

Response Letter

Dear Author,

We have reached a decision for your paper. Please find below the details.

Authors: M. Jaleed Khan, Filip Ilievski, John G. Breslin, Edward Curry

Title: A Survey of Neurosymbolic Visual Reasoning with Scene Graphs and Common Sense Knowledge

Submission Type: 'Regular Paper'

URL: [Click here to follow the link](#)

Tracking number: 689-1669

Assigned editor: Md Kamruzzaman Sarker (mdkamruzzamansarker@gmail.com)

Decision Letter:

Thank you for your submission to Neurosymbolic Artificial Intelligence.

This is to inform you that based on the reviewer's comments, your paper requires **major revisions**. Please carefully take into account the enclosed comments by the reviewers when preparing the revised version. It is incumbent upon you to do so. Please provide punctual responses to the issues raised by the reviewers and prepare a separate text file containing such responses. The revised version of the paper is expected within 20 days.

Please visit [Click here to follow the link](#) to see the reviews of the paper. To submit a new version of the paper, [Click here to follow the link](#) and mention the tracking number of this paper.

Best Regards,

Editors-in-chief, Neurosymbolic Artificial Intelligence

Dear Editors,

We are grateful for the opportunity to revise our paper, "A Survey of Neurosymbolic Visual Reasoning with Scene Graphs and Common Sense Knowledge," and for the insightful comments provided by the reviewers. We have thoroughly revised our manuscript, addressing each point raised in the reviews. We have prepared a detailed response to each review comment in this response letter. This revision process has significantly enhanced the quality and clarity of our paper.

We are excited to resubmit our revised paper for your further consideration. We believe that the improvements made will meet the high standards of the Neurosymbolic Artificial Intelligence Journal and contribute meaningfully to the field.

Thank you for your consideration of our work.

Sincerely,

M. Jaleed Khan, Filip Ilievski, John Breslin, Edward Curry

Review 1

Overall Impression: Good Suggested

Decision: Minor revision

Content:

Technical Quality of the paper: Good

Originality of the paper: Yes

Adequacy of the bibliography: Yes, but see detailed comments

Presentation:

Adequacy of the abstract: Yes

Introduction: background and motivation: Good

Organization of the paper: Satisfactory

Level of English: Satisfactory

Overall presentation: Excellent

Detailed Comments:

Comprehensive review combining NeSy, SGG and Commonsense knowledge. It also includes downstream tasks and a performance evaluation, pointing to state-of-the-art methods and datasets.

[Thank you for your constructive feedback on our paper.](#)

The described KGs used in the literature seem to contain mostly triples without clear formal semantics (e.g. formalised with languages like OWL). With OWL one can represent typical source/domain and target/ranges of a predicate, hierarchy of predicates, inverse predicates, etc. that could potentially enhance the reasoning with a KG (if any). See suggested references [a,b]. Which is the position of the authors in this regard?

[We agree with this observation regarding the potential benefits of using OWL for formalising semantics in KGs. We have revised our paper to include a discussion on this aspect in Sections 2.2.2 \(page 9, lines 50-51 and page 10, lines 1-3\), Section 2.3.2 \(page 11, lines 48-51\) and Section 5.3 \(page 18, lines 34-38\), referencing the suggested literature \[a, b\]. We have elaborated on how OWL's capabilities, such as defining predicate sources/targets and hierarchies and inverse predicates, can enhance reasoning in NeSy systems. We have also added the NeSy4VRD dataset \[a\] to Table 5 \(page 15, line 40\) in the revised manuscript.](#)

I believe the suggested reference [c] is also relevant to the mentioned NeSy methods. For example, [c] uses commonsense knowledge encoded in first-order logic.

[We acknowledge the relevance of \[c\] in the context of NeSy methods using common sense knowledge encoded in first-order logic. We have included it in the revised manuscript in the tightly-coupled NeSy methods in Table 2 \(page 6, line 20\) and Section 2.3.2 \(page 11, lines 41-43\).](#)

Section 5.1. I do not correctly agree with the point as it is not a problem of KGs but in particular KGs that do not include proper semantics. Here a taxonomy of types of bird who can swim and/or fly is key.

Another issue is whether this type of KGs have been applied or not in the literature for SGG, but in principle KGs (especially those using formal logic) have good potential.

