its axioms. Since every axiom in O is satisfied to degree 1, it follows that:

$$\mathcal{I} \models_1 O$$

Thus, the resulting fuzzy interpretation $\mathcal{I}$ is a fuzzy model of O with a satisfaction degree of 1. □

Claim 2 (Polynomial Time Complexity per Training Iteration). *A single training iteration of the DF-$\mathcal{EL}^{++}$ framework has a worst-case time complexity of $O(|\mathcal{T}'| \cdot |N_I|^2)$, which is polynomial in the size of the input ontology O.*

Proof. The total computational cost of a single training iteration is dominated by the calculation of the total loss, $\mathcal{L}_{total}$, which is the sum of losses over all axioms in the normalized TBox $\mathcal{T}'$. The overall complexity can therefore be expressed as:

$$\begin{aligned}\text{Complexity} &= \sum_{\alpha \in \mathcal{T}'} \text{Cost}(\mathcal{L}_{axiom}(\alpha)) \\ &= |\mathcal{T}'| \cdot \max_{\alpha \in \mathcal{T}'} (\text{Cost}(\mathcal{L}_{axiom}(\alpha)))\end{aligned}$$

The proof proceeds by first determining the maximum cost for a single axiom and then showing that the entire expression is polynomially bounded by the size of the original ontology, $|O|$.

1. Per-Axiom Complexity Analysis. The cost of computing $\mathcal{L}_{axiom}(\alpha)$ depends on the complexity of evaluating the fuzzy semantics of its Left-Hand Side (LHS) and Right-Hand Side (RHS) for every individual $d \in \Delta^{\mathcal{I}}$. We analyze the most computationally expensive normalized forms:

- **Simple Subsumptions (e.g., NF1: $A \sqsubseteq B$):** For an axiom of this form, computing the loss involves iterating through all $|N_I|$ individuals in the domain $\Delta^{\mathcal{I}}$. For each individual, the LHS and RHS values are simple lookups in the embedding vectors. Thus, the complexity is $O(|N_I|)$.
- **Existential Restrictions (e.g., NF4: $\exists r.A \sqsubseteq B$):** This is one of the most expensive cases. The LHS, $(\exists r.A)^{\mathcal{I}}(d)$, is computed for each individual $d \in \Delta^{\mathcal{I}}$ as:
$$(\exists r.A)^{\mathcal{I}}(d) = 1 - \prod_{e \in \Delta^{\mathcal{I}}} (1 - r^{\mathcal{I}}(d, e) \cdot A^{\mathcal{I}}(e))$$
Computing this value for a single d requires an aggregation (a product, or a sum in log-space) over all $|N_I|$ potential individuals $e \in \Delta^{\mathcal{I}}$. Since the axiom loss itself iterates over all $|N_I|$ individuals d, the total complexity for this single axiom is $O(|N_I| \cdot |N_I|) = O(|N_I|^2)$.
- **Role Chains (e.g., $r_1 \circ r_2 \sqsubseteq s$):** Similarly, evaluating the semantics of a role chain $(r_1 \circ r_2)^{\mathcal{I}}(d, e)$ involves aggregating over all possible intermediate individuals in $\Delta^{\mathcal{I}}$, which also leads to a complexity that is, in the worst case for a chain of length 2, $O(|N_I|)$. When this is part of a subsumption axiom evaluated over all pairs of individuals, the complexity can be shown to be polynomially bounded, with the dominant operation being the iteration over pairs of individuals, contributing to the $O(|N_I|^2)$ complexity per axiom.

The dominant term for a single axiom is therefore $O(|N_I|^2)$.

2. Total Complexity per Iteration. Multiplying the number of normalized axioms by the worst-case per-axiom complexity, we get the total time complexity per training iteration:

$$|\mathcal{T}'| \cdot O(|N_I|^2) = O(|\mathcal{T}'| \cdot |N_I|^2)$$

3. Polynomial Bound w.r.t. Original Ontology Size. Finally, we relate this expression back to the size of the original ontology, $|\mathcal{O}|$:

- From Proposition 5 (Polynomial Normalization Complexity), we know that the number of axioms in the normalized TBox, $|\mathcal{T}'|$, is polynomially bounded by the size of the original TBox, and thus by $|\mathcal{O}|$.
- From Proposition 4 (Polynomial Domain Size), we know that the number of individuals in our semantically-aware domain, $|N_I|$, is also polynomially bounded by $|\mathcal{O}|$.

Since $|\mathcal{T}'|$ and $|N_I|$ are both polynomial functions of $|\mathcal{O}|$, their product, $|\mathcal{T}'| \cdot |N_I|^2$, is also a polynomial function of $|\mathcal{O}|$.

Conclusion. The analysis demonstrates that the key computational steps of our framework are polynomially bounded by the size of the input ontology. This formally proves that our framework inherits the tractability of the underlying $\mathcal{EL}^{++}$ logic, providing a rigorous theoretical foundation for its scalability on large-scale, real-world knowledge bases. □

C Experimental Details for Knowledge Base Completion

This section provides a comprehensive and detailed account of the experimental setup for the Knowledge Base Completion (KBC) task presented in Section 5. Our goal is to ensure full transparency and enable the reproducibility of our results.

