

Article

Augmented Disentanglement and Aggregation for Nested Named Entity Recognition

Jinjin Zhang ^{1,*}, Kun Zhang ¹, Chengliang Zhong ² and Ruhan A ³¹ Department of Weapon Science and Technology, Xi'an Technological University, Xi'an 710021, China² Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China³ School of Computer Science, Xianyang Normal University, Xianyang 712000, China

* Correspondence: zhangjinjin@xatu.edu.cn

Abstract

Nested named entity recognition (Nested NER) aims to identify and classify all possible span entities within a text. Existing approaches primarily rely on enumeration techniques and span-based methods to address the challenge of overlapping entities. However, these methods often overlook the structural distribution and inherent semantics of entities, making them susceptible to issues such as ambiguous start-end tokens, blurred entity boundaries, and a high degree of token overlap. In this paper, we propose a novel strategy we name **Augmented Disentanglement and Aggregation for Nested Named Entity Recognition** (ada-NER), which employs a series of augmentation strategies to extract nested entities from text. Specifically, we first reformulate the nested NER task as a problem of disentangling and aggregating the relationships between span recognition and type classification. This formulation enables the model to capture fine-grained and comprehensive contextual interactions within sentences. Furthermore, span entities are recognized through the joint modeling of hard boundary encoding and soft edge encoding, while type classification is enhanced by incorporating both intra and inter distribution relationships as well as dependency information. Finally, we introduce a well-designed fusion mechanism to obtain entity representation within a shared space. Extensive experiments on two public datasets demonstrate the effectiveness of our proposed model, which consistently outperforms competitive baseline methods.

Keywords: nested named entity recognition; span-based methods; span detection; boundary detection



Academic Editors: Ivelina Stoyanova and Svetlozara Leseva

Received: 22 May 2026

Revised: 18 June 2026

Accepted: 22 June 2026

Published: 29 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Named Entity Recognition (NER) aims to identify and classify predefined entities, such as persons, locations, organizations, and temporal expressions, from unstructured text. It is widely regarded as a fundamental task in Natural Language Processing (NLP) and plays a crucial role in various downstream applications, including sentiment analysis [1], knowledge graph construction [2], and question answering systems [3]. As a challenging branch of NER, Nested Named Entity Recognition (Nested NER) focuses on identifying and classifying all possible entity spans within a text, where entities may overlap with one another. Recent advances in neural network-based approaches have significantly improved the performance of nested NER by automatically learning rich semantic and contextual representations from text [4].

Existing methods for Nested Named Entity Recognition (Nested NER) can generally be categorized into three groups. (1) Hypergraph-based methods, which employ hypergraphs and hyperedges to represent entities at different hierarchical levels within a sentence [5,6]. (2) Sequence-to-sequence methods, which provide an end-to-end framework that autoregressively generates a sequence of entities [7,8]. (3) Span-based methods, which identify and classify all possible entity spans in a text by leveraging pretrained language models and span representations constructed from boundary tokens [9–11]. Among these approaches, span-based methods have gradually become the dominant paradigm for Nested NER due to their effectiveness and flexibility. In such methods, the performance of entity recognition largely depends on the quality of span representations. To enhance span modeling, recent studies have explored a variety of approaches, including informative span pattern modeling, graph-based structures, contrastive learning strategies, and attention mechanisms. Despite their effectiveness, most existing span-based methods still fail to explicitly capture the structural distribution characteristics and semantic dependencies among neighboring entities, as well as fine-grained token-level interactions. As a result, learned span representations are often sensitive to boundary tokens and prone to ambiguity under conditions of token overlap. As illustrated in Figure 1, different entities may produce highly similar span representations when they share identical boundary tokens or contain the same number of tokens. Moreover, subtle semantic distinctions between overlapping entities are often suppressed in the shared representation space, making it difficult for the model to accurately differentiate entity boundaries and types in complex nested scenarios.

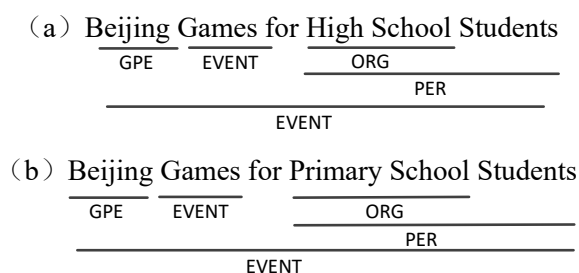


Figure 1. Example of nested entity annotation.

To address the aforementioned challenges, we propose a novel framework, Augmented Disentanglement and Aggregation for Nested Named Entity Recognition (ada-NER), which is designed to effectively handle various nested NER datasets. The core idea of our approach is to explicitly model entity structural distributions and contextual relationships among nested spans through a dynamic adaptive strategy. By jointly exploiting semantic and structural information, the proposed framework enhances the model’s ability to distinguish overlapping entities and accurately classify nested named entities. Specifically, ada-NER consists of two key components: a disentanglement module and an aggregation module. First, to address severe token overlap and complex nested structures, the disentanglement module decouples span detection from boundary relations, thereby reducing the learning complexity of nested NER. In this module, a multi-scale span detection mechanism, incorporating hard boundary encoding and soft edge encoding, is employed to identify entity spans more accurately. Meanwhile, neighboring collaborative entities and contextual dependency semantics are utilized to facilitate type representation. Subsequently, the aggregation module projects entity representation from boundary information and type representation from semantic information into a shared space. This design enables the model to flexibly extract complementary features and establish mutual dependencies between entity boundaries and entity relations, while jointly considering both structural and semantic information. Extensive experiments conducted on the ACE 2005 and GENIA

benchmark datasets demonstrate that the proposed ada-NER consistently outperforms state-of-the-art span-based methods.

In summary, the main contributions of our work are as follows:

- We propose a novel nested NER framework, namely ada-NER, which addresses the challenges of ambiguous entity boundaries and severe token overlap through a disentanglement module and an aggregation module.
- We introduce a multi-scale span detection mechanism that integrates hard boundary encoding and soft edge encoding to alleviate boundary ambiguity and improve span detection. In addition, a multi-level semantic modeling mechanism is proposed to capture neighboring entity information and contextual dependency relationships, thereby learning more discriminative type representation.
- We develop a well-designed fusion mechanism that projects span representation from boundary information and type representation from semantic information into a shared space, enabling effective interaction between structural and semantic information.
- Experimental results on two public datasets reveal that the proposed model outperforms state-of-the-art methods, and some ablation experiments also confirm the effectiveness of our model.

The rest of this paper is structured as follows. Section 2 discusses related work of nested NER. Section 3 presents the proposed ada-NER framework in detail. We conduct comprehensive experiments and present the experimental settings with the corresponding results in Section 4. Finally, we conclude the paper and discuss future works in Section 5.

