

Article

RASD: Relation Aware Spectral Decoupling Attention Network for Knowledge Graph Reasoning

Zheng Wang ¹, Taiyu Li ¹ and Zengzhao Chen ^{1,2,*}¹ Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China² National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079, China

* Correspondence: zzchen@mail.ccnu.edu.cn

Abstract

Knowledge Graph Reasoning (KGR) aims to deduce missing or novel knowledge by learning structured information and semantic relationships within Knowledge Graphs (KGs). Despite significant advances achieved by deep neural networks in recent years, existing models typically extract non-linear representations from explicit features in a relatively simplistic manner and fail to fully exploit semantic heterogeneity of relation types and entity co-occurrence frequencies. Consequently, these models struggle to capture critical predictive cues embedded in various entities and relations. To address these limitations, this paper proposes a relation aware spectral decoupling attention network for KGR (RASD). First, a spectral decoupling attention network module projects joint embeddings of entities and relations into the frequency domain, extracting features across different frequency bands and adaptively allocating attention at the global level to model frequency specific information. Next, a relation-aware learning module employs relation aware filters and an augmentation mechanism to preserve distinct relational properties and suppress redundant features, thereby enhancing representation of heterogeneous relations. Experimental results demonstrate that RASD achieves significant and consistent improvements over multiple leading baseline models on link prediction tasks across five public benchmark datasets.



Academic Editor: Christos Bouras

Received: 19 July 2025

Revised: 11 August 2025

Accepted: 14 August 2025

Published: 16 August 2025

Citation: Wang, Z.; Li, T.; Chen, Z. RASD: Relation Aware Spectral Decoupling Attention Network for Knowledge Graph Reasoning. *Appl. Sci.* **2025**, *15*, 9049. <https://doi.org/10.3390/app15169049>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: knowledge graph reasoning; spectral decoupling attention network; heterogeneous relation; relation-aware

1. Introduction

Knowledge Graphs (KGs), serving as a graph-structured database of human knowledge, have received significant attention and exploration due to their ability to store massive structured concrete facts to boost a wide range of artificial intelligence tasks, including recommendation systems [1], question answering [2], and natural language understanding [3]. KGs store information through entities and relations, which are typically represented as triples (h, r, t) , where h and t denote head and tail entities, respectively, and r represents relation, such as *(Christopher Nolan, Born_in, London)*. However, existing KGs face several challenges, with one of the major issues being their incompleteness and missing facts. Therefore, Knowledge Graph Reasoning (KGR) has received increasing attention due to its strong ability to infer new or missing knowledge.

Existing KGR models can be generally categorized into logic rule-based reasoning [4–6], representation learning-based reasoning [7,8], and deep neural network-based reasoning [9,10].

Logic rule-based reasoning can simulate human reasoning ability by capturing hidden semantic relationships in KGs, providing strong interpretability. However, the heavy reliance on predefined rules and logic limits the ability to generalize. Representation learning-based reasoning, on the other hand, captures implicit associations between entities and relations by mapping entities, relations, and attributes into a low-dimensional real-valued vector space, thus enabling better generalization. Nonetheless, these methods often struggle to learn more expressive features, which can diminish their link prediction capabilities.

With the success of deep neural networks in computer vision and natural language processing, they have increasingly been applied to various fields. ConvE [9], the first to apply convolutional neural networks (CNNs) to link prediction tasks, leverages 2D convolutional neural networks to capture the non-linear features of entity and relation embeddings. Additionally, R-GCN [10], which is based on graph neural networks (GNNs), utilizes convolutional operators to capture positional information in KGs, enabling effective handling of extremely multi-relational data and improving reasoning capabilities. These methods leverage multi-layer non-linear features to capture the underlying connections and semantic relationships between entities and relations, thereby enabling effective reasoning.

However, despite the significant progress KGR has made, KGR continues to be confronted with many challenges. First, as illustrated in Figure 1b, the Laplacian spectral decomposition of the WN18 and YAGO3-10 datasets reveals two types of structural patterns in KGs. On the one hand, there are entities and relations with high connectivity, which correspond to low-frequency components in the spectral domain. On the other hand, there are entities and relations with sparse connections, corresponding to high-frequency components. Current reasoning methods typically extract non-linear representations from explicit features in a rather simplistic way and do not adequately exploit the semantic heterogeneity of different relation types. As a result, they often regard low-connectivity (spectral high-frequency) information as noise and place excessive emphasis on common high-connectivity relations such as *Born_in*, while overlooking semantically critical but rare relations like *Won_Oscar*. Second, these models rarely account for entity co-occurrence frequency—the number of times two entities appear together in the same triple—thereby failing to capture important co-occurrence patterns, which in turn limits their capacity to model complex semantic relationships and to generalize effectively.

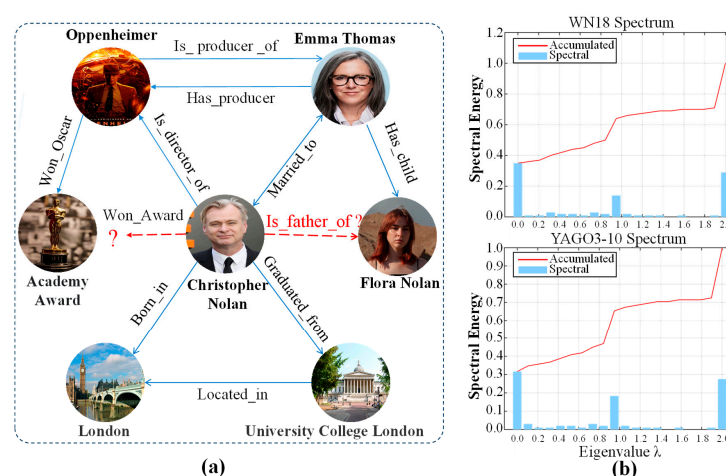


Figure 1. (a) An example of KGs, showcasing their extensive collection of entities and relations. The two red dashed lines represent two distinct link prediction tasks within KGR: One focused on predicting relations, such as (*Christopher Nolan*, ?, *Flora Nolan*), and the other on predicting entities, exemplified by (*Christopher Nolan*, *Won_Award*, ?). (b) The energy distribution of the graph links between WN18 and YAGO3-10.

To address the aforementioned challenges, the probability of co-occurrence of entities as well as the heterogeneity of relations should be comprehensively considered. As shown in Figure 1a, taking the example of (*Christopher Nolan*, *Won_Award*, ?), we can infer that (1) Common relations such as *Born_in* are often assigned disproportionately high weights during reasoning, despite offering little predictive value; conversely, rare but semantically critical relations like *Won_Oscar* may be undervalued, thereby degrading inference efficiency. (2) Additionally, the high co-occurrence frequency between entities Christopher Nolan and the Academy Award for Best Directing makes it easier to predict that Christopher Nolan won the Academy Award for Best Directing when performing the query (*Christopher Nolan*, *Won_Award*, ?) even in the absence of explicit relational paths. (3) Due to the heterogeneity of relations, different relations exhibit various properties. For instance, the relation *Married_to* is symmetric, while *Has_producer* and *Is_producer_of* are asymmetric and inversely related. Similarly, relations such as *Located_in*, *Has_child*, and *Born_in* are asymmetric and non-reflexive. Such diverse relational properties markedly increase the complexity of KGR.

Related studies have demonstrated the significant advantages of Fourier transform-based frequency-domain modeling in natural language processing. Inspired by these findings, we introduce the relation aware spectral decoupling attention network (RASD) for KGR, which captures both entity co-occurrence frequency and relation heterogeneity. RASD comprises three key components: spectral decoupling attention network (SDA), relation-aware learning (RAL), and link scoring. SDA first employs the fast Fourier transform (FFT) to project joint embeddings of entity and relation into the spectral domain and then adaptively allocates attention across different frequency components, which enables comprehensive extraction of fine-grained interactions within high-frequency components while selectively focusing on a small set of critical spectral bands, thereby achieving efficient global reasoning. Then, RAL applies relation-aware filters and an augmentation mechanism to preserve the intrinsic semantic features of each relation and suppress irrelevant signals, facilitating fine-grained modeling of heterogeneous relation properties. Finally, the link scoring module evaluates predicted triples by computing similarity scores between transformed representations and all candidate tail entities. Experimental results demonstrate that RASD achieves state-of-the-art results on most evaluation metrics.