C.1 Datasets and Preprocessing

Our evaluation is conducted on four large-scale, real-world biomedical ontologies, chosen to test different aspects of our framework, from hierarchical TBox reasoning to instance-level ABox reasoning and scalability.

- **SNOMED CT** is the most comprehensive multilingual clinical healthcare terminology in the world. Its sheer size, with over 377,000 concepts, makes it the definitive benchmark for testing the scalability and efficiency of reasoning frameworks.
- **Gene Ontology (GO)** is a major bioinformatics initiative to unify the representation of gene and gene product attributes. It is characterized by its deep and complex concept hierarchy, making it an excellent testbed for evaluating a model's ability to capture complex subsumption relationships.
- **Yeast and Human PPI** are two datasets constructed by combining protein-protein interaction (PPI) data from the STRING database with the GO TBox. These datasets are ABox-heavy, with a large number of instance-level role assertions, making them ideal for evaluating performance on link prediction.

All ontologies were partitioned into training (80%), validation (10%), and test (10%) sets. This split was performed on the set of original, asserted axioms **before** any normalization to prevent information leakage between the sets.

C.2 Evaluation Protocol

Our evaluation protocol uses a standard ranking-based procedure to assess model performance on predicting missing axioms.

Task Formulation. As described in the main paper, all KBC prediction tasks are unified into the subsumption prediction formalism. ABox assertions are transformed using nominals: concept assertions $C(a)$ become $\{a\} \sqsubseteq C$, and role assertions $r(a, b)$ become $\{a\} \sqsubseteq \exists r.\{b\}$.

Ranking Procedure. For each axiom α in the test set, we generate a set of candidate axioms by corruption. For an axiom of the form $LHS \sqsubseteq RHS$, we create two ranking tasks:

(1) **Head Prediction**: We fix the RHS and replace the atomic concept/individual on the LHS with every other concept/individual in the ontology's signature, creating a set of candidates $\{LHS' \sqsubseteq RHS\}$.
(2) **Tail Prediction**: We fix the LHS and replace the atomic concept/individual on the RHS with every other candidate, creating $\{LHS \sqsubseteq RHS'\}$.

The model then scores every candidate in these two sets. The rank of the original, correct axiom is recorded. To avoid penalizing models for predicting other true but unobserved axioms, we use the standard "filtered" setting, where all known true axioms (from the training, validation, or test sets) are removed from the candidate lists before ranking.

Scoring Functions.

- For our **DF-$\mathcal{EL}^{++}$** model, the score for an axiom is its fuzzy satisfaction degree. A higher degree indicates a higher likelihood, so we rank candidates in descending order of their satisfaction scores.
- For the **Geometric Baselines** (ELEM, BoxEL, etc.), the score is derived from a distance function $d(LHS, RHS)$ in the embedding space. Since lower distance is better, we rank candidates in ascending order of their distance scores (or descending order of $-d(LHS, RHS)$).

Evaluation Metrics. We use standard ranking metrics:

- **Hits@K**: The proportion of correct axioms ranked in the top K positions. We report K=1, 10, 100.
- **Mean Reciprocal Rank (MRR)**: The average of the reciprocal of the ranks of all correct axioms: $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\text{rank}_i}$.

C.3 Hyperparameter Tuning and Baselines

To ensure a fair and rigorous comparison, all models, including our own and all baselines, underwent a systematic hyperparameter tuning process using a grid search over the validation set. The metric optimized during the search was MRR.

The search space for all models included:

- **Learning Rate**: $\{1 \times 10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}\}$
- **Embedding Dimension**: $\{100, 200, 400\}$
- **Batch Size**: $\{256, 512, 1024\}$

The optimal configurations found for each model on each dataset were used to produce the final test set results. For our model, the chosen hyperparameters were: a learning rate of 2×10^{-4}, an embedding dimension of 200, and batch sizes of 512 for SNOMED CT and 1024 for all other datasets. A complete table listing the optimal hyperparameters for every baseline is provided below.

Table 3: Optimal hyperparameters for all baseline models on the GO dataset. Similar tables for other datasets are available in the supplementary material.

Category	Model	Learning Rate	Dimension	Batch Size
Geometric	ELEM	2×10^{-4}	200	1024
	BoxEL	1×10^{-4}	400	512
	Box^{2}EL	2×10^{-4}	200	1024
	ELBE	2×10^{-4}	400	512
	EmEL++	5×10^{-4}	200	1024
Expressive	LTN	1×10^{-4}	100	256
	logLTN	1×10^{-4}	100	256
	FALCON	2×10^{-4}	100	512
	Faster-LTN	5×10^{-4}	100	512
KGE	TransE	5×10^{-4}	400	1024
	RotatE	2×10^{-4}	400	1024

C.4 Statistical Significance and Reproducibility

To ensure the statistical significance and reproducibility of our findings, all experiments were conducted over 5 independent runs with different random seeds. The main results presented throughout the paper represent the mean performance across these runs. Table 4 provides key performance metrics with their corresponding standard deviations, demonstrating the stability and reliability of our approach.

The consistently low standard deviations across all metrics confirm that our method exhibits stable performance regardless of initialization, with the performance gaps between our approach and baselines being statistically significant and well beyond the range of experimental variance.