2. Related Work

Nested NER aims to identify and classify entities in complex sentence structures where entities may overlap or be nested within each other. It has attracted substantial research attention, leading to a wide range of modeling approaches in recent years. In this work, we review representative studies selected from major NLP databases using keywords such as “nested named entity recognition”, “span-based NER”. We primarily focus on recent and highly cited peer-reviewed works published in top NLP conferences and journals over the past years. Existing supervised learning approaches for nested NER can be broadly categorized into hypergraph-based methods and span-based methods.

Hypergraph-based methods model the complex structural relationships among entities by representing nested entities through hypergraphs and hyperedges, thereby facilitating the recognition of overlapping and nested entity structures. Huang et al. [5] propose a hypergraph neural network to jointly learn sequence representations and hypergraph structural representations for nested entity recognition. Although this approach effectively captures hierarchical entity structures, it suffers from high computational complexity due to the large number of hyperedges that need to be constructed and processed. To alleviate this issue, Yan et al. [12] introduce the Local Hypergraph Builder Network (LHBN), which represents entity features using multiple local hypergraphs instead of a single large-scale hypergraph. Specifically, entity boundaries are first identified by boundary classifiers, and then candidate subsequences are mapped into local hypergraphs based on the predicted bidirectional boundaries. This design significantly reduces the computational burden while preserving structural information. Furthermore, Yan et al. [13] propose a higher-order hypergraph modeling framework, where nodes correspond to entities and hyperedges encode relationships among them. By explicitly modeling higher-order interactions between entities, the framework effectively alleviates error propagation and enhances the representation of complex entity structures. In addition, Zhang et al. [14] develop a nested NER framework for the Chinese financial domain by integrating span-

based representations with heterogeneous graph structures. This approach leverages both local span semantics and global structural information to improve entity recognition performance. Although hypergraph-based methods have achieved promising results in nested NER, they still face several challenges, including complex graph construction, high computational costs, and potential error propagation during the modeling process.

Span-based methods formulate nested NER as a span classification task by enumerating candidate text spans and assigning entity labels to them. Owing to their flexibility and effectiveness, span-based approaches have become the dominant paradigm for nested entity recognition. To improve span representation learning, numerous studies have explored different modeling strategies. Jana et al. [15] propose two neural network architectures, which include concatenating the nested entity multiple labels into a multilabel task with BILOU encoding and a combination of nested entities encoding and sequence-to-sequence task with BILOU encoding. Zhu and Li [16] introduce a regularization technique for boundary annotation, which redistributes entity probabilities from annotated spans to neighboring spans and alleviates the overconfidence issue of span-based models. Li et al. [17] reformulate nested NER as a segment covering problem and propose a segment enhanced span-based framework. Yan et al. [18] employ convolutional neural networks to capture spatial relationships with score matrices for nested entity recognition. Yu et al. [19] incorporate boundary regression techniques by jointly modeling entity positions and semantic features for full-span NER. Liu et al. [20] propose a two-stage contrastive learning framework that enhances representation learning at both the word embedding stage and the candidate span classification stage. Ning et al. [21] introduce an entity-aware text generation framework and a region-of-interest vertex-based table filling scheme to improve boundary awareness and nested entity recognition. Sun et al. [22] propose the Global Span Semantic Dependency Awareness and Filtering Network, which incorporates both global span dependency awareness and local span dependency enhancement mechanisms to enrich semantic dependency information during entity classification. Mao et al. [23] develop a flat tree-based transformer framework that introduces tree structures into both candidate span generation and span selection stages. Although these methods have achieved remarkable performance improvements, they still suffer from several limitations. Most existing span-based approaches primarily focus on enhancing span representations while overlooking the structural relationships among neighboring entities and token-level interactions. Consequently, they remain vulnerable to boundary ambiguity, semantic inconsistency, and severe token overlap, especially when nested entities share similar boundary tokens or highly overlapping contexts.

3. The Proposed Algorithm

The overall architecture of the **Augmented task Disentanglement and Aggregation for Nested Named Entity Recognition** is shown in Figure 2. For simplicity, we refer to the proposed framework as **ada-NER** throughout this paper. The framework consists of two main components: a disentanglement module and an aggregation module. The disentanglement module is designed to decouple entity span detection from dependency relations. Specifically, it employs a multi-scale span detection mechanism and a multi-level semantic modeling mechanism to learn discriminative span representation and informative type representation. The aggregation module further integrates the outputs of the two subtasks by modeling the interactions between span representation from boundary and type representation from semantic. To this end, a well-designed fusion mechanism is introduced to project span representation from boundary information and type representation from semantic information into a shared entity space, thereby obtaining more comprehensive entity representation.

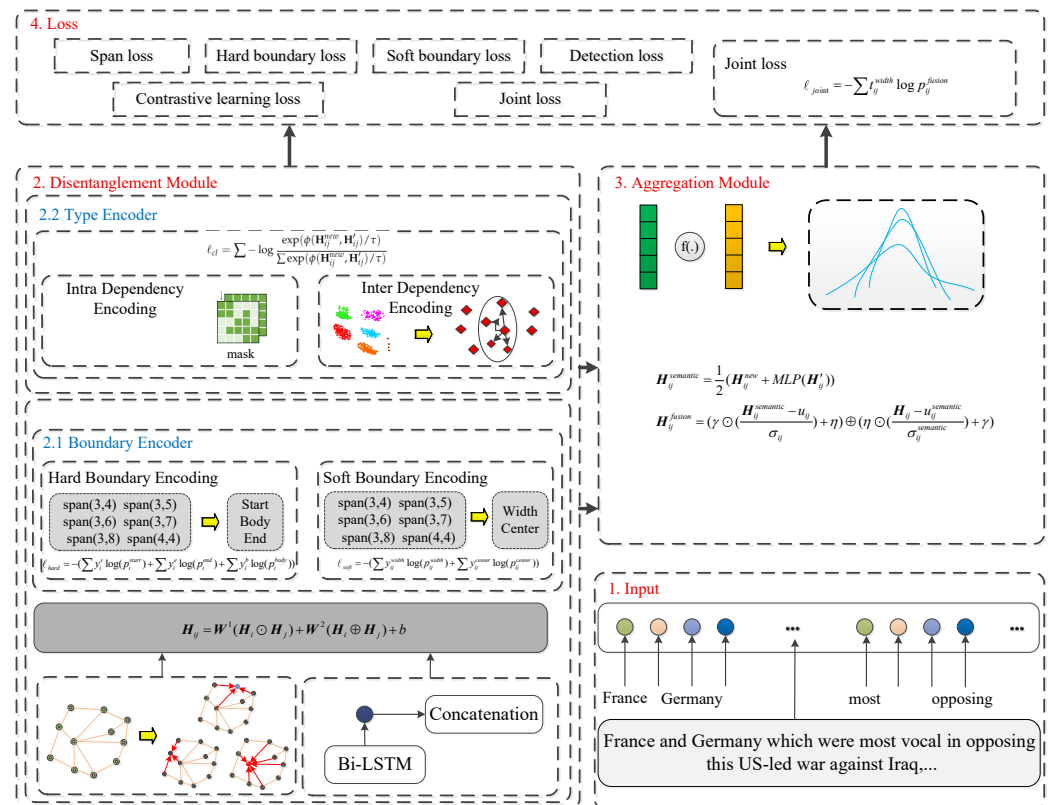


Figure 2. Overview of the proposed ada-NER framework, which consists of input, disentanglement module, aggregation module, and loss for nested entity recognition.