Section 5.2 The NeSy4VRD dataset [a] tried to improve the original dataset with more meaningful and non ambiguous predicates. I agree that this is an important issue if more semantics are to be used.

Thank you for highlighting the need of formal semantics for improving contextual relevance of knowledge and generalizability of scene graph-based visual reasoning. We have revised Section 5.1 (page 17, lines 11-12 and 29-30) to clarify that the limitations discussed include the lack of proper semantics in KGs and to discuss the potential of formal semantics in improving SGG. We have also expanded our discussion in Section 5.2 (page 18, lines 4-6) to emphasise its role in providing more meaningful and non-ambiguous predicates.

Minor:

- i.e. e.g. -> i.e., e.g.,

We have corrected the minor issues as suggested.

- heterogenous KGs -> missing a clear definition. Does it refer to a KG integrating sub-KGs from different domains? Or a KG integrating knowledge of different modality? Or just combining multiple KGs instead of focusing on one?

We have provided a clearer definition of heterogeneous KGs in Section 1 (page 2, lines 37-40), elaborating on their composition and the types of knowledge they integrate.

Suggested literature:

Datasets: [a] NeSy4VRD: A Multifaceted Resource for Neurosymbolic AI Research using Knowledge Graphs in Visual Relationship Detection. CoRR abs/2305.13258 (2023)

Paper2: [b] On the Benefits of OWL-based Knowledge Graphs for Neural-Symbolic Systems. NeSy 2023: 327-335

[c] Scalable Theory-Driven Regularization of Scene Graph Generation Models. AAAI 2023: 6850-6859

Thank you for suggesting the related literature. We have included it in the revised manuscript.

Review 2

Overall Impression: Weak Suggested

Decision: Reject

Content:

Technical Quality of the paper: Average

Originality of the paper: Yes

Adequacy of the bibliography: Yes

Presentation:

Adequacy of the abstract: Yes

Introduction: background and motivation: Good

Organization of the paper: Needs improvement

Level of English: Satisfactory

Overall presentation: Average

Detailed Comments:

This paper presents a survey on rich scene graph visual reasoning. This is an important and interesting topic, particularly in the generative AI era. The paper focuses on two aspects: neuro-symbolic integration and commonsense knowledge infusion.

Pros:

* This is an important topic, and there is no ready survey paper yet. The survey on this topic, if carefully written, will be a positional survey paper.

* The paper is well structured, and most parts are well-written. Some parts have room for improvement in writing, which I will elaborate on as the sequel.

Thank you for the detailed review and the valuable feedback provided. We acknowledge the shortcomings and have made significant revisions to address these concerns.

Cons:

There are different shortcomings with different aspects.

*Scene graph generation part: The survey on the scene graph generation part is insufficient. This is understandable since it is a fast-moving area, and therefore, the authors are suggested to add more in-time literature to enrich this part.

We acknowledge the need to update and enrich the section on scene graph generation (Section 2). The manuscript has been revised to add more recent and related literature, including knowledge-based regularisation for SGG [41], causal inference-based methods [52-53], visual reasoning using knowledge-enriched scene graphs [12], hyperbolic methods [71-73] and the explainability aspect [2,57,60], providing a comprehensive overview of the latest advancements in this rapidly evolving field. We would also like to clarify that this survey paper focuses on SGG works that specifically leverage common sense knowledge, which we have clarified in the revised manuscript (page 2, lines 42-44).

* Some parts are written either too brief or too superficial. 2.1.4 The DQN part covers very "superficially" without a clear picture of how it connects with the visual understanding.

We have thoroughly revised the manuscript and provided further details for more clarity. We have revised the title and content of Section 2.1.4 to provide a clearer connection of DQN with visual understanding (page 7, lines 49-51) and better contextualisation within the scope of our survey. We have also broadened the section to include transformer-based methods (page 7, lines 41-49).

* Commonsense infusion part: for this part, the structure is not clear. It shifts from categorizing different types of priors to the knowledge graph part. If so, it is better to name this section the "Knowledge Graph" part. Also, this section lacks illustrative figures or summarizing tables.

Thank you for pointing out the need to better organise this section. We have reorganised the structure of Section 2.2 and renamed its subsections to better reflect its content. The works reviewed in this section are summarised in Table 2.

* Nesys part: This part is not well written. For example, Both sec 2.3.1 and 2.3.2 only have one single gigantic paragraph. Please carefully re-organize to ease reading.

