

SegDRE: A Salient Entity Guided Approach to Document-Level Relation Extraction

Xing Yang, Chengxiang Tan

School of Computer Science and Technology, Tongji University

{yangxing, jerrytan}@tongji.edu.cn

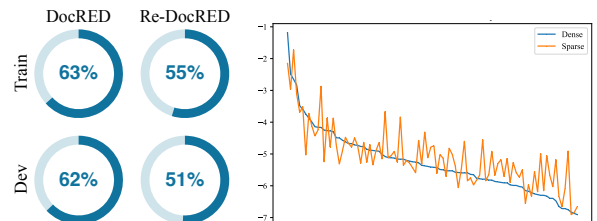
Abstract

In Document-level Relation Extraction (DocRE), relational facts are typically organized around a few salient entities. Explicitly capturing this topological structure is pivotal to addressing the two critical bottlenecks of the task: the extreme class imbalance and the complexity of multi-hop reasoning. Based on this insight, we first introduce the concept of the *salient entity* and propose a novel approach that decouples the extraction space into dense and sparse scenarios. Specifically, our approach restricts the search space for dense pairs to mitigate the dominance of the negative samples, and innovatively injects the rich semantic knowledge of salient entities to explicitly reconstruct the document for bridging disjoint evidence in multi-hop reasoning. Extensive experiments demonstrate that our approach yields consistent improvements over various backbone models and achieves advanced performance compared to existing enhancement methods¹.

1 Introduction

As a key task in natural language understanding, relation extraction aims to identify factual knowledge in the form of relation triples from text, thereby supporting a wide range of downstream applications such as question answering (Luo et al., 2018), knowledge graph alignment (Zhang et al., 2022) and large language models reasoning (Wang et al., 2024a). Prior research has primarily focused on sentence-level extraction (Zhang et al., 2018; Zhu et al., 2019; Sun et al., 2020); however, this simplified setting falls short in practical scenarios where entities are scattered across an entire document. In such cases, multiple relations may exist among entity pairs, which poses new challenges for Document-level Relation Extraction (DocRE). Specifically, existing models struggle to address two major challenges.

¹Our code: <https://github.com/Patrick-Rud/SegDRE-code>



(a) Proportion of relations associated with salient entities. (b) Log-frequency comparison between dense and sparse set of entity pairs over 96 relations.

Figure 1: Statistics of salient entities in document-level relation extraction datasets.

The first challenge lies in the severe imbalance between positive and negative samples. The number of entity pairs grows quadratically with the number of entities, yet most pairs do not hold any relation. In the widely used DocRED dataset, more than 97% of entity pairs are negatives; even in the refined Re-DocRED dataset, this proportion remains as high as 94%. Such a high prevalence of negative samples may easily overwhelm the training process, leading to poor recall of positive relations. While recent works like Wang et al. (2023) attempt to mitigate this via adaptive loss functions, the intrinsic dominance of negatives remains a bottleneck.

The second challenge arises from the complex relational structure in documents: entity pairs may be related across sentences necessitate reasoning over intervening entities, requiring models not only to capture global context but also perform multi-hop reasoning. To this end, Tan et al. (2022a) propose enhancing entity pair representations via axial attention, while Lu et al. (2023) construct dynamic graphs to facilitate reasoning over coreferent mentions and anaphors. Apart from representation learning, Ru et al. (2021) introduce logic rules to explicitly capture long-range dependencies.

Although significant progress has been made, prior studies have largely overlooked an inherent

property of documents: a document typically exhibits a clear thematic focus. As repositories of factual information, documents tend to organize knowledge around one or a few *salient entities*, which we define as the entity involved in the greatest number of relations. Humans, when reading, prioritize processing entity pairs involving salient entities, rather than exhaustively evaluating the set of pairs formed by all entities. This cognitive strategy allows us to rapidly capture core knowledge while reducing the candidate space to a linearly growing subset. Our statistics in Figure 1 (a) show that in DocRED and Re-DocRED, approximately 62.5% and 54.0% of positive relations, respectively, are concentrated in subsets associated with salient entities.

In addition, for complex reasoning, humans often adopt a *two-pass* strategy, integrating initial insights back into the context to facilitate comprehension. Empirically, our statistics on DocRED support this intuition: (1) nearly 89% of relations are reachable within two hops from salient entities (Figure 6 in Appendix A.1); and (2) distinct distribution patterns over relation classes emerge when splitting positive samples into *dense* and *sparse* groups based on their association with salient entities (Figure 1 (b)), which inspire us to treat dense and sparse entity pairs separately and process the document in a progressive manner.

Based on these findings, we propose a **Salient entity guided approach to Document-level Relation Extraction (SegDRE)**. Specifically, SegDRE introduces an identification task to pinpoint the salient entity, thereby decomposing the learning objective into *dense* and *sparse* scenarios. This decomposition yields a two-fold benefit: First, for the dense stage, it restricts the candidate space to a linear subset associated with the salient entity, significantly mitigating the severe positive-negative class imbalance issue. Second, for the sparse stage, it injects the extracted dense relation triples to augment the original text, effectively bridging the semantic gap required for multi-hop reasoning. Our main contributions are summarized as follows:

- We explicitly model salient entities via a count-based prediction head and use them to decompose the DocRE task into complementary scenarios.
- We design a novel progressive approach simultaneously alleviates the impact of negative

samples by narrowing the search scope and enhances the rest relation reasoning via explicit knowledge injection.

- Extensive experiments on two standard benchmarks demonstrate that our method yields consistent improvements over backbone models and outperforms previous state-of-the-art enhancement baselines.

2 Related Work

Existing DocRE methods can be broadly categorized into the following three lines of research.

(1) Enhancing Representation Capacity. This line of work aims at more accurate modeling of semantic interactions across documents by designing novel model architectures or strengthening existing representation mechanisms, thereby improving DocRE performance. Representative approaches include GAIN (Zeng et al., 2020), SIRE (Zeng et al., 2021), ATLOP (Zhou et al., 2021), KD-DocRE (Tan et al., 2022a), DREEAM (Ma et al., 2023), AA (Lu et al., 2023), SRF (Zhang et al., 2024), and GREP (Zhang et al., 2025).

(2) Loss Function Optimization. Another line of research focuses on improving DocRE by devising specialized loss functions. Such methods are particularly effective in handling challenging settings, including long-tailed relation distributions and multi-label classification. For example, ATL (Zhou et al., 2021), Balanced Softmax (Zhang et al., 2021), AML (Wei and Li, 2022), AFL (Tan et al., 2022a), SSR-PU (Wang et al., 2022), NCRL (Zhou and Lee, 2022), PEMSC (Guo et al., 2023), HingeABL (Wang et al., 2023), AMTL (Xu et al., 2025a), and APRDL (Xu et al., 2025b).