3.1. Task Formulation

Following the span-based model works, the input information for ada-NER includes a sentence as $X = [x_1, x_2, \dots, x_n]$ and entity types $T = [t_1, t_2, \dots, t_k]$, where n is the number of tokens in the sentence. The task of nested NER is to extract all entities and assign corresponding possible entity labels. Generally, the entity information is represented as (s_i, e_i, t_i) , where s_i is the start position, e_i is the end position, and t_i is the type information, respectively.

3.2. Disentanglement Module

3.2.1. Boundary Encoder

Accurately identifying entity spans remains challenging due to boundary ambiguity and severe token overlap among nested entities for nested NER. Since the quality of span representations directly affects entity recognition performance, effectively capturing precise entity boundaries is crucial for span-based models. To address these challenges, we propose a multi-scale span detection mechanism that integrates hard boundary encoding and soft edge encoding to improve entity span extraction. Specifically, the hard boundary encoding strategy explicitly models the start, end, and internal body information of an entity span, enabling the generation of more discriminative candidate span representations. In contrast, the soft edge encoding strategy introduces geometric boundary information by jointly modeling the center position and span width of candidate entities. This design captures the uncertainty associated with ambiguous boundaries and allows a token position to belong to multiple entity spans with different probabilities. By combining hard boundary constraints with soft boundary awareness, the proposed mechanism dynamically adapts to annotation ambiguities and boundary uncertainties. As a result, it establishes a robust probabilistic foundation for span detection and enhances the model's ability to capture complex entity structures under fuzzy boundary conditions.

To obtain contextualized token representation, we first employ the pre-trained BERT model $f(\cdot)$ to encode each token into a dense embedding. Given an input sentence $X = [x_1, x_2, \dots, x_n]$, the embedding of the i -th token is computed in Equation (1):

$$\mathbf{x}_i = f(x_i) \quad (1)$$

Then the embedding of sentence is denoted as Equation (2):

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \quad (2)$$

To capture both sequential contextual information and structural dependency information, we employ a Bidirectional Long Short-Term Memory (BiLSTM) network and a Graph Convolutional Network (GCN) as the sequential encoder and structural encoder to map the tokens and obtain the representation vectors for the start and end tokens of a span. The details are shown in Equation (3):

$$\begin{aligned} \mathbf{H}_i^{tim} &= LSTM(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \\ \mathbf{H}_i^{tem} &= \sigma(\mathbf{W}_i \mathbf{x}_i + \sum_{j \in N_i} \frac{1}{\sqrt{|N_i|}} (\mathbf{W}_i \mathbf{x}_i + \mathbf{W}_j \mathbf{x}_j)) \\ \mathbf{H}_j^{tim} &= LSTM(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \\ \mathbf{H}_j^{tem} &= \sigma(\mathbf{W}_j \mathbf{x}_j + \sum_{k \in N_j} \frac{1}{\sqrt{|N_j|}} (\mathbf{W}_j \mathbf{x}_j + \mathbf{W}_k \mathbf{x}_k)) \\ \mathbf{H}_i &= (\mathbf{H}_i^{tim})^T \mathbf{W}_i^t \mathbf{H}_i^{tem} \\ \mathbf{H}_j &= (\mathbf{H}_j^{tim})^T \mathbf{W}_j^t \mathbf{H}_j^{tem} \end{aligned} \quad (3)$$

To effectively integrate sequential and structural information, we take feature interaction and fusion mechanism to build span representation as shown in Equation (4):

$$\mathbf{H}_{ij} = \mathbf{W}^1(\mathbf{H}_i \odot \mathbf{H}_j) + \mathbf{W}^2(\mathbf{H}_i \oplus \mathbf{H}_j) + \mathbf{b} \quad (4)$$

where $\mathbf{W}^1$, $\mathbf{W}^2$ and $\mathbf{b}$ are learnable matrices, $LSTM(\cdot)$ is the function of LSTM model, $\odot$ is the element-wise product, $\oplus$ is element-wise concatenation.

A sigmoid transformation is then employed to predict the likelihood of ij -th tokens being a entity span. The details are shown in Equation (5):

$$\ell_{span} = - \sum y_{ij} \log(\text{sigmoid}(\mathbf{W}_{ij} \mathbf{H}_{ij} + b_{ij})) \quad (5)$$

Hard Boundary Encoding. To improve boundary identification under ambiguous span boundaries, we introduce a hard boundary encoding strategy as an auxiliary span detection task. Specifically, each token is classified as a start token, end token, or body token of an entity span. Given token representation $\mathbf{H}_i$, three independent multilayer perceptrons (MLPs) are employed to estimate the corresponding probabilities. The details are shown in Equation (6):

$$\begin{aligned} p_i^{start} &= \text{Sigmoid}(\text{MLP}_{start}(\mathbf{H}_i)) \\ p_i^{body} &= \text{Sigmoid}(\text{MLP}_{body}(\mathbf{H}_i)) \\ p_i^{end} &= \text{Sigmoid}(\text{MLP}_{end}(\mathbf{H}_i)) \end{aligned} \quad (6)$$

The hard boundary loss is formulated in Equation (7):

$$\ell_{hard} = -(\sum y_i^s \log(p_i^{start}) + \sum y_i^e \log(p_i^{end}) + \sum y_i^b \log(p_i^{body})) \quad (7)$$

This auxiliary task explicitly models entity boundary information and facilitates the generation of more discriminative span representations.

Soft Boundary Encoding. Although hard boundary encoding captures explicit boundary signals, it cannot adequately model boundary uncertainty caused by overlapping entities. To address this issue, we further introduce a soft edge encoding mechanism to characterize geometric relationships among candidate spans. For each candidate span, its center position and span width are extracted from the span representation $\mathbf{H}_{ij}$. The geometric transformation is defined in Equation (8):

$$GT = (W_{width}^1 \cdot \mathbf{H}_{ij}(Width) \cdot W_{width}^2) \oplus (W_{center}^1 \cdot \mathbf{H}_{ij}(Center) \cdot W_{center}^2) \quad (8)$$

Inspired by positional encoding in Transformers, sine and cosine functions are employed to project the geometric features into a high-dimensional representation. The details are shown in Equation (9):

$$\begin{aligned} GT_{2i} &= \sin\left(\frac{GT}{10000^{2i/d}}\right) \\ GT_{2i+1} &= \cos\left(\frac{GT}{10000^{2i+1/d}}\right) \end{aligned} \quad (9)$$

where d denotes the dimensionality of geometric features. Subsequently, the geometric representation is fused with the span representation to obtain an enhanced representation. The details are shown in Equation (10):

$$GT_{emb} = \sigma(W_{emb}^1 \cdot (GT_{2i} \oplus GT_{2i+1}) + W_{emb}^1 \cdot (GT_{2i} \otimes GT_{2i+1}) + b) \quad (10)$$

