Synergizing Sparse Structures and Hyperbolic Geometry for Nested Named Entity Recognition

Chunyue Lu
Faculty of Artificial Intelligence
Shanghai University of Electric Power
Shanghai, China
13916754998@163.com

Xilong Wang*
Faculty of Artificial Intelligence
Shanghai University of Electric Power
Shanghai, China
sallywang@shiep.edu.cn

Weixin Zuo
Faculty of Artificial Intelligence
Shanghai University of Electric Power
Shanghai, China
zuoweixin@mail.shiep.edu.cn

***Abstract*—Nested Named Entity Recognition (NER) involves complex layered entities but is often hindered by boundary ambiguity and intricate dependencies. Current span-based paradigms are frequently limited as they overlook the non-Euclidean nature and discrete structural correlations of nested entities. To address this, we propose a Structure-Geometry Synergistic Framework that integrates sparse structural graphs with hyperbolic mapping mechanisms. Specifically, we employ a structural enhancement layer to construct sparse graphs for dependency learning, and a hyperbolic mapping layer to project representations into the Poincaré ball to explicitly capture hierarchy. Additionally, a calibration strategy guided by boundary priors is introduced to refine decision boundaries. Comprehensive evaluations show that our framework outperforms state-of-the-art (SOTA) baselines.**

***Index Terms*—Nested Named Entity Recognition, Hyperbolic Geometry, Sparse Graph, Boundary Calibration**

I. Introduction

Named Entity Recognition (NER) is a core task in Natural Language Processing, aiming to extract structured information from unstructured text. While sequence labeling methods have achieved human-level performance on flat entities, they face significant representation bottlenecks when dealing with nested structures characterized by boundary ambiguity and multi-layered dependencies. In domains such as biomedicine, entities are often recursively embedded, as shown in Figure 1, where proteins are located within cellular components, forming complex topological hierarchies. Failure to precisely demarcate these overlapping boundaries not only degrades identification accuracy but also propagates errors to downstream tasks.

To address the complexity of hierarchical structures, current mainstream paradigms generally fall into four categories: sequence labeling-based, hypergraph-based, Large Language Model (LLM)-based methods [1], and span-based methods [2], [3]. Among these, span-based approaches have emerged as an effective paradigm by modeling entities through exhaustive span enumeration, successfully bypassing the label collision problem. However, a fundamental geometric limitation persists: most existing models embed span representations in Euclidean space. As the nesting depth increases, the number of nodes grows exponentially, whereas the volume of Euclidean space expands only polynomially. This inherent capacity mismatch forces the model to compress distinct hierarchical levels into crowded representations, causing high distortion and unstable boundary determination for deeply nested entities.

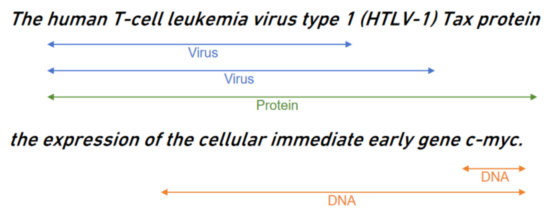

Fig. 1. Examples of nested entities in GENIA

Furthermore, while recent advancements have enhanced semantic capture, standard models fundamentally lack the appropriate geometric inductive bias for hierarchical data. Constrained by the zero-curvature nature of Euclidean space, these models struggle to preserve transitive containment relations. In addition to geometric constraints, current approaches typically employ dense modeling of span neighborhoods. Such indiscriminate aggregation introduces substantial noise from irrelevant spans, diluting precise boundary signals. Consequently, bridging this gap requires a framework that synergizes sparse structural filtering with curvature-adaptive geometric representations.

To bridge the discrepancy between semantic richness and geometric constraints, we propose a novel Structure-Geometry Synergistic Framework that integrates topological denoising with non-Euclidean representation learning. Our approach comprises four synergistic modules. First, a BERT-based encoder establishes the foundational semantic space. Second, we design a Structural Enhancement Layer that constructs a sparse relational graph, explicitly filtering out neighborhood noise to distill robust topological cues. Third, to resolve the Euclidean capacity mismatch, the Hyperbolic Mapping Layer projects span representations into the Poincaré ball. By leveraging adaptive curvature estimation and boundary-driven Möbius fusion, this module explicitly accommodates heterogeneous nesting depths. Finally, an adaptive posterior calibration strategy is employed to refine decision boundaries based on boundary priors.

*Corresponding author.

