

Hierarchical Knowledge Graph Construction from Images for Scalable E-Commerce

Zhantao Yang
Carnegie Mellon University
Pittsburgh, USA
zhantaoy@andrew.cmu.edu

Han Zhang
Carnegie Mellon University
Pittsburgh, USA
hanz3@andrew.cmu.edu

Fangyi Chen
Carnegie Mellon University
Pittsburgh, USA
fangyic@andrew.cmu.edu

Anudeepsekhar Bolimera
Carnegie Mellon University
Pittsburgh, USA
abolimer@andrew.cmu.edu

Marios Savvides
Carnegie Mellon University
Pittsburgh, USA
marios@andrew.cmu.edu

Abstract

Knowledge Graph (KG) is playing an increasingly important role in various AI systems. For e-commerce, an efficient and low-cost automated knowledge graph construction method is the foundation of enabling various successful downstream applications. In this paper, we propose a novel method for constructing structured product knowledge graphs from raw product images. The method cooperatively leverages recent advances in the vision-language model (VLM) and large language model (LLM), fully automating the process and allowing timely graph updates. We also present a human-annotated e-commerce product dataset for benchmarking product property extraction in knowledge graph construction. Our method outperforms our baseline in all metrics and evaluated properties, demonstrating its effectiveness and bright usage potential.

CCS Concepts

• **Information systems** → *Data mining*.

Keywords

Knowledge Graph, Large Language Model, Multimodal Large Language Model, E-Commerce

ACM Reference Format:

Zhantao Yang, Han Zhang, Fangyi Chen, Anudeepsekhar Bolimera, and Marios Savvides. 2024. Hierarchical Knowledge Graph Construction from Images for Scalable E-Commerce. In *Proceedings of the first workshop on Generative AI for E-Commerce 2024, October 25, 2024*. ACM, New York, NY, USA, 6 pages.

1 Introduction

Knowledge graphs (KG), directed graphs representing information and relationships between entities, are commonly used for efficient information processing. In the domain of e-commerce, KGs play a crucial role in scaling up both inventory management and customer service by leveraging various applications [11, 15], including recommendation systems [9, 20, 21, 23, 25], question answering

service [12, 22, 41], information and intention discovery [33], and knowledge completion [29].

Recent studies [7, 24, 34, 40] show improvements extracting information from documents and texts for knowledge graph construction. However, in practice, due to the rapid changes in the fields of e-commerce, informative text descriptions of products are often expensive and time-consuming to acquire through human labeling. In contrast, raw images of products [16, 31] are widely available yet under-explored as sources of automated knowledge graph construction.

In this work, we explore how to establish an automatic process that directly uses product images as the primary sources to construct complex knowledge graphs. Without human-in-the-loop, the process of populating knowledge graphs particularly benefits the fast-paced e-commerce sector, where product catalogs are constantly evolving and expanding, so that timely update is achieved in such an environment. Moreover, product images contain essential information that is language-agnostic, semantic-rich, and involves subtle visual cues, ensuring accurate product representations toward multilingual and multicultural e-commerce platforms.

Despite the advantages, establishing an automatic process for KG with product images is a non-trivial work and faces many challenges. **Firstly**, unlike documents and texts which directly include the entities, properties, and relationships, product images are complex and may contain distractions. Extracting useful information thus requires sophisticated image understanding. **Secondly**, not all relevant information for KG construction is directly visible or explicitly stated in the product image itself. For example, categorizing a chocolate image into candy requires the ability to reason based on common knowledge and contextual understanding. **Thirdly**, unconstrained triple generation may not fully capture the hierarchical nature of product properties in e-commerce. For instance, the category property can have a hierarchical relationship, "chocolate" falls under "candy," which falls under "food". Many previous works construct KGs by either generating [17, 24, 40] or completing [3, 4] triples without additional constraints, which could result in graphs lacking depth and diversity if directly applied to e-commerce products.