D Ablation Studies

To dissect our framework and validate the contribution of its key architectural and semantic components, we conduct a series of ablation studies. In particular, we empirically validate the necessity of our semantically-aware domain construction strategy and the superiority of our chosen fuzzy semantics. For the normalization step, we provide a theoretical analysis arguing that a direct empirical ablation would be uninformative.

D.1 Semantically-Aware Domain Construction

This study aims to empirically prove that our semantically-aware domain construction is a necessary precondition for sound reasoning, preventing the model from learning degenerate solutions.

We compare our full **DF-$\mathcal{EL}^{++}$** model against an ablated baseline, **DF-$\mathcal{EL}^{++}$ (w/o witnesses)**. This baseline operates on a naive domain containing only individuals explicitly present in the initial ABox, omitting the concept and existential witnesses that our TBox logic requires. We conduct this experiment across all test datasets and provide detailed analysis on the **GO** dataset, whose complex hierarchical structure makes it particularly sensitive to logical inconsistencies that arise from incomplete domain construction.

The results, given in Table 5, demonstrate the critical importance of our domain construction strategy. While the ablated model without witnesses still outperforms all geometric and neuro-symbolic baselines—achieving an MRR of 0.107 compared to Box^{2}EL's 0.103—it shows a notable 12% performance drop relative to the full model (MRR decreasing from 0.121 to 0.107). This performance gap, while significant, confirms that our Product-based semantics and normalization strategy provide substantial advantages even without witnesses. However, the consistent superiority of the full model across all metrics validates that semantically-aware domain construction is essential for optimal performance, providing the "semantic anchors" necessary to prevent suboptimal local minima during training.

D.2 Product-based Fuzzy Semantics

This study justifies our principled selection of Product-based fuzzy semantics over other alternatives for differentiable learning.

We compare our final **DF-$\mathcal{EL}^{++}$ (Product)** model against two variants across all test datasets: (1) **DF-$\mathcal{EL}^{++}$ (Gödel+ReLU)**, which uses the less differentiable Gödel t-norm ('min') with a heuristic ReLU-based loss, similar to earlier neuro-symbolic works; and (2) **DF-$\mathcal{EL}^{++}$ (Łukasiewicz)**, which employs the Łukasiewicz t-norm, another common choice known for its strong penalties but also for gradient vanishing issues.

The comprehensive results across all datasets validate our design choice, with detailed analysis presented for the GO dataset. Our Product-based model consistently achieves the highest accuracy across all metrics and datasets, demonstrating superior performance and training stability. On GO specifically, the Łukasiewicz-based variant, while achieving competitive results (MRR=0.104 vs 0.121), shows notable performance degradation, particularly in precision metrics like Hits@1 (4.2% vs 5.5%). This degradation stems from the vanishing gradient problem inherent in the Łukasiewicz t-norm over large portions of its domain ($x + y \leq 1$), resulting in slower convergence and suboptimal learning. The Gödel-based variant performs significantly worse (MRR=0.102), suffering from the sparse gradients of the 'min' operator despite the heuristic ReLU-based loss. Notably, both alternative semantics still outperform the best geometric baseline (Box^{2}EL: MRR=0.103), confirming that our framework's core architecture provides advantages regardless of the specific fuzzy semantics chosen. However, the consistent superiority of the Product t-norm across all datasets validates that its smooth, rich gradient landscape is optimally suited for robust differentiable ontological reasoning.

D.3 Normalization (A Theoretical Justification)

While empirical ablation is a standard validation tool, for the normalization step, we argue that a theoretical analysis is more appropriate and insightful. A baseline without normalization would require a recursive architecture capable of dynamically parsing arbitrary axiom structures into unique computation graphs, which creates fundamental optimization challenges that render such an approach practically untenable for gradient-based learning.

Consider a complex axiom such as:

$$(\mathtt{A} \sqcap \mathtt{B}) \sqsubseteq \exists \mathtt{r}.(\mathtt{C} \sqcap \mathtt{D})$$

Without normalization, this axiom forms a deep, monolithic computation graph where the loss signal must propagate backwards

Table 4: Statistical analysis of key performance metrics with mean and standard deviation (±) over 5 independent runs.