The interactive relations between spans are based on their position information and inherent contextual information. Accordingly, the span representation with interactive relations between spans is denoted in Equation (11):

$$\mathbf{H}_{ij}^{soft} = \mathbf{W}_{soft}(\mathbf{H}_{ij} \oplus GT_{emb}) + b_{soft} \quad (11)$$

The enhanced span representation is further used for span prediction and geometric auxiliary supervision. Specifically, two MLP classifiers are employed to predict span center position and span width. The details are shown in Equation (12):

$$\begin{aligned} p_{ij}^{center} &= \text{Sigmoid}(\text{MLP}_{center}(\mathbf{H}_{ij}^{soft})) \\ p_{ij}^{width} &= \text{Sigmoid}(\text{MLP}_{width}(\mathbf{H}_{ij}^{soft})) \end{aligned} \quad (12)$$

The corresponding soft edge loss is defined as Equation (13):

$$\ell_{soft} = -(\sum y_{ij}^{width} \log(p_{ij}^{width}) + \sum y_{ij}^{center} \log(p_{ij}^{center})) \quad (13)$$

Finally, the overall objective of the multi-scale span detection mechanism is formulated as Equation (14):

$$\ell_{detection} = \lambda \ell_{span} + (1 - \lambda)(\ell_{hard} + \ell_{soft}) \quad (14)$$

3.2.2. Type Encoder

To improve the performance of the classification task, we propose a multi-level semantic modeling mechanism that incorporates intra-dependency relationship encoding and inter-dependency relationship encoding. The intra-dependency encoding aims to capture fine-grained contextual dependencies within each entity by applying a masking strategy, while the inter-dependency encoding leverages collaborative learning to extract global semantic information in the representation space. This mechanism effectively addresses

severe token overlap and complex entity structures, thereby further improving model generalization capability.

Intra-dependency Encoding. Given a sentence $S = [x_1, x_2, \dots, x_n]$, we first employ a MLP to encode the span $\mathbf{H}_{ij} = [\mathbf{H}_i, \mathbf{H}_i + 1, \dots, \mathbf{H}_j]$, thereby obtaining the initial type representation $\mathbf{H}_{ij}$. Then, we construct a masked version of the sentence $S' = [x'_1, x'_2, \dots, x'_n]$, where each token is either kept unchanged or replaced with a special token such as [MASK] according to a masking probability r . Based on the masked sentence, we obtain entity representation $\mathbf{H}'_{ij}$ using an indicator-based mechanism inspired by the transformer architecture. This masking strategy enables the model to capture contextual dependencies within each span by encouraging it to infer missing information from surrounding tokens.

Inter-Dependency Encoding. To generate more contextualized type representation, we introduce a collaborative learning mechanism in span-based models, which incorporates the influence of neighboring entities and models the semantic distribution of entity types in a shared representation space. Specifically, we employ prototypes of each entity type to capture global semantic regularities of different categories. Given a target entity representation $\mathbf{H}_{ij}^{type}$, we first retrieve its K-nearest neighbors in the representation space. Then, the similarity between the target entity and the prototype ϕ_c of entity type c is computed using the dot product. The details are shown in Equation (15):

$$sim_{ij}^c = \mathbf{H}_{ij}^{type} \cdot \phi_c \quad (15)$$

Next, a similarity threshold is defined based on the average similarity score over the K neighbors. The details are shown in Equation (16):

$$sim^T = \frac{1}{K} \sum sim_{ij}^c \quad (16)$$

Neighbors with similarity scores higher than this threshold are selected as valid neighbors. The final collaborative span representation is obtained by averaging the representations of the selected neighbors. The details are shown in Equation (17):

$$\mathbf{H}_{ij}^{new} = \frac{1}{M} \sum \mathbf{H}_{ij}^{type}, \text{ where } sim_{ij}^c > sim^T \quad (17)$$

To further enhance entity type relationships within textual structures, we introduce a contrastive learning-based span representation encoding strategy, which aligns semantic information between contextual dependency and collaborative information. This mechanism is sensitive to subtle entity differences and helps the model adapt to annotation ambiguity and boundary uncertainty. Specifically, we take the collaborative representation $\mathbf{H}_{ij}^{new}$ and the masked span representation $\mathbf{H}'_{ij}$ as inputs to the contrastive learning framework. Positive pairs are formed by representations of the same entity span, while negative pairs are constructed from different entity spans. The objective is formulated as Equation (18):

$$\ell_{cl} = \sum -\log \frac{\exp(\phi(\mathbf{H}_{ij}^{new}, \mathbf{H}'_{ij})/\tau)}{\sum \exp(\phi(\mathbf{H}_{ij}^{new}, \mathbf{H}'_{ij})/\tau)} \quad (18)$$

where $\phi(\cdot)$ denotes the cosine similarity function, and τ is a hyper-parameter that is empirically set to 0.5 in the experiments.

3.3. Aggregation Module

To obtain a more discriminating span representation, we design a well-designed fusion mechanism to integrate span representation from the boundary encoder and semantic encoder. the fusion mechanism adaptively learns the relative importance of boundary-

level and semantic-level representations under different contextual conditions through conditional interactions, which significantly enhances the model's capability to represent complex nested entity structures. The resulting fused representation $\mathbf{H}_{ij}^{fusion}$ is formulated as Equation (19):

$$\begin{aligned}
 H_{ij}^{semantic} &= \frac{1}{2}(H_{ij}^{new} + MLP(H'_{ij})) \\
 \mathbf{H}_{ij}^{fusion} &= (\gamma \odot (\frac{\mathbf{H}_{ij}^{semantic} - \mu_{ij}}{\sigma_{ij}}) + \eta) \oplus (\eta \odot (\frac{\mathbf{H}_{ij} - \mu_{ij}^{semantic}}{\sigma_{ij}^{semantic}}) + \gamma) \\
 \gamma &= \mathbf{W}_{\gamma} \mathbf{H}_{ij} + b_{\gamma} \\
 \eta &= \mathbf{W}_{\eta} \mathbf{H}_{ij}^{semantic} + b_{\eta} \\
 \mu_{ij} &= \frac{1}{d} \sum_{j=1}^d \mathbf{H}_{ij} \\
 \mu_{ij}^{semantic} &= \frac{1}{d} \sum_{j=1}^d \mathbf{H}_{ij}^{semantic} \\
 \sigma_{ij} &= \sqrt{\frac{1}{d} \sum_{j=1}^d (\mathbf{H}_{ij} - \mu_{ij})^2} \\
 \sigma_{ij}^{semantic} &= \sqrt{\frac{1}{d} \sum_{j=1}^d (\mathbf{H}_{ij}^{semantic} - \mu_{ij}^{semantic})^2}
 \end{aligned} \tag{19}$$

As a result, the prediction logits of the aggregation module are calculated by Equation (20):

$$p_{ij}^{fusion} = \text{Sigmoid}(MLP(\mathbf{H}_{ij}^{fusion})) \tag{20}$$

The cross-entropy loss function for the aggregation module is defined in Equation (21):

$$\ell_{joint} = - \sum t_{ij} \log p_{ij}^{fusion} \tag{21}$$

3.4. Training Objectives

The total objective for our framework is defined in Equation (22):

$$\ell_{total} = \eta \ell_{joint} + (1 - \eta)(\kappa \ell_{detection} + (1 - \kappa) \ell_{cl}) \tag{22}$$

where the total loss is minimized during training. η and κ are the hyper-parameters to balance different learning objectives.