(3) Plugin-based Approaches. Plugin-based methods emphasize modularity and generalization, allowing them to be flexibly integrated into different DocRE architectures to further enhance performance. Representative works include LogicRE (Ru et al., 2021), MILR (Fan et al., 2022), BCBR (Liu et al., 2023), P³M (Wang et al., 2024b), JMRL (Qi et al., 2024), VaeDiff-DocRE (Tran et al., 2025), and AMTL (Xu et al., 2025a). Among them, LogicRE, MILR, BCBR and JMRL introduce logical reasoning modules to capture multi-hop relational dependencies, while P³M and VaeDiff-DocRE apply data augmentation techniques to improve model robustness.

In contrast to the aforementioned approaches, we innovatively decompose the DocRE task guided

by the *salient entity*. Building upon this decoupled structure, we leverage knowledge injection to reconstruct the text, effectively mitigating reasoning difficulty. Furthermore, our implementation is orthogonal to existing methods, offering the potential for further combined optimization.

3 Methodology

3.1 Problem Definition

Given a document D consisting of tokens $W_D = \{w_i\}_{i=1}^{|W_D|}$, sentences $X_D = \{x_i\}_{i=1}^{|X_D|}$, and entities $E_D = \{e_i\}_{i=1}^{|E_D|}$, the task of document-level relation extraction is to predict the relations between every ordered entity pair $(e_h, e_t) \in E_D \times E_D, h \neq t$. Each entity $e \in E_D$ is associated with a set of mentions $M_e = \{m_j\}_{j=1}^{|M_e|}$, where each mention is a phrase in the document referring to the entity. The relation label space is defined as $R \cup NA$, where R is a pre-defined set of relation types and NA denotes the absence of any relation. For each entity pair (e_h, e_t) , the model predicts a subset of relations $R_{h,t} \subseteq R \cup NA$, making DocRE a multi-label classification problem. Moreover, evidence sentences $V_{h,t} \subseteq X_D$ may also be provided for entity pairs with valid relations, representing the minimal set of sentences sufficient for human inference.

3.2 Approach Overview

The overall architecture of SegDRE is illustrated in Figure 2, consisting of four sections: (1) **Salient Entity Identification**, which identifies the salient entity in the document, serving as the pivot for subsequent task decomposition; (2) **Dense Relation Extraction**, which performs relation extraction over entity pairs involving the salient entity; (3) **Sparse Relation Extraction**, which targets over entity pairs not directly involving the salient entity; (4) **Extended Inference Stage Fusion**, which integrates predictions from the dense, sparse, and base models to produce the final results.

3.3 Salient Entity Identification

Current DocRE models generally obtain a representation embedding z_h for each entity e_h after encoding the document.

$$\{z_h\}_{e_h \in E_D} = \text{DocREModel}(D) \quad (1)$$

Traditionally, a pair of entity representations is processed by a bilinear classifier² to predict relation scores $y^{(s,o)} \in \mathbb{R}^{|\mathcal{R}|}$:

$$y^{(s,o)} = z_s W_r^T z_o + b_r, \quad (2)$$

where $W_r \in \mathbb{R}^{|\mathcal{R}| \times d \times d}$ and $b_r \in \mathbb{R}^{|\mathcal{R}|}$ are learnable parameters. The probability that relation $r \in \mathcal{R}$ holds between entity e_s and e_o is thus $P(r | e_s, e_o) = \sigma(y_r^{(s,o)})$, where σ is the sigmoid function.

To identify salient entities, we substitute the standard classification head with a count regression head. Given that valid relations between entity pairs are rare in document-level contexts, we employ a Zero-Inflated Poisson (ZIP) model (Lambert, 1992). Specifically, the relation count variable C for an entity pair is modeled as a mixture of a point mass at zero and a Poisson distribution:

$$\begin{aligned} C &\sim 0, && \text{with probability } \pi \\ &\sim \text{Poisson}(\lambda), && \text{with probability } 1 - \pi \end{aligned} \quad (3)$$

where π is the probability for the zero component, and λ is the expected count of the Poisson component. To parameterize this, the model first gets the relation count score $x^{(s,o)} \in \mathbb{R}^{\text{hidden}}$:

$$x^{(s,o)} = z_s W_c^T z_o + b_c, \quad (4)$$

then branches into two heads to predict the probability π and the Poisson rate λ :

$$\begin{aligned} \pi^{(s,o)} &= \sigma(W_\pi x^{(s,o)} + b_\pi) \\ \lambda^{(s,o)} &= \ln(1 + \exp(W_\lambda x^{(s,o)} + b_\lambda)), \end{aligned} \quad (5)$$

where σ denotes the sigmoid function, and the soft-plus function ensures $\lambda > 0$.

The model is optimized by minimizing the negative log-likelihood of the observed counts. The loss function \mathcal{L} is defined as:

$$\begin{aligned} \mathcal{L} = - \sum_{(s,o)} \log &\left[\pi^{(s,o)} \mathbb{I}(y = 0) \right. \\ &\left. + (1 - \pi^{(s,o)}) \frac{e^{-\lambda^{(s,o)}} (\lambda^{(s,o)})^y}{y!} \right], \end{aligned} \quad (6)$$

where $\mathbb{I}(\cdot)$ is the indicator function. During inference, we estimate the scalar count of potential relations for each entity pair using the expected value of the distribution.

$$c^{(s,o)} = (1 - \pi^{(s,o)}) \lambda^{(s,o)} \quad (7)$$

²In practice, a grouped bilinear classifier (Zheng et al., 2019) is applied to save memory.