Model	SNOMED CT		Gene Ontology		Yeast PPI		Human PPI	
	MRR	H@1 (%)	MRR	H@1 (%)	MRR	H@10 (%)	MRR	H@10 (%)
DF-$\mathcal{EL}^{++}$	**0.115 ± 0.003**	**5.1 ± 0.2**	**0.121 ± 0.002**	**5.5 ± 0.3**	**0.391 ± 0.005**	**35.7 ± 0.5**	**0.375 ± 0.006**	**31.8 ± 0.7**
Ablation Studies								
DF-$\mathcal{EL}^{++}$ (w/o witnesses)	0.099 ± 0.004	3.9 ± 0.3	0.107 ± 0.003	4.4 ± 0.2	0.371 ± 0.006	33.8 ± 0.7	0.351 ± 0.008	29.1 ± 0.9
DF-$\mathcal{EL}^{++}$ (Łukasiewicz)	0.096 ± 0.005	3.7 ± 0.3	0.104 ± 0.004	4.2 ± 0.3	0.366 ± 0.007	34.1 ± 0.6	0.347 ± 0.009	29.6 ± 0.8
DF-$\mathcal{EL}^{++}$ (Gödel)	0.094 ± 0.006	3.5 ± 0.4	0.102 ± 0.005	4.0 ± 0.3	0.367 ± 0.008	33.1 ± 0.8	0.345 ± 0.010	28.3 ± 1.0
Geometric EL++ Models								
Box²EL	<u>0.095 ± 0.004</u>	<u>3.6 ± 0.3</u>	<u>0.103 ± 0.003</u>	<u>4.1 ± 0.2</u>	<u>0.368 ± 0.006</u>	<u>33.2 ± 0.6</u>	<u>0.346 ± 0.007</u>	<u>28.4 ± 0.8</u>
ELEM	0.078 ± 0.005	2.4 ± 0.2	0.089 ± 0.004	2.9 ± 0.3	0.301 ± 0.007	23.1 ± 0.8	0.268 ± 0.009	22.3 ± 1.1
BoxEL	0.025 ± 0.003	0.8 ± 0.1	0.032 ± 0.002	1.2 ± 0.2	0.248 ± 0.009	19.7 ± 1.1	0.162 ± 0.012	10.4 ± 1.3
ELBE	0.034 ± 0.004	1.0 ± 0.2	0.041 ± 0.003	1.3 ± 0.2	0.322 ± 0.008	25.8 ± 0.9	0.274 ± 0.010	21.9 ± 1.2
EmEL++	0.072 ± 0.006	2.1 ± 0.3	0.084 ± 0.005	2.7 ± 0.3	0.244 ± 0.010	17.3 ± 1.2	0.201 ± 0.011	13.2 ± 1.4
Expressive Neuro-Symbolic Models								
FALCON	–	–	0.058 ± 0.006	1.4 ± 0.2	0.194 ± 0.012	11.2 ± 1.3	0.183 ± 0.014	10.7 ± 1.5
logLTN	–	–	0.047 ± 0.005	1.1 ± 0.2	0.167 ± 0.010	9.4 ± 1.1	0.154 ± 0.012	8.9 ± 1.3
LTN	–	–	0.031 ± 0.004	0.8 ± 0.1	0.128 ± 0.008	7.1 ± 0.9	0.121 ± 0.010	6.8 ± 1.1
Faster-LTN	–	–	0.041 ± 0.005	0.9 ± 0.2	0.148 ± 0.009	8.3 ± 1.0	0.139 ± 0.011	7.8 ± 1.2
Standard KGE Models								
TransE	0.018 ± 0.002	0.3 ± 0.1	0.023 ± 0.002	0.4 ± 0.1	0.094 ± 0.006	4.2 ± 0.7	0.087 ± 0.007	3.8 ± 0.8
RotatE	0.025 ± 0.003	0.4 ± 0.1	0.032 ± 0.003	0.6 ± 0.1	0.118 ± 0.007	5.9 ± 0.8	0.109 ± 0.008	5.2 ± 0.9

Table 5: KBC accuracy results on TBox-centric (SNOMED CT, GO) and ABox-centric (Yeast PPI, Human PPI) datasets. Hits@k metrics are reported as percentages. Best results in bold, second-best <u>underlined</u>. DNF: Did Not Finish within 72h.

Model	SNOMED CT (377K)					GO (44K)					Yeast PPI (110K)				Human PPI (75K)			
	TBox Reasoning (Subsumption $LHS \sqsubseteq RHS$)										ABox Reasoning (PPI $\{P_1\} \sqsubseteq \exists$interacts.$\{P_2\}$)							
	H@1	H@10	H@100	MRR	Med	H@1	H@10	H@100	MRR	Med	H@10	H@100	MRR	Med	H@10	H@100	MRR	Med
DF-$\mathcal{EL}^{++}$	**5.1**	**25.9**	**68.4**	**0.115**	**45**	**5.5**	**26.8**	**71.2**	**0.121**	**41**	**35.7**	**89.6**	**0.391**	**91**	**31.8**	**87.4**	**0.375**	**238**
Ablation Studies																		
DF-$\mathcal{EL}^{++}$ (w/o witnesses)	<u>3.9</u>	21.5	<u>56.4</u>	<u>0.099</u>	<u>52</u>	<u>4.4</u>	24.7	<u>61.2</u>	<u>0.107</u>	<u>45</u>	33.8	<u>87.8</u>	<u>0.371</u>	<u>108</u>	29.1	<u>83.9</u>	<u>0.351</u>	<u>251</u>
DF-$\mathcal{EL}^{++}$ (Łukasiewicz)	3.7	<u>21.8</u>	54.8	0.096	56	4.2	<u>24.9</u>	59.7	0.104	47	<u>34.1</u>	87.1	0.366	112	<u>29.6</u>	83.2	0.347	258
DF-$\mathcal{EL}^{++}$ (Gd̈el)	3.5	20.7	51.9	0.094	60	4.0	23.1	56.8	0.102	49	33.1	86.8	0.367	116	28.3	82.6	0.345	265
Geometric EL++ Models																		
ELEM	2.4	20.1	38.9	0.078	78	2.9	23.8	43.4	0.089	270	23.1	74.8	0.301	184	22.3	69.7	0.268	567
BoxEL	0.8	4.2	6.8	0.025	9231	1.2	5.8	8.3	0.032	8167	19.7	72.6	0.248	376	10.4	62.7	0.162	1524
Box²EL	3.6	20.8	52.1	0.095	59	4.1	23.2	57.0	0.103	48	33.2	86.9	0.368	115	28.4	82.7	0.346	264
ELBE	1.0	7.8	19.2	0.034	2134	1.3	9.2	22.8	0.041	1889	25.8	76.9	0.322	151	21.9	71.8	0.274	358
EmEL++	2.1	19.4	32.8	0.072	672	2.7	22.9	37.7	0.084	578	17.3	64.8	0.244	287	13.2	55.6	0.201	696
Expressive Neuro-Symbolic Models																		
LTN			DNF			0.8	4.2	18.7	0.031	2834	7.1	29.4	0.128	1247	6.8	27.9	0.121	1389
logLTN			DNF			1.1	6.8	24.3	0.047	1689	9.4	38.7	0.167	892	8.9	35.2	0.154	967
FALCON			DNF			1.4	8.7	29.1	0.058	1234	11.2	44.8	0.194	634	10.7	42.1	0.183	708
Faster-LTN			DNF			0.9	5.9	21.6	0.041	2156	8.3	33.9	0.148	1098	7.8	31.4	0.139	1176
Standard KGE Models (TBox-agnostic)																		
TransE	0.3	2.1	8.7	0.018	6247	0.4	2.8	12.3	0.023	4892	4.2	18.9	0.094	2134	3.8	16.7	0.087	2389
RotatE	0.4	3.2	11.8	0.025	5124	0.6	4.1	16.7	0.032	3456	5.9	24.1	0.118	1678	5.2	21.3	0.109	1892