To address these limitations, we propose a novel method that is equipped with the recent advances in vision-language models (VLMs) and large language models (LLMs), enabling hierarchical

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GenaiEcom '24, October 25, 2024, Boise, ID

© 2024 Copyright held by the owner/author(s).

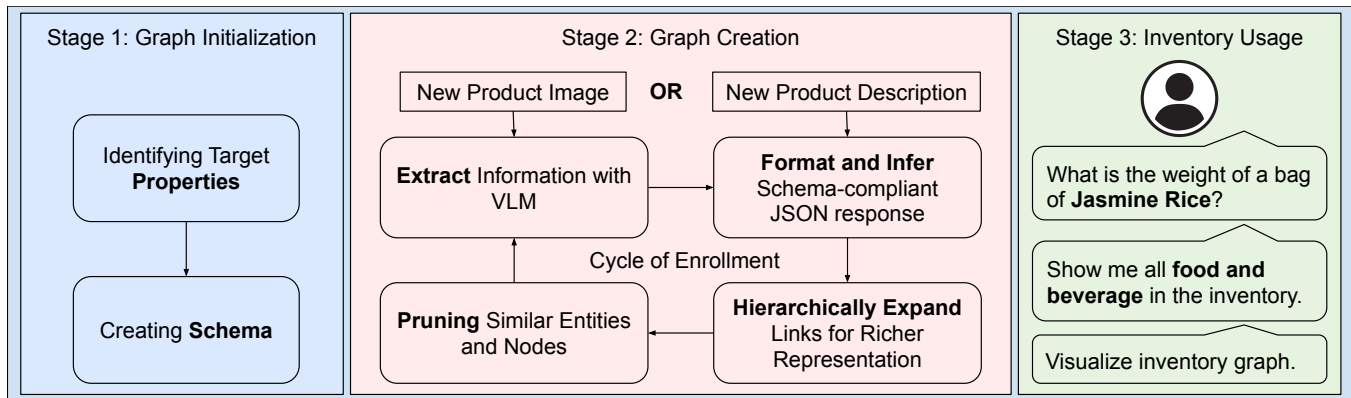


Figure 1: Method Overview. Stage 1: An empty graph is first initialized with target properties and corresponding data types. Stage2: for each product, information is extracted with VLMs, organized and improved by LLM.

knowledge graph generation given any number of product images. The graph follows a schema containing both properties and their corresponding data types. We start with instruction-tuned vision language model (InternVL2-8B [5, 6]) to extract detailed information from product images. Then, we use schema-augmented multi-turn conversation to ensure the description contains more diverse and detailed attributes and relationships. We further use the newest state-of-the-art large language model, Llama3.1-70B [8], to reason and infer KG relevant properties not found in the image, and hierarchically expand the existing links. In the process, SGLang [38] is selected to generate all LLM responses in strictly reliable structured formats, ensuring the output graph follows the schema. Finally, we design programmatical merge to reduce redundancy among similar entities.

Besides the proposed method, we introduce a human-annotated e-commerce product image dataset. It consists of 105 images, each containing a product image, 4 categorical properties, 1 numerical property, and JSON structured metadata. We benchmark our method on the dataset and release the dataset to the community for further research.

Our contribution is in three folds:

- To the best of our knowledge, we are the first to propose a novel fully automatic e-commerce method to generate knowledge graphs from only product images.
- We present a small e-commerce product image dataset for benchmarking the task.
- Our method outperforms the baseline method modified from previous works on multiple metrics.

2 Related Work

2.1 Text Extraction from Images

Image understanding has made significant strides in recent years, particularly in the domain of text extraction from images. Image captioning models aim to generate natural language descriptions of images. Image captioning models like LEMON [10] are often based on encoder-decoder architectures. The encoder projects images into latent space, and a language decoder decodes semantic information into text descriptions. However, their outputs are often general

and lack the specificity required in e-commerce applications. Image tagging models like RAM [37] take a step forward, focusing on identifying and labeling specific objects within an image, providing more fine-grained information, but still lacking the capability of identifying texts and concepts. The emergence of multimodal LLMs [32, 35], especially instruction-tuned large vision language models [13], provides a significant advancement in image understanding. These VLMs can output detailed descriptions based on user prompts, making it possible to align generated texts with desired properties. In this work, we use InternVL2-8B [5, 6], a robust open-source VLM as our image description extractor. While the model provides detailed descriptions, this information is unstructured and may not contain all the required information for KG construction.

