Dual-Perspective Modeling for Chinese NER with Selective SSM and Semantic Difference Convolution

Jianquan Ouyang
Xiangtan University
Xiangtan, China
oyjq@xtu.edu.cn

Xinxin Li
Xiangtan University
Xiangtan, China
xx020403@163.com

***Abstract*—Chinese Named Entity Recognition (Chinese NER) faces challenges in boundary identification due to the absence of explicit word delimiters. While lexicon-based methods improve performance, their high maintenance cost limits domain applicability. Moreover, models often overemphasize high-frequency boundary characters, leading to biased predictions and reduced generalization. To address these issues, we propose a dual-perspective modeling framework that integrates the Selective State Space Model (Selective SSM) and Semantic Difference Convolution (SDC). First, biaffine attention transforms the input sentence into a character-pair grid matrix, capturing inter-character relationships. The Selective SSM then captures long-range dependencies from a global perspective, providing rich contextual information. Concurrently, the SDC enhances local context understanding by modeling subtle semantic variations between adjacent characters, mitigating boundary bias. Ultimately, the fusion of features from both perspectives enables the decoding of entities and categories. Experimental results demonstrate that our method significantly outperforms or matches state-of-the-art approaches on four datasets: Resume, MSRA, Weibo, and OntoNotes 4.0, thereby validating the effectiveness of the proposed dual-perspective paradigm.**

***Index Terms*—Chinese NER, Selective State Space Model, Semantic Difference Convolution, Dual-perspective Modeling**

I. Introduction

Named entity recognition (NER) is a fundamental task in natural language processing (NLP), which aims to identify and classify specific entities, such as person (PER), location (LOC), and organization (ORG), within unstructured text. It plays a vital role in numerous downstream applications, including knowledge graph construction and sentiment analysis. However, compared to English NER, Chinese NER faces more significant challenges. The root cause lies in the lack of natural word boundaries in Chinese text and the flexible, variable nature of character combinations. For example, in the typical sentence **“**尼日尔河流经尼日尔与尼日利亚 **(The Niger River flows through Niger and Nigeria)”**, the character **“**流 **(flow)”** can form **“**河流 **(river)”** with the character **“**河 **(river)”** on the left and **“**流经 **(flow through)”** with the character **“**经 **(through)”** on the right. This ambiguity in local combinations directly complicates the identification of entity boundaries.

Traditional approaches typically rely on Chinese word segmentation to identify word boundaries, but entity boundaries do not always align with word boundaries. For instance, the entity **“**西藏自治区 **(Tibet Autonomous Region)”** may be segmented as **“**西藏自治区 **(Tibet Autonomous Region)”** or as **“**西藏 **(Tibet)”** and **“**自治区 **(Autonomous Region)”** under different segmentation criteria. Researchers have attempted to improve Chinese NER by incorporating external lexicon information, such as Lattice-LSTM [1] and lexicon adapters [2]. While these methods can enhance recognition performance, they incur high maintenance costs for the lexicons. Furthermore, existing models often overlook the entity boundary bias arising from the high mention rate of boundary characters, which severely limits their generalization capability. In recent years, Transformer-based models, such as DUST [3] and LLET [4], have achieved remarkable results in Chinese NER tasks. However, these models have inherent limitations: their fully connected self-attention mechanism causes the query to attend to all characters, potentially integrating irrelevant information and leading to misidentification of entity boundaries. More importantly, they generally struggle to handle the complex long-range dependencies in Chinese text effectively and to capture the subtle semantic differences that determine entity boundaries.

To address these challenges, we proposes a dual-perspective modeling framework that integrates a Selective State Space Model (Selective SSM) with Semantic Difference Convolution (SDC). Our core design principle is to effectively address the shortcomings of existing methods in both global and local modeling through these two complementary modules. Specifically, we utilize the Selective SSM to capture long-range dependencies in sentence efficiently, providing rich global contextual information. Specifically, we utilize Selective SSM to efficiently capture long-range dependencies in text, providing rich global contextual information. This enhances the model's ability to determine entity categories, which in turn offers strong constraints for boundary identification. Simultaneously, we introduce the SDC module to capture subtle semantic

Identify applicable funding agency here. If none, delete this.

