ANAVIT: Enhancing Document-Level Relation Extraction with Anaphor Nodes and Visual Transformation

An Duc Vinh Pham, Quynh-Trang Pham Thi, Thanh-Hai Dang*
Vietnam National University
University of Engineering and Technology
Hanoi, Viet Nam
{25025012,trangptq,hai.dang}@vnu.edu.vn

***Abstract*—Document-level relation extraction is crucial for understanding complex biomedical relationships spanning multiple sentences. While transformer-based approaches like ATLOP have shown promise through adaptive thresholding and localized context pooling, they struggle with capturing anaphoric references and global relational patterns. We propose ANAVIT (Anaphor-Assisted Visual Transformation), a method that enhances document-level relation extraction by: (1) integrating anaphor nodes in document graphs to explicitly model pronoun-entity relationships through Graph Attention Networks, and (2) applying visual transformation to relation matrices using convolutional neural networks and multi-head attention to capture global patterns. Our proposed method introduces graph-based enhancements for anaphora resolution. Experimental results on CDR and GDA biomedical datasets demonstrate that ANAVIT achieves state-of-the-art performance with 79.1% F1 on CDR, significantly outperforming ATLOP (69.4%) and other baselines. The improvements are particularly pronounced in documents with complex coreference chains. Our source-code: https://github.com/vinhpad/anavit**

***Index Terms*—document-level relation extraction, graph neural networks, anaphor resolution, visual transformation, adaptive thresholding**

I. Introduction

Document-level relation extraction (DocRE) aims to identify semantic relationships between entities across multiple sentences within a document, extending beyond traditional sentence-level approaches [1]. This task is particularly crucial in biomedical domains where complex relationships, such as Chemical-Induced Diseases (CID) and Gene-Disease Associations (GDA), often span entire abstracts or articles with intricate interaction patterns.

Traditional relation extraction methods focus on intra-sentence relationships, limiting their ability to capture cross-sentence dependencies essential for comprehensive document understanding [2]. Document-level extraction presents unique challenges: (1) entities may be mentioned multiple times across different sentences, (2) relationships can be expressed through complex reasoning chains involving multiple entities, and (3) anaphoric references (pronouns and definite references) create implicit connections that standard models struggle to resolve.

Recent transformer-based approaches have made significant progress in DocRE. ATLOP [3] introduced adaptive thresholding and localized context pooling, achieving substantial improvements over earlier methods. DocUnet [4] leveraged U-Net architectures for capturing document-level patterns. However, these approaches have limitations: they inadequately handle anaphoric references that are crucial for maintaining entity coherence across sentences, and they lack explicit mechanisms for modeling global relational patterns that emerge from entity interactions throughout the document.

To address these limitations, we propose ANAVIT (Anaphor-Assisted Visual Transformation), a novel framework that enhances document-level relation extraction through two key innovations. First, we construct heterogeneous document graphs that explicitly include anaphor nodes representing pronouns and definite references, enabling our Graph Attention Network (GAT) to model coreference relationships alongside traditional entity mentions. Second, we introduce a visual transformation module that treats entity relation matrices as spatial patterns, applying convolutional neural networks and multi-head attention to capture global interaction patterns.

Our contributions are threefold: (1) We propose a heterogeneous graph construction method that explicitly models anaphoric relationships through dedicated anaphor nodes, improving entity representation quality. (2) We introduce a visual transformation approach that applies spatial processing techniques to relation matrices, enabling the capture of global relational patterns. (3) We achieve state-of-the-art results on CDR (79.1% F1) and competitive performance on GDA (84.7% F1), demonstrating significant improvements over existing methods.

II. Related Work

A. Document-Level Relation Extraction

Early document-level relation extraction methods relied on traditional machine learning with hand-crafted features [5], [6]. The transition to neural approaches began with CNN-

*Corresponding author: hai.dang@vnu.edu.vn

based methods [1] and evolved to LSTM-based architectures [7], [8] that could capture cross-sentence dependencies.

Graph-based methods have shown significant promise by explicitly modeling document structure. Early work by [9] explored distant supervision beyond sentence boundaries. EoG [10] introduced edge-oriented graphs for document-level extraction, while LSR [11] proposed latent structure refinement. The integration of attention mechanisms with graph structures has further improved performance.