2.2 Knowledge Graph Construction

Knowledge graph construction aims to convert less organized raw data into more programmatically processable structured graphs. Recent advancements in deep learning and natural language processing (NLP) [1, 2, 14, 27] have significantly enabled studies in information extraction and knowledge graph construction, leading to improved data management and utilization. Due to the advancements in embedding training, some studies use embeddings to represent and discover complex structures in KGs. ComplEX [26] uses complex-valued embeddings to perform link prediction at a linear time and space complexity. KoPA [36] performs triple classification task by training an LLM adapter for injecting structural embedding. However, these works are limited to individual sub-tasks of completing a KG. Consequently, many works have shifted their attention to constructing knowledge graphs from documents [24, 28, 39] and texts [7, 34, 40]. These recent approaches have leveraged advancements in transformer [27] architecture based large language models (LLMs) trained on internet-scale datasets [18, 19] to enhance or construct knowledge graphs. For example, TKGCon [7] uses GPT-4 [1] to generate theme-related entities and relations from a theme-specific corpus to form KGs. However, few works are leveraging raw images for KG construction. In this work, we use Llama3.1 [8], in collaboration with a VLM to generate high-quality and diverse knowledge graphs for e-commerce product inventory.

3 Constrained Hierarchical Knowledge Graph Generation

In this section, we will show an overview of our Knowledge Graph construction method.

3.1 Method Overview

The knowledge graph construction can be roughly divided into two core stages. An abstract visualization of our method can be found in Figure 1.

- **Graph Initialization.** When creating a new knowledge graph for an inventory or an e-commerce system, two components are used for initializing and preparing the knowledge graph: **identifying target properties** and **creating schema**. An empty inventory knowledge graph will be created once initialization is finished.
- **Cycle of Enrollment.** An e-commerce inventory can grow continuously as more products are added. Therefore, our proposed method treats each product as the fundamental unit of graph creation. Our method cycles through four sequential steps for each product. A product-centric knowledge graph will be generated with four steps: **Extracting, Formatting and Inferring, Hierarchy Expansion, and Graph Pruning**. Each product knowledge graph will be added to the previous inventory KG.
- **Inventory Usage** Once the knowledge graph is initialized, it can be loaded into a graph database and used in various downstream applications.

4 Introduction

4.1 Graph Initialization

Identifying Target Properties. The first step in the initialization phase is to identify and select the properties that will serve as the foundation of the knowledge graph. Not all entities and relationships are essential when utilizing a knowledge graph for e-commerce. The process of identifying target properties aims to determine the most relevant and valuable attributes for products in the inventory. This can be done automatically, manually, or semi-automatically. In automatic mode, the pipeline prompts an instruction-tuned LLM [8, 14] to list the most important properties when describing an e-commerce product. Alternatively, a person can designate the properties that are important for their system.

Creating Schema. While the properties to be generated have been defined, it is important to standardize the structure and format of different products. The schema defines the data types of each property, which serves as a blueprint for each new product subgraph. By enforcing rules and constraints, the schema helps reduce the introduction of redundant or conflicting data when new products are added, which provides better scalability and downstream task efficiency. Enforcing data types also acts as a fail-safe, preventing LLMs from generating invalid information. In this work, we use a prompted autoregressive LLM to find the data type t that maximizes the predicted probability for each given property x :

$$t' = \operatorname{argmax}_{t' \in \{int, float, str, choices\}} P(t'|x) \quad (1)$$

If the data type of a property is identified as int or float, a unit of measurement is similarly predicted with an autoregressive LLM. The model predicts the next token following the prompt "{property} of a product could be 5 ". If the data type is identified as choices, LLM is prompted to generate diverse distinct choices that can generalize to most products, with an additional "Other" choice added.

