Taming the Wild Evolution: Aligning Multi-Modal Temporal Knowledge Graphs (Technical Report)

***Abstract*—This technical report provides additional details on the method, dataset tasks and complete experimental setup for the paper "Taming the Wild Evolution: Aligning Multi-Modal Temporal Knowledge Graphs".**

I. Supplementation of Method

A. Motivation

The MTKGA-Wild task presents two critical properties that distinguish it from traditional entity alignment. First, determining entity correspondence requires examining *collective temporal-modal evolution patterns* rather than isolated fact comparisons. Consider aligning a political figure across two MTKGs: one entity may exhibit rich video coverage during campaign periods (2020-2024) but sparse image data, while its counterpart shows the opposite pattern with extensive historical image archives (1990-2000) but limited recent videos. Traditional pairwise graphs evaluate each candidate independently, failing to capture the high-order dependencies where alignment decisions must simultaneously weigh evidence across multiple temporal scales and modality types. Hypergraphs naturally address this limitation by enabling each source entity to form a hyperedge connecting multiple candidate projections, where the hyperedge structure encodes collective reasoning: "this entity's alignment emerges from jointly comparing these candidates across all available modal-temporal dimensions." Second, the dynamic imbalance challenge creates a constantly shifting decision landscape where modality availability varies dramatically across timestamps. Recent events rely heavily on video and audio modalities due to social media proliferation, while historical events depend predominantly on text and image archives. This temporal heterogeneity in modal distributions demands adaptive strategy selection rather than uniform alignment procedures. Multi-agent systems provide this adaptivity by design: each agent acts as a specialized operator excelling at specific modal-temporal conditions (e.g., video-dominant alignment, text-sparse compensation), while the meta-agent coordinates operator selection based on detected conflicts in the current hypergraph state. This on-demand collaboration mechanism directly mirrors how the alignment strategy must dynamically adjust when facing varying data availability across time periods, transforming the challenge of dynamic imbalance into a collaborative decision-making problem where specialized agents address localized modal-temporal inconsistencies.

These task-driven insights directly inform our algorithmic design in Algorithm 1. Neural retrieval and adaptive decoupling (lines 1-4) decompose complex temporal-modal evolution into comparable projections that enable hyperedge construction, addressing evolutionary diversity through structured dimensionality reduction. Hypergraph construction (lines 5-7) materializes the collective reasoning framework where alignment decisions emerge from hyperedge-level comparisons rather than isolated pairwise evaluations. The MDP-based multi-agent loop (lines 9-19) operationalizes dynamic strategy adaptation: conflict detection identifies regions of modal imbalance (lines 10-14), core block selection activates specialized operators for those scenarios (line 9), agentic execution adjusts hypergraph structures through targeted actions (line 15), and meta evaluation distills successful strategies as reusable signals (lines 16-19). This closed-loop design ensures continuous improvement in handling diverse temporal-modal conditions inherent to MTKGA-Wild, where the hypergraph representation provides the structural foundation for collective reasoning and the multi-agent collaboration supplies the adaptive mechanism for dynamic imbalance resolution.

B. The Workflow of EvoWildAlign

In conclusion, the complete workflow of EvoWildAlign is presented in Algorithm 1. The algorithm takes as input a source MTKG $\mathcal{G}^s$, a target MTKG $\mathcal{G}^t$, and an alignment goal g, and outputs the entity alignment results $\mathcal{A}_{align}$. The algorithm proceeds through two main stages to systematically address the challenges of evolutionary diversity and dynamic imbalance in MTKGA-Wild.

In the *neuro-symbolic evolution hypergraph representation* stage, the algorithm first initializes an empty pre-aligned subgraph set (line 1). For each source entity $e^s \in \mathcal{E}^s$, neural retrieval is performed to identify the top-k most similar target entities using textual embedding similarity, forming pre-aligned subgraphs that are added to the set (lines 2). Subsequently, the algorithm initializes a decoupled projection set and applies adaptive evolution projection to each entity pair in the pre-aligned set (lines 3-4). Specifically, temporal masking projection removes timestamps in target entities that do not appear in source entities, while modal masking projection eliminates absent modality types. Each projection is further decomposed by modality types to ensure single-modality association. The algorithm then constructs modal projection hypergraphs $\mathcal{H}_m$ for each modality $m \in \mathcal{M}$ (lines 5-6), where each source entity forms a hyperedge connecting its filtered target projections. Finally, these modal projection hypergraphs are aggregated into a unified neuro-symbolic evolution hypergraph $\mathcal{H}^{evo}$ that captures comprehensive multi-modal temporal evolution (line 7).

Algorithm 1: EvoWildAlign Framework

Input: Source MTKG $\mathcal{G}^s$, Target MTKG $\mathcal{G}^t$, Alignment goal g

Output: Entity alignment results $\mathcal{A}_{align}$

`// Stage 1: Neuro-symbolic Evolution Hypergraph Representation`

1 Initialize pre-aligned subgraph set $\mathcal{P} \leftarrow \emptyset$; **for** Each entity $e^s \in \mathcal{E}^s$ **do**

2 Perform neural retrieval: $\text{TopK}(e^s) \leftarrow e^t_1, \ldots, e^t_k \subseteq \mathcal{E}^t$; Construct pre-aligned subgraph and add to $\mathcal{P}$;

3 Initialize decoupled projection set $\tilde{\mathbb{P}} \leftarrow \emptyset$; **for** Each entity pair (e^s, e^t) in $\mathcal{P}$ **do**

4 Apply temporal masking projection: $\tilde{\mathbb{P}}^{(e^s,e^t)}_{time} \leftarrow \mathbb{P}_{time}(e^t | \mathcal{T}(e^s))$; Apply modal masking projection: $\tilde{\mathbb{P}}^{(e^s,e^t)}_{modal} \leftarrow \mathbb{P}_{modal}(e^t | \mathcal{M}(e^s))$; Decompose projections by modality types and add to $\tilde{\mathbb{P}}$;

