IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 November 2023, accepted 15 November 2023, date of publication 21 November 2023,
date of current version 29 November 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3335623*

RESEARCH ARTICLE

# A Novel Joint Extraction Model for Entity Relations Using Interactive Encoding and Visual Attention

**YOUREN YU, YANGSEN ZHANG, XUEYANG LIU, AND SIWEN ZHU**
Institute of Intelligent Information Processing, Beijing Information Science and Technology University, Beijing 100101, China

Corresponding author: Yangsen Zhang (zhangyangsen@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 62176023.

**ABSTRACT** Relationship extraction is a fundamental task in natural language processing, with applications ranging from knowledge graph construction to information retrieval. Existing entity-relationship joint extraction models have made significant strides in this field. However, they still face limitations in effectively utilizing interaction information between subjects and objects, as well as capturing the spatial location relationships between entities. In this paper, we propose a novel relationship extraction model that addresses these limitations. Our model introduces innovative techniques to harness interaction information between subjects and objects. We employ subject gates, object gates, entity gates, and relationship gates to partition and filter interaction information between relationship triples during the encoding phase. Additionally, we leverage an attention mechanism inspired by the visual domain to capture spatial location relationships between entities during the decoding phase, transforming the entity-relationship joint extraction task into a table-filling task. To evaluate the effectiveness of our model, we conducted extensive experiments on multiple datasets, including WebNLG, NYT, and ADE. Our model achieved impressive F1 values of 93.65%, 92.58%, and 86.16% on these datasets, respectively, outperforming state-of-the-art models.

**INDEX TERMS** Relation extraction, attention mechanism, knowledge graph construction, natural language processing.

## I. INTRODUCTION

Entity relationship extraction is a fundamental task in natural language processing that aims to identify entities, their types, and specific relationships between them from unstructured natural language text. The applications of this task are numerous, including building knowledge graphs [1], improving the capabilities of intelligent question-answering systems [2], and enhancing the effectiveness of information retrieval [3]. Textual data used for relationship extraction can come from a variety of domains, such as web data, emails, social media, and customer service interactions. In recent years, with the growing size of data, manual relationship extraction methods are no longer sufficient to meet the demand, and hence, automated relationship extraction tasks have gained significant attention.

Traditional relationship extraction models typically employ a pipeline-based approach, which involves dividing relationship extraction into two separate subtasks: named entity recognition and relationship extraction. The approach first performs named entity recognition on a given text and then performs relationship extraction based on the extracted entities. Reference [4] proposed a relationship extraction model based on two independent pretrained encoders, PURE, which uses a pipeline approach to first recognize entities in the input text and then embed them as tokens for relationship extraction. Although the pipeline-based approach is more flexible, it overlooks the interconnection between entities and relationships and may result in problems such as error accumulation. In recent years, joint extraction-based relationship extraction methods have become a new

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Chiara Caschera.

research trend. Reference [5] proposed a single-stage joint extraction model, TPLinker, which uses annotated pairwise concatenation for joint extraction of entity relationships. This approach achieves consistency between training and testing and effectively improves the accuracy of relationship extraction.

Although there have been some advances in existing entity-relationship extraction models, they still have some limitations. Firstly, they fail to make full use of the interaction information between subjects and objects. Secondly, they lack the extraction of spatial location relationship features between entities. To address these challenges, we propose a new entity-relationship extraction model that effectively leverages the interaction information between subjects and objects, entities and relationships.

In this paper, we present a comprehensive overview of our proposed method and its implementation. Our model consists of a novel text encoding unit designed to capture bidirectional interactions within the relational triad extraction task. We achieve this by exploiting the interaction information between subjects and objects, as well as entities and relations. Additionally, we introduce a new decoding unit that transforms the joint entity-relationship extraction task into a table-filling task. Notably, we are the first to introduce a visual domain attention mechanism within this task, enhancing the spatial relationships between relationship triples and thereby improving the accuracy of relationship extraction.

In summary, our research has made the following contributions:

1. We propose a novel text encoding unit that effectively models bidirectional interactions in the relational triad extraction task by harnessing the interaction information between subjects and objects, entities, and relations.

2. We introduce a new decoding unit that converts the joint entity-relationship extraction task into a table-filling task. This novel approach incorporates a visual domain attention mechanism to enhance spatial relationships between relationship triples, leading to improved accuracy in relationship extraction.

3. We provide a detailed implementation and evaluation of our proposed model, showcasing its superiority over other state-of-the-art baseline models on several publicly available datasets.

## II. RELATED WORK

The evolution of the entity-relationship joint extraction task has progressed from an early rule-based approach that used feature engineering and domain knowledge to a traditional machine learning-based approach. With the continued advancements in deep learning, the mainstream approach to entity-relationship joint extraction has evolved into a deep learning-based approach.

Deep learning-based methods for entity-relationship extraction can be mainly divided into pipeline-based methods and joint extraction-based methods. Reference [6] proposed a model RFBFN that exploits the semantic information between relations and can extract both subjects and objects, which models the relation extraction task as a complete fill-in-the-blank task and reduces redundant computations by predicting the types of relations embedded in the text, effectively improving the relation extraction accuracy. Reference [7] proposed a model PL-Maker for subject packing strategy, which uses a padded suspension marker (PL-Marker) to better model entity boundaries and packs each subject and all its objects to model the interrelationships between the same subject span pairs, effectively speeding up the recognition accuracy. Reference [8] proposed a single-module, single-step decoding model, OneRel, which converts joint extraction into a fine-grained trivial classification task that can better capture the interdependencies among triads. Reference [9] proposed a model based on potential relations and global correspondence, PRGC, which reduces the redundant relations by predicting the potential relations embedded in the input text, and effectively reduces the number of parameters and computational effort of the model. Reference [10] proposed a cascaded relation extraction model CASREL, which first extracts subject entities in the input text, and then extracts object entities based on the extracted subjects and relations, effectively solving the overlapping triples problem.