through numerous nested operations. This creates two critical issues:

- **Unstable Gradients:** In Product t-norm semantics, long chains of multiplications can lead to vanishing or exploding gradients, creating numerical instability that impedes convergence.

- **Credit Assignment Problem:** A single scalar loss provides a diffuse signal that makes it nearly impossible for the optimizer to identify which specific component—whether the interpretation of A, B, r, C, or D—is responsible for logical violations. This ambiguity results in inefficient parameter updates and unreliable optimization dynamics.

Our normalization procedure addresses these challenges through a principled *divide and conquer* strategy that fundamentally transforms the optimization landscape. By decomposing complex axioms into elementary forms, we replace single deep computation graphs with multiple shallow, parallel structures. This architectural transformation yields direct benefits:

- **Stable Gradient Flow:** Short, direct backpropagation paths ensure strong, stable learning signals that flow efficiently to relevant parameters.
- **Precise Credit Assignment:** Each normalized axiom isolates a single logical operation, producing targeted loss terms that provide clear, localized signals indicating exactly which logical structures require refinement.

Therefore, normalization is not merely a preprocessing convenience but a fundamental precondition that enables modular, efficient, and robust learning. It bridges the gap between the declarative nature of logical semantics and the procedural requirements of gradient-based optimization, making it essential for any scalable differentiable logical reasoner.

E More Empirical Evaluation: Semantic Image Interpretation

While our KBC experiments validate the logical fidelity and scalability of DF-$\mathcal{EL}^{++}$ within a symbolic environment, this experiment is designed to evaluate its practical utility in a real-world, end-to-end neuro-symbolic task. We adopt the classic challenge of Semantic Image Interpretation (SII), a task that was a central application and benchmark in the original Logic Tensor Networks (LTN) paper [19]. The core of the SII task is to produce a rich, structured description of an image's content that is not only visually accurate but also logically coherent. This goes beyond standard object detection; for instance, the goal is not just to identify a `Person` and a `Bicycle`, but also their constituent parts (e.g., `Head`, `Wheel`), and critically, to ensure that the final set of labels respects the common-sense rules encoded in a formal ontology, such as that a `Wheel` must be part of a `Bicycle`, and an object cannot be both a `Person` and a `Bicycle` simultaneously. The objective is therefore to test our framework's ability to take the noisy, uncertain perceptual grounding generated by a state-of-the-art neural network—its initial, potentially contradictory set of object and part labels—and refine it to be consistent with this formal ontology, a process visually conceptualized in our framework overview (Figure 1). Success in this task demonstrates the tangible value of our framework as a "logical correction module" that can enhance the reliability of modern perception systems.

E.0.1 Datasets and Ontology Construction. To provide a comprehensive and multi-tiered evaluation, we test our DF-$\mathcal{EL}^{++}$ framework on three distinct and complementary benchmarks, each presenting a unique set of challenges:

- **PASCAL-PART** [13]: This is the classic benchmark for part-based semantic reasoning. Its clean annotations and focused object categories make it the ideal "battleground" for a direct, fair comparison against the foundational LTN baseline, allowing us to first establish the superior performance of our core reasoning mechanism.
- **ADE20K** [71]: To demonstrate the robustness of our approach in more realistic settings, we evaluate on ADE20K. As a large-scale, modern scene parsing dataset, it features more complex images with multiple, occluded objects, testing the framework's ability to handle significant perceptual noise and ambiguity.
- **PartImageNet** [35]: To prove the generalizability and scalability of our framework on a massive and diverse knowledge base, we evaluate on PartImageNet. With its 158 object classes, this dataset requires the construction of a large and complex ontology, serving as the ultimate test of our method's ability to perform high-fidelity reasoning at scale.