For reproducibility, we provide a detailed description of the training and inference stage.

During training, the model is optimized using a joint learning objective that consists of a span detection loss and a type classification loss. Specifically, candidate spans are first generated using a maximum span length constraint with exhaustive enumeration over the input sequence. Each candidate span is then assigned a span representation through the boundary encoder, which incorporates both hard boundary encoding and soft boundary encoding. Simultaneously, type representation are constructed using intra- and inter-dependency encoding. The final entity representation is obtained via the well-designed fusion mechanism, which adaptively integrates boundary and semantic information. The overall training objective is defined as a weighted sum of detection and classification losses, enabling the model to jointly learn span localization and entity type prediction.

During inference, the model follows a pipeline consistent with the training process but without parameter updates. Given an input sentence, candidate spans are first generated using the same maximum span length constraint and exhaustive enumeration strategy as in training. For each candidate span, the boundary encoder produces its span representation by integrating hard boundary encoding and soft boundary encoding, while the intra- and inter-dependency encoding modules generate the corresponding semantic type representation. These two representations are then combined through the learned fusion mechanism to obtain the final entity representation for each span. Subsequently, the classification module computes the probability distribution over entity types for each candidate span. During decoding, spans are ranked according to their predicted classification confidence scores. A threshold-based filtering strategy is first applied to remove low-confidence spans. For the remaining spans, final entity predictions are selected based on the highest scoring type assignments, while preserving valid overlapping structures to support nested entity recognition.

For implementation, we use a standard token-level encoder to obtain contextualized embeddings as input representations. The span length is limited to a maximum threshold to reduce computational complexity. All parameters are trained end-to-end using backpropagation with Adam optimization. All hyperparameters are selected based on validation performance.

4. Experiments and Analysis

4.1. Datasets

We evaluate the performance of ada-NER on two publicly available benchmark datasets, namely ACE 2005 and GENIA.

(1) The ACE 2005 dataset (<https://catalog.ldc.upenn.edu/LDC2006T06>) is compiled from newswires, broadcasts, and weblogs, which places these data in the general domain. The dataset encompasses seven entity categories. In addition, the dataset contains seven types of specified entities. A key difference lies in the number of coarse-grained inter-entity relationships defined: ACE 2005 has six types, compared to the seven in ACE 2004. According to these competing methods, the classification criteria for training, development and test sets are 8, 1, 1 respectively. The statistics of the ACE 2005 dataset are summarized in Table 1.

Table 1. Statistics of ACE 2005 dataset.

	Train	Dev	Test	Overall	Nested
Documents	528	30	40	598	–
Sentences	14,668	875	713	16,256	–
PER	25,899	1731	1402	29,032	4043
LOC	1139	98	71	1308	412
ORG	5333	593	649	6575	1746
GPE	7163	684	1179	9026	3203
VEH	905	76	24	1005	207
WEA	819	75	89	983	206
FAC	1481	124	148	1753	301
Overall	42,739	3381	3562	49,682	10,118

(2) The Genia dataset (<https://pubmed.ncbi.nlm.nih.gov/12855455/>) is collected by the GENIA project, and is used in the field of molecular biology. The dataset consists of 1999 MEDLINE abstracts and 5 types of specified entities. According to this method, the

classification criteria for training, development and test sets are 8.1, 0.9, 1, respectively. The statistics of the Genia dataset are summarized in Table 2.

Table 2. Statistics of Genia dataset.

	Train	Dev	Test	Overall	Nested
Documents	1599	189	212	2000	–
Sentences	15,023	1669	1854	18,546	–
Split percentage	81%	9%	10%	100%	–
DNA	7650	1026	1257	9933	1744
RNA	692	132	109	933	407
Protein	28,728	2303	3066	34,097	1902
Cell Line	3027	325	438	3790	347
Cell Type	5832	551	604	6987	389
Overall	45,929	4337	5474	55,740	4789

4.2. Evaluation

We use precision (P), recall (R), and F1-score (F1) as the evaluation metrics for models in this work. In the field of NER, there is an inherent trade-off between precision and recall, and higher values indicate better performance. However, a higher precision often leads to a lower recall in practice, and vice versa. The F-score is a comprehensive evaluation metric that incorporates both precision and recall, addressing the trade-off between precision and recall. The formulations are formulated in Equation (23):

$$\begin{aligned}
 P &= \frac{TP}{TP + FP} * 100\% \\
 R &= \frac{TP}{TP + FN} * 100\% \\
 F1 &= \frac{2P \cdot R}{P + R} * 100\%
 \end{aligned}
 \tag{23}$$

where TP is the number of correctly identified entities, FP is the number of error-identified entities, and FN is the number of unrecognized entities. Therefore, the evaluation of P indicates the ratio of correctly predicted positives to total predicted positive observations, the evaluation of R indicates the ratio of correctly predicted positives to the total actual positives, and the evaluation of F1 combines the precision and recall into one score.

4.3. Baselines

We take several state-of-the-art models to validate the performance of our proposed model. For the baselines, all experiments are conducted under identical settings and reported following common practice in prior NER literature, where statistical significance tests and nested-only subset evaluations are not typically required.

BiLSTM-CRF [24] takes BiLSTM to encode the sequence information and uses CRF as a decoder to capture tag information.

DTDM [25] takes dynamic convolutional kernel and a dynamic scaling factor computation method to extract both local and global contextual semantic feature information for a nested NER task. we re-implement this method under a unified English setting, and replace the original embeddings with the same BERT encoder to ensure a fair comparison under identical experimental conditions.

DAE-NER [26] uses a dual-channel attention-enhancement module to build semantic weights, which can alleviate the interference of contextual differences in semantics. we re-implement this method under a unified English setting, and replace the original

embeddings with the same BERT encoder to ensure a fair comparison under identical experimental conditions.

TSBECL [20] uses a flat NER module and contrastive learning for the identification and classification stages.

GSSDAF [22] introduces semantic information among global spans by using a global span dependency awareness module and local span dependency enhancement module for the identifier stage and classification stage.

Flat Tree-based Transformer [23] uses the constituency parsing tree to assist nested NER.