Following the above procedure, a complete schema can be created. A product subgraph takes the product name as the root node, all edges point to properties starting from the product root node. By default, we use the schema generated fully automatically:

- **Product Name:** *string*
- **Category:** *choices* [Electronics, Fashion, Home and Kitchen, Beauty and Personal Care, Food and Beverages, Sports and Outdoors, Baby and Kids Products, Health and Wellness, Automotive, Arts and Crafts, Pet Products, Office and School Supplies, Industrial and Scientific, Musical Instruments, Toys and Games, Others]
- **Brand:** *string*
- **Price:** *float* (USD)
- **Primary Package Color:** *choices* [White, Black, Gray, Beige, Brown, Tan, Green, Red, Blue, Yellow, Light Blue, Pink, Baby Blue, Mint Green, Silver, Gold, Copper, Purple, Orange, Turquoise, Others]
- **Package Material:** *choices* [Plastic, Paper, Cardboard, Glass, Metal, Wood, Fabric, Foam, Bamboo, Bioplastic, Molded Pulp, Corrugated, Others]
- **Package Shape:** *choices* [Rectangular, Cylindrical, Spherical, Oval, Triangular, Irregular, Flat, Tubular, Conical, Geometric, Others]
- **Weight:** *float* (kg)

4.2 Cycle of Enrollment

With the graph initialized and the schema in place, individual products can be processed and added to the knowledge graph iteratively. The Cycle of Enrollment consists of four key steps: **Extracting, Formatting and Inferring, Hierarchy Expansion, and Graph Pruning**. Additionally, while we primarily focus on studying KG construction from product images, our method inherently supports textual product description as input by skipping the **Extract** phase. Each phase addresses specific challenges in constructing a reliable and hierarchical product knowledge graph for e-commerce.

Extracting product descriptions from the raw image is the first step of enrolling an image. To tackle the challenge of extracting rich information from images, we employed a recent state-of-the-art open-source vision language model, InternVL2 [5, 6]. We first convert the generated schema into text descriptions to augment the original prompt. With schema description embedded in the user prompt, the VLM can generate unstructured or semi-structured text descriptions based on the input product image, ensuring that the generated descriptions cover all relevant product attributes. To maximize information coverage, we employ a multi-turn extraction process, where we prompt VLM to provide additional details in a second turn of the conversation. This will cover more visual cues compared to single-turn extraction to enable more accurate missing information inference in future steps. More turns are possible yet not employed for better speed and scalability.

Table 1: Comparison of our method against the baseline on various properties. Accuracy is reported for categorical properties, while accuracy@threshold is used for numerical properties. All results shown in percentage (%)

Method	Primary Package Color	Package Shape	Package Material	Category	Weight (Acc@0.01)	Weight (Acc@0.05)
Baseline (zero-shot)	26.67	3.81	19.05	0.00	9.78	9.78
Baseline w/ schema	55.24	49.52	54.29	62.86	13.04	16.30
ours w/o reasoning	81.9	76.19	81.9	95.24	55.43	63.04
ours w/o multi-turn	73.33	75.24	79.05	89.52	54.35	69.57
ours	82.86	77.14	86.67	97.14	61.96	73.91

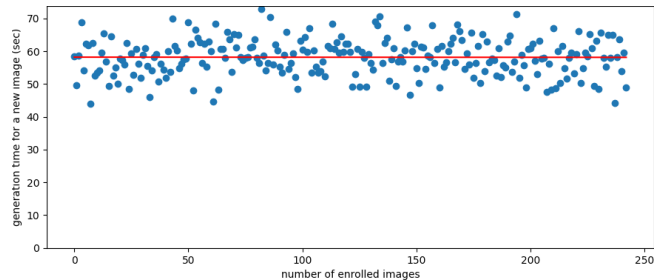
Formatting and Inferring generated description comes after information extraction. While visible information is already included in the text description, the information is not directly usable. Similar to the extracting phase, we use multi-turn conversation to align the response with the predefined schema. Since some information may not be available from the image, we use Llama3.1-70B [8] to first analyze all the extracted text descriptions, encouraging intermediate reasoning steps [30]. Then based on the reasoning, we prompt the model to infer the remaining properties. After the generation in the first turn, we use SGLang [38] for regular expression constrained generation. The output is forced to be generated in JSON format, strictly following the data type and schema structure. This guarantees that the response will always be generated reliably containing all requested properties, and additionally ensures the response can be parsed programmatically.