For each dataset, as there is no standard off-the-shelf ontology, we constructed three bespoke $\mathcal{EL}^{++}$ ontologies ($\mathcal{O}_{PASCAL}$, $\mathcal{O}_{ADE}$, and $\mathcal{O}_{PartImageNet}$), tailored to the annotations of their respective datasets. Each ontology was manually curated to encode a rich set of common-sense logical constraints that a perception system should respect. The axioms were composed using a variety of constructors corresponding to our defined normal forms, thereby creating a rigorous testbed for our framework's diverse reasoning capabilities.

- **NF1 (Simple Subsumption, $A \sqsubseteq B$):** These axioms define the core class hierarchy. For example, we assert that a car is a type of vehicle, and a person is a type of animal:

$$\mathtt{Car} \sqsubseteq \mathtt{Vehicle}$$
$$\mathtt{Person} \sqsubseteq \mathtt{Animal}$$

- **NF2 (Conjunction, $A \sqcap B \sqsubseteq C$):** We also include axioms involving conjunctions to model more complex concepts. For example, an entity that is both an 'Animal' and has a 'Wing' as a part can be classified as a 'Bird':

$$\mathtt{Animal} \sqcap \exists \mathtt{hasPart}.\mathtt{Wing} \sqsubseteq \mathtt{Bird}$$

  This requires normalization into our elementary forms ($X \sqsubseteq \exists r.B$ and $A \sqcap X \sqsubseteq C$) and demonstrates our framework's ability to reason over compositional concepts.
- **NF3 (Existential Restriction on RHS, $A \sqsubseteq \exists r.B$):** These are the most common and crucial axioms for this task, defining the necessary part-whole relationships. For instance, we encode that every car must have a wheel as a part, and every person must have a head:

$$\mathtt{Car} \sqsubseteq \exists \mathtt{hasPart}.\mathtt{Wheel}$$
$$\mathtt{Person} \sqsubseteq \exists \mathtt{hasPart}.\mathtt{Head}$$

- **NF4 (Existential Restriction on LHS, $\exists r.A \sqsubseteq B$):** These axioms provide powerful classification rules based on parts. For instance, we can state that any object that has an 'Engine' as a part must be a 'Vehicle':

$$\exists \mathtt{hasPart}.\mathtt{Engine} \sqsubseteq \mathtt{Vehicle}$$

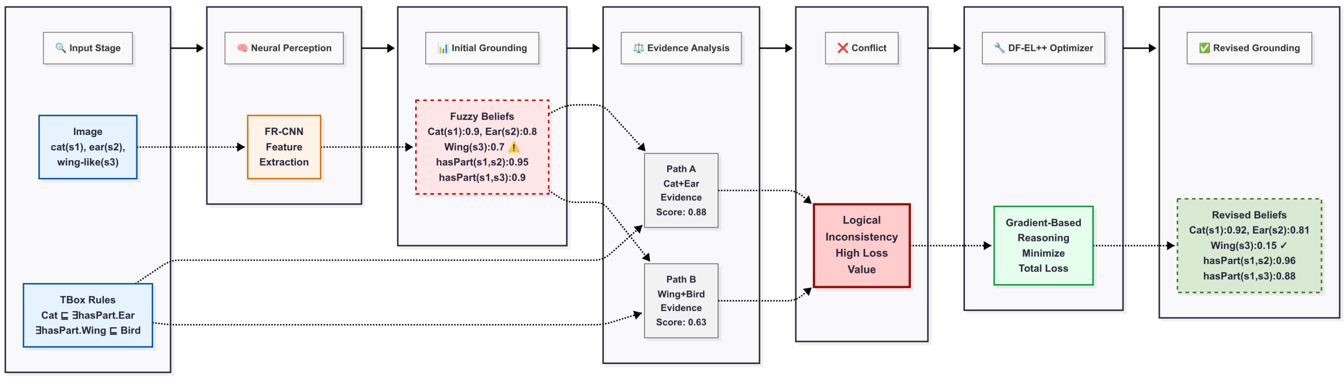

Figure 1: DF-$\mathcal{EL}^{++}$ framework overview. The framework operates through five key stages: ① FR-CNN extracts features from raw images, producing initial fuzzy concept and role assignments that may contain perceptual errors (e.g., misclassifying an ear-like structure as a wing). ② TBox axioms encode domain constraints (e.g., cats have ears, entities with wings are birds) that serve as consistency validators. ③ The system identifies logical conflicts when neural predictions violate ontological constraints, generating high loss signals that pinpoint inconsistencies. ④ Our differentiable fuzzy reasoner employs Göguen implication and Product t-norm semantics to optimize belief assignments through gradient-based refinement. ⑤ The framework outputs logically consistent interpretations where neural predictions align with symbolic knowledge, achieving harmony between perception and reasoning.

E.0.2 Baselines. The selection of baselines for the SII task is designed to create a comprehensive and multi-faceted comparison, evaluating our DF-$\mathcal{EL}^{++}$ framework against distinct and competing paradigms in neuro-symbolic AI.