5 **for** Each modality $m \in \mathcal{M}$ **do**

6 Construct modal projection hypergraph: $\mathcal{H}_m \leftarrow (\mathcal{V}_m, \varepsilon_m)$;

7 Aggregate into neuro-symbolic evolution hypergraph: $\mathcal{H}^{evo} \leftarrow (\mathcal{V}, \varepsilon, \mathcal{L}, \Phi)$;

`// Stage 2: On-demand Agentic Hypergraph Collaboration`

8 Initialize state $s \leftarrow \mathcal{H}^{evo}$, meta-fusion signals $\mathcal{F}_{\text{meta}} \leftarrow \emptyset$; **while** Goal g not achieved and iterations $< max_{iterations}$ **do**

`// Core Block Selection`

9 Select collaboration group: $\mathcal{B}(s, g) \leftarrow \mathcal{B}_{\text{meta}}(s, g, r, \mathcal{F}_{\text{meta}})$;

`// Collaboration Decision-Making`

10 Initialize decision plan $\mathcal{D}(s) \leftarrow \emptyset$;

11 **for** Each source entity $e^s \in \mathcal{E}^s$ **do**

12 Detect cross-layer conflicts: $\text{Conflict}(e^s) \leftarrow$ Eq. (8);

13 **if** $\text{Conflict}(e^s) = 1$ **then**

14 Determine alignment action a_{e^s} and add (e^s, a_{e^s}) to $\mathcal{D}(s)$;

`// Agentic Hypergraph Execution`

15 Execute actions and update state: $s' \leftarrow \mathcal{T}(s, \mathcal{D}(s))$;

`// Meta Evaluation`

16 Evaluate alignment quality: $r \leftarrow \mathcal{R}(s', g)$;

17 Extract meta-fusion signals: $\mathcal{F}_{\text{meta}} \leftarrow \{$ successfully executed actions from $\mathcal{D}(s)\}$;

18 Extract pseudo-labeled pairs: $\mathcal{A}_{\text{seed}} \leftarrow \{(e^s, e^t) \mid \text{Conflict}(e^s) = 0\}$;

19 Update state: $s \leftarrow s'$;

20 Return entity alignment results $\mathcal{A}_{align}$;

In the *on-demand agentic hypergraph collaboration* stage, the algorithm models the alignment process as a Markov decision process. The initial state is set to the neuro-symbolic evolution hypergraph, and meta-fusion signals are initialized as empty (line 8). The algorithm then enters an iterative refinement loop that continues until the alignment goal is achieved or maximum iterations are reached (line 8). Each iteration consists of four key steps: In *core block selection*, a meta-agent dynamically selects appropriate operator agents and assigns role descriptions to form the collaboration group $\mathcal{B}(s, g)$ based on the current state, goal, reward, and meta-fusion signals (line 9). In *collaboration decision-making*, the algorithm initializes a decision plan and examines each source entity for cross-layer conflicts across modality layers (lines 10-14). Entities exhibiting dynamic imbalance—where different modality layers yield inconsistent top-ranked candidates—are identified, and corresponding alignment actions are added to the decision plan. In *agentic hypergraph execution*, agents execute the actions specified in the decision plan to refine the hypergraph structure and update the state (line 15). In *meta evaluation*, the algorithm assesses alignment quality, extracts successfully executed actions as meta-fusion signals for subsequent iterations, and generates pseudo-labeled alignment pairs from conflict-free entities (lines 16-19). If the goal remains unachieved, feedback and meta-fusion signals guide adaptive reconfiguration of the collaboration group in the next iteration. Upon convergence or reaching maximum iterations, the algorithm returns the final entity alignment results (line 20).

II. EXPERIMENTS

In this section, we first introduce the experimental setting in Section II-A. We then empirically evaluate the proposed datasets and framework from various perspectives to address the following research questions:

- **RQ1: How effective is the overall MTKGA-Wild framework?**
- **RQ2: Does the new task offer sufficient research value, and do the `WildMTKGA(W-I)` and `WildMTKGA(Y-I)` datasets provide meaningful insights for the solution to this new task?**
- **RQ3: What is the effectiveness of each component within EvoWildAlign?**
- **RQ4: Does the EvoWildAlign successfully balance alignment accuracy and efficiency in MTKGA-Wild?**
- **RQ5: To what degree is EvoWildAlign adaptable and robust across varying foundation models and more complex scenarios?**

A. Experimental Setting

1) ***MTKGA-Wild Benchmarks and Others:*** Currently, there is no benchmark that effectively studies and evaluates the MTKGA-Wild task. To this end, we construct two new MTKGA-Wild benchmarks, i.e., `WildMTKGA(W-I)` and `WildMTKGA(Y-I)`. As shown in Table I, the construction

TABLE I

DATASET STATISTICS [1]–[4]. "*#Ent*", "*#Rel.*", "*#Facts*": THE NUMBER OF ENTITIES, RELATIONS AND QUADRUPLES IN KG1 (KG2), RESPECTIVELY. "*Temporal.*", "*Multi-modal*": INDICATES WHETHER THE DATASET INCLUDES TEMPORAL KNOWLEDGE INFORMATION AND THE DATASET INCLUDES MULTI-MODAL INFORMATION, RESPECTIVELY.