The main contributions are summarized as follows:

1. We propose an innovative relation aware spectral decoupling attention network for KGR that simultaneously preserves intrinsic relation semantics and effectively models both entity co-occurrence frequencies and relation heterogeneity, thereby addressing data imbalance and relation diversity in KGs. RASD leverages spectral-domain awareness and relation-aware mechanisms to selectively amplify salient signals while suppressing noise, resulting in more accurate relational inference.
2. To our knowledge, RASD is the first model to integrate the FFT into KGR in order to capture the spectral-domain frequency characteristics of entities and relations. By projecting joint embeddings into the frequency domain, RASD deeply extracts frequency features; it balances attention across different frequency bands at the global level, avoiding over- or under-emphasis on high- or low-frequency components and enabling fine-grained spectral modeling.
3. Experiments confirm the effectiveness of RASD on five benchmark datasets. Results show that RASD achieves significant and consistent improvements for most evaluation metrics on link prediction.

The remainder of this paper is arranged as follows. A brief overview and analysis of related work on KGR can be found in Section 2. The research problem is formally defined and the specifics of RASD are presented in Section 3. In Section 4, the experimental

results of our proposed model on five public benchmark datasets are presented. Finally, the conclusions are discussed in Section 5.

2. Related Work

KGR is essentially the process of deducing missing or new knowledge based on existing facts, with its core focus being the prediction of entities and relations within triples. Current KGR models can be broadly categorized into three main types: logic and rule-based reasoning, representation learning-based reasoning, and deep neural network-based reasoning.

2.1. Logic and Rule-Based Models

Early KGR methods primarily relied on logic and rule-based reasoning, drawing from traditional rule inference techniques to extract reliable rules or identify underlying generalizable features within KGs. The main approaches include ontology-based reasoning, rule-based reasoning, random walk methods, and first-order predicate logic algorithms. First, as a domain-independent rule-learning method, FOIL (First Order Inductive Learner) [5] involves a four-step process to derive reasoning rules from text, facilitating accurate and efficient reasoning in small-scale KGs. However, its efficiency and accuracy diminish sharply in large-scale KGs. To address this issue, the AMIE [4] system generates rule libraries based on types of relationships in KGs and then identifies instances satisfying these rules within the graph and computes rule confidence to extract reliable rules, enabling the mining of Horn rules in large RDF (Resource Description Framework) knowledge bases. To further reduce the complexity of rule search, algorithms such as RuleN [11], AnyBURL [12], and their variants, as well as CNN [13], have been developed. These methods sample only triples associated with target relationships in the KG and use this sampled information for rule learning, thereby effectively reducing the search space and simplifying the computation. Although these reasoning methods provide clear and interpretable inference pathways, they face challenges in expressiveness and generalization due to the semantic diversity and data sparsity inherent in KGs.

2.2. Representation Learning-Based Models

The central idea of representation learning-based models is to identify a mapping function that translates entities, relationships, and attributes within a semantic network into a low-dimensional real-valued vector space, thereby obtaining distributed representations that capture implicit associations between entities and relationships. A significant number of inference methods based on representation learning have been proposed, including distance-based models, semantic matching, tensor decomposition, and multi-relational information models. TransE [7], as the first distance-based model, conceptualizes relationships as simple translation operations, represented as $e_h + r \approx e_t$. While it has demonstrated effectiveness, TransE struggles with certain specific relationship types, such as 1-N, N-N, symmetric, and transitive relations. To address these limitations, numerous distance-based models have been developed [14–16]. Tensor decomposition models, on the other hand, represent KGs as three-way tensors, which are decomposed into combinations of low-dimensional vectors for entities and relationships. RESCAL [17], considered a breakthrough tensor decomposition model, effectively captures the latent semantics of entities through vector representations. It further employs matrices to characterize pairwise interactions among these latent factors. However, RESCAL is hindered by significant computational complexity. To mitigate this issue, DistMult [8] introduces bilinear diagonal matrices, thereby reducing the number of parameters for each relation to $\mathcal{O}(d)$. Following this, ComplEx [18] extends the capabilities of DistMult by incorporating complex-valued embeddings, which improve the modeling of asymmetric relationships. Representation

learning-based models focus solely on constraints that satisfy the facts within a KG, neglecting deeper compositional information. This limitation constrains its reasoning capabilities.

2.3. Deep Neural Network-Based Models

Deep neural network (DNN) models have demonstrated exceptional performance in KGR due to their expressive capacity in representation learning over KGs. These DNN-based models are categorized into three primary types: convolutional neural network (CNN) models, recurrent neural network (RNN) models, and graph neural network (GNN) models. Typically, ConvE [9] applies two-dimensional convolutional layers to capture the non-linear features of entity and relation embeddings. In addition, R-GCN [10], a GNN-based model, utilizes convolutional operators to capture structural information within KGs, which allows it to manage highly multi-relational data and enhance reasoning capacity. Additionally, some novel approaches have recently emerged. CircularE [19], for instance, substitutes standard convolutions with circular convolutions, substantially increasing interaction counts. MSHE [20] integrates a self-attention mechanism with multilayer convolutions, which allows CNN-based methods to scale to larger KGs and improves interactions between entity and relation features. Furthermore, the success of Transformers in processing Euclidean-structured data has inspired attempts to model complex non-Euclidean data structures, such as KGs. An example is the StAR algorithm [21], which models embeddings of entities while incorporating their text descriptions. For enabling the model to grasp logical rules, Ruleformer [22] detects symbolic rules in KGs and encodes the logical operations within these rules using Transformer networks. Similarly, N-Former [23] and Knowformer [24] refine multiple knowledge sequences to jointly learn entity and relation embeddings across different sequential data representations using Transformers.

3. Proposed RASD Method

In this section, we formally define the research problem and detail the structure of RASD, which consists of three main components, as illustrated in Figure 2.

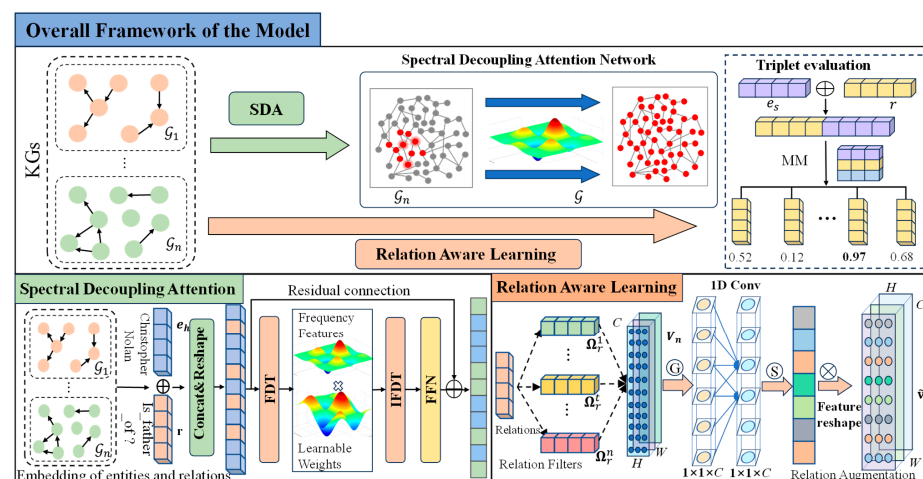


Figure 2. Overview of the proposed RASD that comprises three key components: SDA, RAL, and link scoring. SDA first employs the FFT to project joint embeddings of entity and relation into the spectral domain, and then adaptively allocates attention across different frequency components. Then, RAL applies relation-aware filters and an augmentation mechanism to preserve the intrinsic semantic features of each relation and suppress irrelevant signals, facilitating fine-grained modeling of heterogeneous relation properties. Finally, the link scoring module evaluates predicted triples by computing similarity scores between transformed representations and all candidate tail entities.

3.1. Problem Definition

To represent the set of entities, relations, and facts, $\mathcal{E}$, $\mathcal{R}$ and $\mathcal{T}$ are selected. A KG stores a large number of facts in the form of triples, which can be expressed as follows:

$$\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{T}\} \quad (1)$$

where $\mathcal{E} = \{e_1, e_2, \dots, e_{|\mathcal{E}|}\}$ and $\mathcal{R} = \{r_1, r_2, \dots, r_{|\mathcal{R}|}\}$ denote the collection of entities and relations, respectively. $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ represents the set of triplets.