The main contributions of this work are summarized as follows:

- We introduce hyperbolic geometry into nested NER, demonstrating its pivotal role in providing an appropriate inductive bias for characterizing hierarchical structures and boundary tightness.
- We propose a Structure-Geometry Synergistic Framework that integrates sparse structural graphs with hyperbolic mapping mechanisms to effectively mitigate boundary ambiguity in complex nested scenarios.
- Comprehensive evaluations on four benchmarks demonstrate our model's superiority over existing SOTA methods, while ablation studies rigorously validate the efficacy of each core module.

II. Related Work

A. Nested NER

Nested NER has undergone a significant paradigm shift across several evolutionary stages. Early attempts primarily leveraged sequence labeling with layer-wise decoding [4] or hypergraph-based architectures [5], though they frequently encountered challenges such as cascading error propagation and prohibitive computational complexity. Currently, span-based approaches have emerged as the dominant paradigm [6]. By exhaustively enumerating potential candidates, these methods effectively resolve label collisions while prioritizing boundary precision and inter-span interactions. Key advancements include Biaffine mechanisms for modeling complex boundary dependencies [7] and W2NER for unified relation-based classification [8]. More recent explorations utilize diffusion models for robust boundary refinement [9]. While LLM-based generative approaches [10] demonstrate remarkable zero-shot capabilities, specialized structured models continue to maintain a competitive edge due to LLM-related hallucinations and the inherent difficulty of enforcing rigid structural constraints in complex nesting scenarios.

B. Hyperbolic Geometric Representations

Hyperbolic manifolds, exemplified by the Poincare ball, possess constant negative curvature and exhibit exponential volume growth—properties that render them inherently superior to Euclidean space for representing hierarchical data [11]. This geometric framework was initially introduced to capture intricate lexical sub-relationships [12] and establish fundamental hyperbolic neural operations [13]. Subsequently, it has significantly enhanced performance in knowledge graph completion and relation extraction through the development of adaptive curvature mechanisms [14]. Despite these successes, the application of hyperbolic geometry in NER remains largely confined to lexical or label-level embeddings, leaving span-level structural modeling unexplored. We bridge this gap by synergizing sparse structural filtering with hyperbolic representations to effectively mitigate boundary ambiguity through a non-Euclidean lens.

III. Our Approach

In this section, we introduce the details of our framework as shown in Figure 2. We formulate the task definition of nested NER as follows:

Given a token sequence $X = \{x_1, x_2, \ldots, x_n\}$, any index pair $s = (i, j)$ with $1 \leq i \leq j \leq n$ defines a candidate span. Let $\mathcal{T}$ be the predefined set of entity types and "None" denote the non-entity label. For each span s, the model predicts a label in $\mathcal{T} \cup \{\text{None}\}$; spans assigned labels in $\mathcal{T}$ form the final entity set, where overlaps and containment between spans are allowed to represent nested structures.

A. Encoder

We first utilize a pre-trained BERT model to derive contextual representations $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_n] \in \mathbb{R}^{n \times d_{enc}}$ for an input sequence of length n. To explicitly capture entity boundaries, the encoder generates start and end boundary priors $\mathbf{P}_s, \mathbf{P}_e \in \mathbb{R}^n$ via linear projections:

$$P_s(i) = \sigma(\mathbf{w}_s^\top \mathbf{h}_i + b_s), \quad P_e(i) = \sigma(\mathbf{w}_e^\top \mathbf{h}_i + b_e), \tag{1}$$

where $\mathbf{w}$ and b are learnable parameters. These sequences provide boundary confidence signals, serving as static prior features for subsequent gating mechanisms.

To capture span-level semantics, we employ a Multi-Head Biaffine mechanism [7] for all candidate spans $s = (i, j)$ satisfying length constraints $j - i + 1 \leq L_{\max}$. The feature transformation, biaffine scoring, and final span representation $\mathbf{m}_{ij}$ are formulated as:

$$\begin{gathered} \mathbf{h}_i^s = \mathrm{FFN}_s(\mathbf{h}_i), \quad \mathbf{h}_j^e = \mathrm{FFN}_e(\mathbf{h}_j), \\ \mathbf{m}_{ij} = \mathrm{GELU}\left(\mathbf{W}_m \left[\mathbf{a}_{ij} \oplus \mathbf{h}_i^s \oplus \mathbf{h}_j^e \oplus \mathbf{w}_{\ell_{ij}}\right] + \mathbf{b}_m\right), \end{gathered} \tag{2}$$

where $\mathbf{a}_{ij}$ denotes the concatenation of multi-head biaffine scores $\text{score}_{ij}^{(h)} = (\mathbf{h}_i^s)^\top \mathbf{U}^{(h)} \mathbf{h}_j^e$, and $\mathbf{w}_{\ell_{ij}}$ represents learnable length embeddings. We stack $\mathbf{m}_{ij}$ to form the base semantic matrix $\mathbf{M}_0 \in \mathbb{R}^{S \times d_b}$. Additionally, an initial span score vector $\mathbf{S}^{\text{pre}} \in \mathbb{R}^S$ is generated via a linear layer with Sigmoid activation to regulate the intensity of subsequent structural enhancement.