Hierarchical Expansion attempts to introduce additional entities between the product node and the abstract category node. This phase is crucial for enhancing the knowledge graph's structure and utility in e-commerce applications. An LLM is prompted to analyze and generate an intermediate entity between a category property and the product name. This expansion is repeated several times so that multiple intermediate entities are inserted. For example, the category link starts with "Dark Chocolate Bar → Food and Beverage", an intermediate nodes "Chocolate" and "Dark Chocolate" are sequentially added during the expansion. The resulting link becomes "Dark Chocolate Bar → Dark Chocolate → Chocolate → Food and Beverage". To further improve the diversity of the graph, multiple hierarchical expansions are performed in parallel, each independently choosing intermediate nodes from the predicted tokens with top-k sampling. By introducing multiple levels of abstraction, the system can represent products at various levels of breadth with more fine-grain relationships.

Pruning is the final step of enrolling a product. When properties are created with LLM free-form generation, there can be entities sharing exactly the same or similar meaning, these can be merged into one node. In our method, we applied a simple yet effective method that merges all properties that share the same words in different order or letter cases. This reduces the complexity of the final knowledge graph so that downstream tasks can be performed more efficiently.

This Cycle of Enrollment is executed for each product, allowing for the incremental growth of the knowledge graph and adapting to the rapid changes in the e-commerce domain. Because each product subgraph is generated independent of the size of the existing inventory, as shown in Figure 2. The linear scaling allows our approach

to be scaled to a large inventory size without increasing the cost of generating the graph for each image.

**Figure 2: Time taken to generate KG for an image remains similar as the number of images in the inventory increases.**

By combining advanced VLMs and LLMs with constrained generation, our approach can handle diverse product types and extract rich, consistent structured information from visual data, reliably following the schema. The resulting hierarchical knowledge graph provides a foundation for sophisticated e-commerce applications, including enhanced search capabilities, personalized recommendations, and advanced product analytics.

4.3 Inventory Usage

In the final stage after the enrollment, one can easily leverage the KG to perform user-defined tasks. Apart from the classic product description retrieval and recommendation systems, our method enables intelligent attribute inquiry like packaging materials and weight estimation, offering users a diverse range of applications.

5 Experiment

In this section, we first gather a product image dataset. Then, we use our method to generate a knowledge graph following the method described in the previous section. We compare the ground truth annotation with generation results and evaluate the effectiveness of our method based on several metrics. Unless otherwise specified, we use InternVL2-8B in bfloat16 and Llama3.1-70B in int4. The experiments are conducted on 6 RTX 4090 GPUs.

5.1 Dataset Collection

We collected 120 images and their corresponding metadata using BlueCart Walmart Data Product API¹. The result contains information such as image, product name, and other information displayed

¹<https://www.bluecartapi.com/>

on the Walmart website². Among these, 105 images are valid, we then manually labeled the properties Category, Primary Package Color, Package Material, Package Shape, and Weight based on our generated schema. We perform a series of experiments using only the images resized to 448x448 pixels as inputs.