For each fact, a triplet is represented as (e_h, r, e_t) , where $e_h, e_t \in \mathcal{E}$ are the head and tail entities and $r \in \mathcal{R}$ represents the relation. The bold letters $\mathbf{e}_h, \mathbf{r}, \mathbf{e}_t \in \mathbb{R}^d$ denote the d -dimension embeddings of e_h, r, e_t . The target of the link prediction is to predict $(e_h, r, ?)$ or $(?, r, e_t)$ given an entity with a specific relation. Table 1 offers a concise overview of the principal symbols used in this paper, along with their explanations.

Table 1. Notations and explanations.

Notations	Explanations
$\mathcal{E}, \mathcal{R}, \mathcal{T}$	Set of entity, relation, and triplets
e_h, r, e_t	Head entity, relation, and tail entity
$\mathbf{e}_h, \mathbf{r}, \mathbf{e}_t$	Embedding of head entity, relation, and tail entity
$\mathbf{M}$	Embedding matrix of joint embedding
$\oplus, \odot$	Concatenation operation and Hadamard product
$\mathcal{F}[\cdot] \mathcal{F}^{-1}[\cdot]$	2D FFT and inverse Fourier transform
$\odot, *$	Complex multiplication and convolution operation
$\mathbf{e}_h^c, \mathbf{r}^c, \mathbf{e}_t^c$	Complex space embedding of head entity, relation, and tail entity
$\mathcal{H}_r$	Relation-aware filters
$\mathcal{X}, \tilde{\mathcal{X}}$	The relation map and the augmentation feature map
$\mathcal{L}$	Cross-channel dependency
$\sigma(\cdot)$	Softmax function
$\mathcal{C}\mathcal{W}(\cdot)$	Channel-wise multiplication
$\varphi(e_h, r, e_t)$	Score function
$\hat{y}(e_h, r, e_t)$	Triplet prediction probability

3.2. Overview of RASD

As shown in Figure 2, the proposed RASD comprises three key components: spectral decoupling attention network (SDA), relation-aware learning (RAL), and link scoring. SDA first employs the fast Fourier transform (FFT) to project joint embeddings of entity and relation into the spectral domain, and then adaptively allocates attention across different frequency components, which enables comprehensive extraction of fine-grained interactions within high-frequency components while selectively focusing on a small set of critical spectral bands, thereby achieving efficient global reasoning. Then, RAL applies relation-aware filters and an augmentation mechanism to preserve the intrinsic semantic features of each relation and suppress irrelevant signals, facilitating fine-grained modeling of heterogeneous relation properties. Finally, the link scoring module evaluates predicted triples by computing similarity scores between transformed representations and all candidate tail entities.

3.3. Spectral Decoupling Attention Network

Given an input triplet $(e_h, r, e_t) \in \mathcal{T}$, we use $\mathbf{e}_h, \mathbf{r} \in \mathbb{R}^d$ to represent the embedding of head entity e_h and relation r , respectively, where d is the dimension of the embedding

space. We consider e_h and r as a composition, connecting them to extract their potential features, resulting in the matrix M as follows:

$$M = [e_h \oplus r \in \mathbb{R}^{2 \times d}], \quad (2)$$

where $\oplus$ represents the concatenation operation.

The spectral decoupling attention network module transforms the joint embedding matrix M into the input $\mathfrak{s} \in \mathbb{R}^{H \times W \times D}$ for the spectral gating unit through linear projection, represented as follows:

$$\mathfrak{s} = W \circ M, \quad (3)$$

where W is a linear transformation matrix used to convert the concatenated embeddings.

For the input $\mathfrak{s} \in \mathbb{R}^{H \times W \times D}$, after performing layer normalization, we apply a 2D FFT to transition from the spatial domain to the frequency domain, thereby facilitating the learning of interactions between entities and relations in the frequency domain:

$$\mathcal{S} = \mathcal{F}[\mathfrak{s}] \in \mathbb{R}^{H \times W \times D}, \quad (4)$$

where $\mathcal{F}[\cdot]$ represents the 2D FFT and $\mathcal{S}$ denotes the spectral information of $\mathfrak{s}$.

A learnable weight parameter $W_C \in \mathbb{R}^{H \times W \times D}$ is then employed to assign the contribution of each frequency, thereby capturing a select few key relational frequencies as well as entity co-occurrence frequencies:

$$\tilde{\mathcal{S}} = W_C \odot \mathcal{S}, \quad (5)$$

where $\odot$ is the Hadamard product. The parameter W_C is learned through back-propagation and is specific to each block of the spectral decoupling attention network mechanism.

Finally, after learning the frequency interactions between entities and relations, the inverse FFT is applied to transform the spectral information $\tilde{\mathcal{S}}$ back to the spatial domain:

$$\mathfrak{s} \leftarrow \mathcal{F}^{-1}[\tilde{\mathcal{S}}], \quad (6)$$

where $\mathcal{F}^{-1}[\cdot]$ represents the inverse FFT (IFFT). Following the IFFT, layer normalization and the MLP are applied to capture information across different channels.

3.4. Relation-Aware Learning

After thoroughly learning the interaction information of entities and relations in the frequency domain, the embeddings are re-coupled in the complex space to serve as input for the relation-aware learning:

$$M_c = (e_h + e_h^c \odot r^c) \oplus (r + e_h^c \odot r^c), \quad (7)$$

where $e_h^c = u + \mathfrak{z}i$ and $r^c = y + \mathfrak{z}i$, and $\odot$ represents the complex multiplication.

Considering the specificity of heterogeneous relations, relation-aware filters $\mathcal{H}_r$ are constructed to capture the unique attributes of each relation. And the expression for $\mathcal{H}_r$ is given by the following:

$$\mathcal{H}_r = [\mathcal{H}_r^1, \mathcal{H}_r^2, \mathcal{H}_r^3, \dots, \mathcal{H}_r^n] \in \mathbb{R}^{n \times h \times w}, \quad (8)$$

where n denotes the number of filters. Furthermore, the relation maps $\mathcal{Z} = [\mathfrak{z}_1, \mathfrak{z}_2, \mathfrak{z}_3, \dots, \mathfrak{z}_n] \in \mathbb{R}^{n \times C \times H \times W}$ are derived by convolving the relation filter $\mathcal{H}_r$ with M_c as follows:

$$\mathbf{z}_n = \mathbf{M}_c * \mathcal{H}_r^n = \sum_{i=1}^h \sum_{j=1}^w \mathbf{M}_c \mathcal{H}_r^n(i, j), \quad (9)$$

where $*$ denotes the convolution operation.

To capture more comprehensive information and accentuate salient features, a cross-channel relationship augmentation mechanism is implemented. First, global average pooling (GAP) is employed to extract global features from each channel, thereby enhancing the ability of the model to capture global information while accentuating significant channels and reducing redundant information.

Furthermore, fast 1D convolution of size k is utilized to capture local dependencies among channels, efficiently extracting interaction patterns between features through convolution operations on the feature channels, thereby improving the model's understanding of complex specific features. Finally, the dependencies of the cross-channel feature maps can be represented as follows:

$$\ell_n = \sigma \left(\sum_{i=1}^k \frac{1}{H \times W} \sum_{a=1}^H \sum_{b=1}^W \mathbf{z}_n(a, b) \mathbf{p}(i) \right), \quad (10)$$

where $\mathbf{p}(\cdot)$ refers to 1D convolution, and cross-channel dependencies $\mathcal{L} = [\ell_1, \ell_2, \ell_3, \dots, \ell_n]$ reveal the dependency of each channel on the others.

Finally, we employ channel-wise multiplication to combine the relation maps $\mathbf{z}_n$ with the cross-channel dependencies ℓ_n , resulting in the augmentation feature maps:

$$\tilde{\mathbf{z}}_n = \mathcal{CW}(\mathbf{z}_n, \ell_n) \in \mathbb{R}^{C \times H \times W}, \quad (11)$$

where $\mathcal{CW}(\cdot)$ denotes the channel-wise multiplication. Similar to the relation maps $\mathcal{Z}$, the augmentation feature map is denoted by $\tilde{\mathcal{Z}} = [\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \tilde{\mathbf{z}}_3, \dots, \tilde{\mathbf{z}}_n] \in \mathbb{R}^{n \times C \times H \times W}$. The training procedure of RASD is presented in Algorithm 1.