B. Structural Enhancement Layer

The span structure enhancement layer is designed to simulate the topological dependencies within the span space, transforming the base representation $\mathbf{M}_0$ into the structurally enhanced representation $\widetilde{\mathbf{R}}$.

First, we project $\mathbf{M}_0$ to obtain node representations $\mathbf{B}_s$. To construct the structural graph, we enumerate span pairs (s, t) subject to length constraints and assign geometric relation types [13] (containment, overlap, adjacency, shared boundary) to encode boundary coupling patterns. To circumvent the $\mathcal{O}(S^2)$ complexity of fully connected graphs, we utilize the initial confidence $\mathbf{S}^{\text{pre}}$ from the encoder (stopping gradients) to bucket and prune neighbors, constructing a size-controlled sparse neighborhood $\mathcal{N}(s)$.

On this sparse graph, we employ relation-aware attention. Specifically, via linear projections, we obtain query $\mathbf{q}_s$, value

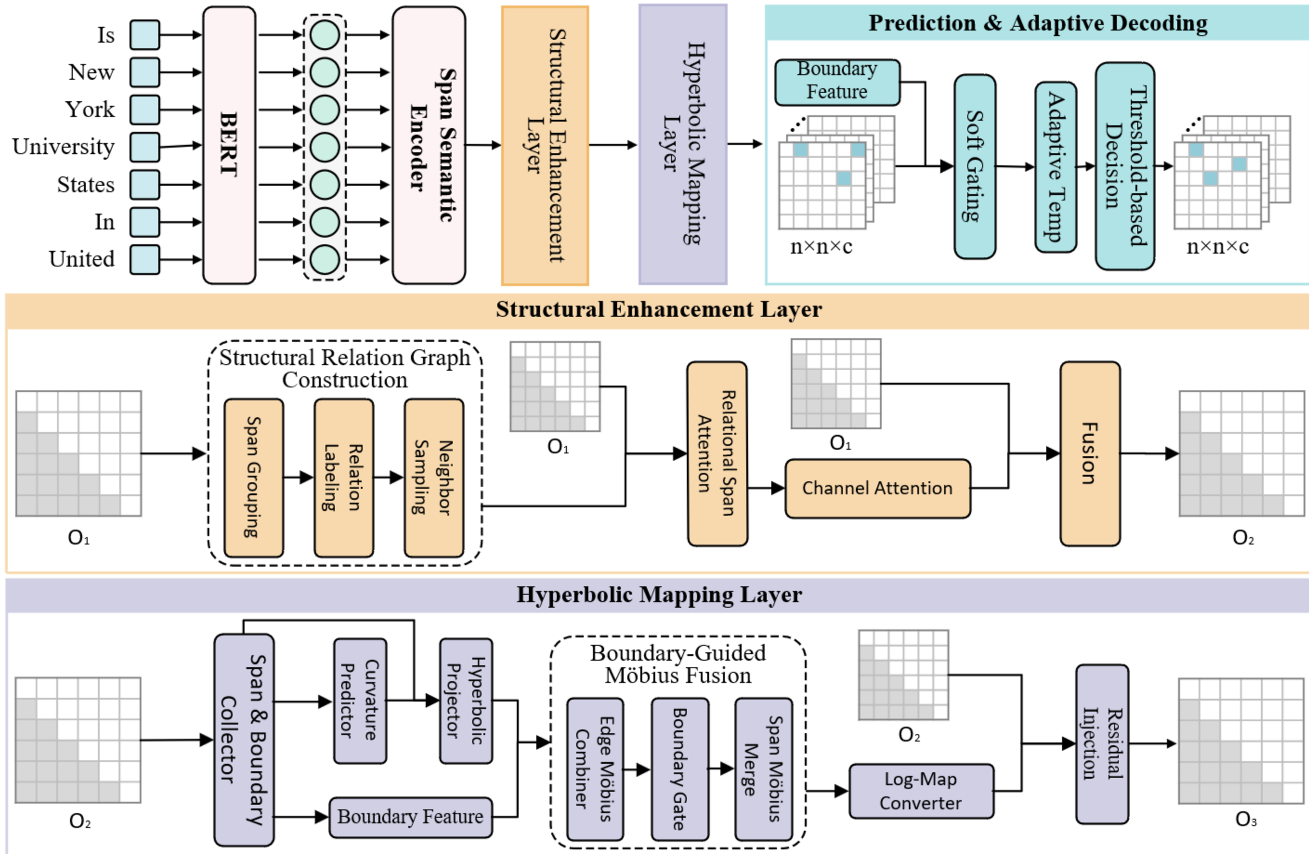

Fig. 2. The overall architecture of the proposed framework.