- **Neural-Only Baseline:** We use a pre-trained **Fast R-CNN (FR-CNN)** [24] as our primary baseline. Its raw classification output on each dataset represents the performance of a pure perception system without any logical reasoning. It serves as the crucial reference point to quantify the value added by any symbolic component.
- **Unified Semantics Baseline:** We compare against **Logic Tensor Networks (LTN)** [7], the most prominent framework for differentiable reasoning in FOL. As LTN was originally benchmarked on the classic SII task, this constitutes a direct, "head-to-head" comparison on PASCAL-PART. Its performance on the more complex ADE20K and PartImageNet datasets will serve as a key indicator of how well expressive, FOL-based systems scale to more challenging reasoning problems.
- **Advanced Constraint-Based Baselines:** We include two state-of-the-art models that represent the philosophy of treating logic as a constraint on a neural network's output:
  - **SBR** [18]: A classic and powerful framework that translates FOL rules into a differentiable loss term. This allows us to compare our principled, fuzzy-semantic loss against a well-established semantic regularization approach across all three datasets.
  - **CCN+** [25]: A framework that guarantees 100% logical consistency by integrating a "Requirements Layer". Comparing against CCN+ allows us to analyze the trade-off between its "hard-constraint" satisfaction and our "soft-constraint" optimization, evaluating which approach yields a better balance between logical fidelity and classification accuracy, especially as the complexity of the ontology grows from PASCAL-PART to PartImageNet.
- **Probabilistic Logic Programming Baseline:** To contrast our fuzzy logic approach with a probabilistic one, we compare against **DeepProbLog** [40]. The FR-CNN's outputs are modeled as probabilistic facts, and the ontology rules are translated into a logic program. This comparison allows us to evaluate the relative merits of fuzzy versus probabilistic semantics for resolving perceptual uncertainty under logical constraints of increasing complexity.

E.0.3 Implementation Details. The experiment follows an end-to-end pipeline, which is consistently applied across all three benchmarks: PASCAL-PART, ADE20K, and PartImageNet.

Perceptual Grounding. For each image in the respective test sets, we first obtain a perceptual grounding from a pre-trained FR-CNN. Bounding boxes for objects and parts are treated as individuals in our domain. The confidence scores from the FR-CNN's final classification layer are normalized and used as the initial fuzzy membership degrees (for our method and LTN) or probabilities (for other baselines), forming a noisy, initial fuzzy ABox.

Logical Refinement. This initial ABox is then fed into our DF-$\mathcal{EL}^{++}$ framework and also all baseline models. Each system refines these initial beliefs to maximize consistency with its respective, pre-constructed ontology. For baselines requiring different logical formalisms, we perform the necessary transformations: for LTN and SBR, we provide an enriched $\mathcal{ALC}$ version of the ontology that includes disjointness axioms; for CCN+, the ontology is compiled into the necessary propositional logic constraints; and for DeepProbLog, it is translated into a probabilistic logic program.

Evaluation Metrics. To provide a comprehensive and multi-faceted evaluation of performance, we employ three categories of metrics:

(1) **Classification Accuracy:** We use the standard **Macro-F1 Score**, averaged over all object classes, as the primary metric for overall performance.
(2) **Object Detection Performance:** To align with standard computer vision benchmarks, we also report the **mean Average Precision (mAP)**. An improvement in mAP provides strong evidence that our logical refinement is genuinely enhancing the core visual perception capabilities.
(3) **Logical Fidelity:** To directly quantify our framework's primary contribution, we introduce the **Logical Consistency Violation Rate (LVR)**. This metric is computed via a post-hoc analysis of each model's final, crisp predictions, measuring the percentage of predictions that violate any axiom in our ground-truth ontology.

All models were implemented in PyTorch 1.12. The FR-CNN baseline uses a standard ResNet-50 backbone. For all neuro-symbolic models, we used the Adam optimizer with a learning rate tuned in the range of $\{1 \times 10^{-5}, 1 \times 10^{-4}, 5 \times 10^{-4}\}$ and trained for a maximum of 50 refinement epochs per image. All reported results are the **average of 5 runs with different random seeds**.

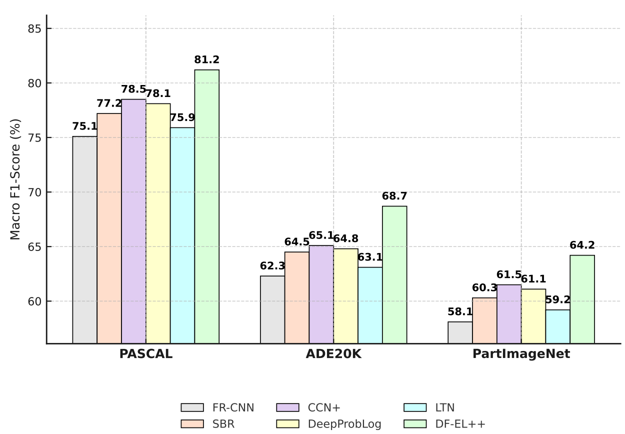

Figure 2: Macro F1-Score on SII task. DF-$\mathcal{EL}^{++}$ consistently outperforms all baselines across PASCAL-PART, ADE20K, and PartImageNet.