In the realm of visual domain, the transformer structure demonstrates commendable performance as well. For instance, [11] introduced a visual transformer-based generative adversarial network known as Transformer-GAN, designed specifically for low-light image enhancement. This model incorporates a multi-head, multi-covariate self-attention, and an optical feature forward module structure, effectively enhancing accuracy in low-light scenarios. Another noteworthy approach is presented by [12] in the form of an integrated learning framework named LightingNet for low-light image enhancement. LightingNet comprises two pivotal components: 1) a complementary learning subnetwork and 2) a visual transformer (VIT) low-light enhancement subnetwork, collectively contributing to the substantial improvement in image quality under low-light conditions.

In previous studies, while there have been methods to extract joint relationships for textual information, they mostly exploit the semantic information between relations but fail to capture the interaction information between subjects and objects in the relationship triad. Moreover, these methods also fail to exploit the spatial location relationships between entities. Thus, in this study, we propose a new text encoding unit that can reasonably model the bidirectional interactions in the joint entity-relationship extraction task. To improve the network's ability to model spatial information, we convert the entity-relationship joint extraction task into a table-filling task and employ an attention mechanism based on visual domain to extract spatial location features from text. This method effectively improves entity-relationship extraction accuracy and enhances the robustness of the trained model, helping the network to learn the correlation between different texts.

## III. METHOD

### A. TASK DEFINITION

Given an input sequence $S = \{x_1, x_2, \ldots, x_n\}$ with n tokens, $x_i$ denotes the i-th token in the sequence S. The task can be decomposed into two subtasks:

#### 1) NAMED ENTITY RECOGNITION

The goal of the NER task is to extract the subject triplet $(x_{si}, e_s, x_{sj}) \in E$, object triples $(x_{oi}, e_o, x_{oj}) \in E$ and all entity triples $(x_i, e, x_j) \in E$, where $x_{si}$ and $x_{sj}$ denote the beginning and end of the subject, $x_{oi}$ and $x_{oj}$ denote the beginning and end of the object, $x_i$ and $x_j$ denote the beginning and end of the entity, $e_s, e_o, e$ are entity types, and E represents all entity triples. If there are two entities that exist in a relationship, we can make the two entities the subject and the object, with the "subject" being the entity or related element that usually performs some action in the text or is central to the action described in the text." Objects" are entities that are related to or acted upon by the subject. They represent the target or receptor of an action or relationship in a text. Relational extraction is the process of finding out the relationship that exists between the subject and the object in unstructured or semi-structured data and representing it as an entity-relationship triad.

#### 2) RELATION EXTRACTION

The goal of the RE task is to extract the relation triples $(x_{ri}, r, x_{rj}) \in R$ where $x_{ri}$ and $x_{rj}$ denote the starting tokens of the subject and object, respectively, r is the relation type, and $R$ represents all the relation triples.

### B. OVERVIEW

This section describes our proposed relationship extraction model with the model structure shown in Figure 1, which consists of two main parts: 1. text encoding unit, which encodes the text using a pre-trained language model and feeds the output of the last two layers of the pre-trained language model into a parallel feature filtering unit to strengthen the correlation relationships between subjects and objects, entities and relationships for feature extraction for subsequent tasks. 2. decoding unit. The relational triad is extracted using a table structure-based decoder, and Shuffle Attention is used to increase the network's ability to model spatial information and help the network learn the correlation between different regions.

### C. TEXT ENCODING UNIT

As to make full use of the connection between subject and object, entity and relation, and reduce the feature parts that affect each other among them, we propose a new parallel feature filtering architecture, which can enhance the feature extraction ability of subject, object and relation triad by modeling the semantic relationship between subject and object, entity and relation synchronously.

Given an input sentence $S$, we first used a pre-trained language model (PLM) to encode the input text to obtain a contextual representation of the sentence. In order to model the semantic relationships between subjects and objects, entities and relations simultaneously, the output of the last two layers of the pre-trained language model is used as the input to the parallel feature filtering unit, as shown in Equation (1).

$$\begin{aligned} \left\{y_1, \ldots, y_{|L|}\right\} &= \mathrm{PLM}^{(-2)}\left(\left\{\mathrm{x}_1, \ldots, \mathrm{x}_{|L|}\right\}\right) \\ \left\{z_1, \ldots, z_{|L|}\right\} &= \mathrm{PLM}^{(-1)}\left(\left\{\mathrm{x}_1, \ldots, \mathrm{x}_{|L|}\right\}\right) \end{aligned} \tag{1}$$

where PLM $^{(-2)}$ is the penultimate layer of the pre-trained model, PLM$^{(-1)}$ is the last layer of the pre-trained model, $x_i$ is the input representation of sentence $L$, and $|L|$ is the sentence length, $y_i$ as the feature for extracting subject and object, and $z_i$ as the feature for extracting relation.

In order to enhance the interconnection between subjects and objects, entities and relations, and reduce the part of features that interact between them, inspired by [13], we propose a parallel feature filtering architecture for feature extraction. The partitioned filtering encoder uses the cumax activation function proposed by [14] to sort and slice the output vectors, dividing the neurons into three partitions: two task partitions storing intra-task information and a shared partition storing inter-task information.