5.2 Results

In this part, we will show our performance on the collected dataset compared to a baseline method. We set our baseline by modifying the zero-shot KG construction method proposed in AutoKG [40] to generate triples from an image using the product name as the subject. Following AutoKG, we use property names as the list of predicates to prompt InternVL2, and the objects of the generated triples are treated as the predicted properties. In addition to the zero-shot baseline, we add our schema-augmented prompt to the start of the baseline prompt, providing additional context for the expected response. Then, we perform an ablation study on our method. First, we evaluate the performance of our complete method. We then conduct ablation studies by removing the multi-turn conversation during VLM information **Extraction** or excluding the intermediate LLM reasoning step during the **Format and Infer** stage, to assess the impact of these components on the overall results. We evaluated the following predictions against the annotation. **Primary Package Color and Package Shape** are categorical properties that can be directly observed from the image with little reasoning and inference. **Package Material** is a categorical property that can be directly observed from the image, but requires some prior knowledge (e.g., material texture) to infer. **Category** is a categorical property that cannot be directly observed from the image, and requires inference with prior knowledge and contextual reasoning on information like brand name. **Weight** is a numerical property that can be found on most of the product images, but due to non-standardized units of measurement across products, calculations are needed to convert to the unit in schema (kg).

For categorical properties, accuracy is used as the metric. For numerical property, we first compute the error ratio e between predicted value v_{pred} and annotated value v_{gt} by

$$e = \frac{|v_{pred} - v_{gt}|}{v_{gt}} \quad (2)$$

Then we use accuracy@threshold as our metrics, where a prediction is considered correct if the error ratio e is strictly lower than the threshold. We report accuracy@0.01 and accuracy@0.05 for the numerical property. Table 1 shows our primary result against baseline and ablation studies. We also show a subgraph containing 3 enrolled products constructed our complete method in Figure 3.

5.3 Analysis

As shown in Table 1, our method exceeds directly prompting VLM for triple generation. By augmenting the baseline with our schema description, predictions for all categorical properties gain large improvements. We notice that this is mainly because when no schema descriptions are embedded in the prompt, VLM tends to give predictions that are not in the choices. Adding schema description provides contextual information for VLM to rectify its answers.

²<https://www.walmart.com/>

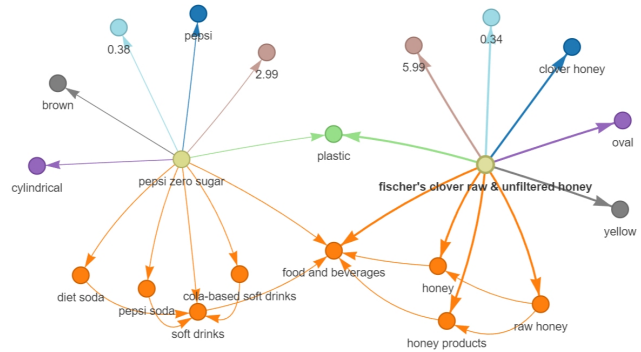


Figure 3: Example KG subgraph of 3 enrolled products.

By removing LLM reasoning from our method, performance on Weight prediction and Package Material drops significantly. Accuracy@0.05 for Weight dropped over 10%. This shows that reasoning is important for analyzing more ambiguous properties that require contextual understanding. With weight information directly shown in most images, our method without reasoning fails to standardize units more frequently than the complete method. The LLM tends to directly provide weight in its original units of measurement, even though it is prompted to respond in kilograms. Our result shows that the reasoning step leverages the LLM’s ability to incorporate external knowledge and perform context-sensitive analysis, which is crucial for property inference.

Furthermore, the drop in performance when removing the multi-turn conversation during VLM information extraction highlights the importance of including diverse additional visual information in image descriptions. The multi-turn process allows the model to progressively add visual cues and details into the descriptions, which could be beneficial for subsequent steps to analyze and dynamically adjust predicted properties based on the context.

Even without reasoning or multi-turn conversation, our method still outperforms the baseline by a large margin, showing the robustness of our method when constructing links from image data.

6 Limitations

While our work shows promising results on various metrics using high-quality images, additional work may be required for low-resolution images.

7 Conclusion

In this paper, we propose a novel method that fully automatically generates a knowledge graph from scratch using only image data. We propose several collaborative components to analyze and infer schema-compliant properties from each product image. We propose a benchmark for knowledge graph generation from images, with emphasis on the correctness of generated properties. We compare our method against an adaptation of a previous work [40], and perform ablation studies, showing the effectiveness of our method and several key features.