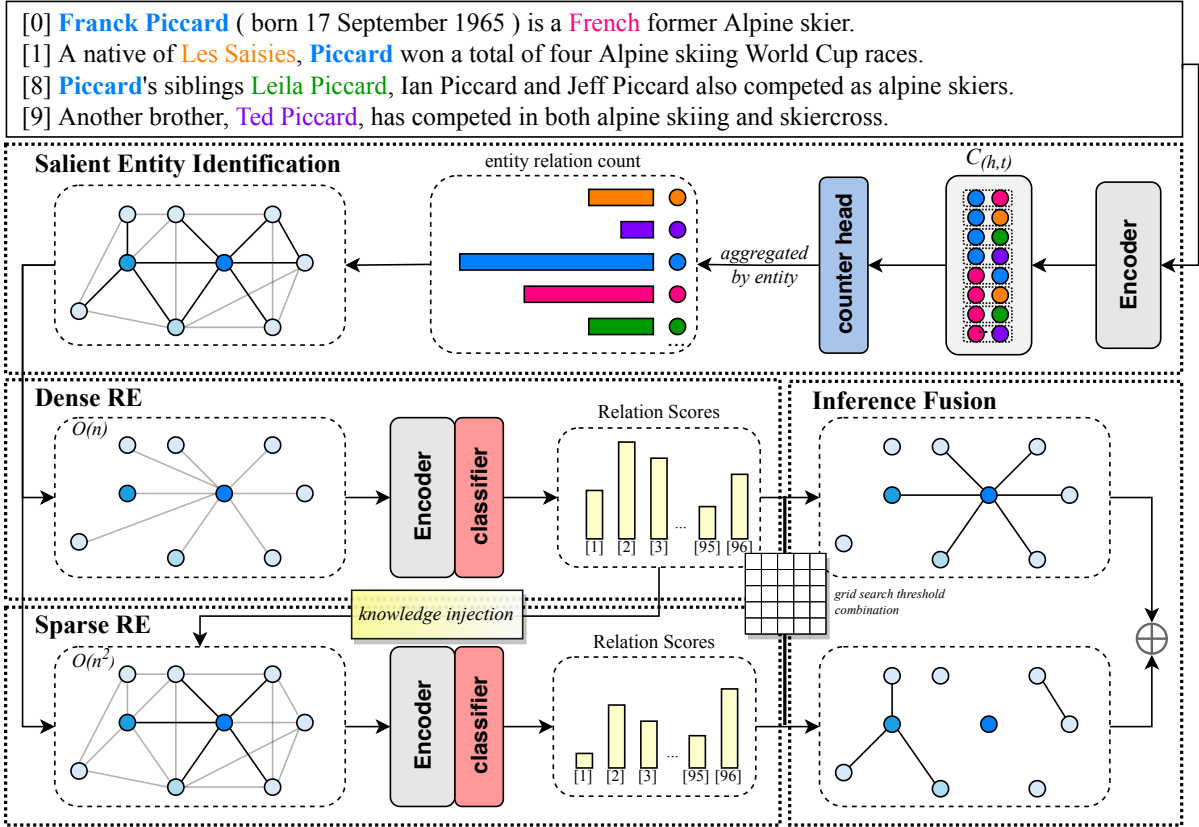


Figure 2: The overall architecture of SegDRE. It identifies salient entities to segregate the graph and injects triples from the dense section to the sparse section for relation extraction.

We then aggregate $c^{(s,o)}$ by entity. The entity with the highest aggregated relation count is identified as the *salient entity*, denoted as e_{sal} . To evaluate the capability of the identified e_{sal} to enrich relational information, in addition to accuracy, we introduce a metric named **Saliency**, defined as:

$$\text{Saliency} = \frac{\text{Count}(e_{sal})}{\text{Count}(e_{gold})}, \quad (8)$$

where e_{gold} denotes the ground-truth salient entity.

We set the *hidden* dimension of the count head to 768 and initialize it with shared parameters from a trained DocRE model for faster convergence. The model is trained for 10 epochs with a warm-up ratio of 0.3, while all other hyper-parameters are kept consistent with those of the base DocRE model. For model selection, we retain the checkpoint achieving the highest saliency score.

3.4 Dense Relation Extraction

Based on the identified salient entity e_{sal} , we decouple the relation triples in document D into two categories: *dense* triples that involve e_{sal} , and *sparse* triples that do not. For dense relation extraction, the candidate entity pairs number is sig-

nificantly reduced to $2 \times (|E_D| - 1)$ with over half of the original DocRED relation triples included, thereby guiding the model to focus on learning interactions around the salient entity without being overwhelmed by numerous negative samples. The extracted dense relation triples are denoted as $T_{dense} = \{(e_{sal}, r, e_j), (e_j, r, e_{sal}) \mid e_j \neq e_{sal}, r \in R\}$.

3.5 Sparse Relation Extraction

The sparse relation extraction stage targets relation triples that are complementary to those identified in the dense stage. Given that sparse triples may use salient entity as bridge to perform multi-hop reasoning, we leverage the dense predictions as prior knowledge to reduce the semantic difficulty. Technically, we adopt the injection mechanism from K-Bert (Liu et al., 2020) to incorporate these dense triples into the original document. The transformation could be formulated as:

$$D_{trans} = \text{K}_{\text{Inject}}(D, T_{dense}, \alpha), \quad (9)$$

where D_{trans} denotes the transformed document enriched with salient knowledge, and α controls

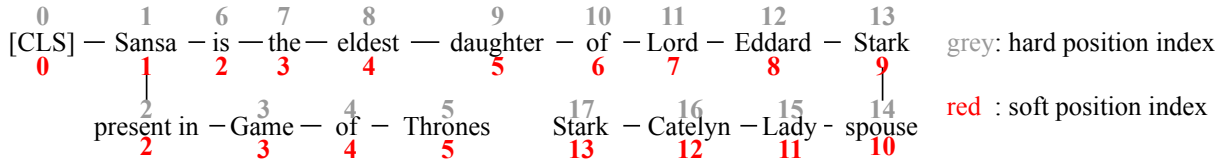


Figure 3: Illustration of knowledge injection in K-Bert(Liu et al., 2020).

the proportion of top-ranked triples from T_{dense} utilized for document reconstruction.

Specifically, the original document is restructured into a sentence tree, where the relation and tail entity are appended as additional branches to the head entity. To preserve the original word order while accommodating the expanded structure, K-Bert adopt a *soft-position index* mechanism, as illustrated in Figure 3, which rearranges the position indices of tokens to align with the original sequence. Meanwhile, a *visible matrix* is employed to restrict the attention scope of each token, defined as:

$$M_{ij} = \begin{cases} 0, & w_i \bowtie w_j \\ -\infty, & w_i \not\bowtie w_j, \end{cases} \quad (10)$$

where $w_i \bowtie w_j$ indicates that tokens w_i and w_j are located in the same branch of the sentence tree. This design prevents irrelevant entities from incorporating injected knowledge while facilitating mutual semantic interaction between head and tail entities. For example, in Figure 3, the relation *present in* between *Eddard Stark* and *Game of Thrones* could be easily extracted with the reconstructed context. The extracted sparse relation triples in this stage are denoted as $T_{sparse} = \{(e_i, r, e_j) \mid e_i, e_j \neq e_{sal}, r \in R\}$.

In documents where an entity appears with multiple mentions, we insert the extracted relation triples after a salient entity node and allow all its mentions to be mutually visible. This strategy ensures consistent information propagation across different mentions and prevents the transformed text from becoming excessively large for the encoder when numerous salient relations are present.