As shown in Equations (2), a sequence of subject vectors $s$, a sequence of object vectors $o$, and a sequence of relationship vectors $r$ are obtained using a linear transformation.

$$\begin{aligned} s &= W_s[y_t; h_{so,t-1}] + b_s \\ o &= W_o[y_t; h_{so,t-1}] + b_o \\ r &= W_r[z_t; h_{re,t-1}] + b_r \end{aligned} \tag{2}$$

where $W_s$, $W_o$ and $W_r$ re weight matrices, $b_s$, $b_o$ and $b_r$ are bias terms, $y_t$ and $z_t$ represent the vectors at the tth time step, $h_{so,t-1}$ represents the state of the hidden layer used to extract subject and object features at the previous moment, and $h_{re,t-1}$ represents the state of the hidden layer used to extract entity and relationship features at the previous moment.

Then, we use the activation function cumax to convert the pre-trained model output vectors $s, o, r$ to subject gate $g_s$, object gate $g_o$, relation gate $g_r$ and entity gate $g_e$, and using these gates, the subject partition $p_s$, object partition $p_o$, entity partition $p_e$, relation partition $p_r$, entity relation sharing partition $p_{re}$ and subject-object sharing partition $p_{so}$ are formed, as shown in Equation (3)(4).

$$\begin{aligned} g_s &= cumax(s) \\ g_o &= 1 - cumax(o) \\ g_e &= cumax(W_{ge}[s; o] + b_{ge}) \\ g_r &= 1 - cumax(r) \end{aligned} \tag{3}$$

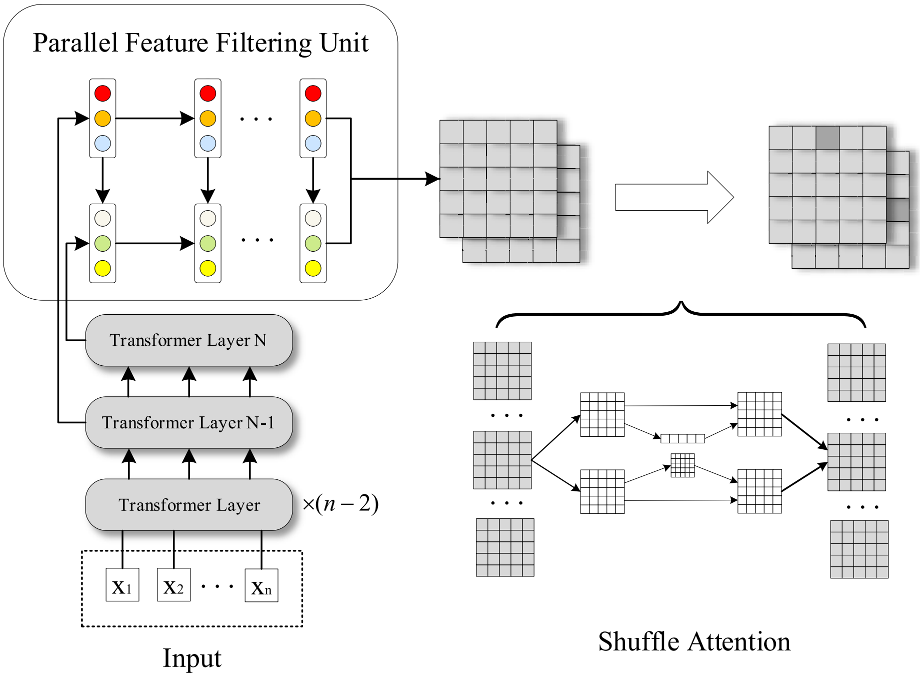

**FIGURE 1.** Model structure. In the parallel feature filtering unit, red, orange, and blue indicate subject-related, subject-object shared, and object-related features, respectively, and white, green, and yellow indicate entity-related, entity-relationship shared, and relationship-related features, respectively. In the decoder using shuffle attention, for the subject, object and entity extraction tasks, the location of the darkened squares in the table [i, j] indicates the i-th word to the j-th word constituent entity in the input sequence; for the relationship extraction task, the location of the darkened squares in the grid [i, j] indicates the relationship triad with the i-th word as the subject starting word and the j-th word as the object starting word in the input sequence.

$$\begin{aligned} p_{re/so,t-1} &= g_{e/s,t-1} \circ g_{r/o,t-1} \\ p_{e/s,t-1} &= g_{e/s,t-1} - p_{re/so,t-1} \\ p_{r/o,t-1} &= g_{r/o,t-1} - p_{re/so,t-1} \end{aligned} \tag{4}$$

where $\circ$ indicates that the vector elements are multiplied by their corresponding positions. $p_{re/so,t-1}$ denotes $p_{re,t-1}$ or $p_{so,t-1}$, $g_{e/s,t-1}$ denotes $g_{e,t-1}$ or $g_{s,t-1}$, and the rest are similar.