$\mathbf{v}_{s,k}$, and key $\mathbf{k}_{s,k}$, where $\mathbf{k}_{s,k}$ explicitly fuses the neighbor projection with a learnable relation encoding. The normalized weights $\alpha_{s,k}$ and structural context $\mathbf{g}_s$ are computed as:

$$
\begin{aligned}
\alpha_{s,k} &= \frac{\exp\left(\mathbf{q}_s^\top \mathbf{k}_{s,k}/\sqrt{d_a}\right)}{\sum_{t_l \in \mathcal{N}(s)} \exp\left(\mathbf{q}_s^\top \mathbf{k}_{s,l}/\sqrt{d_a}\right)}, \\
\mathbf{g}_s &= \sum_{t_k \in \mathcal{N}(s)} \alpha_{s,k} \mathbf{v}_{s,k},
\end{aligned}
\tag{3}
$$

where d_a is the scaling factor.

To further highlight discriminative features and suppress noise, we introduce a channel re-weighting mechanism. A gating network projects the aggregated context $\mathbf{g}_s$ to obtain channel-wise importance weights $\mathbf{w}_{\text{gate}}$, which are then used to adaptively rescale the original features via element-wise multiplication, yielding the refined structural representation $\hat{\mathbf{g}}_s$.

Finally, to inject structural information while maintaining semantic robustness, we adopt a boundary-prior-driven gated residual fusion. We concatenate the encoder-derived boundary priors $P_s(i_s), P_e(j_s)$, normalized length $\hat{\ell}_s$, and initial confidence S_s^{pre} as input to a gating network, adaptively computing the injection coefficient γ_s to update the span representation:

$$
\begin{aligned}
\gamma_s &= \sigma\left(\text{MLP}_{\text{gate}}([P_s(i_s), P_e(j_s), \hat{\ell}_s, S_s^{\text{pre}}])\right), \\
\tilde{\mathbf{R}}_s &= \mathbf{B}_s + \gamma_s \cdot \hat{\mathbf{g}}_s.
\end{aligned}
\tag{4}
$$

This mechanism facilitates dynamic information regulation: when boundary priors and initial confidence are high, the model increases γ_s to incorporate more structural refinement; conversely, it prioritizes preserving base semantics, thereby enhancing robustness against boundary noise.

C. Hyperbolic Mapping Layer

Leveraging the exponential volume growth of hyperbolic space, we project $\tilde{\mathbf{R}}_s$ into a d_h-dimensional negative curvature manifold to model nested hierarchies. This encoding captures tree-structured dependencies with minimal distortion, effectively alleviating the Euclidean bottleneck inherent in deep nesting.

Adhering to the manifold learning paradigm, the model adaptively estimates the local curvature κ_s conditioned on span semantics. To accommodate the heterogeneity of nesting depths, this dynamic curvature mechanism projects complex spans into high-curvature subspaces to secure adequate embedding margins, while maintaining flat entities within near-Euclidean regions:

$$
\kappa_s = \kappa_{\min} + \sigma\left(\text{MLP}_\kappa([\tilde{\mathbf{R}}_s \oplus \mathbf{h}_{i_s} \oplus \mathbf{h}_{j_s}])\right)(\kappa_{\max} - \kappa_{\min}). \tag{5}
$$

Subsequently, the exponential map $\exp_0^{\kappa_s}$ projects Euclidean vectors onto the Poincaré ball. As shown in Figure 3, this pushes deeply nested entities towards the boundary where

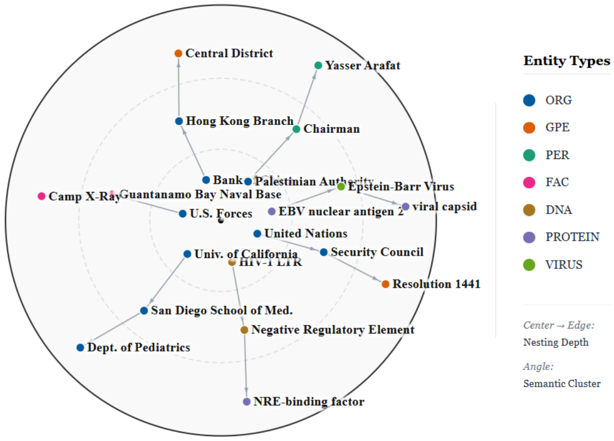

Fig. 3. Poincaré embeddings of nested entities.