4.4. Model Comparisons

To evaluate the effectiveness and robustness of our proposed model, we compare ada-NER with several baseline methods on the ACE 2005 and GENIA datasets. The experimental results are reported in Table 3. From the results, we can make the following observations.

Table 3. Performance comparison of ada-NER with baseline methods on two datasets.

Models	ACE 2005			Genia		
	P	R	F1	P	R	F1
BiLSTM-CRF	74.54	70.56	72.50	75.71	72.03	72.98
DTDM	85.47	83.68	85.35	76.24	77.74	77.13
DAE-NER	86.59	86.15	86.49	79.32	79.89	79.11
TSBECL	87.84	87.10	87.52	82.51	82.97	82.76
GSSDAF	87.13	87.65	87.71	81.94	80.55	81.08
Flat Tree-based Transformer	89.01	88.22	88.36	84.18	83.41	83.97
ada-NER	89.10	89.83	89.54	84.66	84.46	84.23

First, DTDM consistently outperforms BiLSTM-CRF on both datasets, demonstrating the effectiveness of jointly modeling intra- and inter-dependency information. This is because DTDM captures both intra- and inter-semantic features, which highlights the importance of leveraging multi-level contextual information for named entity recognition. This further indicates that nested NER models heavily rely on both intra- and inter-dependency context to learn rich type representation.

Second, GSSDAF achieves better performance than several semantic-enhanced models such as DTDM and DAE-NER. This improvement can be attributed to its stronger span-based modeling capability, which better captures span-level semantic dependencies.

Finally, our proposed ada-NER achieves the best F1-scores on both the ACE 2005 and GENIA datasets for the nested NER task. Specifically, ada-NER achieves an F1-score of 89.54% on ACE 2005, outperforming the best baseline Flat Tree-based Transformer by 1.18%, and 84.23% on GENIA, surpassing Flat Tree-based Transformer by 0.26%. Compared with existing span-based methods, ada-NER consistently demonstrates superior performance, indicating its stronger ability in identifying and classifying all possible entity spans in text.

The performance gains can be attributed to the following causes. First, most existing methods rely primarily on either boundary information or shallow semantic features, while neglecting intra-entity and inter-entity dependency relationships among spans. In contrast, our model introduces a multi-scale span detection mechanism and a multi-level semantic modeling mechanism to capture richer structural and semantic information, which is particularly beneficial for nested entity recognition. Second, although methods such as TSBECL, GSSDAF and Flat Tree-based Transformer consider both recognition and classification stages, they do not explicitly model the disentanglement between these two tasks in a

unified framework. In contrast, our disentanglement-and-aggregation architecture explicitly separates span detection and type classification while enabling effective interaction between them. Overall, the experimental results validate the importance of modeling boundary information, dependency relationships, and a disentanglement-and-aggregation architecture for improving performance on nested NER tasks.

4.5. Influence of Disentanglement Module and Aggregation Module

To investigate the effectiveness of the proposed ada-NER framework for nested NER, we conduct experiments at three different settings: the disentanglement structure, the aggregation structure, and the joint disentanglement module and aggregation module. The corresponding results are reported in Table 4. Specifically, the disentanglement module follows a span-based paradigm by performing span detection and classification for nested NER. The aggregation module is inspired by sequence tagging models and directly utilizes pre-trained embeddings as span representation for entity recognition. The joint disentanglement module and aggregation module form the complete ada-NER framework. As shown in Table 4, the disentanglement module consistently outperforms the aggregation module, indicating that the former is more susceptible to error propagation due to its multi-scale boundary recognition and type classification. When further integrating the aggregation structure into the disentanglement module to form the full ada-NER model, we observe a substantial performance improvement, achieving gains of 1.63% on ACE 2005 and 3.16% on GENIA in terms of F1-score. This result demonstrates that although span-based methods are effective in capturing fine-grained span representation and classification, their performance is still limited by their sensitivity to the integration of boundary information and semantic information in nested NER scenarios. Moreover, these findings further validate the effectiveness of the proposed coupling of the disentanglement module and aggregation module in enhancing entity representation learning and improving nested NER performance.

Table 4. Performance comparison of disentanglement structure and aggregation structure on two datasets.

Models	ACE 2005			Genia		
	P	R	F1	P	R	F1
disentanglement module	87.98	87.23	87.91	81.22	80.88	81.07
aggregation module	86.24	87.12	86.47	78.95	79.81	79.36
ada-NER	89.10	89.83	89.54	84.66	84.46	84.23

4.6. Ablation Studies

In this section, we investigate the effectiveness of different components in the multi-scale span detection mechanism and multi-level semantic modeling mechanism within the disentanglement module. Specifically, we denote the hard-soft boundary encoding in the multi-scale span detection module as SSE, the intra-dependency encoding in the multi-level semantic modeling mechanism as IDR, and the inter-dependency encoding as EDR. We remove each component individually and evaluate the model on the ACE 2005 and GENIA datasets. The experimental results are reported in Table 5, where mSSE, mIDR, and mEDR represent the variants of ada-NER without SSE, IDR, and EDR, respectively.

First, removing SSE leads to a significant performance drop, with F1-scores decreasing by 2.28% on ACE 2005 and 4.11% on GENIA. This indicates that SSE plays a crucial role in accurately modeling entity boundaries. By incorporating positional information and sequential contextual cues, SSE enriches span representations and explicitly encodes start, end, and boundary-aware features, thereby effectively improving span boundary identification.

Second, when IDR is removed, the performance in terms of F1-score decreases by 0.44% and 0.22% on ACE 2005 and GENIA, respectively. This demonstrates that modeling intra-entity contextual dependencies via the masking strategy is beneficial for improving type representation within semantic information. Similarly, removing EDR results in performance drops of 0.39% and 0.13%, respectively, indicating that inter-entity collaborative relationships also contribute positively to model performance by capturing contextual semantic structure among entity types. Furthermore, we observe that the performance drop caused by removing IDR is slightly larger than that of EDR, suggesting that intra-entity dependency modeling is more critical than inter-entity dependency modeling for nested NER. Compared with both mIDR and mEDR, mSSE exhibits a more substantial degradation, highlighting that boundary modeling plays the most important role among the three components.

Finally, the full ada-NER model consistently achieves the best performance across all settings, outperforming mSSE, mIDR, and mEDR. These results demonstrate the effectiveness of jointly modeling multi-scale span detection module and multi-level semantic modeling mechanism, and confirm their complementary contributions to the overall performance of nested NER.

Table 5. Performance comparison of ablation studies.