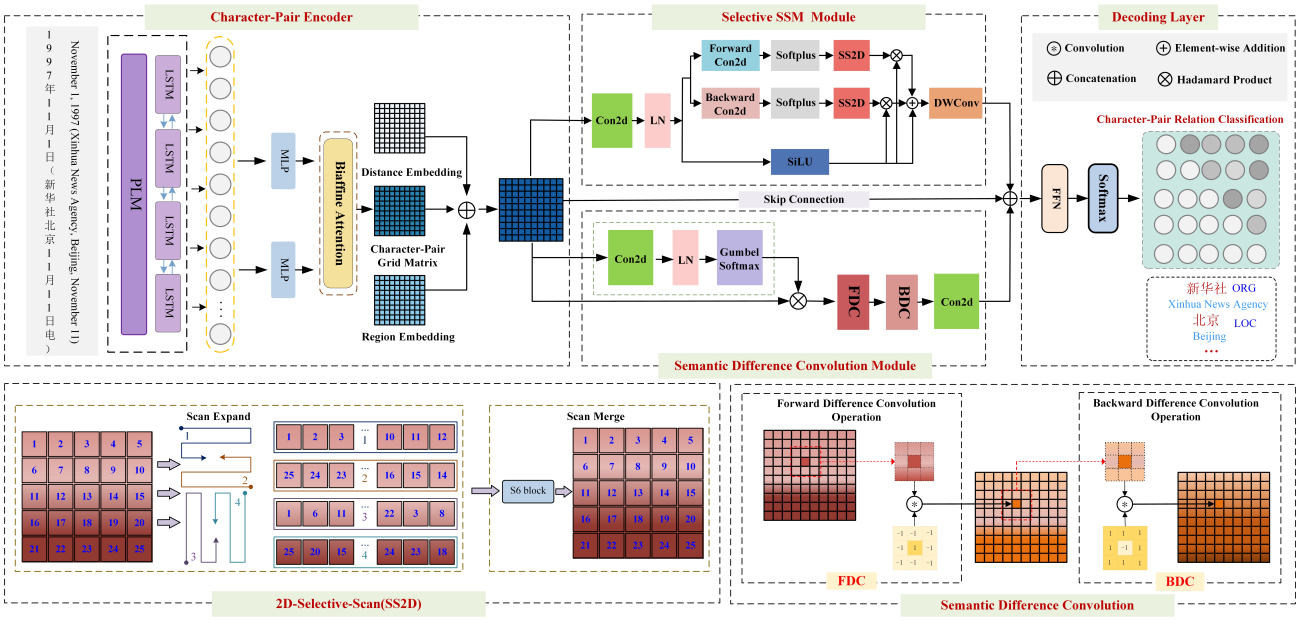

Fig. 1. The top figure displays overview architecture of the proposed model. Below it, the left diagram is a schematic of SS2D, illustrating its four-way scanning for efficient global context acquisition. The right diagram shows the SDC module, detailing how it uses an adaptive mask and difference operators to model local semantic differences.

changes between adjacent characters. Its unique convolutional differential operation precisely focuses on key semantic jumps at entity boundaries, effectively alleviating the recognition bias caused by the high mention rate of boundary characters. Our main contributions are as follows:

- We apply the Selective State Space Model to the Chinese NER task, leveraging its selective state update mechanism to address the shortcomings of existing models in efficiently modeling long-range dependencies.
- We introduce the Semantic Difference Convolution module, designing a convolutional differential operation to model subtle semantic differences between adjacent characters explicitly. This module effectively addresses the limitations of the Selective State Space Model in capturing local boundary details, thereby improving the performance of entity boundary and category recognition
- We conduct extensive experiments on four widely used Chinese NER benchmark datasets: Resume, MSRA, Weibo, and OntoNotes 4.0. Experimental results show that the proposed method outperforms or is comparable to existing state-of-the-art methods, verifying its effectiveness.