Then, using the above partitioning unit, neuron $c$ is divided into entity memory $q_e$, relation memory $q_r$, subject memory $q_s$, object memory $q_o$, entity-relationship shared memory $q_{re}$, and subject-object shared memory $q_{so}$ as shown in Equations (5)-(8).

$$\begin{aligned} \widehat{c}_{so,t} &= \tanh(W_{so}[y_t; h_{so,t-1}] + b_{so}) \\ \widehat{c}_{re,t} &= \tanh(W_{re}[z_t; h_{re,t-1}] + b_{re}) \end{aligned} \tag{5}$$

$$\begin{aligned} p_{re/so} &= p_{re/so,t-1} \circ c_{re/so,t-1} + p_{re/so,t} \circ \widehat{c}_{re/so,t} \\ p_{r/o} &= p_{r/o,t-1} \circ c_{re/so,t-1} + p_{r/o,t} \circ \widehat{c}_{re/so,t} \\ p_{e/s} &= p_{e/s,t-1} \circ c_{re/so,t-1} + p_{e/s,t} \circ \widehat{c}_{re/so,t} \end{aligned} \tag{6}$$

$$q_e = p_e + p_{re};\ q_r = p_r + p_{re};\ q_{re} = p_{re} \tag{7}$$

$$q_s = p_s + p_{so};\ q_o = p_o + p_{so};\ q_{so} = p_{so} \tag{8}$$

where $W_{so}$ and $W_{re}$ are weight matrices, $b_{so}$ and $b_{re}$ are bias terms, tanh is the activation function.

After that, the information in each memory is updated and the corresponding memory is used to generate subject feature$h_s$, object feature $h_o$, entity feature $h_e$, relationship feature $h_r$, subject-object shared feature $h_{so}$, and entity-relationship shared feature $h_{re}$. The formula is shown in (9)(10).

$$h_{re} = \tanh(q_{re});\ h_r = \tanh(q_r);\ \mathrm{h}_e = \tanh(q_e) \tag{9}$$

$$h_{so} = \tanh(q_{so});\ h_s = \tanh(q_s);\ \mathrm{h}_o = \tanh(q_o) \tag{10}$$

Finally, after the segmentation and filtering steps, the information in all three memories is used to form the entity-relationship cell state $c_{re,t}$ and the subject-object cell state $c_{so,t}$, which are then used to generate the entity-relationship hidden state $h_{re,t}$ and the subject-object hidden state $h_{so,t}$ (the hidden state and cell state of time step t are the input to the next time step). The formula is shown in (11)-(14).

$$c_{re,t} = W_{cre}[q_{e,t}; q_{r,t}; q_{re,t}] + b_{cre} \tag{11}$$

$$h_{re,t} = \tanh(c_{re,t}) \tag{12}$$

$$c_{so,t} = W_{cso}[q_{s,t}; q_{o,t}; q_{so,t}] + b_{cso} \tag{13}$$

$$h_{so,t} = \tanh(c_{so,t}) \tag{14}$$

### D. DECODING UNIT

Our model comprises four task units: a subject extraction unit, an object extraction unit, an entity extraction unit, and a relationship extraction unit. The subject and object extraction unit aims to identify and classify all subjects and objects present in a sentence's relational triad. The entity extraction unit aims to identify and classify all entities present in a given sentence. The relationship extraction unit aims to identify the starting positions of the subject and object of a sentence's relational triad and to classify their relations.

To capture the spatial location relationships of entities in text, we convert the entity relationship extraction task into a table filling task and propose a novel decoder using the attention mechanism in the visual domain.

Specifically, for the entity extraction unit, the parallel feature filtering unit is used to output entity features $h_e$ and entity-relationship shared features $h_{re}$ for computing the globalized representation $h_{ge}$, as shown in Equation (15)(16).

$$h_{ge,t} = \tanh(W_{ge}[h_e; h_{re}] + b_{ge}) \tag{15}$$

$$h_{ge} = \mathrm{meanpool}(h_{ge,1}, \ldots, h_{ge,L}) \tag{16}$$

where $W_{ge}$ is the weight matrice, $b_{ge}$ is the bias terms, *meanpool* is the average pooling layer.

As shown in Equation (17), for each word pair $(w_i; w_j)$, we connect the word-level entity features $h_{e,i}$ and $h_{e,j}$, and the sentence-level global feature $h_{ge}$, and then input them to the fully connected layer to obtain the entity span representation $X_{ij}$.

$$X_{ij} = W_{he}[h_{e,i}; h_{e,j}; h_{ge}] + b_{he} \tag{17}$$

Then, we introduce Shuffle Attention [15] in the visual domain to enhance the spatial location relationships of entities in the text. Shuffle Attention is a novel attention mechanism for the visual domain that uses a more efficient computational approach. Traditional attention mechanisms usually obtain attention weights by calculating the importance of each position in the input sequence. In contrast, Shuffle Attention first disrupts and reorders the input sequence, and then calculates the attention weights based on the importance of each position. This mechanism can combine local feature mappings to enhance the modeling of spatial information for the whole network.

The Shuffle Attention mechanism consists of two key steps: channel shuffling and feature reorganization. In the channel shuffle step, the input features are divided into groups, each group containing a portion of channels that will be randomly reassembled together. In the feature reorganization step, the features within each group will be reorganized according to some pattern, thus enhancing the network's ability to model spatial information. This reorganization process helps the network to learn the correlation between different regions.

In the Shuffle Attention module, the global information is first embedded using global average pooling (GAP) as shown in Equation (18).

$$s = \frac{1}{N * N} \sum_{i=1}^{N} \sum_{j=1}^{N} X(i, j) \tag{18}$$

where N is the length of the input text.