Dataset		#Ent.	#Rel.	Temporal.	Multi-modal	#Seed	#Facts	Image	Text	Audio	Video
`DBP15K(EN-FR)`	EN FR	15,000 15,000	193 166	✗ ✗	✗ ✗	15,000	96,318 80,112	✗	✓	✗	✗
`DBP-WIKI`	DBP WIKI	100,000 100,000	413 261	✗ ✗	✗ ✗	100,000	293,990 251,708	✗	✓	✗	✗
`ICEWS-WIKI`	ICEWS WIKI	11,047 15,896	272 226	✓ ✓	✗ ✗	5,058	3,527,881 198,257	✗	✓	✗	✗
`ICEWS-YAGO`	ICEWS YAGO	26,863 22,734	272 41	✓ ✓	✗ ✗	18,824	4,192,555 107,118	✗	✓	✗	✗
`YAGO-WIKI50K`	YAGO WIKI	49,629 49,222	11 30	✓ ✓	✗ ✗	49,172	221,050 317,814	✗	✓	✗	✗
`BETA`	WIKI YAGO	42,666 42,297	257 45	✓ ✓	✗ ✗	40,364	199,879 162,320	✗	✓	✗	✗
`FB15K-DB15K`	Freebase DB	14,951 12,842	1,345 279	✗ ✗	✓ ✓	12,846	621,608 137,277	✓	✓	✗	✗
`WildMTKGA(W-I)`	WIKI ICEWS	13,194 10,052	224 268	✓ ✓	✓ ✓	4,978	196,117 3,527,053	✓	✓	✓	✓
`WildMTKGA(Y-I)`	YAGO ICEWS	21,098 25,316	41 272	✓ ✓	✓ ✓	17,895	107,018 4,192,212	✓	✓	✓	✓

of `WildMTKGA(W-I)` and `WildMTKGA(Y-I)` follows the established construction guidelines of predecessors [1].

- **MTKGA-Wild Benchmarks Construction Process:** The `WildMTKGA(W-I)` and `WildMTKGA(Y-I)` integrate the event knowledge graph derived from the Integrated Crisis Early Warning System (ICEWS) with representative general KGs (i.e., *WIKIDATA* and *YAGO*), incorporating rich multi-modal and temporal information. There exists considerable demand to align these KGs in real-world scenarios. *ICEWS* is a representative domain-specific KG containing political events, news, and other content with temporal annotations that embody the multi-modal and dynamic interactions between politically-related entities. *WIKIDATA* and *YAGO* are two common KGs with extensive multi-modal and time-evolving commonsense knowledge that provide comprehensive background information. Aligning them enables a more comprehensive view for understanding events and supports multi-modal temporal knowledge reasoning tasks. For ease of explanation, we describe the construction process using `WildMTKGA(W-I)` as an example, while the process for `WildMTKGA(Y-I)` follows similar procedures. Notably, since multi-modal information in YAGO primarily links to corresponding entities in WIKIDATA, we obtain YAGO's multi-modal information through these entity mappings. Following standard construction procedures [1], we first pre-processed raw event data from *ICEWS* spanning 1995-2021, obtained from the official website[1]. Simultaneously, we extracted multi-modal information (images, text, audio, video) from source news articles to provide multi-modal temporal support, transforming the raw data into KG format comprising entities, relations, and facts. Specifically, we pre-processed entity names in *ICEWS* and retrieved corresponding entities along with their multi-modal temporal information in WIKIDATA through the official Wikidata API. Next, we reviewed candidate entity pairs and their associated modal information, manually filtering unrealistic data to yield 23,912 high-quality entity pairs across *ICEWS* and *WIKIDATA*. We then sampled neighbors of these entity pairs from the original data without enforcing the 1-to-1 assumption. To maintain distributional similarity with the original KGs, we adopted the *Iterative Degree-based Sampling (IDS)* algorithm [4], which simultaneously removes entities from both KGs guided by the original degree distributions until reaching the desired size. Finally, the `WildMTKGA(W-I)` dataset was obtained through this sampling process.
- **Other non-MTKGA Benchmarks:** Additionally, we comprehensively evaluate the MTKGA-Wild benchmarks and our proposed approach using seven non-MTKGA benchmarks, which are summarized in Table I. For uni-modal temporal KGA benchmarks, we employ four representative datasets, i.e., `ICEWS-WIKI`, `ICEWS-YAGO`, `BETA`, and `YAGO-WIKI50K`. `ICEWS-WIKI` and `ICEWS-YAGO` [1] exemplify highly heterogeneous temporal knowledge graph alignment datasets[2], characterized by substantial differences between their constituent KGs. These differences manifest not only in the quantities of entities, relations, and facts, but also in structural characteristics (e.g., density and structural similarity) and their low overlapping ratios. Notably, the number of anchor links does not equal the entity count, rendering the alignment task more challenging on these complex HHEA datasets. `BETA` [5] is designed for more complex temporal EA scenarios, comprising six realistic complex cases that better reflect real-world situations with multi-granularity temporal features. The most widely adopted dataset, `YAGO-WIKI50K`, is derived from YAGO and Wikidata. Specifically, the `YAGO-WIKI50K` datasets are constructed by first selecting the top 50,000 most frequent entities from a Wikidata subset extracted by [6], then linking these entities to their counterparts in

[1]https://dataverse.harvard.edu/dataverse/icews

[2]https://github.com/IDEA-FinAI/Simple-HHEA

YAGO [3]. Temporal facts are subsequently incorporated to form two TKGs for alignment. For uni-modal static KGA benchmarks, we adopt two representative datasets, i.e., `DBP15K(EN-FR)` and `DBP-WIKI`. `DBP15K(EN-FR)` and `DBP-WIKI` [4] are well-established benchmarks in entity alignment. `DBP15K(EN-FR)` specifically targets cross-lingual entity alignment, while `DBP-WIKI` represents a larger-scale dataset for heterogeneous knowledge graphs. Nevertheless, these KGs exhibit similar structural properties, including comparable density and structural similarity, with 100% overlapping ratios and similar scales in terms of entities, relations, and facts. Furthermore, for multi-modal static KGA benchmarks, we select the most representative dataset, i.e., `FB15K-DB15K`. For the monolingual setting, we utilize `FB15K-DB15K` from MMKG [7], [8] with the standard data split, specifically employing 20% training data.

2) ***Benchmark Configurations****:* Currently, there is no specific solution for the MTKGA-Wild task. To establish a comprehensive evaluation framework for MTKGA-Wild, we systematically propose 27 representative and classic benchmark configurations for extensive comparison. These configurations can be divided into uni-modal static KGA configurations, uni-modal temporal KGA, and multi-modal static KGA configurations.