Transformer-based approaches have achieved state-of-the-art results. BRAN [12] was among the first to apply BERT [13] for document-level relation extraction but struggled with long-range dependencies. Subsequent work explored fine-tuning strategies. ATLOP [3] introduced adaptive thresholding and localized context pooling, significantly improving performance on biomedical datasets. Recent advances include semantic segmentation approaches [4].

B. Graph Neural Networks for Relation Extraction

Graph Convolutional Networks (GCNs) [14] have been widely adopted for relation extraction tasks. Variants include Graph Attention Networks (GATs) [15] that learn attention weights between nodes, and deeper architectures [16]. ResGCN [17] introduced residual connections for better gradient flow, while DropEdge [18] addressed overfitting in deep graph networks. CID-GCN [19] specifically targeted chemical-induced disease relation extraction using graph structures.

C. Biomedical Relation Extraction

The biomedical domain presents unique challenges due to complex entity relationships and specialized terminology. The BioCreative V CDR task [20] established important benchmarks for chemical-disease relation extraction. SciBERT [21] provided domain-specific pre-trained models that significantly improved biomedical NLP tasks.

D. Anaphora Resolution in Relation Extraction

Anaphora resolution has been recognized as crucial for document understanding, yet few relation extraction methods explicitly address this challenge. Traditional coreference resolution systems operate independently from relation extraction, leading to error propagation. Recent work has begun integrating coreference information directly into relation extraction models, but comprehensive approaches that explicitly model anaphoric relationships in graph structures remain limited.

E. Visual Processing for NLP

The application of computer vision techniques to natural language processing has gained attention, particularly for document-level tasks. U-Net models have been applied to document-level relation extraction [4]. Visual processing approaches treat textual relationships as spatial patterns, enabling the use of convolutional neural networks for capturing local and global dependencies. Our work extends this concept by applying visual transformation specifically to relation matrices enhanced with anaphoric information.

III. Methodology

Our ANAVIT model consists of two main modules: GAT-based entity enhancement and Visual Transformation. The GAT module processes heterogeneous document graphs containing both entity mentions and anaphor nodes, generating enhanced entity representations that capture coreference relationships. The Visual Transformation module treats enhanced entity relationships as spatial patterns, applying convolutional operations and multi-head attention to capture global relational dependencies.

A. Problem Formulation

Given a document d with entities $\{e_i\}_{i=1}^n$, document-level relation extraction aims to predict relations from $\mathcal{R} \cup \{\text{NA}\}$ for all entity pairs $(e_s, e_o)_{s,o=1...n;s\neq o}$, where $\mathcal{R}$ is a predefined relation set and NA indicates no relation. Each entity e_i may have multiple mentions $\{m_i^j\}_{j=1}^{N_{e_i}}$. A relation exists between entities (e_s, e_o) if it is expressed by any pair of their mentions. Entity pairs that do not express any relation are labeled NA. At test time, the model predicts labels for all entity pairs $(e_s, e_o)_{s,o=1...n;s\neq o}$ in document d.

B. Encoder

The encoder learns contextualized word representations using pre-trained BERT [13], [22]. For biomedical domains, we utilize SciBERT [21], which has been specifically pre-trained on biomedical literature. Given a document $d = [x_t]_{t=1}^l$ where l is document length, the encoder processes the entire document once, following the approach established by [12]:

$$H = [h_1, h_2, ..., h_l] = \text{BERT}([x_1, x_2, ..., x_l]) \tag{1}$$

For entity representation, since entity e_i may appear multiple times through mentions $\{m_j^i\}_{j=1}^{N_{e_i}}$, we apply logsumexp pooling [23], a smoother version of max pooling:

$$h_{e_i} = \log \sum_{j=1}^{N_{e_i}} \exp(h_{m_j^i}) \tag{2}$$

We employ localized context pooling [3] to enhance entity pair embeddings with additional local context. After BERT processing, we obtain attention matrix $A \in \mathbb{R}^{H \times l \times l}$ where A_{ijk} represents attention from token j to token k in the i^{th} attention head. Entity-level attention $A_i^E \in \mathbb{R}^{H \times l}$ denotes attention from the i^{th} entity to all tokens. For entity pair (e_s, e_o), we locate important local context by:

$$A^{(s,o)} = A_s^E \cdot A_o^E \tag{3}$$

$$q^{(s,o)} = \sum_{i=1}^{H} A_i^{(s,o)} \tag{4}$$

$$a^{(s,o)} = q^{(s,o)} / \mathbf{1}^T q^{(s,o)} \tag{5}$$

$$c^{(s,o)} = H^T a^{(s,o)} \tag{6}$$

where H is the contextual embedding and H is the number of attention heads.