3.6 Extended Inference Stage Fusion

For DocRE model extracting evidence sentences simultaneously, Xie et al. (2022) proposed Inference Stage Fusion (ISF) to reconcile predictions from original document D and pseudo document D' , the concatenation of evidence sentences. Specifically, the two sets of prediction scores are fused via a blending layer (Wolpert, 1992) to form the final

prediction:

$$P(r \mid e_s, e_o) = P(r \mid e_s, e_o; D) + P(r \mid e_s, e_o; D') - \tau, \quad (11)$$

where τ is a threshold tuned on the development set.

We extend this fusion strategy to integrate T_{dense} and T_{sparse} into T_{base} , the base DocRE model’s predictions. Concretely, the threshold τ tuned on the development set is treated as an anchor point, around which a grid search is performed to determine the optimal offsets for τ_{base} , τ_{dense} and τ_{sparse} on the development set. For models that do not employ the ISF mechanism, such as ATLOP, we set the anchor threshold $\tau = 0$. The search range is set to $[\tau - 1, \tau + 4]$ with a step size of 0.1 for DocRED, and $[\tau - 2, \tau + 5]$ with a step size of 0.1 for Re-DocRED.

Datasets	DocRED		Re-DocRED		
	Train	Dev	Train	Dev	Test
#Docs	3,053	1,000	3,053	500	500
Avg. #Entities	19.5	19.6	19.4	19.4	19.6
Avg. #Entity Pairs	392.6	392.2	390.8	386.5	397.3
Avg. #Triples	12.5	12.3	28.1	34.6	34.9
Avg. #Sentences	7.9	8.1	7.9	8.2	7.9
Avg. #Relations	5.4	5.3	8.6	10.0	9.5

Table 1: Statistics of DocRED and Re-DocRED. Avg.# denotes the average number of items per document.

4 Experiments Settings

4.1 Datasets and Evaluation Metrics

We conduct experiments with our proposed method on two benchmark datasets: DocRED (Yao et al., 2019) and Re-DocRED (Tan et al., 2022b), a re-labeled version of DocRED.

Following Yao et al. (2019), we employ F1 and Ign-F1 as evaluation metrics. The Ign-F1 metric excludes relational facts that are already present in the training set. We also report Intra-F1 for relations within a single sentence and Inter-F1 for relations that span multiple sentences.

Method	Dev				Test	
	Ign-F1	F1	Intra-F1	Inter-F1	Ign-F1	F1
<i>implemented with BERT_{base}</i>						
ATLOP (Zhou et al., 2021)	59.22	61.09	67.26*	53.20*	59.31	61.30
ATLOP-Seg (this work)	59.52±0.06	61.53±0.07	67.80±0.12	53.86±0.18	59.22	61.49
DREEAM (Ma et al., 2023)	60.51	62.55	68.39 [†]	55.22 [†]	60.03	62.49
DREEAM-Seg (this work)	60.94±0.05	63.00±0.03	68.93±0.08	55.73±0.09	60.61	62.88
<i>implemented with RoBERTa_{large}</i>						
ATLOP (Zhou et al., 2021)	61.32	63.18	69.61*	55.01*	61.39	63.40
ATLOP-Seg (this work)	61.68±0.14	63.68±0.16	69.99±0.23	56.01±0.15	61.61	63.79
DREEAM (Ma et al., 2023)	62.29	64.20	70.16 [†]	56.57 [†]	62.12	64.27
DREEAM-Seg (this work)	62.78±0.13	64.73±0.14	70.72±0.12	57.30±0.23	62.63	64.80

Table 2: Performance of base models with and without SegDRE on the DocRED dataset. Results with* are referred from Xie et al. (2022) and with[†] are our re-implementations. Others are reported in original papers.

Method	Dev		Test	
	Ign-F1	F1	Ign-F1	F1
ATLOP (Zhou et al., 2021)	76.88*	77.63*	76.94*	77.73*
ATLOP-HingeABL(Wang et al., 2023)	-	-	78.82	79.79
ATLOP-PEMSCL(Guo et al., 2023)	79.02	79.89	79.01	79.86
ATLOP-P ³ M(Wang et al., 2024b)	-	-	78.82	<u>80.02</u>
ATLOP-APRDL(Xu et al., 2025b)	78.54	79.40	<u>79.10</u>	79.90
ATLOP-AMTL(Xu et al., 2025a)	79.15	80.09	79.04	79.97
ATLOP-Seg (this work)	79.48±0.13	80.16±0.11	79.63	80.25
DREEAM (Ma et al., 2023)	-	-	77.39	77.94
DREEAM-JMRL(Qi et al., 2024)	-	-	77.98	78.61
DREEAM-AMTL(Xu et al., 2025a)	-	-	<u>79.40</u>	<u>79.98</u>
DREEAM-Seg (this work)	79.30±0.16	79.97±0.13	79.68	80.25
DREEAM w. ISF(Ma et al., 2023)	-	-	79.66	80.73
DREEAM-APRDL w. ISF(Xu et al., 2025b)	80.38	81.05	<u>80.66</u>	<u>81.28</u>
DREEAM-Seg w. ISF (this work)	80.13±0.24	80.83±0.23	80.67	81.32

Table 3: Performance of base models with different augmentation on the Re-DocRED dataset using RoBERTa_{large}. Results with* are sourced from Tan et al. (2022b). Others are reported in original papers. The best results under the same setting are highlighted in **bold**, and the second-best results are underlined.

4.2 Baselines

We organize baselines along two comparison axes. First, we instantiate our method with ATLOP and DREEAM as base methods to assess its generality across implementations. We further compare against representative loss-based optimization methods and plugin-based approaches, including HingeABL, PEMSCL, P³M, APRDL, AMTL, and JMRL, under the same setting.

4.3 Implementation Details

We implemented our approach by PyTorch (Paszke et al., 2019) and HuggingFace Transformers (Wolf et al., 2019). Experiments using BERT_{base} were conducted on a single NVIDIA 5060Ti 16GB GPU, while those using RoBERTa_{large} were conducted on a single NVIDIA 4090 24GB GPU. We utilized

the public repositories of ATLOP³ and DREEAM⁴ and set all hyper-parameters according to their the original papers throughout training and evaluation. To mitigate potential bias, we present the average results and standard deviations of our proposed methods over 5 independent runs. The results on the DocRED test set were obtained by submitting the predictions to CodaLab⁵.