II. Related Work

Early research in Chinese NER mainly focused on character-based approaches [5]–[7]. However, the lack of explicit word boundary identifiers in Chinese text makes it difficult to determine entity boundaries. Developing deep learning frameworks has brought Chinese NER to a new breaking point. Lattice-LSTM [1] utilized lexicon information to improve the recognition of entity boundaries.CNN [8] with rethinking mechanism solved the problem of inter-lexical conflicts. Sui et al. [9] used GNN to model the relationship between characters and words to enhance the expressiveness of the model. Li et al. [10] simplified the complex lattice structure into a spanning sequence, simplifying the model and improving performance. Recently, researchers have proposed new methods such as MCL [11], NerCo [12], SSMI [13], ATSSA [14], and LADA-trans-NER [15], which have improved the recognition effect but still suffer from the problems of inaccurate word boundaries, insufficient context utilization, and insufficient ability to handle complex semantic relations. To address these issues, we propose using Selective SSM and SDC to improve the accuracy of Chinese entity recognition.

III. Methodology

In this paper, we introduce a a dual-perspective modeling framework for Chinese NER, integrating a Selective SSM module and an SDC module. As illustrated in Fig. 1, the model first encodes the input sentence into a character-pair grid matrix. The Selective SSM module captures long-range dependencies to extract global contextual information, while the SDC module enhances local understanding by modeling semantic differences between adjacent characters. Finally, the model fuses the features from both branches and performs entity recognition through a decoding layer.

A. Task Formulation

In this paper, following the approach of Li et al. [16], we formalize Chinese NER as a relation classification task over character pairs (x_i, x_j) where $i < j$. Given a sentence $X = \{x_1, \ldots, x_N\}$ and a set of entity types E, the model aims to

predict a relation $r \in R$ for each character pair. The predefined relation set R consists of the following:

- NNC: The NNC relationship indicates that the character-pair belongs to the same entity mention and that the two characters are adjacent in the entity mention.
- THC-Type: The THC-Type relationship indicates a character-pair, one of which is the end of the entity mention and the other is the head of the entity mention. "Type" indicates one of the entity types E.
- NONE: indicates that the character-pair does not belong to any entity mention.

B. Character-Pair Encoder

Given an input sentence $X = \{x_1, x_2, \ldots, x_N\}$ composed of Chinese characters, where x_i denotes the i-th character, we first encode it using a pre-trained language model (PLM) followed by a Bi-LSTM to obtain character representations $H_X \in \mathbb{R}^{N \times d_h}$. Next, two multi-layer perceptrons (MLPs) are applied to generate the start and end representations of each character-pair:

$$H_s = \mathrm{MLP}_{\mathrm{start}}(H_X), \quad H_e = \mathrm{MLP}_{\mathrm{end}}(H_X) \tag{1}$$

For each character-pair (x_i, x_j) with $i < j$, we employ a biaffine attention to compute its feature representation $H^x_{i,j}$:

$$H^x_{i,j} = h^\top_{s,i} W_1 h_{e,j} + W_2(h_{s,i} \oplus h_{e,j}) + v \tag{2}$$

where $W_1 \in \mathbb{R}^{d_b \times d_b}$ and $W_2 \in \mathbb{R}^{d_b \times 2d_b}$ are weight matrices, $v \in \mathbb{R}^{d_b}$ is a bias vector, and $\oplus$ denotes vector concatenation.

Finally, to incorporate positional and structural information, we fuse $H^x \in \mathbb{R}^{N \times N \times d_b}$ with distance embeddings H^{dis} and region embeddings H^{reg}, yielding the final character-pair grid matrix:

$$H_b = H^x \oplus H^{dis} \oplus H^{reg} \in \mathbb{R}^{N \times N \times d_c} \tag{3}$$

C. Selective SSM Module

To efficiently model long-range dependencies within the character-pair grid, we introduce VMamba's 2D Selective Scan (SS2D) to build the Selective SSM module [17]. SS2D achieves a global receptive field through four-way scanning while transmitting contextual information with linear time complexity, thereby preserving dynamic weighting properties similar to those of self-attention.