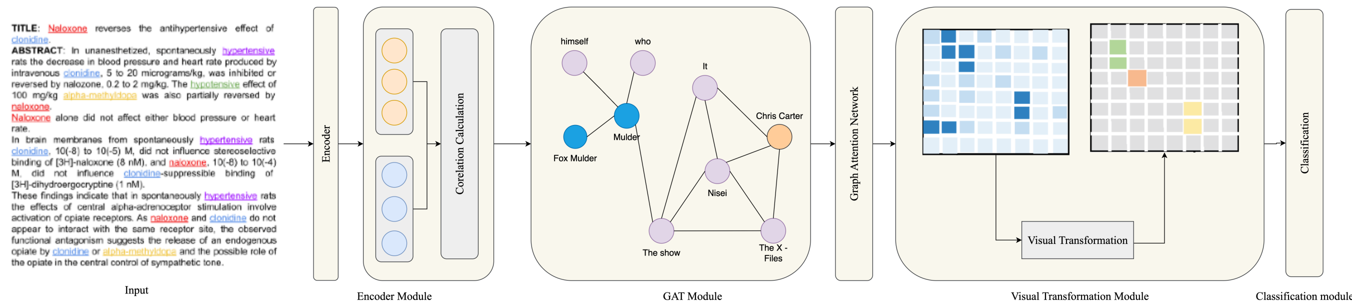

Fig. 1: ANAVIT method overview showing the complete pipeline from input document through GAT-based entity enhancement and visual transformation to final relation predictions.

C. GAT-based Entity Enhancement

Recent work by Lu et al. (2023) [24] incorporated anaphors into document graphs for relation extraction. While their method only uses graph processing, we propose ANAVIT that combines graph-based anaphor modeling with visual transformation of relation matrices, leading to improved performance.

1) Anaphor Extraction: Anaphor detection uses spaCy's POS tagger and dependency parser to identify two types of anaphoric expressions: (1) pronouns tagged as PRON, and (2) definite noun phrases spanning from determiner 'the' (with dependency 'det') to their syntactic heads.

2) Document Graph Construction: We construct a heterogeneous document graph $G = (V, E)$ that explicitly models both entity mentions and anaphoric expressions. The graph incorporates two node types:

- **Mention nodes**: Each entity mention is represented by a mention node
- **Anaphor nodes**: Pronouns and definite referents that may refer to entities are represented by dedicated anaphor nodes

To capture diverse linguistic relationships, we establish four types of directed edges:

- **Self-loop edges**: maintaining node identity
- **Inter-entity edges**: connecting different entity mentions within the same sentence
- **Co-referent edges**: linking mentions referring to the same entity
- **Anaphor-entity edges**: connecting anaphoric expressions to potential entity referents

This formulation enables explicit modeling of co-reference chains, anaphora resolution patterns, and local entity interactions.

3) Graph Attention Networks: We apply Graph Attention Networks [15] over the document graph to model node interactions, extending approaches from [25], [26]. Unlike traditional GCNs [14], GAT computes attention coefficients for edge (i, j):

$$\alpha_{ij} = \frac{\exp(W_q h_i (W_k h_j)^T)}{\sum_{k \in \mathcal{N}_i} \exp(W_q h_i (W_k h_k)^T)} \tag{7}$$

Let g_i^{k-1}, g_i^k denote input and output representations of the k^{th} GAT layer for node v_i:

$$g_i^k = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W^k g_j^{k-1} + b^k\right) \tag{8}$$

where W^k and b^k are learnable parameters, and σ is ReLU activation.