5 Main Results and Ablation Study

5.1 Main Results

Table 2 presents the performance of our proposed SegDRE framework applied to the baselines, ATLOP and DREEAM, on the DocRED dataset. As observed, equipping base models with SegDRE

³<https://github.com/wzhoudad/ATLOP>

⁴<https://github.com/YoumiMa/dreeam>

⁵<https://codalab.lisn.upsaclay.fr/competitions/365#results>

consistently yields improvements across most metrics. Specifically, with BERT_{base}, SegDRE improves ATLOP and DREEAM by 0.19, 0.39 in F1 score, respectively. Similarly, when encoded with the RoBERTa_{large} encoder, our approach improves the F1 scores of ATLOP and DREEAM by 0.39 and 0.53, respectively. Notably, SegDRE boosts Inter-F1 scores of ATLOP and DREEAM by 1.00 and 0.73 on RoBERTa_{large}, demonstrating its strength in complex multi-hop reasoning.

Table 3 shows our results on the Re-DocRED dataset, where the performance gaps between base models with and without our augmentation are more pronounced. Our approach elevates the F1 score of ATLOP, DREEAM w.o. ISF (Inference Stage Fusion) and DREEAM w. ISF by 2.52, 2.31 and 0.59, respectively. Corresponding improvements of 2.69, 2.29, and 1.01 are also observed in the Ign-F1 scores. Our approach exhibits superior performance compared to competitive enhancements such as JMRL, APRDL, and AMTL. In the ATLOP setting, ATLOP-SegDRE achieves a Test F1 of 80.25, outperforming the previous best variant ATLOP-P³M (80.02) and ATLOP-AMTL (79.97). Similarly, in the DREEAM setting, our method demonstrates significant advantages, surpassing DREEAM-AMTL with a Test F1 of 80.25. When combined with ISF, DREEAM-SegDRE w. ISF achieves a Test F1 of 81.32, slightly outperforming the state-of-the-art loss-based approach, APRDL.

5.2 Ablation Study

To examine the effectiveness of different components in our approach, we conduct a series of ablation studies on DocRED using DREEAM encoded by BERT_{base}, and the corresponding results are presented in Table 4. The detailed analysis is outlined below:

w/o salient entity identification. Removing the salient entity identification task and assigning random entity as salient results in a decrease in performance, indicating the necessity of accurately identifying the salient entity for effective goal decomposition.

w/o dense or sparse. Excluding either the dense or sparse relation extraction components leads to a degradation in performance. This demonstrates that both components contribute significantly to overall performance.

w/o knowledge injection. We remove the knowledge injection mechanism, directly feeding the

original document text to train the model and infer on sparse triples. This leads to a significant ~ 9 drop in F1 score on the sparse dev set (see Table 7), which in turn imposes a bottleneck on the overall results, thereby demonstrating that the knowledge injection is indispensable for both sparse and final fused predictions.

w/o threshold grid search. Directly merging dense and sparse predictions with positive scores leads to a slight performance drop. This indicates that the proposed approach already benefits from the dense/sparse decomposition under a simple fusion scheme, while threshold grid search further refines the final predictions by calibrating the confidence of different components.

naive ensemble. We compare SegDRE with a naive ensemble of two independent base models. While this ensemble outperforms a single model, it still lags behind SegDRE. This confirms that our performance gains stem from the explicit decomposition into heterogeneous learning objectives (dense vs. sparse), rather than the mere combination of homogeneous predictions.

Model	Ign-F1	F1
DREEAM-Seg with BERT _{base}	60.98	63.06
w/o salient entity identification	60.62	62.45
w/o dense	60.84	62.82
w/o sparse	60.81	62.80
w/o knowledge injection	60.83	62.83
w/o threshold grid search	60.92	62.84
naive ensemble	60.72	62.79
DREEAM with BERT _{base}	60.60	62.50

Table 4: Ablation study on DocRED dev set.

6 Further Analysis

6.1 On Selecting Salient Entity

As the pivotal entry point of the SegDRE, salient entity plays a dual role: it determines whether the dense stage is assigned the correct scope of sample pairs, and it dictates the quality of the intermediate knowledge injected into the subsequent sparse stage. To evaluate the effectiveness of our proposed ZIP-based identifier, we compare it against four alternative selection strategies: (1) **Max Mention**: selecting the entity with the highest frequency of occurrence; (2) **Edit Distance**: selecting the entity whose name has the smallest Levenshtein distance to the document title; (3) **Aggregation**: counting the number of relations predicted by a standard relation extractor to identify the entity with the most

relations; and (4) **Regression (linear)**: predicting relation counts with a standard linear regression objective.

Selection Method	Accuracy	Saliency	Dev	Dev _{dense}	Dev _{sparse}
Regression (linear)	73.86	84.52	62.95	71.81	40.13
Regression (ZIP)	74.35	85.76	63.06	71.84	40.79
Aggregation	69.14	83.61	62.93	70.66	39.46
Max Mention	38.06	54.67	62.68	61.03	24.92
Edit Distance	36.29	55.53	62.59	66.04	23.75

Table 5: Performance comparison on DocRED dev set across different selection methods.

The comparative results are presented in Table 5. We observe that heuristic rules (Max Mention and Edit Distance) exhibit poor identification accuracy ($< 40\%$). Consequently, this misidentification leads to a performance drop in the downstream extraction, particularly in the sparse stage (Dev_{sparse} drops from 40.79 to ~ 24). This phenomenon strongly validates our motivation: the reasoning of sparse relations relies heavily on the injection of high-quality semantic information derived from the correct salient entity. When the salient entity is misidentified, the bridge for knowledge transfer is broken.

Furthermore, although linear count regression is competitive, it still underperforms ZIP-based regression, indicating that explicitly modeling zero inflation is beneficial for salient-entity estimation under highly sparse DocRE labels. The same trend holds against the aggregation strategy, confirming that the ZIP design is a statistically aligned choice for our decomposition pipeline.

6.2 Impact of Ratio α

In our proposed approach, α determines the percentage of top-ranked T_{dense} injected into the document during inference to balance error propagation and semantic richness. We observe that the final extraction performance exhibits an increasing trend as α increases. Therefore, we fix α at 100% for our approach. Additionally, we conducted an analysis on the impact of knowledge injection from prediction of base model and ground-truth relation triples. Please refer to Appendix A.2 for more details.