- **Uni-modal Static KGA Methods:** We collect 11 state-of-the-art uni-modal static KGA methods as benchmark configurations for comparison. These methods include MTransE [9], AlignE [10], BootEA [10], GCN-Align [11], RDGCN [12], Dual-AMN [13], BERT [14], FuAlign [15], BERT-INT [16], PARIS [17], [18] and Naive RAG [19]–[21].
  - MTransE [9] is a translation-based entity alignment method that introduces translation vectors to align entity embeddings across different languages. It extends the TransE model by learning cross-lingual transformations that map entities from one knowledge graph to another, enabling alignment between multilingual KGs.
  - AlignE [10] focuses on capturing essential relational patterns that can serve as alignment signals, improving the precision of entity matching.
  - BootEA [10] utilizes a bootstrapping approach that iteratively expands alignment seeds, progressively improving alignment quality through self-training on high-confidence predictions.
  - GCN-Align [11] trains Graph Convolutional Networks (GCNs) to embed entities from each language into a unified vector space. By leveraging the graph structure through neighborhood aggregation, it captures both structural and semantic information for cross-lingual entity alignment.
  - RDGCN [12] leverages GCNs for modeling structural information within knowledge graphs. It explicitly incorporates relational information through dual graph structures, enhancing the representation of both entities and their relationships.
  - Dual-AMN [13] jointly captures intra-graph and cross-graph dependencies through attention mechanisms. It models both the internal structure of individual KGs and the cross-KG correspondences simultaneously, enabling more comprehensive entity alignment.
  - BERT [14] is utilized as a pretrained language model to initialize entity embeddings using name-based features. By leveraging BERT's powerful contextual representations, it captures semantic information from entity names and descriptions to improve alignment performance.
  - FuAlign [15] incorporates auxiliary information to address knowledge graph heterogeneity. It fuses multiple types of information sources, including structural, textual, and attribute data, to handle the diverse characteristics of different KGs.
  - BERT-INT [16] combines BERT-based augmentation with auxiliary cues for improved alignment. It integrates pre-trained language model representations with additional contextual information, creating richer entity embeddings for cross-KG matching.
  - PARIS [17], [18] is a probabilistic iterative method capable of aligning entities without requiring prior alignments. It uses a holistic approach that simultaneously aligns entities, relations, and schema elements through probabilistic reasoning.
  - Naive RAG [19]–[21] is a basic LLM-based Retrieval-Augmented Generation approach that first retrieves relevant information based on a user query and then generates answers using the retrieved content. It represents a straightforward application of large language models to the entity alignment task.
- **Uni-modal Temporal KGA Methods:** We collect 11 state-of-the-art uni-modal temporal KGA methods as benchmark configurations for comparison. These methods include TEA-GNN [3], TREA [22], STEA [23], LightTEA [24], Dual-Match [25], Simple-HHEA [1], ChatEA [2], MGTEA [5], AdaCoAgentEA [20], Self-Consistency [20], [21], and Self-RAG [20], [21].
  - TEA-GNN [3] treats timestamps as link attributes, using a time-aware attention mechanism to enrich entity and relation representations. It explicitly models temporal dynamics to capture how entities evolve over time.
  - TREA [22] enhances training using neighborhood aggregation and margin-based multi-class loss. It incorporates temporal information into the relational structure, enabling more accurate alignment in temporal knowledge graphs.
  - STEA [23] utilizes a temporal dictionary to guide temporal alignment. It maintains explicit temporal mappings that help identify corresponding entities across different time periods in temporal KGs.

TABLE II

MAIN EXPERIMENT RESULTS ON MTKGA-WILD AND TKGA DATASETS. BEST RESULTS ARE HIGHLIGHTED IN **bold**, WHILE RUNNER-UP RESULTS ARE <u>underlined</u>. "*Struc.*, *Text.*, *Imag.*, *Audi.*, *Vide.*" INDICATES THAT THE METHOD CAN PROCESS STRUCTURAL INFORMATION, TEXT SEMANTICS, IMAGE INFORMATION, AUDIO INFORMATION, AND VIDEO INFORMATION, RESPECTIVELY. DENOTES WHETHER THE MODEL IS CAPABLE OF PERCEIVING TEMPORAL INFORMATION. AND INDICATE WHETHER THE MODEL UTILIZES LARGE LANGUAGE MODELS, RESPECTIVELY.THESE CONVENTIONS APPLY TO ALL SUBSEQUENT TABLES.