Algorithm 1. Relation Aware Spectral Decoupling Attention Network for Knowledge Graph Reasoning

Input: Knowledge Graph $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{T}\}$;

Epoch number B ;

Output: Link prediction result $\hat{y}(x) \in (0, 1)$

1: **for** $epoch \leftarrow 1$ to B **do**:

2: $P_{batch} \leftarrow \emptyset$; // Training set initialization.

3: **for** (e_h, r, e_t) in $\mathcal{T}$ **do**:

4: $(e_r, r^{-1}, e_h) \leftarrow (e_h, r, e_t)$; // Reciprocal triplet set generation.

5: $P_{batch} \leftarrow P_{batch} \cup \{(e_r, r^{-1}, e_h), (e_h, r, e_t)\}$; // Training set generation.

6: **end for**

7: **for** (e_h, r, e_t) in P_{batch} **do**:

8: $\mathbf{e}_h \leftarrow e_h, \mathbf{r} \leftarrow r, \mathbf{M} \leftarrow \mathbf{e}_h, \mathbf{r}$; // Feature embedding and stacking.

9: $\mathcal{S} \leftarrow \mathcal{S} \leftarrow \mathbf{M}$; // Spectral Decoupling Attention

10: $\mathbf{M}_c \leftarrow \mathbf{e}_h, \mathbf{r}, \mathbf{e}_h^c, \mathbf{r}^c$; // Complex space re-coupled

11: $\tilde{\mathbf{z}}_n \leftarrow \mathbf{z}_n, \mathcal{L} \leftarrow \mathbf{M}_c, \mathcal{H}_r$; // Construction of relation enhancement feature map

12: $\varphi(e_h, r, e_t) = f(v(f(\mathcal{CW}(\mathbf{M}_c * \mathcal{H}_r, \mathcal{L})))\mathcal{S})e_t$; // Calculate triplet score

13: $\hat{y}(e_h, r, e_t) = \sigma(\varphi(e_h, r, e_t) + b) \in (0, 1)$; // The probabilistic prediction

14: **end for**

15: **end for**

16: **return** $\hat{y}$

3.5. Link Scoring and Prediction

To evaluate the predictive performance of RASD, we multiply the hidden layer vector by the embeddings of all candidate tail entities, with the score function represented as follows:

$$\varphi(e_h, r, e_t) = \text{ReLU}\left(\left(v\left(f\left(\tilde{\mathcal{Z}}\right)\right)\mathcal{S}\right)\right)e_t, \quad (12)$$

where $v(\cdot)$ denotes the flatten operation. Building on the aforementioned formulas, the score function can be represented as follows:

$$\varphi(e_h, r, e_t) = f(v(f(\mathcal{E}\mathcal{W}(\mathbf{M}_c * \mathcal{H}_r, \mathcal{Z})))\mathcal{S})e_t, \quad (13)$$

where $f(x)$ is the rectified linear unit. Finally, the triplet prediction function can be defined as follows:

$$\hat{y}(e_h, r, e_t) = \sigma(\varphi(e_h, r, e_t) + b) \in (0, 1). \quad (14)$$

In RASD, we aim to maximize the likelihood function $p(\mathcal{G}|\Theta)$ by adjusting the parameters Θ within the method. For a valid triplet (e_h, r, e_t) , its scoring $\varphi(e_h, r, e_t)$ is equal to 1; otherwise, the scoring is equal to 0. Therefore, the likelihood function is defined as follows:

$$p(\mathcal{G}|\Theta) = \prod_{\substack{(e_h, r, e_t) \in \mathcal{T} \\ (e_h, r, e'_t) \in \mathcal{T}'}} (\hat{y}(e_h, r, e_t))^y (1 - \hat{y}(e_h, r, e'_t))^{1-y}, \quad (15)$$

in which

$$y = \begin{cases} 1 & \text{if } (e_h, r, e_t) \in \mathcal{T} \\ 0 & \text{if } (e_h, r, e'_t) \in \mathcal{T}' \end{cases} \quad (16)$$

where $\mathcal{T}'$ represents the set of the corrupt triplet generated by altering a set of correct triplets set $\mathcal{T}$, and the loss function for RASD is as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{\substack{(e_h, r, e_t) \in \mathcal{T} \\ (e_h, r, e'_t) \in \mathcal{T}'}} (y \log \hat{y}(e_h, r, e_t) + (1 - y) \log (1 - \hat{y}(e_h, r, e'_t))). \quad (17)$$

During training, we employ the efficient Adam optimizer to minimize the loss function and apply both Batch Normalization and Dropout for regularization. In addition, we incorporate Label Smoothing to further enhance the model generalization ability.

4. Experiment and Discussion

4.1. General Setting

4.1.1. Datasets

In these experiments, five benchmark datasets are adopted: WN18, FB15k, WN18RR, FB15k-237, and YAGO3-10. The statistical information of the five datasets is summarized in Table 2, and a brief description is denoted as follows:

1. FB15k is a subset of the Freebase knowledge graph, comprising 14,951 entities and 1345 relation types. It has become a de facto standard benchmark for evaluating embedding-based knowledge representation and link prediction methods.
2. WN18 is derived from WordNet and contains 40,943 entities linked by 18 semantic relation types. It is routinely used to assess the capacity of the model to capture lexical polysemy, hierarchical structure, and complex semantic patterns.
3. FB15k-237 is a refined version of FB15k from which all inverse-relation “shortcuts” have been removed to eliminate test-set leakage. By preserving 14,541 entities and

237 relations, it presents a more challenging and representative benchmark for KGE and link prediction tasks.

4. WN18RR is an improved variant of WN18 in which all direct inverse-relation pairs have been excised. It retains 40,943 entities and 11 carefully vetted relations, serving as a rigorous testbed for deep semantic reasoning and model generalization.
5. YAGO3-10 is a high-connectivity subset of the multilingual YAGO3 knowledge base, limited to entities that participate in at least ten distinct relations. It comprises approximately 123,182 entities, 37 relation types, and 1.18 million triples, and is widely employed for knowledge graph completion and link prediction benchmarks.

Table 2. Statistical information of five benchmark datasets.

Datasets	# Entities	# Relations	# Triplets		
			Train	Valid	Test
YAGO3-10	123,182	37	1,079,040	5000	5000
FB15k	14,951	1345	483,142	50,000	59,071
WN18	40,943	18	141,442	5000	5000
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3034	3134

4.1.2. Evaluation Metrics

Two evaluation indicators, involving the mean reciprocal rank (MRR) and Hits@k (Hits@1, Hits@3, and Hits@10), are utilized to reflect the accuracy of link prediction.

4.1.3. Comparison Models

Various state-of-the-art baseline methods are compared with our proposed model to verify the effectiveness, which are represented as follows: TransE, TransR, DistMult, TorusE, ComplEx, RotatE, Rot-Pro, R-GCN, ConvE, InteractE, ComplexGCN, RotatPRH, CirlularE, Rotate4D, Rule, CompoundE3D WDS, SDFormer, and MSHE.

4.1.4. Implementation Details

We performed a grid search to identify well-performing hyperparameters and used the Adam optimizer for all experiments. For all five datasets, the embedding dimension is set to 200, the batch size to 128, and the label smoothing to 0.1. From the grid search, the best learning rates are 0.001 for YAGO3-10, WN18, and WN18RR; 0.0001 for FB15k; and 0.0005 for FB15k-237. The augmentation convolution kernel size is set to 7 for YAGO3-10, FB15k-237, and WN18RR, and to 3 for FB15k and WN18. Dropout rates for the input, feature-map, and hidden layers are chosen from {0.1, 0.2, 0.3, 0.4, 0.5, 0.6} and the best value for each dataset is selected based on validation performance. A concise summary of the final hyperparameter settings used in our experiments is provided in Table 3.

Table 3. Hyperparameter settings on various datasets.