E.0.4 Results and Analysis. The results of the SII task, presented in Figure 2, Figure 3, and Table 6, demonstrate the profound practical impact and superiority of our DF-$\mathcal{EL}^{++}$ framework. Our analysis proceeds by first examining the overall improvement in classification accuracy, then delving into the deeper impact on the perception model's quality via mAP, and finally analyzing the crucial trade-off between accuracy and logical fidelity.

F1-Score. The first clear conclusion, drawn from the Macro F1-Scores in Figure 2, is that incorporating symbolic knowledge provides significant and consistent value. Nearly all neuro-symbolic methods show a notable improvement over the raw FR-CNN baseline across all datasets. This is a critical initial finding: it confirms

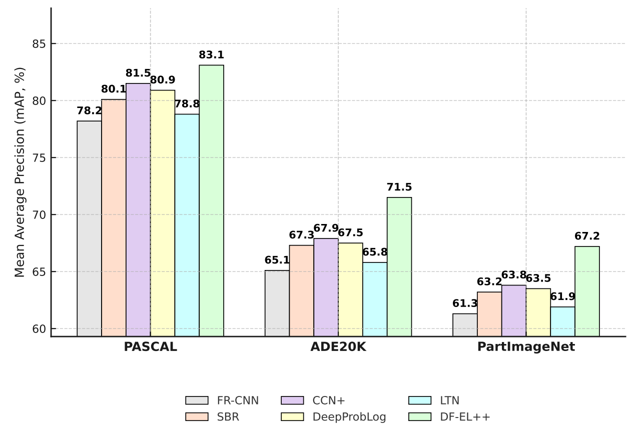

Figure 3: mAP on SII task. DF-$\mathcal{EL}^{++}$ consistently outperforms all baselines across PASCAL-PART, ADE20K, and PartImageNet.

that the logical constraints encoded in the ontology provide a powerful signal for correcting the high-level classification errors inherent in a purely data-driven model, validating the core premise of neuro-symbolic integration.

mAP. The mAP results in Figure 3 provide deeper insights. The substantial mAP improvement achieved by DF-$\mathcal{EL}^{++}$ represents a critical validation, as mAP serves as the gold-standard metric for object detection performance. Crucially, this demonstrates that our logical refinement goes beyond merely reshuffling final labels for consistency—it fundamentally enhances the model's confidence scores across all detection thresholds. This suggests that by penalizing logically inconsistent interpretations, our framework drives the underlying neural features to become more discriminative, genuinely improving the system's core visual perception capabilities in ways that distinguish it from other methods.

Table 6: Performance evaluation on SII task: accuracy (mAP) vs. logical consistency (LVR) across three datasets

Method	Mean Average Precision (%)			Logical Violation Rate (%)		
	PASCAL	ADE20K	PartImg	PASCAL	ADE20K	PartImg
FR-CNN	78.2	65.1	61.3	12.4	18.9	22.1
SBR	80.1	67.3	63.2	5.8	9.1	11.4
CCN+	81.5	67.9	63.8	**0.0**	**0.0**	**0.0**
DeepProbLog	80.9	67.5	63.5	3.1	5.2	6.8
LTN	78.8	65.8	61.9	4.1	7.2	9.8
DF-$\mathcal{EL}^{++}$	**83.1**	**71.5**	**67.2**	**0.3**	**0.8**	**1.1**

Summary: **Best accuracy** (red), **zero violations** (blue), **near-zero violations** (green). Our method achieves optimal balance between accuracy and logical consistency.

The Optimal Balance of Accuracy and Logical Fidelity. The most compelling story emerges from the trade-off between classification accuracy (from the figures) and logical fidelity (from Table 6). While hard-constraint methods like CCN+ by design achieve a perfect zero violation rate, this rigid enforcement appears to come at the cost of both F1-score and mAP. Conversely, other soft-constraint methods improve accuracy but still permit a significant number of logical contradictions (4-9% violation rates).

Our DF-$\mathcal{EL}^{++}$ framework uniquely excels on all fronts. Through its principled fuzzy optimization, it drives the violation rate to a near-zero minimum (0.3% on PASCAL-PART) while simultaneously attaining the highest F1-score and mAP. This demonstrates a superior balance, achieving the near-perfect logical consistency of hard-constraint methods without sacrificing the perceptual accuracy of the underlying neural network.

The superior performance of DF-$\mathcal{EL}^{++}$, even when compared against an LTN armed with a more expressive $\mathcal{ALC}$ ontology, is particularly insightful. While LTN's FOL foundation allows it to model disjointness, its highly expressive nature also creates a much more complex and potentially rugged optimization landscape. In contrast, our framework's focus on the simpler, more constrained structure of $\mathcal{EL}^{++}$ appears to provide a stronger inductive bias, guiding the optimizer towards more robust and generalizable solutions for this specific task. This suggests that for neuro-symbolic tasks like SII, the theoretical expressivity of the logic may be less critical than the stability and efficiency of the learning dynamics that its structure enables.

Finally, the results on the more challenging ADE20K and PartImageNet datasets highlight the robustness of our approach. While all methods experience a performance drop, the performance gap between DF-$\mathcal{EL}^{++}$ and the baselines widens across all metrics. This demonstrates that our framework's ability to leverage the deeper and more complex logical constraints of the larger ontologies allows it to resolve ambiguities that overwhelm other methods, confirming the generalizability of our principled semantic approach.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009