After that, the features are weighted by channel attention by using linear transformation with sigmoid activation function $\sigma$ as shown in Equation (19).

$$X_{channel}\text{=X} * \sigma(W_c s + b_c) \tag{19}$$

In addition, feature extraction is performed using spatial attention for the vectors, which complements the channel attention by calculating the corresponding attention weights for each spatial location, thus making the neural network pay more attention to the important spatial locations and improving the representational power and performance of the network. We use Group Norm (GN) to obtain spatial statistics on $X_{ij}$ and employ a fully connected layer and a sigmoid activation function to enhance the representation of $h_{r,ij}$ as shown in Equation (20).

$$x_{spatial} = X^{*}_{ij}\sigma(W_{spatial}(GroupNorm(X_{ij})) + b_{spatial}) \tag{20}$$

Finally, the feature vectors extracted by the channel attention mechanism and the spatial attention mechanism are connected, and the "channel shuffle" operation is applied to make the information across groups flow along the channel dimension. With the span representation, we can predict whether the span is an entity of type k by feeding the span into the feedforward neural layer, as shown in Equation (21).

$$e^{k}_{ij} = \sigma(\mathrm{LE}(W_e([X_{channel}; X_{spatial}]) + b_e)) \tag{21}$$

where $W_e$ is the weight matrix, $b_e$ is the bias term, LE is a linear layer with ELU activation function, and $\sigma$ is the sigmoid activation function.

The subject and object extraction units bear similarity to the entity extraction units. Concerning the relationship extraction unit, we solely predict the initial word positions of the subject and object. Subsequently, we search for entities within the subject and object extraction units that have the same initial word position. If an entity is not found within the subject or object extraction units, we then seek the corresponding entity position within the entity extraction unit. Equation (22) depicts the ultimate score function $r^{m}_{ij}$ for the relationship.

$$r^{m}_{ij} = \sigma(\mathrm{LE}(W_r([X_{channel}; X_{spatial}]) + b_r)) \tag{22}$$

where $r^{m}_{ij}$ represents the score with subject starting word position $i$ and object starting word position $j$ in the relation triad, and the relation type between them is $m$.

## IV. EXPERIMENTS

### A. DATASETS AND EXPERIMENTAL SETTINGS

#### 1) DATASETS

We performed comparative experiments using three publicly available datasets: NYT [16], WebNLG [17] and ADE. The NYT dataset contains text from the New York Times and named entities in the text are annotated using the Stanford NER tool in combination with the Freebase knowledge base. The relationship pairs in the text were annotated using a remotely supervised approach. The WebNLG dataset

**Algorithm 1** Model Pseudo-Algorithm

```
Input: Sentence
Output: Relational triad
# Step 1: Encode the input text using a pre-trained language model encoded_text = PLM(input_text)
# Step 2: Parallel Feature Filtering subject_features, object_features, relation_features = ParallelFeatureFiltering(encoded_text) entity_features = concat(subject_features, object_features)
# Step 3: Partitioned Filtering Encoder subject_ features, object_ features, relation_features, entity_ features = PartitionedFilteringEncoder(subject_features, object_features, relation_features, entity_features)
# Step 4: Shuffle Attention subject_ features, object_ features, relation_ features, entity_features = ShuffleAttention(subject_features, object_features, relation_features, entity_features)
# Step 5: Decoding Unit subject_span, object_span, relation_span, entity_span = DecodingUnit(subject_features, object_features, relation_features, entity_features)
```

was originally created for the Natural Language Generation (NLG) task and contains 171 predefined relation types with 5019 texts in the training set, 500 texts in the validation set, and 703 texts in the test set. The ADE dataset is derived from information on medical descriptions of adverse effects of drug use. It consists of two predefined entity types (Adverse-Effect and Drug) and a single relationship type (Adverse-Effect).

#### 2) EVALUATION METRICS

Following previous work, we evaluate our model using partial matching on the NYT and WebNLG datasets, exact matching on the ADE datasets, and using the micro-F1 metric as an evaluation metric on the NYT and WebNLG datasets, and the macro-F1 metric as an evaluation metric on the ADE dataset.

#### 3) IMPLEMENTATION DETAILS

For the relational extraction task, we used bert-base-uncased [18] as the base encoder on the WebNLG and NYT datasets, Albert-xxlarge-v1 as the base encoder on the ADE dataset in order to provide a fair comparison with previous work [5], [10]. The specific parameters are shown in Table 1.

**TABLE 1. Experimental parameter settings.**

| Parameters | Value |
|---|---|
| Batch Size | 8 |
| Learning rate | 2e-5 |
| Dropout rate | 0.1 |
| Max len | 128 |
| Optimizer | AdamW |

### *B. EXPERIMENTAL RESULTS*

In this section, we compare our model experimentally with several state-of-the-art models, which are: CGT [19], Casrel [10], TpLinker [5], CasDE [20], SPN [21], TriMF [22], ERIGAT [23], Rel-Metric [24], SpERT [25], eRPR MHS [26], SCDM [27], LAPREL [28], ESEI [29], RS-TTS [30].

CGT is a model for contrast triad extraction using a generative transformer, which first introduces a shared converter module based on encoder-decoder generation and proposes a novel triad contrast training object that effectively improves the model's ability to capture long-term text dependencies.

CASREL is a cascaded relationship extraction model, which first extracts subject entities from the input text, and then extracts object entities according to the extracted subject and relationship, effectively solving the overlapping triad problem.

TPLinker is a single-stage joint extraction model, which uses labeled pairwise concatenation for joint extraction of entity relationships, achieving consistency between training and testing, and effectively improving the accuracy of relationship extraction.