Method	Modality					WildMTKGA(W-I)			WildMTKGA(Y-I)			ICEWS-WIKI			ICEWS-YAGO		
	Struc.	*Text.*	*Imag.*	*Audi.*	*Vide.*	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
Uni-modal Static KGA Methods																	
MTransE						0.015	0.155	0.061	0.008	0.081	0.032	0.021	0.158	0.068	0.012	0.084	0.040
AlignE	✓					0.054	0.261	0.119	0.010	0.112	0.046	0.057	0.261	0.122	0.019	0.118	0.055
BootEA	✓					0.053	0.261	0.123	0.009	0.099	0.046	0.072	0.275	0.139	0.020	0.120	0.056
GCN-Align	✓					0.043	0.183	0.086	0.016	0.075	0.037	0.046	0.184	0.093	0.017	0.085	0.038
Dual-AMN *(basic)*	✓					0.065	0.282	0.130	0.012	0.134	0.063	0.077	0.285	0.143	0.032	0.147	0.069
Dual-AMN *(semi-supervised)*	✓					0.026	0.175	0.084	0.015	0.080	0.023	0.037	0.188	0.087	0.020	0.093	0.045
RDGCN	✓	✓				0.050	0.191	0.088	0.016	0.079	0.031	0.064	0.202	0.096	0.029	0.097	0.042
Dual-AMN *(name)*	✓	✓				0.064	0.243	0.130	0.012	0.112	0.038	0.083	0.281	0.145	0.031	0.144	0.068
BERT		✓				0.341	0.503	0.480	0.360	0.512	0.451	0.546	0.687	0.596	0.749	0.845	0.784
FuAlign	✓	✓				0.252	0.466	0.329	0.299	0.571	0.414	0.257	0.570	0.361	0.326	0.604	0.423
BERT-INT	✓	✓				0.424	0.510	0.486	0.373	0.572	0.488	0.561	0.700	0.607	0.756	0.859	0.793
PARIS	✓	✓				0.431	-	-	0.513	-	-	0.672	-	-	0.687	-	-
Naive RAG *	✓	✓				0.463	-	-	0.545	-	-	0.723	-	-	0.736	-	-
Uni-modal Temporal KGA Methods																	
TEA-GNN	✓					0.043	0.240	0.114	0.014	0.104	0.063	0.063	0.253	0.126	0.025	0.135	0.064
TREA	✓					0.070	0.270	0.129	0.019	0.138	0.036	0.081	0.302	0.155	0.033	0.150	0.072
STEA	✓					0.072	0.249	0.126	0.019	0.119	0.035	0.079	0.292	0.152	0.033	0.147	0.073
Dual-Match	✓	✓				0.035	0.200	0.096	0.018	0.111	0.049	0.043	0.204	0.096	0.022	0.114	0.053
Simple-HHEA		✓				0.509	0.641	0.573	0.517	0.675	0.597	0.720	0.872	0.754	0.847	0.915	0.870
Simple-HHEA *(structure)*	✓	✓				0.514	0.640	0.567	0.534	0.648	0.561	0.639	0.812	0.697	0.749	0.864	0.775
ChatEA	✓	✓				0.590	0.685	0.611	<u>0.622</u>	0.713	<u>0.699</u>	<u>0.880</u>	<u>0.945</u>	<u>0.912</u>	<u>0.935</u>	<u>0.955</u>	<u>0.944</u>
MGTEA	✓	✓				0.421	0.500	0.471	0.417	0.516	0.442	0.534	0.693	0.627	0.673	0.780	0.704
AdaCoAgentEA *(supervised)*	✓	✓				<u>0.625</u>	<u>0.736</u>	<u>0.683</u>	0.621	<u>0.794</u>	0.685	0.844	0.882	0.861	0.872	0.914	0.889
Self-Consistency *	✓	✓				0.372	-	-	0.419	-	-	0.718	-	-	0.725	-	-
Self-RAG *	✓	✓				0.476	-	-	0.592	-	-	0.772	-	-	0.762	-	-
Multi-modal Static KGA Methods																	
EVA	✓	✓	✓			0.443	0.546	0.475	0.391	0.471	0.421	-	-	-	-	-	-
MEAformer	✓	✓	✓			0.534	0.621	0.583	0.503	0.657	0.564	-	-	-	-	-	-
MMKG-CoT *	✓	✓	✓			0.501	-	-	0.530	-	-	0.618	-	-	0.626	-	-
MMKG-RAG *	✓	✓	✓			0.551	-	-	0.588	-	-	0.682	-	-	0.697	-	-
EvoWildAlign (Ours)	✓	✓	✓	✓	✓	**0.818**	**0.871**	**0.836**	**0.829**	**0.897**	**0.861**	**0.988**	**0.997**	**0.991**	**0.979**	**0.997**	**0.987**

- LightTEA [24] is designed as a lightweight temporal knowledge graph alignment model. Although it aims for efficiency, its temporal component has shown limited improvements on existing datasets, suggesting room for enhancement in temporal modeling.
- Dual-Match [25] employs a temporal encoder for unsupervised layer-wise propagation. It processes temporal information through multiple layers, gradually refining entity representations by incorporating temporal patterns without requiring extensive supervision.
- Simple-HHEA [1] is a representation learning-based approach tailored for aligning heterogeneous and temporal knowledge graphs. It addresses both structural heterogeneity and temporal dynamics in a unified framework.
- ChatEA [2] applies large language models with fine-tuning to perform advanced knowledge graph alignment. It leverages the reasoning capabilities of LLMs and adapts them specifically for the entity alignment task through targeted fine-tuning.
- MGTEA [5] proposes a simple yet effective multi-granularity approach for temporal alignment. It captures temporal information at different levels of granularity, from fine-grained timestamps to coarse-grained temporal periods.
- AdaCoAgentEA [20] represents an adaptive coordination-based agent approach for entity alignment, incorporating both structural and temporal information through coordinated multi-agent learning strategies.
- Self-Consistency [20], [21] is a chain-of-thought baseline that produces multiple reasoning paths and selects the most frequent answer as the final output. In the implementation, it is enhanced by using the top-1 most similar entity from the similarity matrix produced by Simple-HHEA as a preprocessing step for the knowledge graph.
- Self-RAG [20], [21] is a self-reflective Retrieval-Augmented Generation method aimed at improving the generation quality of large language models. It incorporates self-reflection mechanisms that enable the model to critique and refine its own outputs during the alignment process.

- **Multi-modal Static KGA Methods:** We collect 5 state-of-the-art uni-modal static KGA methods as benchmark configurations for comparison. These methods include EVA [26], MMEA [8], MEAformer [8], MMKG-CoT [27], and MMKG-RAG [19].

- EVA [26] is a multi-modal entity alignment method that incorporates visual information alongside structural and textual features, enabling alignment across knowledge graphs with rich multi-modal attributes.
- MMEA [8] leverages multiple modalities including text, images, and structure to perform entity alignment, addressing the challenge of aligning KGs with diverse information types.
- MEAformer [8] is a transformer-based multi-modal entity alignment framework that uses attention mechanisms to effectively fuse information from different modalities for improved alignment performance.
- MMKG-CoT [27] applies chain-of-thought reasoning to multi-modal knowledge graph alignment, enabling step-by-step reasoning over multi-modal entity features.
- MMKG-RAG [19] combines multi-modal knowledge graphs with Retrieval-Augmented Generation, retrieving relevant multi-modal information to support LLM-based entity alignment decisions.