Specifically, we first apply a 1×1 Convolution Layer (Conv2d) and Layer Normalization (LN) to the input H_b, yielding the preprocessed feature $\overline{H}_b$:

$$\overline{H}_b = \mathrm{LN}\left(\mathrm{Conv2d}(H_b)\right) \tag{4}$$

$\overline{H}_b$ then enters two parallel branches. The main branch passes through a 3×3 convolution, SoftPlus, and SiLU activations before being input to the SS2D module, which generates forward features $\overline{H}_{fb}$ and reverse features $\overline{H}_{bb}$. The skip-connection branch directly applies a SiLU activation to $\overline{H}_b$ to obtain $\overline{H}_{sb}$:

$$\overline{H}_{fb} = \mathrm{SS2D}\left(\mathrm{SiLU}\left(\mathrm{SoftPlus}\left(\mathrm{Conv2d}(\overline{H}_b)\right)\right)\right) \tag{5}$$

$$\overline{H}_{bb} = \mathrm{SS2D}\left(\mathrm{SiLU}\left(\mathrm{SoftPlus}\left(\mathrm{Conv2d}(\overline{H}_b)\right)\right)\right) \tag{6}$$

$$\overline{H}_{sb} = \mathrm{SiLU}\left(\overline{H}_b\right) \tag{7}$$

We fuse these three feature paths using a Hadamard product to get the fused feature $\overline{H}_{mb}$:

$$\overline{H}_{mb} = \left(\overline{H}_{sb} \odot \overline{H}_{fb}\right) + \left(\overline{H}_{sb} \odot \overline{H}_{bb}\right) + \overline{H}_{sb} \tag{8}$$

where $\odot$ represents the Hadamard product.

Finally, a 3×3 Depthwise Separable Convolution (DWConv2d) is applied to $\overline{H}_{mb}$ to output the final global character-pair representation, H_{gc}:

$$H_{gc} = \mathrm{DWConv2d}\left(\overline{H}_{mb}\right) \tag{9}$$

D. Semantic Difference Convolution Module

To enhance the model's ability to perceive local semantic differences between character pairs, we introduce the SDC module. The SDC module first applies a 3×3 Convolution Layer and Layer Normalization to the input character-pair grid H_b:

$$H_{\mathrm{ln}} = \mathrm{LN}\left(\mathrm{Conv2d}(H_b)\right) \tag{10}$$

we then use the Gumbel-Softmax estimator to generate an adaptive mask matrix M, which is combined with H_b via the Hadamard product to focus on features in key regions:

$$M = \mathrm{GumbelSoftmax}(H_{\mathrm{ln}}) \tag{11}$$

$$H_m = M \odot H_b \tag{12}$$

Inspired by classical difference convolution operators such as Sobel and LoG [18] in computer vision, We design a forward difference operator w_f and a reverse difference operator w_b to capture local semantic contrast:

$$w_f = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & -1 \end{bmatrix}, \quad w_b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{13}$$

We apply w_f to H_m, followed by a GELU activation and Layer Normalization to get H_{fd}. Then, we apply w_b to H_{fd}, again performing activation and normalization to get H_{bd}. This two-stage process extracts multi-level local contrast features.

$$H_{\mathrm{fd}} = \mathrm{LN}\left(\mathrm{GELU}\left(w_f * H_m\right)\right) \tag{14}$$

$$H_{\mathrm{bd}} = \mathrm{LN}\left(\mathrm{GELU}\left(w_b * H_{\mathrm{fd}}\right)\right) \tag{15}$$

where $*$ denotes a convolution operation.

Finally, a 1×1 Convolution Layer integrates the features to produce the enhanced local context representation:

$$H_{lc} = \mathrm{Conv2d}(H_{\mathrm{bd}}) \tag{16}$$

E. Decoding and Training

During decoding and training, we fuse the representations H_b, H_{gc}, and H_{lc} by concatenating them along the feature dimension. A feedforward network (FFN) then transforms the fused features into a relation probability distribution:

$$P = \mathrm{Softmax}\left(\mathrm{FFN}\left(H_b \oplus H_{gc} \oplus H_{lc}\right)\right) \tag{17}$$

where each entry $p_{i,j} \in P$ denotes the probability distribution over the relation set $R = \{\text{NNC}, \text{THC-Type}, \text{NONE}\}$ for the character pair (x_i, x_j) at position (i, j). During inference, the model decodes the probability distribution P to jointly predict entity boundaries and types based on predefined character-pair relations. The **NNC** relation links consecutive characters within the same entity, forming contiguous spans. In contrast, the **THC-Type** relation captures non-consecutive associations that propagate entity type information, enabling accurate identification of both entity spans and their types.