CasDE is a model that uses a cascaded dual decoder approach to extract overlapping relational triples. The model uses a text-specific relational decoder to detect relationships in a sentence based on the textual semantics of the sentence, and uses a span-based tagging scheme to detect the corresponding head and tail entities, effectively solving the overlapping triples problem.

SPN is a model that uses a transformer characterized by non-autoregressive parallel decoding, which treats the joint extraction of entities and relations as an ensemble prediction problem and lets the model directly output the final set of relational triples, and designs an ensemble-based loss function for prediction by bilateral matching, which can achieve more accurate training results by ignoring the order of relational triples.

TriMF is a model that employs a multi-level memory flow attention mechanism to bolster the bidirectional interaction between entity recognition and relation extraction. The model constructs a memory module to retain the category representations acquired from the entity recognition and relation extraction tasks. In situations where manual annotation is lacking, the model can augment the relation trigger information in sentences by activating the sensor module. Consequently, this improves model performance and facilitates enhanced interpretation of model predictions.

ERIGAT is a model that employs graph attention networks for entity and relation extraction. It introduces a novel multi-headed attention mechanism and utilizes shared attention module parameters to improve model performance.

Rel-Metric is a relationship extraction model that utilizes table structure. It employs two-dimensional convolution to capture local feature dependencies and effectively enhance the model's performance.

SpERT is a span-based joint entity-relationship extraction model that effectively enhances model performance. It achieves this by integrating the embedding layer of the BERT model into the inference model and leveraging contextual information for relationship classification.

The eRPR MHS is a relationship extraction model that employs relative representation of entity positions. It leverages distance information between entities and contextual tokens, and introduces an auxiliary global relationship classification to enhance the learning of local contextual features.

SCDM is a joint entity-relationship extraction model based on span and cascade double decoding, which divides word vectors according to spans, forms span sequences and decodes the relationships between span sequences to get the relationship types in the span sequences, and fuses the span sequences and relationship types obtained from relationship decoding at the entity decoding layer, which effectively improves the effect of recognizing overlapping relationships.

LAPREL is a label-aware parallel network model for relation extraction. The model embeds a priori labeling information into the tag embeddings and adjusts the sentence embeddings for each relation type, and applies a parallel network to solve the problem of addressing exposure bias and effectively improves the modeling results.

ESEI is a joint entity and relation extraction model based on efficient sampling and explicit interaction. The model divides negative samples into sentences that overlap with positive samples or not, improves the model's ability to extract entity boundary information by controlling the sampling ratio, and introduces a heterogeneous graphical neural network (GNN) to the model, which significantly improves the model's discriminative ability on the relationship extraction task.

RS-TTS is a relation-specific ternary labeling and scoring model. The model uses a relation judgment module to predict all potential relations and introduces a boundary smoothing mechanism for entity pair extraction, effectively mitigating the computational redundancy of the model as well as the overconfidence problem.

The results presented in Tables 2-4 demonstrate that our proposed model outperforms other comparative models in terms of entity-level F1 scores. Specifically, our model achieves F1 scores of 93.65%, 92.58%, and 86.16% on the publicly available datasets of WebNLG, NYT, and ADE, respectively. Moreover, our proposed model achieves F1 scores that are 1.75% and 0.68% higher than the F1 scores obtained by the previously proposed relationship extraction SOTA model SCDM and TpLinker on the WebNLG dataset and the NYT dataset, respectively. Furthermore, our model

**TABLE 2.** Experimental results on the WebNLG dataset. P, R and F1 represent precision, recall and F1 relation scores.

| Model | WebNLG | | |
|---|---|---|---|
| | P | R | F1 |
| CGT[19] | 92.9 | 75.6 | 83.4 |
| Casrel[10] | 93.4 | 90.1 | 91.8 |
| TpLinker[5] | 91.8 | 92.0 | 91.9 |
| CasDE[20] | 90.3 | 91.5 | 90.9 |
| SCDM[27] | 91.6 | 92.2 | 91.9 |
| LAPREL[28] | 91.7 | 91.5 | 91.6 |
| RS-TTS[30] | 90.7 | 89.7 | 90.2 |
| **Ours** | **94.07** | **93.23** | **93.65** |

**TABLE 3.** Experimental results on the NYT dataset.

| Model | NYT | | |
|---|---|---|---|
| | P | R | F1 |
| CGT[19] | **94.7** | 84.2 | 89.1 |
| Casrel[10] | 89.7 | 89.5 | 89.6 |
| TpLinker[5] | 91.3 | 92.5 | 91.9 |
| CasDE[20] | 90.2 | 90.9 | 90.5 |
| SCDM[27] | 89.8 | **92.7** | 91.2 |
| LAPREL[28] | 90.7 | 91.4 | 91.1 |
| **Ours** | 92.97 | 92.20 | **92.58** |

**TABLE 4.** Experimental results on the ADE dataset.

| Model | ADE | | |
|---|---|---|---|
| | P | R | F1 |
| TriMF[22] | 74.22 | 83.43 | 80.66 |
| ERIGAT[23] | 84.81 | 75.86 | 80.09 |
| Rel-Metric[24] | 77.36 | 77.25 | 77.29 |
| SpERT[25] | 78.09 | 80.43 | 79.24 |
| eRPR MHS[26] | 74.35 | 86.12 | 79.80 |
| ESEI[29] | 80.37 | 85.40 | 82.88 |
| Ours | **85.89** | **86.44** | **86.16** |

achieves F1 scores that surpass those of recently proposed models, including TriMF, Casrel, CasDE, RS-TTS, ESEI and LAPREL, by margins of 1.85%, 2.08%, and 3.28% on the WebNLG, NYT, and ADE public datasets, respectively. These results demonstrate the superior performance of our model across all three datasets.