volume grows exponentially, explicitly preserving the hierarchical structure. Given that the Poincaré ball model lacks the linear structure of traditional vector spaces, non-Euclidean operations are necessary. We employ Möbius addition [13] to non-linearly synthesize start and end features on the hyperbolic manifold, deriving a geometric center representation that encapsulates entity boundary constraints:

$$\begin{aligned} \mathbf{z}_s^{\text{span}} &= \exp_0^{\kappa_s}(\tilde{\mathbf{R}}_s), \\ \mathbf{z}_s^{\text{edge}} &= \exp_0^{\kappa_s}(\mathbf{h}_{i_s}) \oplus_{\kappa_s} \exp_0^{\kappa_s}(\mathbf{h}_{j_s}). \end{aligned} \tag{6}$$

Building upon this, we introduce a gating coefficient β_s, governed by boundary priors, to dynamically fuse semantic points with boundary points on the manifold. This gating mechanism is designed to minimize the geodesic distance between the span center and boundary anchors, thereby suppressing geometric deviations induced by boundary noise:

$$\begin{aligned} \beta_s &= \sigma\big(\text{MLP}_{\text{edge}}([P_s(i_s), P_e(j_s), \hat{\ell}_s])\big), \\ \mathbf{z}_s &= \mathbf{z}_s^{\text{span}} \oplus_{\kappa_s} (\beta_s \otimes_{\kappa_s} \mathbf{z}_s^{\text{edge}}), \end{aligned} \tag{7}$$

where $\otimes_{\kappa_s}$ denotes Möbius scalar multiplication.

Finally, to ensure compatibility with Euclidean classifiers, we project $\mathbf{z}_s$ back to the tangent space via the logarithmic map. The resulting geometric residual $\mathbf{r}_s^{\text{hyp}}$ is weighted by λ and fused with $\tilde{\mathbf{R}}_s$ to yield the geometry-enhanced representation $\mathbf{R}_s^{\text{hyp}}$:

$$\begin{aligned} \mathbf{r}_s^{\text{hyp}} &= \log_0^{\kappa_s}(\mathbf{z}_s), \\ \mathbf{R}_s^{\text{hyp}} &= \text{LayerNorm}(\tilde{\mathbf{R}}_s + \lambda \cdot \mathbf{r}_s^{\text{hyp}}). \end{aligned} \tag{8}$$

This adaptive mechanism effectively modulates the geometric inductive bias, optimizing the balance between semantic features and hierarchical constraints.

D. Decoder and Training

The decoder generates class logits $\mathbf{o}_s$ from $\mathbf{R}_s^{\text{hyp}}$, concatenated with span length $\hat{\ell}_s$ and boundary confidence $\mathcal{P}_s$. To rectify biases from long-tail distributions, we introduce a prior-guided posterior calibration using learnable temperature τ_c and dynamic bias $b_c(\mathcal{P}_s)$:

$$P(y=c|s) = \frac{\exp\big((o_{s,c} + b_c(\mathcal{P}_s))/\tau_c\big)}{\sum_{k\in\mathcal{T}'} \exp\big((o_{s,k} + b_k(\mathcal{P}_s))/\tau_k\big)}, \tag{9}$$

where $\mathcal{T}' = \mathcal{T} \cup \{\text{None}\}$.

For optimization, we adopt a generalized multilabel cross-entropy loss $\mathcal{L}_{\text{span}}$ for calibrated logits. For span s with positive/negative label sets $\mathcal{C}_s^+/\mathcal{C}_s^-$, the loss is formulated as:

$$\mathcal{L}_{\text{span}}(s) = \log\Big(1 + \sum_{j\in\mathcal{C}_s^-} e^{o_{s,j}}\Big) + \log\Big(1 + \sum_{k\in\mathcal{C}_s^+} e^{-o_{s,k}}\Big). \tag{10}$$

An auxiliary boundary detection task $\mathcal{L}_{\text{boundary}}$ via standard BCE is added to supervise priors $\mathbf{P}_s$ and $\mathbf{P}_e$. The total objective is:

$$\mathcal{L} = \frac{1}{S}\sum_s \mathcal{L}_{\text{span}}(s) + \lambda\mathcal{L}_{\text{boundary}}, \tag{11}$$

where λ balances the boundary term, and all calibration parameters are updated end-to-end.

IV. Experiments

A. Dataset

To evaluate the model, we conducted experiments on four publicly available English nested datasets: ACE2004, ACE2005, GENIA, and KBP17. The dataset divisions are shown in Table I. ACE2004 [15] and ACE2005 [16] contain news articles, broadcasts, and blog posts, featuring seven general entity types. GENIA [17] contains annotated abstracts of biomedical papers, including five hierarchical nested entity categories. KBP17 [18] contains news and discussion forum documents, featuring multiple entity types and complex nested references, and was originally used for knowledge base construction.