Datasets	Aug Value	Batch Size	Learning Rate	Label Smoothing	Dropout		
					Input	Feature Map	Hidden Layer
YAGO3-10	7	128	0.001	0.1	0.2	0.2	0.3
FB15k	3	128	0.0001	0.1	0.2	0.2	0.3
WN18	3	128	0.001	0.1	0.2	0.2	0.3
FB15k-237	7	128	0.0005	0.1	0.5	0.3	0.6
WN18RR	7	128	0.001	0.1	0.4	0.3	0.4

4.2. Results on Entity Prediction

Tables 4–6 present a comprehensive comparison of our approach against state-of-the-art models on five benchmark datasets for entity prediction. Notably, RASD⁺ attains the top position across all evaluation metrics on WN18 and FB15k, and ranks first on three out of four metrics on WN18RR and YAGO3-10. Remarkably, on FB15k, both RASD and its enhanced variant RASD⁺ occupy the first and second places; compared with the strongest baseline, RASD⁺ alone yields relative improvements of 2.1% in MRR, 1.0% in Hits@10, 1.4% in Hits@3, and 2.5% in Hits@1. Furthermore, by introducing frequency partitioning to learn multi-scale representations of global and local spectral components, RASD⁺ further outperforms RASD under identical evaluation settings. Overall, by integrating frequency-domain aware attention and relation-aware augmentation, the RASD framework effectively captures complex semantic patterns and multi-scale structural features in KGs, thereby more accurately capturing entity–relation interactions and achieving leading results on both entity prediction and link prediction tasks. The superior performance of RASD⁺ further corroborates the scalability of frequency-domain feature learning for KGR.

Table 4. Link prediction results for WN18RR and WN18 datasets.

Models	WN18RR				WN18			
	MRR	Hits			MRR	Hits		
		@10	@3	@1		@10	@3	@1
TransE [7]	0.182	0.444	0.295	0.027	0.454	0.934	0.823	0.089
DistMult [8]	0.430	0.490	0.440	0.390	0.822	0.936	0.914	0.728
ComplEx [18]	0.440	0.510	0.460	0.410	0.941	0.947	0.936	0.936
ConvE [9]	0.430	0.520	0.440	0.400	0.943	0.956	0.946	0.935
R-GCN [10]	0.226	0.376	0.269	0.157	0.814	0.955	0.928	0.686
TorusE [25]	0.453	0.512	0.464	0.422	0.947	0.954	0.950	0.943
InteractE [26]	0.467	0.529	0.481	0.436	0.950	0.958	0.952	0.946
Rot-Pro [27]	0.457	0.577	0.482	0.397	0.949	<u>0.958</u>	0.951	0.944
ComplexGCN [28]	0.455	0.516	0.468	0.423	-	-	-	-
SDFormer [29]	0.458	0.528	0.471	0.425	0.948	0.957	0.951	0.944
MSHE [20]	0.461	0.530	0.473	0.429	0.948	0.957	0.951	0.943
RASD	<u>0.476</u>	0.538	<u>0.490</u>	<u>0.445</u>	<u>0.950</u>	0.956	<u>0.953</u>	<u>0.947</u>
RASD ⁺	0.479	<u>0.541</u>	0.495	0.448	0.952	0.959	0.955	0.949

Table 5. Link prediction results for FB15k-237 and FB15k datasets.

Models	FB15k-237				FB15k			
	MRR	Hits			MRR	Hits		
		@10	@3	@1		@10	@3	@1
TransE [7]	0.257	0.420	0.284	0.174	0.380	0.641	0.472	0.231
TransR [16]	0.263	0.428	0.267	0.168	0.346	0.582	0.404	0.218
DistMult [8]	0.241	0.419	0.263	0.155	0.654	0.824	0.733	0.546
ComplEx [18]	0.247	0.428	0.275	0.158	0.692	0.840	0.759	0.599
ConvE [9]	0.325	0.501	0.356	0.237	0.657	0.831	0.723	0.558
R-GCN [10]	0.249	0.417	0.263	0.151	0.696	0.842	0.760	0.601
RotatE [30]	0.338	0.533	0.375	0.241	0.797	0.884	0.830	0.746
InteractE [26]	0.348	0.531	0.383	0.258	0.780	0.884	0.825	0.718
TorusE [25]	0.316	0.484	0.335	0.217	0.733	0.832	0.771	0.674
Rot-Pro [27]	0.344	0.540	0.383	0.246	0.767	0.868	0.808	0.712
RotatPRH [31]	0.343	0.538	0.381	0.246	0.793	0.886	0.833	0.736
ComplexGCN [28]	0.338	0.524	0.371	0.245	-	-	-	-

Table 5. Cont.

Models	FB15k-237				FB15k			
	MRR	Hits			MRR	Hits		
		@10	@3	@1		@10	@3	@1
CirlularE [19]	0.339	0.533	0.374	0.243	0.792	0.887	0.836	0.730
Rotate4D [32]	0.349	0.545	0.388	0.252	0.790	0.890	0.836	0.727
SDFormer [29]	0.356	<u>0.541</u>	0.390	0.264	0.692	0.802	0.732	0.628
RASD	0.343	0.529	0.379	0.251	<u>0.808</u>	<u>0.894</u>	<u>0.844</u>	<u>0.757</u>
RASD ⁺	<u>0.352</u>	0.538	<u>0.389</u>	<u>0.260</u>	0.814	0.896	0.848	0.765

Table 6. Link prediction results for YAGO3-10 datasets.

Models	YAGO3-10			
	MRR	Hits		
		@10	@3	@1
TransE [7]	0.238	0.447	0.361	0.212
TransR [16]	0.256	0.478	0.356	0.223
DistMult [8]	0.340	0.540	0.380	0.240
ComplEx [18]	0.360	0.550	0.400	0.260
R-GCN [10]	0.235	0.421	0.376	0.212
ConvE [9]	0.440	0.620	0.490	0.350
TorusE [25]	0.481	0.640	0.521	0.387
InteractE [26]	0.541	0.688	0.588	0.461
Rot-Pro [27]	0.542	0.699	0.596	0.443
Rule [6]	0.535	0.694	0.588	0.447
MSHE [20]	0.537	0.682	-	0.460
CompoundE3D WDS [33]	0.551	0.703	<u>0.608</u>	0.463
RASD	<u>0.558</u>	0.693	0.607	<u>0.482</u>
RASD ⁺	0.562	<u>0.702</u>	0.609	0.486

4.3. Heterogeneous Relation Modeling

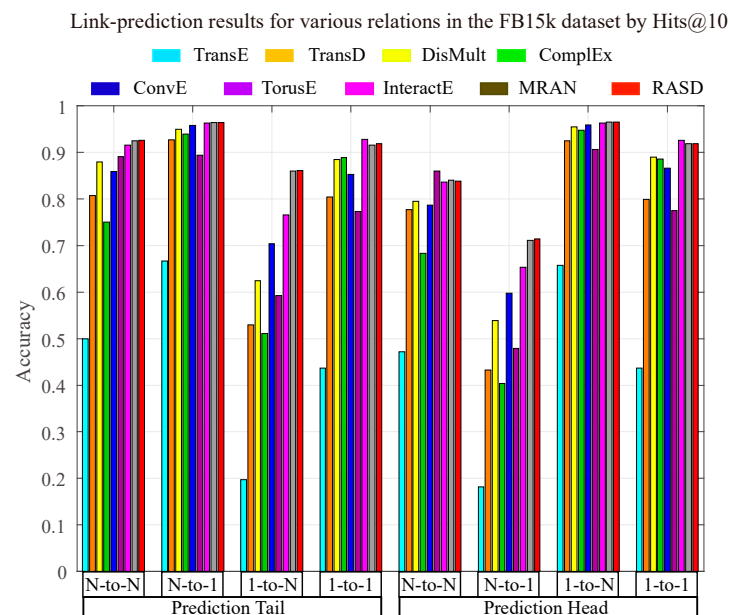
Table 7 and Figure 3 report the link prediction performance of the RASD framework on the FB15k and WN18 datasets when modeling complex heterogeneous relations. Figure 3 examines four relation patterns in FB15k (1-to-1, 1-to-N, N-to-1, and N-to-N), evaluating both tail- and head-entity prediction using Hits@10, thereby alleviating path-dependence effects inherent in FB15k. Against eight competing approaches, RASD achieves the highest Hits@10 in three of the four categories for tail-entity prediction and demonstrates highly competitive performance on head-entity prediction.