*3) **Evaluation Protocol and Implementation Details**:* To evaluate the quality of MTKGA-Wild produced by EvoWildAlign and the benchmark configurations, we follow established conventions [1], [4] and use Hits@1/5/10 and Mean Reciprocal Rank (MRR) as evaluation metrics. Since some models provide only final alignment results, Hits@1 is replaced with precision for evaluation consistency. Additionally, we use runtime (in seconds) as an efficiency metric. For the multi-modal static KGA benchmark dataset, we use 20% as the training set. For all other benchmark datasets, we follow the 3:7 splitting ratio for training/testing data. All baseline methods with * follow the same preprocessing procedure to obtain pseudo-labels for use as input by *neural retrieval*.

All LLMs reported in Table II and Table III are implemented using the same model version, GPT-4 (gpt-4-0125-preview). For subsequent experiments, unless otherwise specified, GPT-3.5 (gpt-3.5-turbo-1106) is used as the default LLM configuration due to its lower computational cost. All experiments are conducted on a server equipped with four NVIDIA GeForce RTX 4090 graphics cards, each featuring 24GB of GDDR6X memory. The system is powered by a 64-core processor and 480GB of system RAM. For storage, the server employs a 30GB system disk and a 50GB solid-state drive (SSD) for data storage. All experiments are implemented using the PyTorch framework. Note that we do not employ multi-modal LLMs due to cost considerations. In our methodology, multi-modal information is converted into textual format for processing, enabling unified handling through text-based LLMs. For the *neural retrieval* component, we set the retrieval depth to top-$k = 5$. The codebase is implemented in Python 3.8 with PyTorch 1.12.0, LangChain 0.1.0, and FAISS for vector indexing. More details can be found in the code.

B. Main Experiment Results (RQ1 & RQ2)

*1) **Comparison with All Configurations**:* As shown in Table II and Table III, to address **RQ1**, we have performed a comprehensive comparison. The experiments demonstrate that EvoWildAlign significantly outperforms all benchmark configurations across all datasets, achieving up to 33.2% improvement in Hits@1. These results yield several key observations:

- **Comparison with Uni-modal Static KGA Configurations.** First, we compared methods within uni-modal static KGA configurations that utilize structural information, structural and semantic information, or structural information alone. Notably, EvoWildAlign's performance significantly surpasses all these configurations. Specifically, it improves the overall performance on the `WildMTKGA(W-I)` and `WildMTKGA(Y-I)` datasets, achieving up to 76.7% relative improvement in Hits@1 compared to the state-of-the-art baseline configuration Naive RAG. This indicates that EvoWildAlign effectively addresses the evolutionary diversity challenge by jointly modeling temporal dynamics and multi-modal information, which static uni-modal methods fail to capture.
- **Comparison with Uni-modal Temporal KGA Configurations.** Subsequently, we compared methods within uni-modal temporal KGA configurations that leverage structural information, structural and semantic information, or structural information alone. EvoWildAlign's performance likewise significantly outperforms all these configurations. Specifically, it improves the overall performance on the `WildMTKGA(W-I)` and `WildMTKGA(Y-I)` datasets, achieving up to 33.2% relative improvement in Hits@1 compared to the state-of-the-art baseline configuration AdaCoAgentEA. This demonstrates that EvoWildAlign effectively handles the dynamic imbalance challenge through adaptive fusion of diverse evolving modalities, which uni-modal temporal methods cannot adequately address.
- **Comparison with Multi-modal Static KGA Configurations.** We also compared current state-of-the-art and classical Multi-modal Static KGA Configurations. EvoWildAlign's performance significantly exceeds all these configurations. Specifically, it improves the overall performance on the `WildMTKGA(W-I)` and `WildMTKGA(Y-I)` datasets, achieving up to 48.5% relative improvement in Hits@1 compared to the state-of-the-art baseline configuration MMKG-RAG. This indicates that EvoWildAlign successfully integrates temporal evolution patterns with multi-modal characteristics, overcoming the limitations of static multi-modal methods that ignore temporal dynamics.
- **Comparing with LLM-based Configurations.** Finally, we compared seven advanced LLM-based methods (e.g., Naive RAG, ChatEA, AdaCoAgentEA, MMKG-RAG) under different configurations. We observe that our method consistently outperforms all LLM-based approaches, including

TABLE III
MORE EXPERIMENT RESULTS ON DBP15K(EN-FR), DBP-WIKI, BETA AND YAGO-WIKI50K-1K(YW1K) DATASETS.