The training objective is the cross-entropy loss between the predicted probabilities $p^r_{i,j}$ and the true labels $y^r_{i,j}$, given by:

$$\mathcal{L} = -\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{r=1}^{|R|} y^r_{i,j} \log p^r_{i,j} \tag{18}$$

where r indexes the relationship type and $|R|$ is the size of the relation set.

IV. Experiments

We evaluate the performance of our proposed method using the F1 score as the primary metric, adhering to a strict matching criterion. A predicted entity is considered correct only if both its boundaries and type are an exact match with the ground truth. All experiments were implemented in PyTorch and run on an NVIDIA A100 GPU.

A. Datasets

We conduct experiments on six widely used Chinese NER benchmark datasets: **MSRA** [19], a standard news corpus for Chinese NER; **OntoNotes 4.0** [20], a multi-genre news dataset with diverse linguistic styles; **Weibo** [21], a social media corpus from Sina Weibo featuring informal language and slang; **Resume** [1], a domain-specific dataset of executive biographies with rich professional entity types; and the nested NER datasets **ACE 2004-zh** and **ACE 2005-zh**, which contain approximately 10–11% nested entities across seven types (e.g., PER, ORG, LOC) and are used to evaluate modeling of complex entity structures [22]. We strictly follow the data splits from prior work.

B. Baselines

We systematically compare our method with three categories of advanced baselines to evaluate its effectiveness. **Character-based methods**, such as CAN-NER [23], RICON [24], W2NER [16], DiffusionNER [25], and BOPN [26], improve recognition performance by modeling character sequences or span relationships. **Lexicon-enhanced methods**, including Lattice-LSTM [1], FLAT [10], LEBERT [2], MFB-NER [27], LLET [4], and HiNER [28], enhance semantic representation by integrating external lexical knowledge. **Other approaches**, such as ATSSA [14], MCL [11], MSE [29], MGBERT [30], and DuST [3], explore effective strategies in attention modeling, contrastive learning, and multi-granularity representation learning. Through comprehensive comparisons with these representative baselines, the performance advantages of our method are clearly demonstrated.

TABLE I
Experiment results on four flat Chinese NER datasets.

Models	OntoNotes 4.0	Weibo	MSRA	Resume
Character-based Methods				
CAN-NER [23]	73.64	59.31	92.97	94.94
RICON [24]	83.33	-	96.14	-
W2NER [16]	83.08	72.32	96.10	96.65
DiffusionNER [25]	81.34	70.43	94.91	95.08
BOPN [26]	-	72.92	96.39	96.78
Lexicon-enhanced Methods				
Lattice-LSTM [1]	73.88	58.79	93.18	94.46
FLAT [10]	82.85	68.68	95.86	96.10
LEBERT [2]	82.08	70.75	95.70	96.08
MFB-NER [27]	81.67	71.57	96.26	96.49
LLET [4]	81.85	68.64	96.11	95.94
HiNER [28]	83.17	72.68	96.56	96.85
Other Innovative Methods				
ATSSA [14]	83.31	72.53	96.45	96.73
MCL [11]	82.96	73.08	96.11	96.46
MSE [29]	82.97	72.09	96.21	96.83
MGBERT [30]	83.05	72.22	96.10	96.23
DuST [3]	83.67	70.57	96.37	96.63
Ours	**83.58**	**74.51**	**96.58**	**96.95**

C. Implementation Details

In all experiments, we use **RoBERTa-wwm-ext-large** [31] as the PLM for our character-pair encoder, obtaining character embeddings with a hidden dimension of 1024. During training, we optimize the model parameters using the **AdamW optimizer** [32] and a **linear warmup-decay** learning rate schedule [33]. The learning rate is set to 2e-5 or 5e-6 for PLM fine-tuning, and the learning rate for the other parts is 0.001, with models trained for 10 to 30 epochs. The **Bi-LSTM** has a single hidden layer, a hidden size ranging from [512, 1024], and a dropout rate of 0.5. The output dimension d_c of the **biaffine attention** ranges from 64 to 256. The convolution layers in both the **Selective SSM** and **SDC** modules have a hidden size in the range of [64, 128] and a dropout rate of 0.5. Finally, the **FFN** in the decoding layer has a hidden size in the range of [64, 256] and a dropout rate of 0.33.