TABLE I
Nested Datasets Statistics

Dataset	Train		Test		Dev	
	Total	Nested	Total	Nested	Total	Nested
ACE2004	22231	10176	3036	1422	2514	1092
ACE2005	25300	10005	3099	1096	3321	1214
GENIA	47006	8382	5596	1212	4461	818
KBP17	31236	8773	12601	3707	1879	605

B. Main Results

In the comprehensive experiment, we compared the model proposed in this study with various baseline models. There are Biaffine [7] and Triaffine [20] which adopt multi-dimensional affine attention for boundary interaction in the early stage, as well as Sequence2Set [19] which models nested NER as unordered set prediction. Additionally, there are W2NER [8] (based on word-word relation classification), CNNNER [21] (using grid convolution for local encoding), and Diffusion-NER [9] (based on the denoising diffusion process). We

TABLE II
RESULTS ON ACE2004, ACE2005, GENIA, AND KBP17 DATASETS

Model	ACE2004			ACE2005			GENIA			KBP17		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Biaffine (2020) [7]	87.30	86.00	86.70	85.20	85.60	85.40	81.80	79.30	80.50	–	–	–
Sequence2Set (2021) [19]	88.46	86.10	87.26	87.48	86.63	87.05	82.30	78.70	80.40	84.91	83.04	83.96
Triaffine (2022) [20]	87.13	87.68	87.40	86.70	86.94	86.82	80.42	82.06	81.23	89.42	85.22	87.27
CNNNER (2023) [21]	87.33	87.29	87.31	86.70	86.16	87.42	83.19	79.70	81.40	–	–	–
DiffusionNER (2023) [9]	87.32	87.52	87.42	85.04	88.42	86.70	81.85	79.59	80.70	–	–	–
W2NER (2022) [8]	87.19	87.72	87.45	85.77	87.80	86.76	83.10	79.76	81.39	–	–	–
DSpERT (2023) [22]	88.29	88.32	88.31	87.01	87.84	87.42	82.31	81.49	81.90	87.37	87.93	87.65
DiFiNet (2024) [23]	88.64	88.32	88.48	**88.62**	88.17	88.39	83.01	80.80	81.87	–	–	–
RLAN (2024) [24]	87.47	86.18	86.82	87.41	86.72	87.06	**84.07**	82.83	83.44	–	–	–
GSSDAF (2025) [25]	88.39	**88.83**	88.61	87.26	**89.44**	88.33	83.20	81.11	82.14	**89.22**	87.16	88.18
Ours	**88.75**	88.82	**88.78**	87.98	89.31	**88.64**	82.94	**84.23**	**83.58**	87.58	**88.91**	**88.24**

also compared the performance with recent excellent models, such as DSpERT [22] (deep span encoder representations), DiFiNet [23] (boundary-aware semantic differentiation and filtration), RLAN [24] (recursive label attention mechanism), and GSSDAF [25] (global span semantic dependency awareness and filtering).

As shown in Table II, our model achieved the best F1 score in all benchmark tests: ACE2004 (88.78%), ACE2005 (88.64%), GENIA (83.58%), and KBP17 (88.24%). On ACE2004, we obtained the highest precision rate (88.75%). Additionally, this model achieved higher recall rates on GENIA (84.23%) and KBP17 (88.91%), indicating its strong retrieval ability for entities that are difficult to find. In summary, these results demonstrate that this model exhibits excellent stability and performance across different domains.

TABLE III
ABLATION STUDY ON ACE2005 AND GENIA

Setting	ACE2005 F1	Δ	GENIA F1	Δ
Default (Ours, $K = 8$)	88.64	–	83.58	–
w/o SEL	87.33	-1.31	82.93	-0.65
w/o HML	88.15	-0.49	82.60	-0.98
w/o AD	87.50	-1.14	83.00	-0.58
$K = 4$	88.12	-0.52	83.25	-0.33
$K = 16$	88.35	-0.29	83.42	-0.16
$K = 32$	87.90	-0.79	82.80	-0.78

C. Ablation Study

Table III evaluates the contribution of each core component and the sensitivity of neighborhood size K. The complete model ($K = 8$) achieves the best performance, as removing any module incurs a significant F1 drop. Specifically, removing the structural enhancement layer (w/o SEL) or hyperbolic mapping layer (w/o HML) leads to substantial declines (e.g., -1.31% on ACE2005 and -0.98% on GENIA, respectively), validating the importance of topological cues and non-Euclidean representations for deep nesting. Furthermore, excluding adaptive decoding (w/o AD) confirms that boundary-prior-guided calibration is essential for precise identification.