Table 7 summarizes the link-prediction results on WN18. Relations are partitioned into symmetric, asymmetric, and inverse groups. Best and second-best scores are highlighted in bold and underlined, respectively. RASD attains either the top or runner-up MRR in 17 distinct relations, excelling particularly on symmetric relations, securing three first-place and one second-place results among the four symmetric relation types. Notably, RASD reaches an MRR of 1.000 on the symmetric relation *similar_to* and the asymmetric relation *synset_domain_usage_of*, and achieves a 3.67% relative improvement over the prior state-of-the-art on the *member_of_domain_topic* relation.

These findings demonstrate that RASD effectively models complex heterogeneous relations. By integrating a frequency-aware attention mechanism, RASD more accurately discriminates among relation properties and realizes finer-grained modeling in the frequency domain, thereby substantially improving link prediction accuracy.

Table 7. Link prediction results for relations of different properties contained in the WN18 by MRR.

Relation Properties	Relation Names	Models						
		RASD	InteractE	ConvE	ComplEx	DistMult	TransD	TransE
Symmetric Relation	verb_group	0.975	<u>0.973</u>	0.805	0.936	0.782	0.746	0.283
	similar_to	1.000	1.000	<u>0.917</u>	1.000	0.832	0.724	0.242
	derivationally_related_form	0.958	<u>0.953</u>	0.821	0.946	0.726	0.596	0.362
	also_see	<u>0.666</u>	0.675	0.634	0.603	0.571	0.462	0.257
Asymmetric and Inverse Relation	synset_domain_usage_of	1.000	0.952	<u>0.960</u>	1.000	0.927	0.883	0.182
	synset_domain_region_of	0.959	0.959	<u>0.946</u>	0.919	0.918	0.783	0.197
	member_of_domain_region	0.942	0.942	<u>0.923</u>	0.865	0.901	0.872	0.358
	member_of_domain_usage	0.924	0.953	<u>0.928</u>	0.917	0.912	0.879	0.270
	instance_hypernym	0.965	<u>0.963</u>	0.948	0.965	0.922	0.854	0.680
	instance_hyponym	<u>0.966</u>	0.971	0.957	0.945	0.924	0.726	0.626
	Hypernym	<u>0.957</u>	0.958	0.891	0.953	0.854	0.652	0.376
	Hyponym	<u>0.953</u>	0.954	0.882	0.946	0.861	0.653	0.379
	member_meronym	0.938	<u>0.933</u>	0.816	0.921	0.783	0.678	0.433
	member_holonym	0.954	<u>0.949</u>	0.837	0.946	0.814	0.753	0.438
	part_of	0.943	<u>0.940</u>	0.881	<u>0.940</u>	0.842	0.684	0.415
	has_part	0.951	<u>0.946</u>	0.861	0.933	0.854	0.733	0.417
	synset_domain_topic_of	0.940	0.925	0.890	<u>0.930</u>	0.813	0.693	0.536
	member_of_domain_topic	0.958	0.923	0.922	<u>0.924</u>	0.899	0.783	0.502

**Figure 3.** Link-prediction results for relations of different properties contained in the FB15k.

4.4. Visualizations

Dropout study. Dropout is employed to enhance model generalization. We conducted a grid search over hidden-dropout and feature map dropout rates in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6] on the WN18RR and FB15k-237 datasets to assess their impact on four link prediction metrics (MRR, Hits@1, Hits@3, Hits@10). As shown in Figure 4, the results indicate that, on FB15k-237, setting the hidden dropout rate to 0.6 and the feature map dropout rate to 0.3 yields optimal performance across all metrics. Conversely, on WN18RR, the best outcomes are achieved with a hidden dropout rate of 0.4 and a feature map dropout rate of 0.3. These findings suggest that a moderately elevated hidden-dropout rate effectively mitigates overfitting, while a mid-range feature map dropout rate facilitates the extraction

of more robust local features, thereby improving the generalization of the model across diverse data distributions.

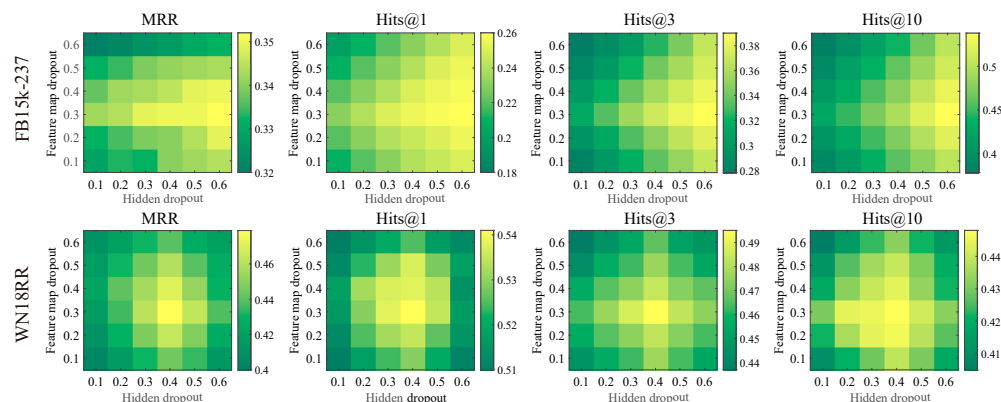


Figure 4. Performance of various hidden and feature map dropouts on FB15k-237 and WN18RR datasets.

Augmentation value study. Augmentation value k determines the size of the fast 1D convolution kernel, a parameter that not only influences the receptive field of the convolution, but also directly relates to the efficiency of feature integration across channels. To evaluate the effect of this parameter on the performance of the model, we conducted systematic experiments on the FB15k and YAGO3-10 datasets. In Figure 5, Figure 5a corresponds to the YAGO3-10 dataset and Figure 5b corresponds to the FB15k dataset; the results show that setting $k = 3$ yields optimal outcomes across all four evaluation metrics on both datasets. A smaller augmentation value weakens the capacity to capture cross-channel interactions and thus limits representational power, whereas an excessively large value may introduce redundant information and reduce precision of feature representations. Taken together, $k = 3$ achieves the best balance by capturing essential channel dependencies while suppressing extraneous noise, enabling RASD to achieve peak performance under this configuration.

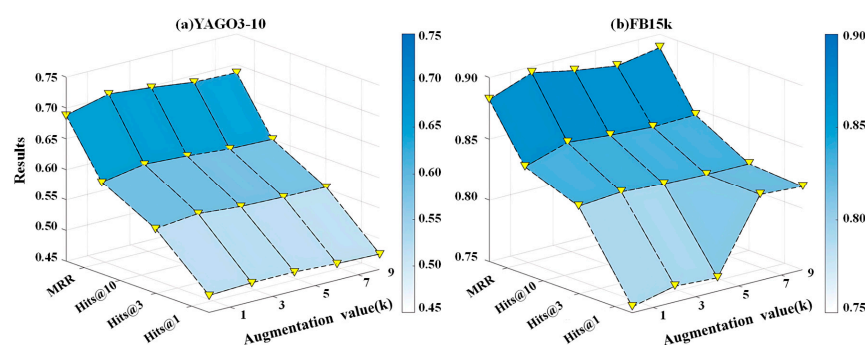


Figure 5. Performance of various augmentation values on YAGO3-10 and FB15k datasets.

Impact of batch size and label smoothing. Batch size determines the number of samples used to compute gradients during each forward and backward pass, and the proper selection of this parameter is critical for training efficiency and model performance. Label smoothing adjusts ground truth labels in loss computation, enhancing model robustness to noise and data bias while improving the stability of gradient updates. To assess the impact of these two hyperparameters on model performance, extensive experiments were conducted on the WN18 and YAGO3-10 datasets with the mean reciprocal rank as an evaluation metric, as shown in Figure 6.