### C. ABLATION STUDY

Ablation experiments were conducted on each component of the model to evaluate their effectiveness. The detailed results of these experiments are presented in Table 5-7, which were based on three publicly available datasets.

**TABLE 5.** Ablation studies on the WebNLG dataset.

| | WebNLG | | |
|---|---|---|---|
| | P | R | F1 |
| **Ours** | **94.07** | **93.23** | **93.65** |
| -PFFN | 93.25 | 92.60 | 92.92 |
| -SA | 92.71 | 94.12 | 93.41 |
| -Layer | 93.38 | 93.67 | 93.53 |
| -Cumax | 92.38 | 93.61 | 92.99 |
| -Meanpool | 92.93 | 93.99 | 93.46 |

**TABLE 6.** Ablation studies on the NYT dataset.

| | NYT | | |
|---|---|---|---|
| | P | R | F1 |
| **Ours** | **92.97** | **92.20** | **92.58** |
| -PFFN | 91.18 | 92.17 | 91.67 |
| -SA | 91.27 | 92.60 | 91.93 |
| -Layer | 91.47 | 92.50 | 91.98 |
| -Cumax | 90.99 | 92.23 | 91.60 |
| -Meanpool | 91.77 | 91.57 | 91.67 |

**TABLE 7.** Ablation studies on the ADE dataset.

| | ADE | | |
|---|---|---|---|
| | P | R | F1 |
| **Ours** | **85.89** | **86.44** | **86.16** |
| -PFFN | 84.42 | 85.49 | 84.95 |
| -SA | 86.29 | 84.38 | 85.33 |
| -Layer | 85.01 | 86.75 | 85.87 |
| -Cumax | 84.29 | 85.49 | 84.89 |
| -Meanpool | 84.30 | 87.72 | 85.74 |

**-PFFN** We omitted the parallel feature filtering network and encoded the input using only the pre-trained language model to verify the effectiveness of the module. According to the F1 scores in Table 5-7, the model without the parallel feature filtering network substantially decreased. These results suggest that the method is effective in extracting the interaction information among entities, relations, and subjects/objects.

**-SA** We performed experiments with linear layers instead of Shuffle Attention modules to verify the effectiveness of the Shuffle Attention module. According to the F1 scores in Table 5-7, the models without Shuffle Attention modules had lower F1 scores. These results suggest that the use of the Shuffle Attention module can enhance the network's ability to model spatial information and help the network to learn inter-regional correlations.

**-Layer** We utilized only the last layer of output from the pre-trained language model to verify the validity of using the penultimate layers of the pre-trained language model for output. According to the F1 scores in Table 5-7, using the penultimate layers of the pre-trained language model as input to the parallel feature filtering unit leads to a reduction in the F1 score compared to using only the last layer of the pre-trained language model. This suggests that the method is effective in terms of feature extraction.

**-Cumax**We conducted experiments without the cumax activation function to verify the effectiveness of the cumax activation function. According to the F1 scores in Table 5-7, the model without cumax activation function has lower F1 scores. These results show that using cumax activation function can enhance the interconnection between subject and object, entity and relationship, and reduce the part of features that influence each other among them.

**-Meanpool**We conducted experiments using the maxpool layer instead of the meanpool layer to verify the effectiveness of the meanpool layer. According to the F1 scores in Tables 5-7, the model using maxpool layer instead of meanpool layer has a lower F1 score. These results indicate that the meanpool layer enhances the model effectiveness.

## V. CONCLUSION

In this paper, we present a relationship extraction model that takes into account the interaction information between the subject and object, as well as the spatial location relationship between entities in the relationship triad. We address the problem that existing relationship extraction models fail to make effective use of this information. First, we encode the text using a pre-trained language model. To better represent the corpus, we model bidirectional inter-task interactions using the last two output layers of the model. Second, we use subject gates, object gates, entity gates, and relation-ship gates to partition and filter the interaction information between the relational triads. This approach strengthens the connections between the subjects and objects, as well as entities and relationships in the relational triads. Finally, we convert the joint entity-relationship extraction task into a table-filling task. To capture the spatial location relationships between entities, we introduce an attention mechanism from the visual domain. This approach increases the network's ability to model spatial information and helps the network learn the correlations between different regions. We evaluate our proposed model on three public relational extraction datasets and compare it with existing models such as LAPREL, CasRel, CGT, and TPLinker. Our results demonstrate that our model outperforms the comparison models in terms of accuracy, recall, and F1 value, achieving optimal performance.

## REFERENCES

[1] D. Yu, C. Zhu, Y. Yang, and M. Zeng, "JAKET: Joint pre-training of knowledge graph and language understanding," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 10, pp. 11630–11638.

[2] W. Wu, Z. Zhu, J. Qi, W. Wang, G. Zhang, and P. Liu, "A dynamic graph expansion network for multi-hop knowledge base question answering," *Neurocomputing*, vol. 515, pp. 37–47, Jan. 2023.

[3] Q. Xia, B. Zhang, R. Wang, Z. Li, Y. Zhang, F. Huang, L. Si, and M. Zhang, "A unified span-based approach for opinion mining with syntactic constituents," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2021, pp. 1795–1804.

[4] Z. Zhong and D. Chen, "A frustratingly easy approach for entity and relation extraction," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Human Lang. Technol.*, 2021, pp. 50–61.