Performance regarding K peaks at $K = 8$. Smaller values ($K = 4$) miss critical structural contexts, while larger values ($K \geq 16$) introduce span noise that dilutes salient patterns. This supports our dynamic sparse graph hypothesis: a moderate window optimally balances dependency coverage and noise suppression.

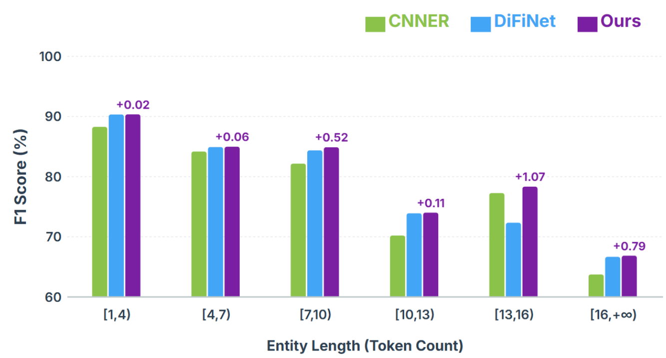

Fig. 4. Entity length-wise results on ACE2004

D. Performance on Long Entities

In the NER task, long entities pose greater challenges as they are more likely to contain deeper nested structures, which exacerbates boundary insensitivity. Moreover, identifying long entities is a typical long-tail problem. To evaluate robustness, we divide entities into six length categories and conduct a detailed comparison with CNNNER and SOTA DiFiNet on ACE2004, as illustrated in Fig. 4. Results show our model achieves highly competitive performance in the short-entity interval [1, 4), obtaining a 90.36% F1 score, which is on par with the strongest baseline, DiFiNet (90.34%). This confirms the solidity of our fundamental feature extraction. However, as entity length increases, the superiority of our approach becomes increasingly evident. Notably, in ultra-long intervals [13, 16) and $[16, +\infty)$, where baseline models typically suffer significant performance degradation, our model demonstrates remarkable stability with absolute F1 gains of 1.07% and 0.79%, respectively. These results validate the model's outstanding ability in handling complex boundary tasks.

E. Case Study

Table IV presents a complex example from GENIA, where a long protein structure is surrounded by a virus and its abbreviation. The baseline model is unable to handle this long entity nesting structure. It mistakenly shortens the outer part to "Tax protein" and mistakenly identifies the virus as "DNA" due to the confusion of overlapping. Then our model precisely identified the nested entities through structural enhancement and hyperbolic embedding and accurately identified the boundaries. The model correctly extracted all three entities, thereby proving its ability to handle extremely long, nested structure entities and to accurately classify them.

TABLE IV
CASE STUDY ON NESTED ENTITY RECOGNITION

	Long-Span & Deeply Nested Boundaries
Text	*Expression of the human T-cell leukemia virus type 1 (HTLV-1) Tax protein represses the cellular immediate early gene c-myc.*
Gold/Ours	{Protein: human T-cell leukemia virus type 1 (HTLV-1) Tax protein}, {Virus: human T-cell leukemia virus type 1 (HTLV-1)}, {Virus: HTLV-1}
Baseline	{Protein: Tax protein}, {DNA: human T-cell leukemia virus type 1}

V. CONCLUSION

In this paper, we propose a Structure-Geometry Synergistic Framework to address the challenge of boundary ambiguity in nested NER. Specifically, we introduced a Structural Enhancement Layer to construct sparse relational graphs, effectively distilling robust topological cues from candidate spans. Furthermore, we designed a Hyperbolic Mapping Layer to project nested entities into the Poincaré ball. By leveraging adaptive curvature and boundary-driven gating mechanisms, we successfully improved the geometric compactness of span representations within the continuous manifold. Extensive experiments on four benchmark datasets demonstrate that our model achieves state-of-the-art performance. The ablation studies and specific evaluations on long entities further validate the effectiveness of our dual-space design. In future work, we plan to extend this framework to other complex information extraction tasks. Moreover, we envision this framework serving as a geometric plug-and-play module for Large Language Models (LLMs), enhancing their capability in precise boundary prediction and understanding of recursive hierarchical structures.

REFERENCES

[1] D. Xu, Y. Chen, S. Yuan *et al.*, "Large language models for information extraction: A survey," *arXiv preprint arXiv:2401.03666*, 2024.

[2] Y. Wang, Q. Liu, T. Dai, J. Lang, L. Lu, and Y. Chen, "Contrastive learning for nested chinese named entity recognition via template words," *CAAI Transactions on Intelligence Technology*, vol. 10, no. 2, pp. 450–459, 2025.