References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altmenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- [2] Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint ArXiv:2005.14165* (2020).
- [3] Jiao Chen, Luyi Ma, Xiaohan Li, Nikhil Thakurdesai, Jianpeng Xu, Jason HD Cho, Kaushiki Nag, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. 2023. Knowledge graph completion models are few-shot learners: An empirical study of relation labeling in e-commerce with llms. *arXiv preprint arXiv:2305.09858* (2023).
- [4] Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022. Knowprompt: Knowledge-aware prompting with synergistic optimization for relation extraction. In *Proceedings of the ACM Web conference 2022*. 2778–2788.
- [5] Zhe Chen, Weiyun Wang, Hao Tian, Shenglong Ye, Zhangwei Gao, Erfei Cui, Wenwen Tong, Kongzhi Hu, Jiapeng Luo, Zheng Ma, et al. 2024. How far are we to gpt-4v? closing the gap to commercial multimodal models with open-source suites. *arXiv preprint arXiv:2404.16821* (2024).
- [6] Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. 2024. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 24185–24198.
- [7] Linyi Ding, Sizhe Zhou, Jinfeng Xiao, and Jiawei Han. 2024. Automated Construction of Theme-specific Knowledge Graphs. *arXiv preprint arXiv:2404.19146* (2024).
- [8] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783* (2024).
- [9] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2020. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering* 34, 8 (2020), 3549–3568.
- [10] Xiaowei Hu, Zhe Gan, Jianfeng Wang, Zhengyuan Yang, Zicheng Liu, Yumao Lu, and Lijuan Wang. 2022. Scaling up vision-language pre-training for image captioning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 17980–17989.
- [11] Mayank Kejriwal. 2022. Knowledge graphs: A practical review of the research landscape. *Information* 13, 4 (2022), 161.
- [12] Feng-Lin Li, Weijia Chen, Qi Huang, and Yikun Guo. 2019. Alime kbqa: Question answering over structured knowledge for e-commerce customer service. In *Knowledge Graph and Semantic Computing: Knowledge Computing and Language Understanding: 4th China Conference, CCKS 2019, Hangzhou, China, August 24–27, 2019, Revised Selected Papers 4*. Springer, 136–148.
- [13] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2024. Visual instruction tuning. *Advances in neural information processing systems* 36 (2024).
- [14] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* 35 (2022), 27730–27744.
- [15] Ciyuan Peng, Feng Xia, Mehdi Naseriparsa, and Francesco Osborne. 2023. Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review* 56, 11 (2023), 13071–13102.
- [16] Jingtian Peng, Chang Xiao, and Yifan Li. 2020. RP2K: A large-scale retail product dataset for fine-grained image classification. *arXiv preprint arXiv:2006.12634* (2020).
- [17] Bartosz Przytycka, Paweł Kaleta, Artur Dmowski, Jacek Piwkowski, Piotr Czarnecki, and Tomasz Cieplak. 2024. Product knowledge graphs: creating a knowledge system for customer support. (2024).
- [18] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.
- [19] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research* 21, 140 (2020), 1–67.
- [20] André Gomes Regino, Rodrigo Oliveira Caus, Victor Hochgreb, and Julio Cesar Dos Reis. 2022. Knowledge Graph-based Product Recommendations on e-Commerce Platforms. In *KDIR*. 32–42.
- [21] André Gomes Regino, Rodrigo Oliveira Caus, Victor Hochgreb, and Julio Cesar dos Reis. 2023. Leveraging Knowledge Graphs for E-commerce Product Recommendations. *SN Computer Science* 4, 5 (2023), 689.