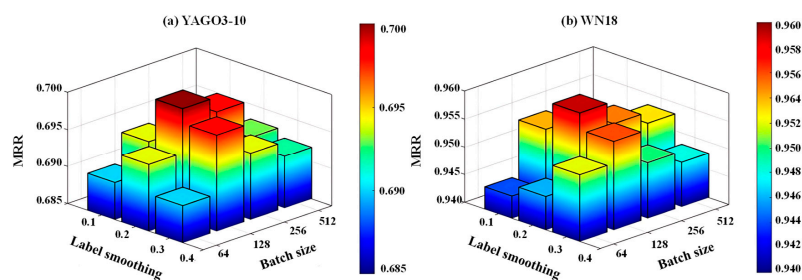


Figure 6. Performance of different batch sizes and label smoothing on YAGO3-10 and WN18 datasets.

Figure 6a corresponds to the YAGO3-10 dataset and Figure 6b corresponds to the WN18 dataset. Results indicate that the optimal performance of RASD is achieved when label smoothing equals 0.2 and batch size equals 128, whereas performance is lowest when both label smoothing and batch size adopt low values. These findings demonstrate that the appropriate choice of label smoothing and batch size is essential for optimizing model performance; however, excessively large batch size or overly strong label smoothing may induce underfitting and reduce model accuracy.

Spectral visualization of relation heterogeneity. Figure 7 presents relation-level average frequency-domain gain heatmaps for the WN18RR dataset. Each row corresponds to a relation; *similar_to* and *also_see* are symmetric relations, whereas *hypernym* and *member_of_domain_region* are asymmetric relations. Color intensity denotes the average amplification factor assigned by SDA at each frequency band, with brighter colors indicating stronger weighting of the corresponding spectral component. The results reveal a consistent spectral pattern: symmetric relations exhibit pronounced gains in low-frequency components, which indicates smoother and more globally consistent vector representations; asymmetric relations show stronger responses in mid-to-high frequency components, which indicates representations with greater local variation or directional features. These observations suggest that spectral decoupling separates interaction patterns across semantic scales in feature space and that SDA assigns differentiated frequency-band weights to different relation types, thereby improving link prediction performance.

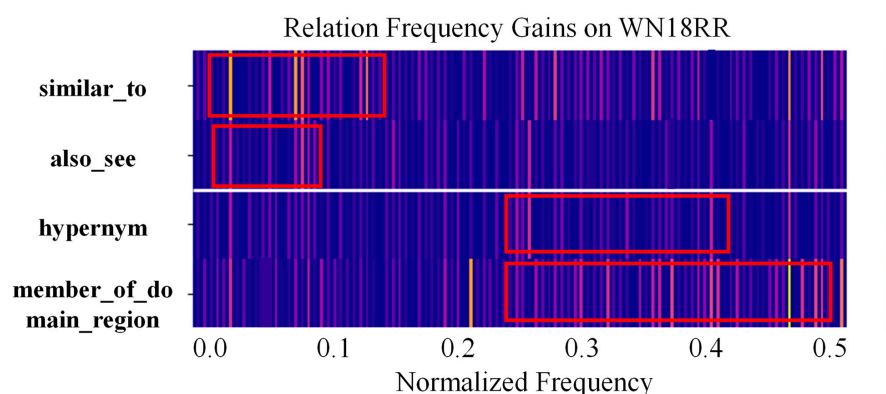


Figure 7. Spectral visualization of relation heterogeneity on WN18RR, where *similar_to* and *also_see* are symmetric relations, and *hypernym* and *member_of_domain_region* are asymmetric relations.

4.5. Ablation Study

To evaluate the specific contributions of each module in the model, we conducted ablation studies on three datasets to investigate the impact of different components on the performance of the model. Specifically, we designed six versions of our model for comparative analysis: the complete version of variant RASD+, as well as five modified versions with individual modules removed; *w/o* RAL retains only the spectral decoupling attention network module for KGR, while the *w/o* RFM and *w/o* RAM versions exclude

the relational filters and relation augmentation module, respectively. The *w/o* FFT version replaces FFT with the wavelet transform for frequency-domain transformation and *w/o* SDA removes the spectral decoupling attention network module and relies solely on the relation-aware learning module to capture additional relational information. As shown in Table 8, the variant RASD+ consistently outperforms the other versions across all datasets, demonstrating the effectiveness of the collaborative integration of its modules. Furthermore, compared with *w/o* RAL, the *w/o* RFM and *w/o* RAM versions achieve better results, indicating that relational filtering and relation augmentation can effectively capture relational patterns and improve model performance. In addition, the results of *w/o* FFT indicate that the wavelet transform is less effective than the Fourier transform in extracting frequency-domain features, confirming the advantage of FFT in modeling frequency information. Finally, the results of *w/o* SDA demonstrate that the spectral decoupling attention module can effectively enhance the modeling of both high- and low-frequency features, thereby significantly improving overall model performance.

Table 8. Link prediction results of different versions of RASD.

Models	WN18		FB15k		YAGO3-10	
	MRR	Hits@1	MRR	Hits@1	MRR	Hits@1
RASD+	0.952	0.949	0.814	0.765	0.562	0.486
<i>w/o</i> RAL	0.949	0.945	0.799	0.742	0.556	0.476
<i>w/o</i> RFM	0.950	0.946	0.804	0.747	0.560	0.481
<i>w/o</i> RAM	0.949	0.946	0.801	0.745	0.559	0.479
<i>w/o</i> FFT	0.948	0.944	0.797	0.739	0.553	0.474
<i>w/o</i> SDA	0.950	0.945	0.795	0.736	0.555	0.473

4.6. Model Efficiency Evaluation

Table 9 reports the per-epoch training time and space complexity of RASD and RASD+ on four benchmark datasets. As shown in Table 9, the space complexity of RASD is $(|\mathcal{E}| + |\mathcal{R}|)d$, where d denotes the embedding dimensionality. In terms of per-epoch runtime, InteractE is the slowest across all datasets, whereas RASD achieves runtimes comparable to those of three relatively shallow baselines (DisMult, ComplEx, ConvE), indicating that RASD maintains competitive computational efficiency while providing richer representations. As a variant of RASD, RASD+ further reduces per-epoch time by processing features in blocks in the frequency domain and using multi-channel parallelism; this parallelized design also contributes to good scalability on larger graphs, facilitating effective deployment on large-scale knowledge-graph scenarios.

Table 9. Training time comparison for each epoch of various datasets.

Models	Running Time (s)				Space Complexity $\mathcal{O}$
	FB15k-237	WN18RR	FB15k	WN18	
DisMult	26.07	22.68	48.19	35.90	$(\mathcal{E} + \mathcal{R})d$
ComplEx	20.29	22.26	38.60	37.25	$(\mathcal{E} + \mathcal{R})d$
ConvE	20.36	26.06	39.54	54.66	$(\mathcal{E} + \mathcal{R})d$
InteractE	93.06	111.65	100.37	92.17	$(\mathcal{E} + \mathcal{R})d$
RASD	22.05	23.09	41.38	39.67	$(\mathcal{E} + \mathcal{R})d$
RASD+	21.64	22.75	40.07	38.68	$(\mathcal{E} + \mathcal{R})d$

5. Conclusions and Future Discussion

Our work presents a novel relation aware spectral decoupling attention network (RASD) to address the limitations of existing KGR models in capturing relational semantic

heterogeneity and entity co-occurrence frequency. By integrating a spectral decoupling attention network module with a relation-aware learning module, RASD is capable of jointly modeling global interaction patterns and local relational features. The spectral decoupling attention network module projects joint embeddings of entities and relations into the frequency domain, enabling the extraction of multi-band spectral features and adaptive allocation of attention across frequency components. This facilitates the effective modeling of both high-frequency and low-frequency information. In parallel, the relation-aware learning module incorporates relation-specific filters and an augmentation mechanism to preserve the intrinsic semantic characteristics of diverse relations while suppressing irrelevant signals, thereby enhancing the ability of the model to represent heterogeneous relations. Experimental results on multiple benchmark datasets demonstrate that RASD consistently achieves substantial and robust improvements in link prediction performance, validating its effectiveness in modeling complex semantic structures.

In practice, RASD can support a wide range of cross-domain KG downstream tasks, providing a solid technical foundation for policy-making and decision-support applications. Its enhanced multi-scale relational modeling capacity and efficient reasoning performance enable a more accurate identification of potential research hypotheses in scientific discovery. In medical KGs, RASD can improve the efficiency and accuracy of the disease–drug association prediction, thereby accelerating the integration of clinical knowledge. Furthermore, in knowledge-driven domain-specific question answering systems, it facilitates more precise and rapid multi-hop reasoning, while in recommendation systems it better captures diverse entity relations to enhance personalized recommendation quality. These application domains typically involve large-scale, complex, and dynamically evolving KGs, and the efficiency and scalability of RASD make it well suited for deployment in such real-world scenarios.