D. Main Results

1) Performance on Flat Entities: The experiment results of our proposed method and baselines on the four flat Chinese NER datasets are shown in Table I. Our method achieves state-of-the-art performance on Weibo and MSRA, with F1-scores of **74.51%** and **96.58%**, respectively, representing improvements of +1.43% over MCL and +0.02% over HiNER, demonstrating strong robustness in informal and general-domain text. On OntoNotes 4.0, our method obtains **83.58%**, outperforming all character-based and lexicon-enhanced models. Although DuST achieves 83.67%, a margin of -0.09%, our model remains competitive without relying on external syntactic features, indicating strong cross-domain generalization. On Resume, our method achieves **96.95%**, surpassing HiNER by +0.10% and setting a new state-of-the-art, confirming superior accuracy in professional texts. Overall, our method

outperforms existing approaches on three datasets and remains highly competitive on OntoNotes 4.0. The Selective SSM captures long-range dependencies for global context modeling, while the SDC module refines local representations through fine-grained character semantics.

2) Performance on Nested Entities: Since the recently proposed Chinese NER methods are generally not evaluated on the ACE 2004-zh and ACE 2005-zh datasets and lack a unified experimental setting, to ensure a fair comparison, we select five recent advanced nested NER methods, including DiffusionNER [25], DiFiNet [34], CNNNER [35], W2NER [16], and Planarized [36], based on the same pre-processed dataset, and reproduce all models using their public codes. All reproduced results are marked with †. Table II presents the performance of various models on the nested Chinese NER datasets. On ACE 2004-zh, our method achieves 88.49%, surpassing the best-performing baseline Planarized with 87.43% by a relative improvement of 1.06%. On ACE 2005-zh, our method achieves 87.78%, significantly outperforming the best-performing baseline DiFiNet with 87.20% by a relative improvement of 0.58%. The experimental results show that our method achieves the best performance on the two nested Chinese NER datasets and is capable of accurately recognizing nested entity structures.

TABLE II
THE PERFORMANCE OF VARIOUS MODELS ON NESTED CHINESE NER DATASETS.† INDICATES RESULTS REPRODUCED USING THE SAME PREPROCESSING AND PUBLICLY AVAILABLE CODE.

Models	ACE 2004-zh	ACE 2005-zh
W2NER [16]†	87.07	85.52
DiffusionNER [25]†	85.30	86.30
CNNNER [35]†	86.74	85.31
Planarized [36] †	87.43	86.59
DiFiNet [34]†	87.38	87.20
Ours	**88.49**	**87.78**

E. Ablation Studies

As shown in Table III, we conduct ablation studies on the Weibo and Resume datasets to evaluate the contribution of each component. The performance degradation when removing (**w/o Distance Emb**) decreases F1 by -0.87% on Weibo and -0.33% on Resume, confirming its role in encoding positional patterns within entity spans. Similarly, removing (**w/o Region Emb**) leads to a drop of -0.90% on Weibo and -0.46% on Resume, demonstrating that sample distribution features enhance character boundary representation. Disabling (**w/o Selective SSM Module**) results in a decline of -1.71% on Weibo and -0.92% on Resume, verifying its necessity for capturing long-range contextual dependencies. In contrast, removing (**w/o SDC Module**) causes a degradation of -2.22% on Weibo and -0.89% on Resume, highlighting the importance of local semantic modeling for fine-grained boundary detection. Most critically, eliminating (**w/o Skip Connection**), which refers to the skip connection in the character-pair grid matrix $\mathbf{H}_b$, incurs a severe drop of -5.03% on Weibo and -1.04% on Resume, underscoring the vital role of the biaffine attention mechanism in structuring pairwise character interactions, particularly in noisy and informal text.

TABLE III
ABLATION STUDY OF THE PROPOSED MODEL ON WEIBO AND RESUME DATASETS.