[5] Y. Wang, B. Yu, Y. Zhang, T. Liu, H. Zhu, and L. Sun, "TPLinker: Single-stage joint extraction of entities and relations through token pair linking," in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 1572–1582.

[6] Z. Li, L. Fu, X. Wang, H. Zhang, and C. Zhou, "RFBFN: A relation-first blank filling network for joint relational triple extraction," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics, Student Res. Workshop*, 2022, pp. 10–20.

[7] D. Ye, Y. Lin, P. Li, and M. Sun, "Packed levitated marker for entity and relation extraction," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 4904–4917.

[8] Y. Shang, M. Huang, and H. Mao, "OneRel: Joint entity and relation extraction with one module in one step," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, no. 10, pp. 11285–11293.

[9] H. Zheng, R. Wen, X. Chen, Y. Yang, Y. Zhang, Z. Zhang, N. Zhang, B. Qin, X. Ming, and Y. Zheng, "PRGC: Potential relation and global correspondence based joint relational triple extraction," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics, 11th Int. Joint Conf. Natural Lang. Process.*, 2021, pp. 6225–6235.

[10] Z. Wei, J. Su, Y. Wang, Y. Tian, and Y. Chang, "A novel cascade binary tagging framework for relational triple extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 1476–1488.

[11] S. Yang, D. Zhou, J. Cao, and Y. Guo, "Rethinking low-light enhancement via transformer-GAN," *IEEE Signal Process. Lett.*, vol. 29, pp. 1082–1086, 2022.

[12] S. Yang, D. Zhou, J. Cao, and Y. Guo, "LightingNet: An integrated learning method for low-light image enhancement," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 29–42, 2023.

[13] Z. Yan, C. Zhang, J. Fu, Q. Zhang, and Z. Wei, "A partition filter network for joint entity and relation extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2021, pp. 185–197.

[14] Y. Shen, S. Tan, and A. Sordoni, "Ordered neurons: Integrating tree structures into recurrent neural networks," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–14.

[15] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2235–2239.

[16] S. Riedel, L. Yao, and A. McCallum, "Modeling relations and their mentions without labeled text," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Berlin, Germany: Springer, 2010, pp. 148–163.

[17] X. Zeng, D. Zeng, S. He, K. Liu, and J. Zhao, "Extracting relational facts by an end-to-end neural model with copy mechanism," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 506–514.

[18] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. North Amer. Chapter Assoc. Comput. Linguistics*, pp. 4171–4186, 2019.

[19] H. Ye, N. Zhang, and S. Deng, "Contrastive triple extraction with generative transformer," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 16, pp. 14257–14265.

[20] L. Ma, H. Ren, and X. Zhang, "Effective cascade dual-decoder model for joint entity and relation extraction," 2021, *arXiv:2106.14163*.

[21] D. Sui, Y. Chen, K. Liu, J. Zhao, X. Zeng, and S. Liu, "Joint entity and relation extraction with set prediction networks," 2020, *arXiv:2011.01675*.

[22] Y. Shen, X. Ma, Y. Tang, and W. Lu, "A trigger-sense memory flow framework for joint entity and relation extraction," in *Proc. Web Conf.*, Apr. 2021, pp. 1704–1715.

[23] Q. Lai, Z. Zhou, and S. Liu, "Joint entity-relation extraction via improved graph attention networks," *Symmetry*, vol. 12, no. 10, p. 1746, Oct. 2020.

[24] T. Tran and R. Kavuluru, "Neural metric learning for fast end-to-end relation extraction," 2019, *arXiv:1905.07458*.

[25] M. Eberts and A. Ulges, "Span-based joint entity and relation extraction with transformer pre-training," 2019, *arXiv:1909.07755*.

[26] T. Zhao, Z. Yan, Y. Cao, and Z. Li, "Entity relative position representation based multi-head selection for joint entity and relation extraction," in *Chinese Computational Linguistics*. Hainan, China: Springer, 2020, pp. 184–198.

[27] T. Liao, H. Sun, and S. Zhang, "A joint extraction model for entity relationships based on span and cascaded dual decoding," *Entropy*, vol. 25, no. 8, p. 1217, Aug. 2023.

[28] X. Li, Y. Li, J. Yang, H. Liu, and P. Hu, "A relation aware embedding mechanism for relation extraction," *Appl. Intell.*, vol. 52, pp. 10022–10031, Jan. 2022.

[29] Q. Li, N. Yao, N. Zhou, J. Zhao, and Y. Zhang, "A joint entity and relation extraction model based on efficient sampling and explicit interaction," *ACM Trans. Intell. Syst. Technol.*, vol. 14, no. 5, pp. 1–18, Oct. 2023.

[30] J. Zhang, X. Jiang, Y. Sun, and H. Luo, "RS-TTS: A novel joint entity and relation extraction model," in *Proc. 26th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2023, pp. 71–76.

**YOUREN YU** was born in 1998. He is currently pursuing the M.S. degree. His research interests include natural language processing and artificial intelligence.

**YANGSEN ZHANG** was born in 1962. He received the Ph.D. degree. He is currently a Professor. His research interests include Chinese information processing, artificial intelligence, and web content security.

**XUEYANG LIU** was born in 1998. He is currently pursuing the M.S. degree. His research interests include natural language processing and artificial intelligence.

**SIWEN ZHU** was born in 1998. He is currently pursuing the M.S. degree. His research interests include natural language processing and artificial intelligence.