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

Funding: This work has been supported by the National Natural Science Foundation of China (Grant No. 62477022) and the Fundamental Research Funds for the Central Universities at Central China Normal University: Artificial Intelligence + (No. CCNU24ai013).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in the GitHub repository KGDatasets at <https://github.com/ZhenfengLei/KGDatasets> (accessed on 13 August 2025).

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

References

1. Liu, R.; He, L.; Zheng, J. GLARA: A Global–Local Attention Framework for Semantic Relation Abstraction and Dynamic Preference Modeling in Knowledge-Aware Recommendation. *Appl. Sci.* **2025**, *15*, 6386. [CrossRef]
2. Chen, Y.; Sun, S.; Hu, X. DRKG: Faithful and Interpretable Multi-Hop Knowledge Graph Question Answering via LLM-Guided Reasoning Plans. *Appl. Sci.* **2025**, *15*, 6722. [CrossRef]
3. Wang, Y.; Yu, Z.; Wang, Z.; Yu, Z.; Wang, J. Multi-Examiner: A Knowledge Graph-Driven System for Generating Comprehensive IT Questions with Higher-Order Thinking. *Appl. Sci.* **2025**, *15*, 5719. [CrossRef]

4. Galárraga, L.A.; Teflioudi, C.; Hose, K.; Suchanek, F. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 413–422.
5. Schoenmackers, S.; Davis, J.; Etzioni, O.; Weld, D.S. Learning first-order horn clauses from web text. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA, 9–11 October 2010; pp. 1088–1098.
6. Tang, X.; Zhu, S.-C.; Liang, Y.; Zhang, M. RuleE: Knowledge Graph Reasoning with Rule Embedding. In Proceedings of the Findings of the Association for Computational Linguistics ACL 2024, Bangkok, Thailand, 11–16 August 2024; pp. 4316–4335.
7. Bordes, A.; Usunier, N.; Garcia-Durán, A.; Weston, J.; Yakhnenko, O. Translating embeddings for modeling multi-relational data. In Proceedings of the 26th International Conference on Neural Information Processing Systems—Volume 2, Lake Tahoe, Nevada, 5–10 December 2013; pp. 2787–2795.
8. Yang, B.; Yih, W.-t.; He, X.; Gao, J.; Deng, L. Embedding entities and relations for learning and inference in knowledge bases. *arXiv* **2014**, arXiv:1412.6575.
9. Dettmers, T.; Minervini, P.; Stenetorp, P.; Riedel, S. Convolutional 2d knowledge graph embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
10. Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; van den Berg, R.; Titov, I.; Welling, M. Modeling Relational Data with Graph Convolutional Networks. In Proceedings of the European Semantic Web Conference, Crete, Greece, 3–7 June 2018; pp. 593–607.
11. Meilicke, C.; Fink, M.; Wang, Y.; Ruffinelli, D.; Gemulla, R.; Stuckenschmidt, H. Fine-grained evaluation of rule-and embedding-based systems for knowledge graph completion. In Proceedings of the Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, 8–12 October 2018; Proceedings, Part I 17. pp. 3–20.
12. Meilicke, C.; Chekol, M.W.; Fink, M.; Stuckenschmidt, H. Reinforced anytime bottom up rule learning for knowledge graph completion. *arXiv* **2020**, arXiv:2004.04412. [[CrossRef](#)]
13. Ferré, S. Link prediction in knowledge graphs with concepts of nearest neighbours. In Proceedings of the Semantic Web: 16th International Conference, ESWC 2019, Portorož, Slovenia, 2–6 June 2019; Proceedings 16. pp. 84–100.
14. Wang, Z.; Zhang, J.; Feng, J.; Chen, Z. Knowledge graph embedding by translating on hyperplanes. In Proceedings of the AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014.
15. Ji, G.; He, S.; Xu, L.; Liu, K.; Zhao, J. Knowledge graph embedding via dynamic mapping matrix. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 687–696.
16. Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; Zhu, X. Learning entity and relation embeddings for knowledge graph completion. In Proceedings of the AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
17. Nickel, M.; Tresp, V.; Kriegel, H.-P. A three-way model for collective learning on multi-relational data. In Proceedings of the ICML, Bellevue, WA, USA, 28 June–2 July 2011; pp. 3104482–3104584.
18. Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; Bouchard, G. Complex embeddings for simple link prediction. In Proceedings of the 33rd International Conference on International Conference on Machine Learning—Volume 48, New York, NY, USA, 19–24 June 2016; pp. 2071–2080.
19. Fang, Y.; Lu, W.; Liu, X.; Pedrycz, W.; Lang, Q.; Yang, J.; Jing, W. CircularE: A Complex Space Circular Correlation Relational Model for Link Prediction in Knowledge Graph Embedding. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2023**, *31*, 3162–3175. [[CrossRef](#)]
20. Jiang, D.; Wang, R.; Xue, L.; Yang, J. Multisource hierarchical neural network for knowledge graph embedding. *Expert Syst. Appl.* **2024**, *237*, 121446. [[CrossRef](#)]
21. Wang, B.; Shen, T.; Long, G.; Zhou, T.; Wang, Y.; Chang, Y. Structure-augmented text representation learning for efficient knowledge graph completion. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 1737–1748.
22. Xu, Z.; Ye, P.; Chen, H.; Zhao, M.; Chen, H.; Zhang, W. Ruleformer: Context-aware rule mining over knowledge graph. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 2551–2560.
23. Liu, Y.; Sun, Z.; Li, G.; Hu, W. I know what you do not know: Knowledge graph embedding via co-distillation learning. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 1329–1338.
24. Li, G.; Sun, Z.; Hu, W.; Cheng, G.; Qu, Y. Position-aware relational transformer for knowledge graph embedding. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 11580–11594. [[CrossRef](#)] [[PubMed](#)]
25. Ebisu, T.; Ichise, R. Generalized Translation-Based Embedding of Knowledge Graph. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 941–951. [[CrossRef](#)]
26. Vashishth, S.; Sanyal, S.; Nitin, V.; Agrawal, N.; Talukdar, P.P. InteractE: Improving Convolution-Based Knowledge Graph Embeddings by Increasing Feature Interactions. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 3009–3016.

27. Song, T.; Luo, J.; Huang, L. Rot-Pro: Modeling Transitivity by Projection in Knowledge Graph Embedding. In Proceedings of the Thirty-Fifth Annual Conference on Advances in Neural Information Processing Systems (NeurIPS), Online, 6–14 December 2021; pp. 24695–24706.
28. Zeb, A.; Saif, S.; Chen, J.; Haq, A.U.; Gong, Z.; Zhang, D. Complex graph convolutional network for link prediction in knowledge graphs. *Expert Syst. Appl.* **2022**, *200*, 116796. [[CrossRef](#)]
29. Li, D.; Xia, T.; Wang, J.; Shi, F.; Zhang, Q.; Li, B.; Xiong, Y. SDFormer: A shallow-to-deep feature interaction for knowledge graph embedding. *Knowl.-Based Syst.* **2024**, *284*, 111253. [[CrossRef](#)]
30. Sun, Z.; Deng, Z.-H.; Nie, J.-Y.; Tang, J. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In Proceedings of the 7th International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019.
31. Le, T.; Huynh, N.; Le, B. Knowledge graph embedding by projection and rotation on hyperplanes for link prediction. *Appl. Intell.* **2023**, *53*, 10340–10364. [[CrossRef](#)]
32. Le, T.; Tran, H.; Le, B. Knowledge graph embedding with the special orthogonal group in quaternion space for link prediction. *Knowl.-Based Syst.* **2023**, *266*, 110400. [[CrossRef](#)]
33. Ge, X.; Wang, Y.C.; Wang, B.; Kuo, C.-C.J. Knowledge graph embedding with 3d compound geometric transformations. *APSIPA Trans. Signal Inf. Process.* **2024**, *13*, e4. [[CrossRef](#)]

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