Models	ACE 2005			Genia		
	P	R	F1	P	R	F1
mSSE	87.81	87.15	87.26	81.06	80.73	80.12
mIDR	88.31	89.69	89.10	83.83	84.11	84.01
mEDR	88.42	89.77	89.15	84.04	84.19	84.10
ada-NER	89.10	89.83	89.54	84.66	84.46	84.23

4.7. Influence of Fusion Mechanism

To assess the effectiveness of the well-designed fusion mechanism in the aggregation module, we further evaluate the performance of the fusion module with the concatenation, addition, and fusion gate strategies for feature fusion, and the experimental results are shown in Table 6. As shown in the table, compared with the proposed conditional normalization fusion strategy, the F1-scores obtained using the concatenation, addition, and fusion-gate strategies decrease by 2.09%, 2.12%, and 1.53% on the ACE 2005 dataset. Similarly, on the Genia dataset, the corresponding F1-score drops are 2.24%, 2.62%, and 2.07%. These results demonstrate the superiority of the conditional normalization fusion strategy. The main reason is that this strategy effectively captures fine-grained and complementary interactions among span representation and type representation. By dynamically integrating information from multiple sources, it enhances the model's ability to learn more discriminating entity representation and better understand entity semantics, thereby improving overall NER performance.

Table 6. Performance comparison of fusion mechanism with F1 scores.

Models	ACE 2005	Genia
add	87.42	81.61
concatenation	87.45	81.99
fusion gate	88.01	82.16
ada-NER	89.54	84.23

4.8. Case Study

To further demonstrate the superiority of the proposed ada-NER model, we conduct a case study on a subset of samples selected from the ACE 2005 dataset. Specifically, we compare the nested NER results produced by TSBECL and our ada-NER model. The objective of this case study is to evaluate the models' ability to correctly identify entity boundaries and predict entity types. In Table 7, green, blue, and red labels represent correctly recognized entities, missed entities, and incorrectly predicted entity types, respectively. As shown in the first sentence, TSBECL fails to effectively capture span distribution patterns and comprehensive semantic dependencies, resulting in missed entity recognition and incorrect entity type predictions. In contrast, the proposed ada-NER model successfully identifies all entities and accurately predicts their corresponding types, owing to its effective exploitation of both boundary information and semantic information. In the second sentence, TSBECL also exhibits limitations in handling complex nested entity structures, leading to inaccurate predictions for several nested entities. However, our model demonstrates robust and superior performance by accurately recognizing all nested entities and assigning the correct entity types. These case study results further highlight the importance of modeling span distribution and comprehensive semantic dependencies in nested NER. Moreover, they provide strong evidence for the effectiveness of the proposed ada-NER model in accurately recognizing nested entities and capturing complex semantic relationships.

Table 7. Examples from ACE 2005 dataset.

No.	TSBECL	ada-NER
1	[France] _{GPE} and [Germany] _{GPE} [which] _{GPE} were most vocal in opposing this [US] _{GPE} -led war against [Iraq] _{GPE} , have been insisting that the [United Nations] _{GPE} should play a central role in the phase of post-war [Iraq] _{GPE} .	[France] _{GPE} and [Germany] _{GPE} which were most vocal in opposing this [US] _{GPE} -led war against [Iraq] _{GPE} , have been insisting that the [United Nations] _{ORG} should play a central role in the phase of post-war [Iraq] _{GPE} .
2	During an interview, [[[[Washington] _{GPE} Post] _{ORG} reporter] _{PER} John Miles] _{PER} revealed that the [[[[European] _{GPE} Union's] _{ORG} recent policy] _{WEA} has been criticized by senior [[[[White House] _{FAC} official] _{PER} Robert Chan] _{PER} , who believes it could harm [trans-[Atlantic] _{LOC} trade] _{GPE} relations.	During an interview, [[[[Washington] _{GPE} Post] _{ORG} reporter] _{PER} John Miles] _{PER} revealed that the [[[[European] _{GPE} Union] _{ORG} 's recent policy] _{WEA} has been criticized by senior [[[[White House] _{FAC} official] _{PER} Robert Chan] _{PER} , who believes it could harm [trans-[Atlantic] _{LOC} trade] _{GPE} relations.

5. Discussion

To address the challenges of entity boundary ambiguity and a high degree of token overlap in nested named entity recognition, we propose an Augmented Task Disentanglement and Aggregation framework that jointly integrates a disentanglement module and an aggregation module. The experimental results demonstrate consistent improvements across benchmark datasets, indicating that explicitly modeling boundary information and semantic dependencies can effectively enhance entity representation robustness and improve recognition accuracy. In particular, the incorporation of a multi-scale span detection mechanism helps alleviate boundary uncertainty in complex nested structures, while the multi-level semantic modeling mechanism further strengthens intra- and inter-dependency information, thereby reducing ambiguity in overlapping entities. Although the proposed model achieves competitive performance compared with strong baselines, a limitation still remains in that the current framework relies on predefined span enumeration strategies, which may introduce additional computational overhead and limit scalability when dealing with extremely long sequences or densely distributed entity structures. Future work can

explore more efficient span generation strategies, such as adaptive pruning mechanisms or reinforcement learning-based span selection, to improve efficiency while maintaining coverage. Overall, this study provides an effective solution for modeling multi-scale boundary information and hierarchical semantic interactions in nested NER, offering a solid foundation for more accurate, robust, and generalizable entity recognition systems.

6. Conclusions

In this paper, we propose an Augmented Task Disentanglement and Aggregation framework, termed ada-NER, which jointly leverages disentanglement and aggregation mechanisms to learn more expressive and discriminative entity representation. Specifically, the proposed framework first introduces a disentanglement module that decouples span detection and type classification tasks. For span detection, a multi-scale span detection mechanism is designed, which incorporates both hard boundary encoding and soft boundary encoding to capture comprehensive boundary-aware information. For type classification, a multi-level semantic modeling strategy is developed to capture both intra- and inter-dependencies, thereby enriching contextual semantic representations. These representations are further integrated through a well-designed fusion mechanism that effectively combines boundary-aware and semantic information in a unified framework. Finally, the outputs of the disentanglement and aggregation modules are jointly optimized to obtain more robust and informative entity representations. Extensive experiments on two public benchmark datasets demonstrate that the proposed ada-NER model consistently outperforms strong baseline methods, validating the effectiveness of the proposed framework for nested NER. In the future, we plan to explore more efficient span generation strategies to reduce computational overhead while preserving coverage and performance.

Author Contributions: Conceptualization, J.Z.; methodology, J.Z.; software, K.Z.; validation, K.Z., C.Z. and J.Z.; formal analysis, R.A.; investigation, J.Z. and R.A.; resources, K.Z.; data curation, K.Z. and R.A.; writing—original draft preparation, J.Z. and K.Z.; writing—review and editing, J.Z. and K.Z.; visualization, J.Z. and K.Z.; supervision, J.Z. and C.Z.; project administration, C.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the support of the Shaanxi Fundamental Science Research Project for Mathematics and Physics (Grant No. 23JSQ045) and Scientific Research Program Funded by Shaanxi Provincial Education Commission (Program No. 24JK0494).

Data Availability Statement: The original contributions presented in this study are included in the article.