[3] Y. Gan, Y. Liu, Y. Cai, R. Lin, S. Yang, Q. Liu, Y. Wang, and X. Shi, "Optimizing boundary dynamics for nested named entity recognition via semantic refinement and trimming," *Neural Networks*, p. 108218, 2025.

[4] M. Ju, M. Miwa, and S. Ananiadou, "A neural layered model for nested named entity recognition," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 1446–1459.

[5] W. Lu and D. Roth, "Joint mention extraction and classification with mention hypergraphs," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 857–867.

[6] M. G. Sohrab and M. Miwa, "Deep exhaustive model for nested named entity recognition," in *Proceedings of the 2018 conference on empirical methods in natural language processing*, 2018, pp. 2843–2849.

[7] J. Yu, B. Bohnet, and M. Poesio, "Named entity recognition as dependency parsing," *arXiv preprint arXiv:2005.07150*, 2020.

[8] J. Li, H. Fei, J. Liu, S. Wu, M. Zhang, C. Teng, D. Ji, and F. Li, "Unified named entity recognition as word-word relation classification," in *proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 10, 2022, pp. 10 965–10 973.

[9] Y. Shen, K. Song, X. Tan, D. Li, W. Lu, and Y. Zhuang, "Diffusionner: Boundary diffusion for named entity recognition," *arXiv preprint arXiv:2305.13298*, 2023.

[10] W. Zhou, S. Zhang, Y. Gu, M. Chen, and H. Poon, "UniversalNER: Targeted distillation from large language models for open named entity recognition." International Conference on Learning Representations, 2024.

[11] M. Nickel and D. Kiela, "Poincaré embeddings for learning hierarchical representations," *Advances in neural information processing systems*, vol. 30, 2017.

[12] A. Tifrea, G. Bécigneul, and O.-E. Ganea, "Poincaré glove: Hyperbolic word embeddings," *arXiv preprint arXiv:1810.06546*, 2018.

[13] O. Ganea, G. Bécigneul, and T. Hofmann, "Hyperbolic neural networks," *Advances in neural information processing systems*, vol. 31, 2018.

[14] Y. Zhang, X. Wang, C. Shi, N. Liu, and G. Song, "Lorentzian graph convolutional networks," in *Proceedings of the web conference 2021*, 2021, pp. 1249–1261.

[15] G. R. Doddington, A. Mitchell, M. A. Przybocki, L. A. Ramshaw, S. Strassel, and R. M. Weischedel, "The automatic content extraction (ACE) program-tasks, data, and evaluation," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, 2004.

[16] C. Walker, S. Strassel, J. Medero, and K. Maeda, "ACE 2005 multilingual training corpus," Linguistic Data Consortium, Philadelphia, 2005, lDC2006T06.

[17] J.-D. Kim, T. Ohta, Y. Tateisi, and J. Tsujii, "Genia corpus-a semantically annotated corpus for bio-textmining," *Bioinformatics*, vol. 19, no. suppl_1, pp. i180–i182, 2003.

[18] H. Ji, X. Pan, B. Zhang, J. Nothman, J. Mayfield, P. McNamee, C. Costello *et al.*, "Overview of TAC-KBP2017 13 languages entity discovery and linking," in *Proceedings of the Text Analysis Conference (TAC)*, 2017.

[19] Z. Tan, Y. Shen, S. Zhang, W. Lu, and Y. Zhuang, "A sequence-to-set network for nested named entity recognition," *arXiv preprint arXiv:2105.08901*, 2021.

[20] Z. Yuan, C. Tan, S. Huang, and F. Huang, "Fusing heterogeneous factors with triaffine mechanism for nested named entity recognition," in *Findings of the association for computational linguistics: ACL 2022*, 2022, pp. 3174–3186.

[21] H. Yan, Y. Sun, X. Li, and X. Qiu, "An embarrassingly easy but strong baseline for nested named entity recognition," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2023, pp. 1442–1452.

[22] E. Zhu, Y. Liu, and J. Li, "Deep span representations for named entity recognition," in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 10 565–10 582.

[23] Y. Cai, Q. Liu, Y. Gan, R. Lin, C. Li, X. Liu, D. Luo, and J. Yang, "DiFiNet: Boundary-aware semantic differentiation and filtration network for nested named entity recognition," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024, pp. 6455–6471.

[24] H. Kim and H. Kim, "Recursive label attention network for nested named entity recognition," *Expert Systems with Applications*, vol. 249, p. 123657, 2024.

[25] Y. Sun, X. Wang, H. Wu, and M. Hu, "Global span semantic dependency awareness and filtering network for nested named entity recognition," *Neurocomputing*, vol. 617, p. 129035, 2025.