6.3 Threshold Shift Analysis

To better understand the behavior of our inference stage fusion strategy, we record the distribution of threshold shifts for each component after grid search on the development set in Figure 5. We observe that the optimal thresholds for the dense

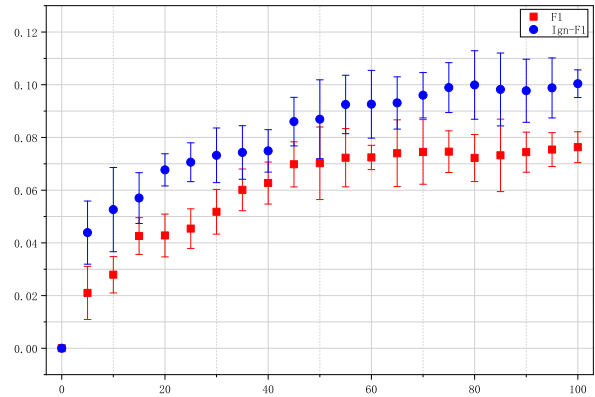


Figure 4: F1 gains on dev set with varying α (dense predictions) compared to the $\alpha = 0$ setting. Thresholds are fixed to save time.

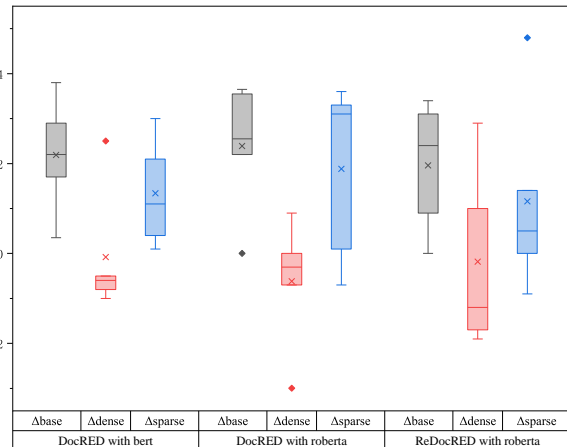


Figure 5: Threshold shift distribution for each part.

part are generally lower than the anchor threshold, while the thresholds associated with the base and sparse parts are mostly shifted upward. This asymmetric shift pattern suggests that the fusion process implicitly favors dense relations as the primary source of reliable predictions, while selectively incorporating complementary information from base and sparse parts.

6.4 Case Study

We conduct a case study to qualitatively analyze the behavior of different stages in our approach. The results show that the dense stage focuses on accurately extracting relations centered on the salient entity, while the sparse stage improves the recall of multi-hop relations. Their combination leads to more complete and accurate predictions. Detailed examples are provided in the Appendix A.3.

7 Conclusions

In this paper, we propose a model-agnostic approach that decomposes the complex extraction task into distinct sub-objectives driven by salient entity identification. By explicitly leveraging these entities to decouple both training and inference process, our approach significantly narrows the search space for salient entity relations and mitigates the difficulty of multi-hop reasoning via reconstructing the document context. Extensive experiments verify that our approach yields consistent improvements across various backbone models and achieves state-of-the-art performance among competitive enhancement methods.

Acknowledgements

This work is supported by National Natural Science Foundation of China (under project No. 72274138).

Limitations

Our method has certain limitations that should be acknowledged. First, our method assumes the existence of a single salient entity to guide task decomposition. Although valid for the datasets used in our experiments, this assumption may be restrictive for documents that contain multiple entity clusters with densely interconnected relations. Second, for simplicity and fair comparison, we adopt unified hyper-parameter settings, training schemes, and model architectures across different stages of SegDRE. While this design choice highlights the generality of SegDRE, allowing each component to employ task-specific modeling strategies could further improve the overall extraction performance. We leave the exploration of such task-specific combinations to future work.

References

- Shengda Fan, Shasha Mo, and Jianwei Niu. 2022. [Boosting document-level relation extraction by mining and injecting logical rules](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10311–10323, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Jia Guo, Stanley Kok, and Lidong Bing. 2023. [Towards integration of discriminability and robustness for document-level relation extraction](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*,

pages 2606–2617, Dubrovnik, Croatia. Association for Computational Linguistics.

- Diane Lambert. 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.
- Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. [K-bert: Enabling language representation with knowledge graph](#). In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 2901–2908.
- Yichun Liu, Zizhong Zhu, Xiaowang Zhang, Zhiyong Feng, Daoqi Chen, and Yaxin Li. 2023. [Document-level relationship extraction by bidirectional constraints of beta rules](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2256–2266, Singapore. Association for Computational Linguistics.
- Chonggang Lu, Richong Zhang, Kai Sun, Jaemin Kim, Cunwang Zhang, and Yongyi Mao. 2023. [Anaphor assisted document-level relation extraction](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15453–15464, Singapore. Association for Computational Linguistics.
- Kangqi Luo, Fengli Lin, Xusheng Luo, and Kenny Zhu. 2018. [Knowledge base question answering via encoding of complex query graphs](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2185–2194, Brussels, Belgium. Association for Computational Linguistics.
- Youmi Ma, An Wang, and Naoaki Okazaki. 2023. [DREEM: guiding attention with evidence for improving document-level relation extraction](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2023, Dubrovnik, Croatia, May 2-6, 2023*, pages 1963–1975. Association for Computational Linguistics.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. [Pytorch: An imperative style, high-performance deep learning library](#). In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 8024–8035.
- Kunxun Qi, Jianfeng Du, and Hai Wan. 2024. [End-to-end learning of logical rules for enhancing document-level relation extraction](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7247–7263, Bangkok, Thailand. Association for Computational Linguistics.