Acknowledgments: The authors want to thank all the students who participated in the preparation process of this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang, L.; Wang, Z.Z.; Li, Z.Y.; Na, J.C.; Yu, J.F. An empirical study of Multimodal Entity-Based Sentiment Analysis with ChatGPT: Improving in-context learning via entity-aware contrastive learning. *Inf. Process. Manag.* **2024**, *61*, 103724. [\[CrossRef\]](#)
2. Huang, X.Y.; Cheng, L.L.; Deng, J.F.; Wang, T. Binocular attention-based stacked BiLSTM NER model for Supply chain management event knowledge graph construction. In *ICMLC '23: Proceedings of the 2023 15th International Conference on Machine Learning and Computing*; Association for Computing Machinery: New York, NY, USA, 2023; pp. 40–46. [\[CrossRef\]](#)
3. Arora, J.; Park, X.J. Split-NER: Named Entity Recognition via Two Question-Answering-based Classifications. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, ON, Canada, 9–14 July 2023; pp. 416–426. [\[CrossRef\]](#)
4. Geng, R.S.; Chen, Y.P.; Huang, R.Z.; Qin, Y.B.; Zheng, Q.H. Planarized sentence representation for nested named entity recognition. *Inf. Process. Manag.* **2023**, *60*, 103352. [\[CrossRef\]](#)

5. Huang, H.Y.; Lei, M.; Feng, C. Hypergraph network model for nested entity mention recognition. *Neurocomputing* **2021**, *423*, 200–206. [\[CrossRef\]](#)
6. Shang, Y.M.; Mao, H.L.; Tian, T.; Huang, H.Y.; Mao, X.L. From local to global: Leveraging document graph for named entity recognition. *Knowl. Based Syst.* **2025**, *312*, 113017. [\[CrossRef\]](#)
7. Yan, H.; Gui, T.; Dai, J.Q.; Guo, Q.P.; Zhang, Z.; Qiu, X.P. A Unified Generative Framework for Various NER Subtasks. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online, 1–6 August 2021; pp. 5808–5822. [\[CrossRef\]](#)
8. Ming, H.; Yang, J.Y.; Liu, S.; Jiang, L.L.; An, N. Mitigating prototype shift: Few-shot nested named entity recognition with prototype-attention contrastive learning. *Expert Syst. Appl.* **2025**, *268*, 126293. [\[CrossRef\]](#)
9. Li, Y.H.; Qi, S.; Zhou, C.J.; Lu, L.; Tian, Y.C. Named entity recognition based on anchor span for manufacturing knowledge extraction. *Eng. Appl. Artif. Intell.* **2025**, *162*, 112693. [\[CrossRef\]](#)
10. Wang, Y.L.; Gan, J.X.; Fang, R.X.; Lai, P.C. Span-based sliding window feature extraction and label knowledge for few-shot Named Entity Recognition. *Eng. Appl. Artif. Intell.* **2025**, *162*, 112702. [\[CrossRef\]](#)
11. Zhu, S.C.; Yuan, Y.; Shi, L.; Xu, S.K. Advancements in few-shot nested Named Entity Recognition: The efficacy of meta-learning convolutional approaches. *Neurocomputing* **2025**, *635*, 129893. [\[CrossRef\]](#)
12. Yan, Y.K.; Cai, B.L.; Song, S. Nested Named Entity Recognition as Building Local Hypergraphs. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; pp. 13878–13886. [\[CrossRef\]](#)
13. Yan, Z.H.; Yang, S.L.; Liu, W.; Tu, K.W. Joint Entity and Relation Extraction with Span Pruning and Hypergraph Neural Networks. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023; pp. 7512–7526. [\[CrossRef\]](#)
14. Zhang, H.; Dang, Y.P.; Zhang, Y.Z.; Liang, S.Y.; Liu, J.X.; Ji, L.X. Chinese nested entity recognition method for the finance domain based on heterogeneous graph network. *Inf. Process. Manag.* **2024**, *61*, 17. [\[CrossRef\]](#)
15. Jana, S.; Milan, S.; Jan, H. Neural architectures for nested NER through linearization. In Proceedings of the 57th Conference of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5326–5331. [\[CrossRef\]](#)
16. Zhu, E.W.; Li, J.P. Boundary Smoothing for Named Entity Recognition. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022; pp. 7096–7108. [\[CrossRef\]](#)
17. Li, F.; Wang, Z.; Hui, S.C.; Liao, L.J.; Zhu, X.H.; Huang, H.Y. A segment enhanced span-based model for nested named entity recognition. *Neurocomputing* **2021**, *465*, 26–37. [\[CrossRef\]](#)
18. Yan, H.; Sun, Y.; Li, X.N.; Qiu, X.P. An Embarrassingly Easy but Strong Baseline for Nested Named Entity Recognition. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, ON, Canada, 9–14 July 2023; pp. 1442–1452. [\[CrossRef\]](#)
19. Yu, J.H.; Chen, Y.P.; Zheng, Q.H.; Wu, Y.F.; Chen, P. Full-span named entity recognition with boundary regression. *Connect. Sci.* **2023**, *35*, 2181483. [\[CrossRef\]](#)
20. Liu, Y.D.; Zhang, K.; Tong, R.; Cai, C.X.; Chen, D.Y.; Wu, X.H. A Two-Stage Boundary-Enhanced Contrastive Learning approach for nested named entity recognition. *Expert Syst. Appl.* **2025**, *271*, 126707. [\[CrossRef\]](#)
21. Ning, J.Z.; Sun, Y.Y.; Yang, Z.H.; Wang, Z.J.; Luo, L.; Lin, H.F.; Zhang, Y.J. GenEn-MNER: Enhancing Nested Chinese NER with Multimodal Fusion and Alignment via Speech-to-Text Generation. *IEEE Trans. Audio Speech Lang. Process.* **2025**, *33*, 1628–1640. [\[CrossRef\]](#)
22. Sun, Y.L.; Wang, X.Y.; Wu, H.S.; Hu, M. Global Span Semantic Dependency Awareness and Filtering Network for nested named entity recognition. *Neurocomputing* **2025**, *617*, 129035. [\[CrossRef\]](#)
23. Mao, H.L.; Mao, X.L.; Tang, H.L.; Gao, X.Y.; Xu, C.; Huang, H.Y. A Flat Tree-based Transformer for Nested Named Entity Recognition. *Knowl.-Based Syst.* **2025**, *318*, 113405. [\[CrossRef\]](#)
24. Huang, Z.H.; Xu, W.; Yu, K. Bidirectional LSTM-CRF models for sequence tagging. *arXiv* **2015**, arXiv:1508.01991. [\[CrossRef\]](#)
25. Huang, Y.; Li, Y.X.; Zhang, X.Y. DTDM: Dynamic Temporal Convolutional Network and Dynamic Multihead Attention for Chinese Named Entity Recognition. *IEEE Access* **2023**, *11*, 128153–128161. [\[CrossRef\]](#)
26. Liu, J.X.; Sun, M.Z.; Zhang, W.H.; Xie, G.Q.; Jing, Y.X.; Li, X.L.; Shi, Z.X. DAE-NER: Dual-channel attention enhancement for Chinese named entity recognition. *Comput. Speech Lang.* **2024**, *85*, 101581. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.