After multiple graph convolutions, we apply logsumexp pooling on mention node embeddings belonging to the same entity:

$$h_{e_i}^{\text{enhanced}} = \log \sum_{j=1}^{N_{e_i}} \exp(g_{m_j^i}^K) \tag{9}$$

The enhanced entity pair feature vector is computed as:

$$M_{s,o} = \text{FFN}([\tanh(W_s[h_s; c^{(s,o)}]); \tanh(W_o[h_o; c^{(s,o)}])]) \tag{10}$$

where W_s and W_o are learnable weight matrices.

D. Visual Transformation Module

Previous studies [4] considered only initial reasoning patterns. Following the multi-scale representation learning approach of [23], we identify three reasoning patterns that encompass more relationship triples, enabling deeper fine-grained reasoning:

TABLE I: Three reasoning patterns for relation extraction

Type	**Reasoning Pattern**
1	$(e_h, r, e_i), (e_i, r, e_t) \rightarrow (e_h, r, e_t)$
2	$(e_i, r, e_h), (e_i, r, e_t) \rightarrow (e_h, r, e_t)$
3	$(e_h, r, e_i), (e_t, r, e_i) \rightarrow (e_h, r, e_t)$

We input the entity pair feature matrix M into a feature processing unit consisting of 2D convolution (kernel size 5×5) and ReLU activation, inspired by CNN architectures for text processing [27]:

$$M_c = \text{ReLU}(\text{Conv2D}(M)) \tag{11}$$

For self-attention, we reshape the input tensor treating each spatial location as a token:

$$x = \text{Reshape}(M_c, [B, H \times W, C]) \tag{12}$$

$$x_{\text{proj}} = \text{Linear}(x) \tag{13}$$

The transformer block calculates attention weights between all tokens:

$$Q = K = V = x_{\text{proj}} \tag{14}$$

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{15}$$

$$A_{\text{out}} = \text{Attention}(Q, K, V) \tag{16}$$

$$M_1 = \text{LayerNorm}(x_{\text{proj}} + \text{Dropout}(A_{\text{out}})) \tag{17}$$

The output is processed by a Feed-Forward Network with residual connections:

$$M_2 = \text{LayerNorm}(M_1 + \text{FFN}(M_1)) \tag{18}$$

$$M_{\text{out}} = \text{Reshape}(M_2, [B, H, W, D_{\text{hidden}}]) \tag{19}$$

E. Adaptive Thresholding

Traditional RE models use fixed global thresholds, failing when models have varying confidence across entity pairs. We introduce a learnable **threshold class (TH)** [3] that adapts to each entity pair.

The adaptive-thresholding loss combines two components:

$$\mathcal{L}_1 = -\sum_{r \in \mathcal{P}_T} \log\left(\frac{\exp(\text{logit}_r)}{\sum_{r' \in \mathcal{P}_T \cup \{\text{TH}\}} \exp(\text{logit}_{r'})}\right) \tag{20}$$

$$\mathcal{L}_2 = -\log\left(\frac{\exp(\text{logit}_{\text{TH}})}{\sum_{r' \in \mathcal{N}_T \cup \{\text{TH}\}} \exp(\text{logit}_{r'})}\right) \tag{21}$$

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 \tag{22}$$

where $\mathcal{P}_T$ and $\mathcal{N}_T$ represent positive and negative relation classes for entity pair T.

IV. Experiments and Results

A. Datasets

We evaluate ANAVIT on two widely-used biomedical document-level relation extraction datasets: CDR and GDA, representing different scales and annotation approaches.

1) CDR (Chemical-Disease Relations): CDR [20] originates from the BioCreative V challenge, comprising 1,500 carefully curated PubMed abstracts with manual annotations by domain experts. This dataset has become a standard benchmark for chemical-disease relation extraction. The corpus is uniformly partitioned into 500 documents each for training, development, and testing.

2) GDA (Gene-Disease Associations): GDA [20] represents a large-scale corpus for gene-disease association extraction using distant supervision techniques. The corpus contains 23,353 abstracts for training, 5,839 for development, and 1,000 manually verified abstracts for testing.