Component	settings	Weibo	Resume
Embedding	Default	74.51	96.95
	w/o Distance Emb	73.64	96.62
	w/o Region Emb	73.61	96.49
Modules	Default	74.51	96.95
	w/o Selective SSM	72.80	96.03
	w/o SDC	72.29	96.06
	w/o Skip Connection	69.48	95.91

F. Analysis of Computational Efficiency.

To evaluate computational efficiency, we compare parameter count and inference speed on the Weibo dataset. As shown in Table 5, our model achieves an F1 of 74.51% with only 327.21MB parameters, and runs at 32.12 sents/s, 12.55 times faster than Lattice-LSTM. This demonstrates a strong balance between performance and efficiency compared to existing methods.

TABLE IV
EFFICIENCY COMPARISON OF DIFFERENT METHODS.

Model	Parameters	F1 (%)	Speed (sents/s)	SpeedUp
Lattice-LSTM	132M	58.79	2.56	1.00
DiffusionNER	381M	70.43	6.16	2.41
W2NER	222M	72.32	15.42	6.02
BOPN	105M	72.92	47.68	18.63
HiNER	290M	72.68	21.03	8.22
Ours	**327.21M**	**74.51**	**32.12**	**12.55**

V. CONCLUSION

In this paper, we present a novel dual-perspective modeling approach for Chinese NER. The method integrates Selective SSM to capture long-range dependencies and provide rich global context, and SDC to model fine-grained semantic variations for enhanced local representation. This synergy effectively addresses the challenge of character-level boundary ambiguity in Chinese text. Extensive experiments on four benchmark datasets, namely Resume, MSRA, Weibo, and OntoNotes, show that our approach achieves state-of-the-art or highly competitive performance, demonstrating its effectiveness and practical value in NER applications.

REFERENCES

[1] Y. Zhang and J. Yang, "Chinese ner using lattice lstm," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1554–1564.

[2] W. Liu, X. Fu, Y. Zhang, and W. Xiao, "Lexicon enhanced chinese sequence labeling using bert adapter," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2021, pp. 5847–5858.

[3] Y. Xiao, Z. Ji, and J. Li, "Dust: Dual-grained syntax-aware transformer network for chinese named entity recognition," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 12 717–12 721.

[4] Z. Ji and Y. Xiao, "Llet: Lightweight lexicon-enhanced transformer for chinese ner," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 12 677–12 681.

[5] J. He and H. Wang, "Chinese named entity recognition and word segmentation based on character," in *Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing*, 2008.

[6] Z. Liu, C. Zhu, and T. Zhao, "Chinese named entity recognition with a sequence labeling approach: based on characters, or based on words?" in *International Conference on Intelligent Computing*. Springer, 2010, pp. 634–640.

[7] H. Li, M. Hagiwara, Q. Li, and H. Ji, "Comparison of the impact of word segmentation on name tagging for chinese and japanese," in *9th International Conference on Language Resources and Evaluation, LREC 2014*. European Language Resources Association (ELRA), 2014, pp. 2532–2536.

[8] T. Gui, R. Ma, Q. Zhang, L. Zhao, Y.-G. Jiang, and X. Huang, "Cnn-based chinese ner with lexicon rethinking," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 2019.

[9] D. Sui, Y. Chen, K. Liu, J. Zhao, and S. Liu, "Leverage lexical knowledge for chinese named entity recognition via collaborative graph network," in *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, 2019, pp. 3830–3840.

[10] X. Li, H. Yan, X. Qiu, and X.-J. Huang, "Flat: Chinese ner using flat-lattice transformer," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 6836–6842.

[11] S. Zhao, C. Wang, M. Hu, T. Yan, and M. Wang, "Mcl: Multi-granularity contrastive learning framework for chinese ner," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 14 011–14 019.

[12] Z. Zhang, B. Shi, H. Zhang, H. Xu, Y. Zhang, Y. Wu, B. Dong, and Q. Zheng, "Nerco: a contrastive learning based two-stage chinese ner method," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 2023, pp. 5287–5295.

[13] P. Qi and B. Qin, "Ssmi: semantic similarity and mutual information maximization based enhancement for chinese ner," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 13 474–13 482.