- [22] Diogo Teles Sant’Anna, Rodrigo Oliveira Caus, Lucas dos Santos Ramos, Victor Hochgreb, and Julio Cesar dos Reis. 2020. Generating Knowledge Graphs from Unstructured Texts: Experiences in the E-commerce Field for Question Answering. In *ASLD@ ISWC*. 56–71.
- [23] Bilin Shao, Xiaojun Li, and Genqing Bian. 2021. A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. *Expert Systems with Applications* 165 (2021), 113764.
- [24] Qiang Sun, Yuanyi Luo, Wenxiao Zhang, Sirui Li, Jichunyang Li, Kai Niu, Xiangrui Kong, and Wei Liu. 2024. Docs2KG: Unified Knowledge Graph Construction from Heterogeneous Documents Assisted by Large Language Models. *arXiv preprint arXiv:2406.02962* (2024).
- [25] Zhu Sun, Qing Guo, Jie Yang, Hui Fang, Guibing Guo, Jie Zhang, and Robin Burke. 2019. Research commentary on recommendations with side information: A survey and research directions. *Electronic Commerce Research and Applications* 37 (2019), 100879.
- [26] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*. PMLR, 2071–2080.
- [27] Ashish Vaswani. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762* (2017).
- [28] Yu Wang, Nedim Lipka, Ryan A Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. 2024. Knowledge graph prompting for multi-document question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 19206–19214.
- [29] Yuqi Wang, Zeqiang Wang, Wei Wang, Qi Chen, Kaizhu Huang, Anh Nguyen, and Suparna De. 2023. Zero-Shot Medical Information Retrieval via Knowledge Graph Embedding. In *International Workshop on Internet of Things of Big Data for Healthcare*. Springer, 29–40.
- [30] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.
- [31] Xiu-Shen Wei, Quan Cui, Lei Yang, Peng Wang, and Lingqiao Liu. 2019. RPC: A large-scale retail product checkout dataset. *arXiv preprint arXiv:1901.07249* (2019).
- [32] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. 2023. A survey on multimodal large language models. *arXiv preprint arXiv:2306.13549* (2023).
- [33] Changlong Yu, Weiqi Wang, Xin Liu, Jiabin Bai, Yangqiu Song, Zheng Li, Yifan Gao, Tianyu Cao, and Bing Yin. 2022. FolkScope: Intention knowledge graph construction for E-commerce commonsense discovery. *arXiv preprint arXiv:2211.08316* (2022).
- [34] Bowen Zhang and Harold Soh. 2024. Extract, Define, Canonicalize: An LLM-based Framework for Knowledge Graph Construction. *arXiv preprint arXiv:2404.03868* (2024).
- [35] Duzhen Zhang, Yahan Yu, Chenxing Li, Jiahua Dong, Dan Su, Chenhui Chu, and Dong Yu. 2024. Mm-llms: Recent advances in multimodal large language models. *arXiv preprint arXiv:2401.13601* (2024).
- [36] Yichi Zhang, Zhuo Chen, Wen Zhang, and Huajun Chen. 2023. Making large language models perform better in knowledge graph completion. *arXiv preprint arXiv:2310.06671* (2023).
- [37] Youcai Zhang, Xinyu Huang, Jinyu Ma, Zhaoyang Li, Zhaochuan Luo, Yanchun Xie, Yuzhuo Qin, Tong Luo, Yaqian Li, Shilong Liu, et al. 2024. Recognize anything: A strong image tagging model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1724–1732.
- [38] Lianmin Zheng, Liangsheng Yin, Zhiqiang Xie, Jeff Huang, Chuyue Sun, Cody Hao Yu, Shiyi Cao, Christos Kozyrakis, Ion Stoica, Joseph E Gonzalez, et al. 2023. Efficiently programming large language models using sglang. *arXiv preprint arXiv:2312.07104* (2023).
- [39] Jie Zhou, Xin Chen, Hang Zhang, and Zhe Li. 2024. Automatic Knowledge Graph Construction for Judicial Cases. *arXiv preprint arXiv:2404.09416* (2024).
- [40] Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. 2023. Lms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *arXiv preprint arXiv:2305.13168* (2023).
- [41] Zicheng Zuo, Zhenfang Zhu, Wenqing Wu, Wenling Wang, Jiangtao Qi, and Linghui Zhong. 2023. Improving question answering over knowledge graphs with a chunked learning network. *Electronics* 12, 15 (2023), 3363.