TABLE II: Statistics of the datasets in experiments

Statistics	CDR	GDA
# Train	500	23,353
# Dev	500	5,839
# Test	500	1,000
# Relations	2	2
Avg # entities per Doc.	7.6	5.4

B. Experimental Settings

Our model is implemented in PyTorch with SciBERT [21] as the encoder, which has demonstrated superior performance on biomedical text mining tasks. We use a learning rate of 2×10^{-5}, batch size of 4, and train for 30 epochs. AdamW [28] optimizer is used with learning rate 1×10^{-4}, with linear warmup for the first 6% of total steps following best practices for transformer fine-tuning [22]. Training is conducted on NVIDIA GeForce RTX 3090 GPU.

C. Results and Analysis

We compare against several baseline methods spanning different architectural approaches. Graph-based models include EoG [10], GCNII-ATLOP [29], which uses edge-oriented graphs, LSR [11] with latent structure refinement, and GCNN [26] for inter-sentence relation extraction. Transformer-based models include CNN [30], BRAN [12] which applies BERT to full abstracts, ATLOP [3] with adaptive thresholding.

TABLE III: Test F1 scores (%) on CDR and GDA datasets. ANAVIT with SciBERT encoder achieves state-of-the-art results

Model	CDR	GDA
BRAN [12]	62.1	-
CNN [30]	62.3	-
EoG [10]	63.6	81.5
LSR [11]	64.8	82.2
ATLOP-SciBERT [3]	69.4	83.9
GCNII-ATLOP-SciBERT [29]	70.0	-
ANAVIT (ours)	$\mathbf{79.1 \pm 0.6}$	$\mathbf{84.7 \pm 0.3}$

ANAVIT achieves significant improvements: 9.7 percentage points over ATLOP on CDR (79.1% vs 69.4%) and maintains competitive performance on GDA (84.7% vs 83.9%). The substantial CDR improvement demonstrates the effectiveness of explicit anaphora modeling and visual transformation for complex biomedical relation extraction.

TABLE IV: Ablation study results showing contribution of each component

Model Variant	CDR
ANAVIT	79.1
- w/o GAT	78.6
- w/o Visual Transformation	69.5

To validate the contribution of each component, we conduct comprehensive ablation studies by systematically removing key modules and analyzing performance degradation.

V. Conclusion

We presented ANAVIT, enhancing document-level relation extraction through anaphor nodes and visual transformation. By explicitly modeling anaphoric relationships via graph structures and applying visual processing to relation matrices, we achieve state-of-the-art 79.1% F1 on CDR, significantly outperforming ATLOP (69.4%) and other baselines.

Our work demonstrates that combining linguistic insights (anaphora resolution) with computational techniques (visual transformation) while building on strong foundations (ATLOP) yields substantial improvements. The success of integrating multiple modalities and explicit linguistic structures opens new directions for document understanding. Future work will explore extending this approach to other domains and investigating more sophisticated anaphora resolution techniques.

References

[1] Y. Yao, D. Ye, P. Li, X. Han, Y. Lin, Z. Liu, Z. Liu, L. Huang, J. Zhou, and M. Sun, "Docred: A large-scale document-level relation extraction dataset," *arXiv preprint arXiv:1906.06127*, 2019.

[2] J. A. Diaz-Garcia and J. A. D. Lopez, "A survey on cutting-edge relation extraction techniques based on language models," *Artificial Intelligence Review*, vol. 58, no. 9, p. 287, 2025.

[3] W. Zhou, K. Huang, T. Ma, and J. Huang, "Document-level relation extraction with adaptive thresholding and localized context pooling," in *AAAI Conference on Artificial Intelligence*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:225039888

[4] N. Zhang, X. Chen, X. Xie, S. Deng, C. Tan, M. Chen, F. Huang, L. Si, and H. Chen, "Document-level relation extraction as semantic segmentation," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, Z.-H. Zhou, Ed. International Joint Conferences on Artificial Intelligence Organization, 8 2021, pp. 3999–4006, main Track.