[14] B. Hu, Z. Huang, M. Hu, Z. Zhang, and Y. Dou, "Adaptive threshold selective self-attention for chinese ner," in *Proceedings of the 29th International Conference on Computational Linguistics*, 2022, pp. 1823–1833.

[15] J. Liu, C. Liu, N. Li, S. Gao, M. Liu, and D. Zhu, "Lada-trans-ner: adaptive efficient transformer for chinese named entity recognition using lexicon-attention and data-augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 13 236–13 245.

[16] 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.

[17] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, J. Jiao, and Y. Liu, "Vmamba: Visual state space model," *Advances in neural information processing systems*, vol. 37, pp. 103 031–103 063, 2024.

[18] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, "Searching central difference convolutional networks for face anti-spoofing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5295–5305.

[19] G.-A. Levow, "The third international chinese language processing bakeoff: Word segmentation and named entity recognition," in *Proceedings of the Fifth SIGHAN workshop on Chinese language processing*, 2006, pp. 108–117.

[20] R. Weischedel, S. Pradhan, L. Ramshaw, J. Kaufman, M. Franchini, M. El-Bachouti, N. Xue, M. Palmer, M. Marcus, A. Taylor *et al.*, "Ontonotes release 4.0," 2011.

[21] N. Peng and M. Dredze, "Named entity recognition for chinese social media with jointly trained embeddings," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 548–554.

[22] G. R. Doddington, A. Mitchell, M. A. Przybocki, L. A. Ramshaw, S. M. Strassel, R. M. Weischedel *et al.*, "The automatic content extraction (ace) program-tasks, data, and evaluation." in *Lrec*, vol. 2, no. 1. Citeseer, 2004, pp. 837–840.

[23] Y. Zhu and G. Wang, "Can-ner: Convolutional attention network for chinese named entity recognition," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 3384–3393.

[24] Y. Gu, X. Qu, Z. Wang, Y. Zheng, B. Huai, and N. J. Yuan, "Delving deep into regularity: A simple but effective method for chinese named entity recognition," in *Findings of the Association for Computational Linguistics: NAACL 2022*, 2022, pp. 1863–1873.

[25] Y. Shen, K. Song, X. Tan, D. Li, W. Lu, and Y. Zhuang, "Diffusionner: Boundary diffusion for named entity recognition," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 3875–3890.

[26] M. Tang, Y. He, Y. Xu, H. Xu, W. Zhang, and Y. Lin, "A boundary offset prediction network for named entity recognition," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 14 834–14 846.

[27] X. Ke, X. Wu, Z. Ou, and B. Li, "Chinese named entity recognition method based on multi-feature fusion and biaffine," *Complex & Intelligent Systems*, vol. 10, no. 5, pp. 6305–6318, 2024.

[28] S. Hou, Y. Qian, J. Chen, J. Zhao, H. Lv, J. Zhang, H. Leng, and M. Ma, "Hiner: Hierarchical feature fusion for chinese named entity recognition," *Neurocomputing*, vol. 611, p. 128667, 2025.

[29] Z. Li, S. Cao, M. Zhai, N. Ding, Z. Zhang, and B. Hu, "Multi-level semantic enhancement based on self-distillation bert for chinese named entity recognition," *Neurocomputing*, vol. 586, p. 127637, 2024.

[30] L. Zhang, P. Xia, X. Ma, C. Yang, and X. Ding, "Enhanced chinese named entity recognition with multi-granularity bert adapter and efficient global pointer," *Complex & Intelligent Systems*, vol. 10, no. 3, pp. 4473–4491, 2024.

[31] Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang, "Pre-training with whole word masking for chinese bert," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3504–3514, 2021.

[32] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:53592270

[33] J. Ma and D. Yarats, "On the adequacy of untuned warmup for adaptive optimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 10, 2021, pp. 8828–8836.

[34] Y. Cai, Q. Liu, Y. Gan, R. Lin, C. Li, X. Liu, D. Luo, and J. JiayeYang, "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.

[35] 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.

[36] R. Geng, Y. Chen, R. Huang, Y. Qin, and Q. Zheng, "Planarized sentence representation for nested named entity recognition," *Information processing & management*, vol. 60, no. 4, p. 103352, 2023.