- Dongyu Ru, Changzhi Sun, Jiangtao Feng, Lin Qiu, Hao Zhou, Weinan Zhang, Yong Yu, and Lei Li. 2021. [Learning logic rules for document-level relation extraction](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1239–1250, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2020. [Recurrent interaction network for jointly extracting entities and classifying relations](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3722–3732, Online. Association for Computational Linguistics.
- Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng. 2022a. [Document-level relation extraction with adaptive focal loss and knowledge distillation](#). In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1672–1681, Dublin, Ireland. Association for Computational Linguistics.
- Qingyu Tan, Lu Xu, Lidong Bing, Hwee Tou Ng, and Sharifah Mahani Aljunied. 2022b. [Revisiting docred - addressing the false negative problem in relation extraction](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 8472–8487. Association for Computational Linguistics.
- Khai Phan Tran, Wen Hua, and Xue Li. 2025. [VaeDiff-DocRE: End-to-end data augmentation framework for document-level relation extraction](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7307–7320, Abu Dhabi, UAE. Association for Computational Linguistics.
- Jianing Wang, Qiushi Sun, Xiang Li, and Ming Gao. 2024a. [Boosting language models reasoning with chain-of-knowledge prompting](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4958–4981, Bangkok, Thailand. Association for Computational Linguistics.
- Jize Wang, Xinyi Le, Xiaodi Peng, and Cailian Chen. 2023. [Adaptive hinge balance loss for document-level relation extraction](#). In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3872–3878, Singapore. Association for Computational Linguistics.
- Ye Wang, Xinxin Liu, Wenxin Hu, and Tao Zhang. 2022. [A unified positive-unlabeled learning framework for document-level relation extraction with different levels of labeling](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4123–4135, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Ye Wang, Huazheng Pan, Tao Zhang, Wen Wu, and Wenxin Hu. 2024b. [A positive-unlabeled metric learning framework for document-level relation extraction with incomplete labeling](#). In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 19197–19205.
- Ying Wei and Qi Li. 2022. [Sagdre: Sequence-aware graph-based document-level relation extraction with adaptive margin loss](#). In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '22*, page 2000–2008, New York, NY, USA. Association for Computing Machinery.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. [Huggingface’s transformers: State-of-the-art natural language processing](#). *arXiv preprint arXiv:1910.03771*.
- David H Wolpert. 1992. Stacked generalization. *Neural networks*, 5(2):241–259.
- Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, and Jiawei Han. 2022. [Eider: Empowering document-level relation extraction with efficient evidence extraction and inference-stage fusion](#). In *Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 257–268. Association for Computational Linguistics.
- Huangming Xu, Fu Zhang, and Jingwei Cheng. 2025a. [An adaptive multi-threshold loss and a general framework for collaborating losses in document-level relation extraction](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20996–21007, Vienna, Austria. Association for Computational Linguistics.
- Huangming Xu, Fu Zhang, Jingwei Cheng, and Xin Li. 2025b. [Arpdl: adaptive relational prior distribution loss as an adapter for document-level relation extraction](#). In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI '25*.
- Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 2019. [DocRED: A large-scale document-level relation extraction dataset](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 764–777, Florence, Italy. Association for Computational Linguistics.
- Shuang Zeng, Yuting Wu, and Baobao Chang. 2021. [SIRE: Separate intra- and inter-sentential reasoning for document-level relation extraction](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 524–534, Online. Association for Computational Linguistics.
- Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. 2020. [Double graph based reasoning for document-level relation extraction](#). In *Proceedings of the 2020*

Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1630–1640, Online. Association for Computational Linguistics.

Fu Zhang, Qi Miao, Jingwei Cheng, Hongsen Yu, Yi Yan, Xin Li, and Yongxue Wu. 2024. [SRF: Enhancing document-level relation extraction with a novel secondary reasoning framework](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15426–15439, Miami, Florida, USA. Association for Computational Linguistics.

Fu Zhang, Yi Yan, and Jingwei Cheng. 2025. [Document-level relation extraction with global relations and entity pair reasoning](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 19556–19567, Vienna, Austria. Association for Computational Linguistics.

Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. 2021. [Document-level relation extraction as semantic segmentation](#). In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 3999–4006. ijcai.org.

Rui Zhang, Bayu Distiawan Trisedya, Miao Li, Yong Jiang, and Jianzhong Qi. 2022. [A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning](#). *The VLDB Journal*, 31:1143–1168.

Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. [Graph convolution over pruned dependency trees improves relation extraction](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215, Brussels, Belgium. Association for Computational Linguistics.

Heliang Zheng, Jianlong Fu, Zheng-Jun Zha, and Jiebo Luo. 2019. [Learning deep bilinear transformation for fine-grained image representation](#). In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. [Document-level relation extraction with adaptive thresholding and localized context pooling](#). In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 14612–14620.

Yang Zhou and Wee Sun Lee. 2022. [None class ranking loss for document-level relation extraction](#). In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 4538–4544. ijcai.org.

Hao Zhu, Yankai Lin, Zhiyuan Liu, Jie Fu, Tat-Seng Chua, and Maosong Sun. 2019. [Graph neural networks with generated parameters for relation extraction](#). In *Proceedings of the 57th Annual Meeting of*

the Association for Computational Linguistics, pages 1331–1339, Florence, Italy. Association for Computational Linguistics.

A Appendix

A.1 On Defining Salient Entity

Salient Standard In addition to **Max Relation**, we consider two alternatives: **Max Mention** and **Max Intra-relation**. The Max Mention standard assumes that the most frequently mentioned entity represents the semantic focus. Max Intra-relation generally aligns with Max Relation but may better capture the specific topic in cases of disagreement. For instance, in the document *Sansa Stark*, Max Intra-relation correctly identifies the character as the focus, whereas Max Relation points to the broader parent entity *A Song of Ice and Fire*.

However, quantitative analysis in Table 6 reveals that Max Mention lacks sufficient coverage in both 1-hop (essential for positive enrichment) and 2-hop (crucial for bridging) settings. While Max Intra-relation is thematically more precise, it offers lower structural coverage and yielded slightly inferior performance in our experiments compared to the one guided by Max Relation. Consequently, we retain Max Relation as the defining standard.

Selection Criterion	DocRED		Re-DocRED	
	1-hop	2-hop	1-hop	2-hop
Max Relation	61.12	87.27	54.09	90.89
Max Intra-relation	52.83	83.10	49.51	89.15
Max Mention	31.09	67.45	27.09	76.82

Table 6: Average coverage (%) of relations per document within 1-hop and 2-hops from salient entities.

Number of Salient Entities Our decision to select a single salient entity is primarily driven by the specific characteristics of the DocRED dataset. Constructed from Wikipedia, each document in DocRED is naturally grounded by a title entity, making a unique semantic anchor appropriate.

Empirically, we justified this choice by analyzing coverage efficiency. We constructed a greedy entity chain by iteratively selecting the entity covering the most remaining relations. As illustrated in Figure 6, the number of covered relations follows an approximate exponential decay as the chain order increases. However, the search space in the dense phase grows linearly with the number of salient entities. Consequently, selecting only the

top-ranked entity achieves the highest positive sample density, while avoiding the scarcity of relation labels in the sparse relation extraction that leads to insufficient supervision.

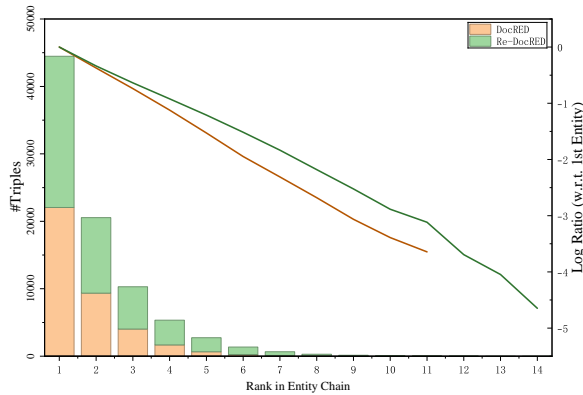


Figure 6: Relation coverage statistics across the greedy entity chain. The left axis displays the absolute number of covered relations, while the right axis shows the logarithmic ratio relative to the top-ranked entity.