[5] H. Zhou, H. Deng, and J. He, "Chemical-disease relations extraction based on the shortest dependency path tree," 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:17054387

[6] J. Gu, L. Qian, and G. Zhou, "Chemical-induced disease relation extraction with various linguistic features," *Database: The Journal of Biological Databases and Curation*, vol. 2016, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:543003

[7] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W. tau Yih, "Cross-sentence n-ary relation extraction with graph lstms," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 101–115, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:2797612

[8] L. Song, Y. Zhang, Z. Wang, and D. Gildea, "N-ary relation extraction using graph-state lstm," in *Conference on Empirical Methods in Natural Language Processing*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:52115592

[9] C. Quirk and H. Poon, "Distant supervision for relation extraction beyond the sentence boundary," *ArXiv*, vol. abs/1609.04873, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:15359942

[10] F. Christopoulou, M. Miwa, and S. Ananiadou, "Connecting the dots: Document-level neural relation extraction with edge-oriented graphs," in *Conference on Empirical Methods in Natural Language Processing*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202541610

[11] G. Nan, Z. Guo, I. Sekulic, and W. Lu, "Reasoning with latent structure refinement for document-level relation extraction," in *Annual Meeting of the Association for Computational Linguistics*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:218613850

[12] P. Verga, E. Strubell, and A. McCallum, "Simultaneously self-attending to all mentions for full-abstract biological relation extraction," *ArXiv*, vol. abs/1802.10569, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:3576631

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *North American Chapter of the Association for Computational Linguistics*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:52967399

[14] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *ArXiv*, vol. abs/1609.02907, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:3144218

[15] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio', and Y. Bengio, "Graph attention networks," *ArXiv*, vol. abs/1710.10903, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:3292002

[16] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, "Simple and deep graph convolutional networks," in *International Conference on Machine Learning*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:220363476

[17] Y. Pei, T. Huang, W. van Ipenburg, and M. Pechenizkiy, "Resgcn: attention-based deep residual modeling for anomaly detection on attributed networks," *Machine Learning*, vol. 111, pp. 519 – 541, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:222067153

[18] Y. Rong, W. bing Huang, T. Xu, and J. Huang, "Dropedge: Towards deep graph convolutional networks on node classification," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:212859361

[19] D. Zeng, C. Zhao, and Z. Quan, "Cid-gcn: An effective graph convolutional networks for chemical-induced disease relation extraction," *Frontiers in Genetics*, vol. 12, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:231858422

[20] C.-H. Wei, Y. Peng, R. Leaman, A. P. Davis, C. J. Mattingly, J. Li, T. C. Wiegers, and Z. Lu, "Assessing the state of the art in biomedical relation extraction: overview of the biocreative v chemical-disease relation (cdr) task," *Database: The Journal of Biological Databases and Curation*, vol. 2016, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:2527024

[21] I. Beltagy, K. Lo, and A. Cohan, "Scibert: A pretrained language model for scientific text," in *Conference on Empirical Methods in Natural Language Processing*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202558505

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *North American Chapter of the Association for Computational Linguistics*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:52967399

[23] R. Jia, C. Wong, and H. Poon, "Document-level n-ary relation extraction with multiscale representation learning," *ArXiv*, vol. abs/1904.02347, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:102353905

[24] C. Lu, R. Zhang, K. Sun, J. Kim, C. Zhang, and Y. Mao, "Anaphor assisted document-level relation extraction," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, H. Bouamor, J. Pino, and K. Bali, Eds. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 15 453–15 464. [Online]. Available: https://aclanthology.org/2023.emnlp-main.955

[25] Z. Guo, Y. Zhang, and W. Lu, "Attention guided graph convolutional networks for relation extraction," *ArXiv*, vol. abs/1906.07510, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:190001673

[26] S. K. Sahu, F. Christopoulou, M. Miwa, and S. Ananiadou, "Inter-sentence relation extraction with document-level graph convolutional neural network," in *Annual Meeting of the Association for Computational Linguistics*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:184487889

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, pp. 2278–2324, 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:14542261

[28] Loshchilov and F. Hutter, "Decoupled weight decay regularizatio," 2017. [Online]. Available: https://arxiv.org/pdf/1711.05101

[29] N.-H. Ngo, A.-D. Nguyen, Q.-T. P. Thi, and T. H. Dang, "Integrating graph and transformer-based models for enhanced chemical-disease relation extraction in document-level contexts," in *International Symposium on Information and Communication Technology*. Springer, 2024, pp. 174–187.

[30] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:6916627