Method	Modality					DBP15K(EN-FR)			DBP-WIKI			BETA			YAGO-WIKI50K-1K(YW1K)		
	Struc.	*Text.*	*Imag.*	*Audi.*	*Vide.*	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
Uni-modal Static KGA Methods																	
MTransE						0.247	0.577	0.360	0.281	0.520	0.363	0.033	0.108	0.060	0.037	0.122	0.067
AlignE	✓					0.481	0.824	0.599	0.566	0.827	0.655	0.229	0.424	0.296	0.552	0.704	0.605
GCN-Align	✓					0.411	0.772	0.530	0.494	0.756	0.590	0.121	0.270	0.172	0.046	0.140	0.079
Dual-AMN *(basic)*	✓					0.756	0.948	0.827	0.786	0.952	0.848	0.496	0.695	0.566	0.757	0.895	0.806
Dual-AMN *(name)*	✓	✓				0.954	0.994	0.970	0.983	0.996	0.991	0.507	0.729	0.571	0.766	0.923	0.807
BERT		✓				0.937	0.985	0.956	0.941	0.980	0.963	0.526	0.753	0.632	0.810	0.934	0.848
FuAlign	✓	✓				0.936	0.988	0.955	0.980	0.991	0.986	0.552	0.774	0.691	0.856	0.929	0.877
Naive RAG *	✓	✓				0.643	-	-	0.696	-	-	0.588	-	-	0.102	-	-
Uni-modal Temporal KGA Methods																	
TEA-GNN	✓					-	-	-	-	-	-	0.557	0.709	0.611	0.723	0.870	0.775
STEA	✓					-	-	-	-	-	-	0.506	0.637	0.553	0.887	0.965	0.916
LightTEA	✓	✓				-	-	-	-	-	-	0.615	0.727	0.654	0.966	0.988	0.975
Dual-Match	✓	✓				-	-	-	-	-	-	0.650	0.765	0.690	0.945	0.984	0.960
Simple-HHEA		✓				0.948	0.991	0.960	0.967	0.988	0.979	0.762	0.835	0.797	0.156	0.368	0.209
Simple-HHEA *(structure)*	✓	✓				0.959	0.995	0.972	0.975	0.991	0.988	0.730	0.808	0.751	0.121	0.303	0.170
ChatEA	✓	✓				0.990	**1.000**	0.995	0.995	**1.000**	**0.998**	0.798	0.883	0.851	0.657	0.893	0.737
MGTEA	✓	✓				0.432	0.762	0.551	0.552	0.783	0.626	0.686	0.795	0.724	0.947	0.982	0.960
AdaCoAgentEA *(supervised)*	✓	✓				0.986	0.995	0.992	0.987	0.993	0.990	0.819	0.893	0.875	0.901	0.964	0.924
Self-Consistency *	✓	✓				0.678	-	-	0.695	-	-	0.456	-	-	0.094	-	-
Self-RAG *	✓	✓				0.712	-	-	0.752	-	-	0.633	-	-	0.175	-	-
EvoWildAlign (Ours)	✓	✓	✓	✓	✓	**0.992**	**1.000**	**0.996**	**0.997**	0.999	**0.998**	**0.953**	**0.972**	**0.961**	**0.982**	**0.996**	**0.985**

CoT-based methods, retrieval-augmented generation methods, self-feedback methods, and multi-agent collaboration methods. This finding underscores the importance of formulating the MTKGA-Wild task as an agentic hypergraph collaboration problem to enhance overall performance through structured neuro-symbolic reasoning and adaptive multi-agent coordination.

*2) **Evaluation of Task Value and Dataset Challenge**:* Furthermore, we conducted a comprehensive analysis of the WildMTKGA(W-I) and WildMTKGA(Y-I) datasets, focusing on their challenges and practical utility to validate the value of the MTKGA-Wild task. As shown in Table II and Table III, the results indicate substantial room for improvement in overall performance, as evidenced by the Hits@1 of the state-of-the-art benchmark configurations, with Hits@1 scores of only 0.625 and 0.622 for WildMTKGA(W-I) and WildMTKGA(Y-I), respectively. These results also demonstrate that current benchmark configurations, whether uni-modal static KGA, uni-modal temporal KGA, or multi-modal static KGA methods, are unable to address the challenges of *evolutionary diversity* and *dynamic imbalance* posed by the new MTKGA-Wild task. Notably, in related research tasks such as uni-modal static KGA, uni-modal temporal KGA, and multi-modal static KGA, the performance of most models exceeds 0.9 in many cases [2], [5], [8], [20], and even approaches 100%. These findings highlight the intrinsic difficulty and value of this task, indicating that numerous unresolved challenges remain.

REFERENCES

[1] X. Jiang, C. Xu, Y. Shen, Y. Wang, F. Su, Z. Shi, F. Sun, Z. Li, J. Guo, and H. Shen, "Toward practical entity alignment method design: Insights from new highly heterogeneous knowledge graph datasets," in *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*, T. Chua, C. Ngo, R. Kumar, H. W. Lauw, and R. K. Lee, Eds. ACM, 2024, pp. 2325–2336. [Online]. Available: https://doi.org/10.1145/3589334.3645720

[2] X. Jiang, Y. Shen, Z. Shi, C. Xu, W. Li, Z. Li, J. Guo, H. Shen, and Y. Wang, "Unlocking the power of large language models for entity alignment," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, L. Ku, A. Martins, and V. Srikumar, Eds. Association for Computational Linguistics, 2024, pp. 7566–7583. [Online]. Available: https://aclanthology.org/2024.acl-long.408

[3] C. Xu, F. Su, and J. Lehmann, "Time-aware graph neural network for entity alignment between temporal knowledge graphs," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2021, pp. 8999–9010. [Online]. Available: https://doi.org/10.18653/v1/2021.emnlp-main.709

[4] Z. Sun, Q. Zhang, W. Hu, C. Wang, M. Chen, F. Akrami, and C. Li, "A benchmarking study of embedding-based entity alignment for knowledge graphs," *Proc. VLDB Endow.*, vol. 13, no. 11, pp. 2326–2340, 2020. [Online]. Available: http://www.vldb.org/pvldb/vol13/p2326-sun.pdf

[5] W. Zeng, J. Zhou, and X. Zhao, "Benchmarking challenges for temporal knowledge graph alignment," *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:273501043

[6] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor decompositions for temporal knowledge base completion," in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[7] Y. Liu, H. Li, A. García-Durán, M. Niepert, D. Oñoro-Rubio, and D. S. Rosenblum, "MMKG: multi-modal knowledge graphs," in *The Semantic Web - 16th International Conference, ESWC 2019, Portorož, Slovenia, June 2-6, 2019, Proceedings*, ser. Lecture Notes in Computer Science, P. Hitzler, M. Fernández, K. Janowicz, A. Zaveri, A. J. G. Gray, V. López, A. Haller, and K. Hammar, Eds., vol. 11503. Springer, 2019, pp. 459–474. [Online]. Available: https://doi.org/10.1007/978-3-030-21348-0_30

[8] Z. Chen, J. Chen, W. Zhang, L. Guo, Y. Fang, Y. Huang, Y. Zhang, Y. Geng, J. Z. Pan, W. Song, and H. Chen, "Meaformer: Multi-modal entity alignment transformer for meta modality hybrid," in *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023, Ottawa, ON, Canada, 29 October 2023- 3 November 2023*, A. El-Saddik, T. Mei, R. Cucchiara, M. Bertini, D. P. T. Vallejo, P. K. Atrey, and M. S. Hossain, Eds. ACM, 2023, pp. 3317–3327. [Online]. Available: https://doi.org/10.1145/3581783.3611786

[9] M. Chen, Y. Tian, M. Yang, and C. Zaniolo, "Multilingual knowledge graph embeddings for cross-lingual knowledge alignment," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Aug. 2017.