A.2 Knowledge Injection

Impact of injecting base prediction When injecting T_{base} as intermediate knowledge, we observe that the performance initially improves but subsequently declines as the injection ratio α increases in Figure 7. This trend suggests that knowledge injection requires a careful balance between knowledge richness and error propagation.

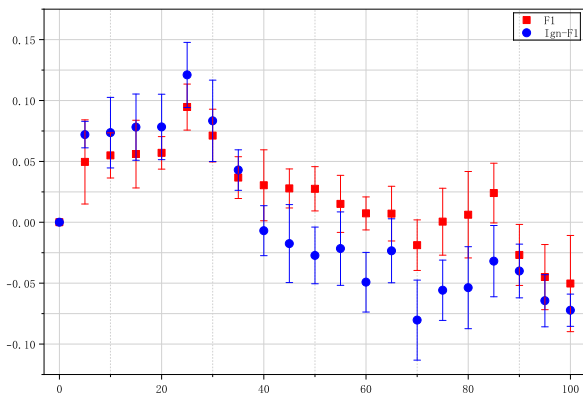


Figure 7: F1 gains on dev set with varying α (base predictions) compared to the $\alpha = 0$ setting. Thresholds are fixed to save time.

Impact of injecting ground-truth We also examine the performance at the sparse stage when ground-truth labels related to salient entities are applied to enrich the original text. Table 7 demonstrates that the performance gap between using predicted versus ground-truth dense relations is

minimal. This indicates that our sparse extraction module is highly robust: it effectively leverages the semantic guidance from the dense stage even when the intermediate predictions contain noise, validating the practicality of our pipeline approach.

Standard	Ign-F1	F1
DocRED with BERT _{base}		
ground-truth	40.55±0.22	41.12±0.15
dense inference	40.31±0.29	41.05±0.23
original text	30.97	31.63
Re-DocRED with RoBERTa _{large}		
ground-truth	64.77±0.15	65.20±0.13
dense inference	64.74±0.22	65.06±0.27

Table 7: Performance comparison of knowledge injection on (Re-)DocRED sparse dev set: ground-truth vs dense inference. The baseline performance on the original text (*w/o* injection) is included for reference.

A.3 Case Study

Figure 8 presents one of case studies, which illustrates how our approach enhances the base method DREEAM. Firstly, in the dense RE stage, the model identifies more accurate relations involving the salient entity "Franck". Then in the sparse RE stage, by transforming the original document with these triples as intermediate knowledge, the model successfully extracts multi-hop relations like ("Les Saisies", "French", country) and ("Leila", "Ted", sibling). Above all, SegDRE achieves better performance on this example.

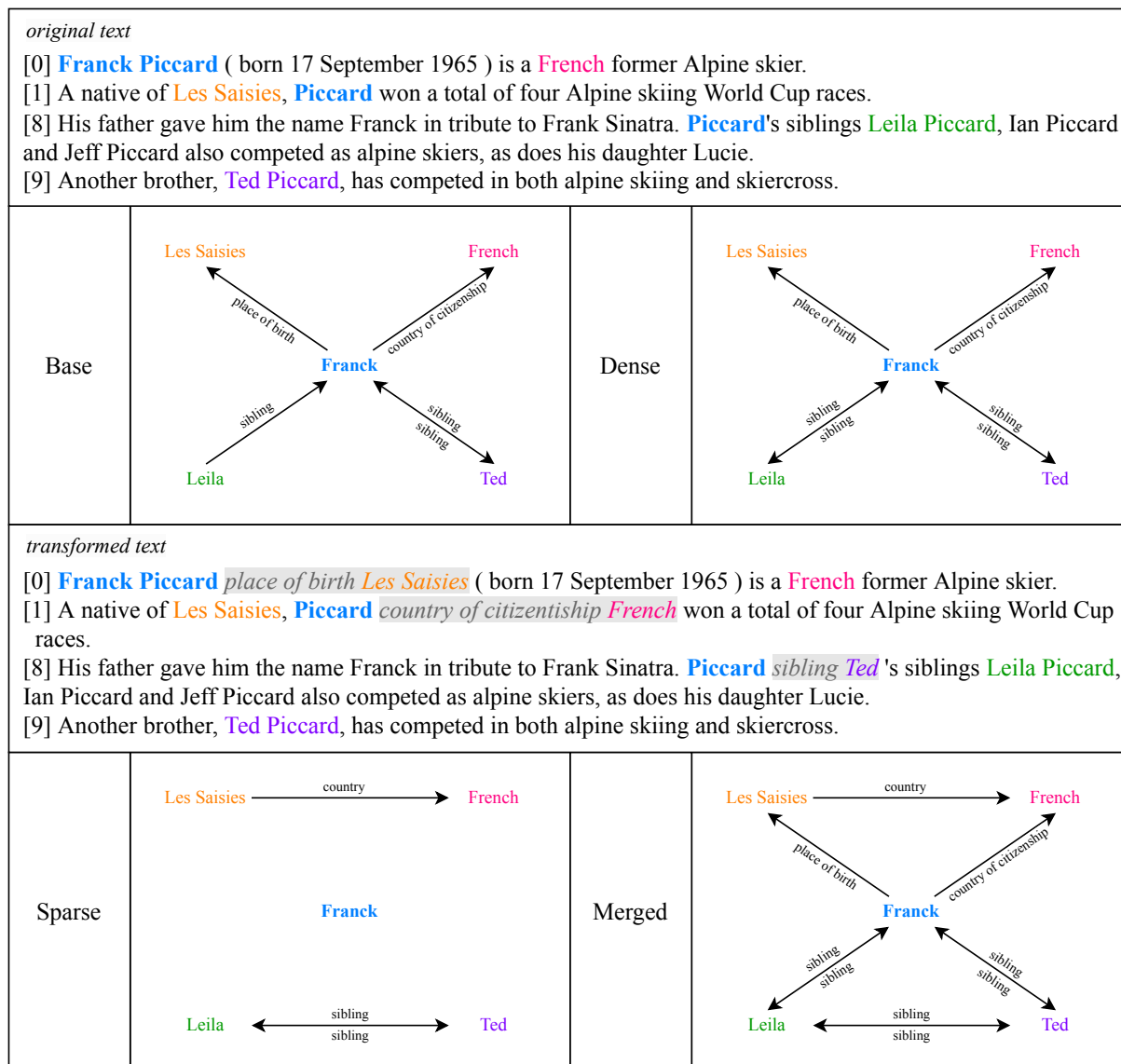


Figure 8: A case study comparing DREEAM and DREEAM-Seg on a document from DocRED dev set. Part of entities, relations and sentences are displayed due to space limit.