[10] Z. Sun, W. Hu, Q. Zhang, and Y. Qu, "Bootstrapping entity alignment with knowledge graph embedding," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, vol. 18. International Joint Conferences on Artificial Intelligence Organization, Jul. 2018, pp. 4396–4402. [Online]. Available: https://doi.org/10.24963/ijcai.2018/611

[11] Z. Wang, Q. Lv, X. Lan, and Y. Zhang, "Cross-lingual knowledge graph alignment via graph convolutional networks," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2018, pp. 349–357. [Online]. Available: https://doi.org/10.18653/v1/d18-1032

[12] Z. Chen, Y. Wu, Y. Feng, and D. Zhao, "Integrating manifold knowledge for global entity linking with heterogeneous graphs," *Data Intelligence*, vol. 4, no. 1, pp. 20–40, 2022.

[13] X. Mao, W. Wang, Y. Wu, and M. Lan, "Boosting the speed of entity alignment 10 ×: Dual attention matching network with normalized hard sample mining," in *Proceedings of the Web Conference 2021*. ACM, Apr. 2021, pp. 821–832. [Online]. Available: https://doi.org/10.1145/3442381.3449897

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[15] C. Wang, Z. Huang, Y. Wan, J. Wei, J. Zhao, and P. Wang, "FuAlign: Cross-lingual entity alignment via multi-view representation learning of fused knowledge graphs," *Inform. Fusion*, vol. 89, pp. 41–52, Jan. 2023. [Online]. Available: https://doi.org/10.1016/j.inffus.2022.08.002

[16] X. Tang, J. Zhang, B. Chen, Y. Yang, H. Chen, and C. Li, "BERT-INT: A bert-based interaction model for knowledge graph alignment," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, C. Bessiere, Ed. ijcai.org, 2020, pp. 3174–3180. [Online]. Available: https://doi.org/10.24963/ijcai.2020/439

[17] F. M. Suchanek, S. Abiteboul, and P. Senellart, "Paris: Probabilistic alignment of relations, instances, and schema," *Proceedings of the VLDB Endowment*, vol. 5, no. 3, 2011.

[18] M. Leone, S. Huber, A. Arora, A. García-Durán, and R. West, "A critical re-evaluation of neural methods for entity alignment," *Proc. VLDB Endow.*, vol. 15, no. 8, pp. 1712–1725, 2022. [Online]. Available: https://www.vldb.org/pvldb/vol15/p1712-arora.pdf

[19] G. Dong, Y. Zhu, C. Zhang, Z. Wang, J. Wen, and Z. Dou, "Understand what LLM needs: Dual preference alignment for retrieval-augmented generation," in *Proceedings of the ACM on Web Conference 2025, WWW 2025, Sydney, NSW, Australia, 28 April 2025- 2 May 2025*, G. Long, M. Blumestein, Y. Chang, L. Lewin-Eytan, Z. H. Huang, and E. Yom-Tov, Eds. ACM, 2025, pp. 4206–4225. [Online]. Available: https://doi.org/10.1145/3696410.3714717

[20] R. Zhao, W. Zeng, J. Tang, Y. Li, G. Ye, J. Du, and X. Zhao, "Towards unsupervised entity alignment for highly heterogeneous knowledge graphs," in *41st IEEE International Conference on Data Engineering, ICDE 2025, Hong Kong, May 19-23, 2025*. IEEE, 2025, pp. 3792–3806. [Online]. Available: https://doi.org/10.1109/ICDE65448.2025.00283

[21] R. Zhao, W. Zeng, W. Zhang, X. Zhao, J. Tang, and L. Chen, "Towards temporal knowledge graph alignment in the wild," *CoRR*, vol. abs/2507.14475, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2507.14475

[22] C. Xu, F. Su, B. Xiong, and J. Lehmann, "Time-aware entity alignment using temporal relational attention," in *Proceedings of the ACM Web Conference 2022*. ACM, Apr. 2022, pp. 788–797. [Online]. Available: https://doi.org/10.1145/3485447.3511922

[23] L. Cai, X. Mao, M. Ma, H. Yuan, J. Zhu, and M. Lan, "A simple temporal information matching mechanism for entity alignment between temporal knowledge graphs," in *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*. International Committee on Computational Linguistics, 2022, pp. 2075–2086. [Online]. Available: https://aclanthology.org/2022.coling-1.181

[24] L. Cai, X. Mao, Y. Xiao, C. Wu, and M. Lan, "An effective and efficient time-aware entity alignment framework via two-aspect three-view label propagation," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China*. ijcai.org, 2023, pp. 5021–5029.

[25] X. Liu, J. Wu, T. Li, L. Chen, and Y. Gao, "Unsupervised entity alignment for temporal knowledge graphs," in *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, Y. Ding, J. Tang, J. F. Sequeda, L. Aroyo, C. Castillo, and G. Houben, Eds. ACM, 2023, pp. 2528–2538. [Online]. Available: https://doi.org/10.1145/3543507.3583381

[26] F. Liu, M. Chen, D. Roth, and N. Collier, "Visual pivoting for (unsupervised) entity alignment," in *Proceedings of AAAI*, 2021, pp. 4257–4266.

[27] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., 2022. [Online]. Available: http://papers.nips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html