We appreciate highlighting this issue. We have rewritten Sections 2.3.1 and 2.3.2 for better readability. We have restructured both sections, breaking the content into smaller, more digestible paragraphs.

* Some categories are not explicit, arguable, and lack sufficient review. For example, what is the relation between "explainable AI" and "Nesys"? This is a very relevant topic, and the survey paper is supposed to provide some review and insights on applying "explainable AI" and "Nesys" to visual reasoning. Overall, this survey could be much improved after a careful revision.

We acknowledge the importance of the explainability aspect in relation to NeSy. We have revised Section 2.3 and provided additional insights into the relationship between explainable AI and NeSy in the context of visual reasoning (page 10, lines 30-36).

Review 3

Overall Impression: Good

Suggested Decision: Major revision

Content:

Technical Quality of the paper: Average

Originality of the paper: Yes

Adequacy of the bibliography: Yes, but see detailed comments

Presentation:

Adequacy of the abstract: Yes

Introduction: background and motivation: Good

Organization of the paper: Needs improvement

Level of English: Satisfactory

Overall presentation: Average

Detailed Comments:

The paper presents an extensive survey on the works regarding visual reasoning with scene graphs and common-sense knowledge, classify them w.r.t architecture used, tasks, knowledge graphs and the loose and tight coupling, evaluation metrics . The paper is relevant since this is certainly an important gap to fill in the literature when it comes to make a survey. Moreover, I find the tight-coupling and loose-coupling classification useful.

[Thank you for your insightful and constructive feedback on our paper. We have carefully considered your comments and have accordingly made substantial revisions to the manuscript.](#)

- Language: Manuscript language is certainly satisfactory (no narrative mistakes or typos, in general), but in general style-wise quite dry. In many sections, it mentions a single sentence per citation about what it does, and moves to the next, with the monotone same structure e.g., 3.1. This could be easily fixed.

[We have proofread the manuscript, including Section 3.1, restructured and rewritten several parts of the manuscript for better readability and revised the language to make it more engaging and less monotonous. We also expanded on many of the key ideas and methods, such as the SGG task definition \(page 4, lines 33-38\), latest knowledge-based SGG methods and DNN architectures involved \(Section 2\), explainability in NeSy visual reasoning \(Section 2.3\) and motivational example \(page 2, lines 5-15\). We appreciate highlighting this important aspect.](#)

- Technical style: I believe that the paper's style could be improved if it did introduce the problem technically or half-formally under definitions or boxes. I think this is a must as it would provide substance: What is the task, "Input" , "Output". Moreover, major network architectures like RNN, GNN etc all lack the citation of the paper that introduces it. (Ideally, I would also suggest to put a figure, or input - output schema for each of them. But this one is surely optional.) Again, try to define knowledge graph instead of giving just a verbal example. When you say "For instance, a KG can provide information that "a bird is likely to be found in a tree"", the word "likely" is not natural for a knowledge graph,

triggers a statistical prior instead where you have a section for. (the word "in general" would serve better.)

We agree to the importance of better style, defining the key concepts and referencing the architectures. We have rewritten several parts of the manuscript with improved writing style and added citations for deep neural network architectures used in SGG and visual reasoning such as CNN, Faster-RCNN, RNN, LSTM, GRU, Bi-LSTM, GNN, GCN, GAT, etc. We have also defined the key concepts in the revised text, including a clearer definition of knowledge graphs and examples in the context of SGG (Section 2.2.2, page 9, lines 20-23), avoiding ambiguous terms. We have elaborated on the SGG task, its input and output, and the overall process in Section 2 (page 4, lines 33-38).

Also start the text with an example in the introduction if possible. It would really help and keep the reader.

Thanks for the useful suggestion. We have included motivational scenarios (page 2, lines 5-15) in the introduction to better engage readers and provide a clearer context for the survey.

- Missing major Literature: The survey disregards two important directions of literature completely : 1) Causality-based approaches. I think causality needs its own part under section 2.2. or next to statistical priors as a tool for common-sense reasoning or knowledge (e.g., Liu et al 2022, Cross-Modal Causal Relational Reasoning for Event-Level Visual Question Answering, or . Liu et al. Show, Deconfound and Tell: Image Captioning with Causal Inference. Zhou and Yang 2021, Relation Network and Causal Reasoning for Image Captioning . There are others for other tasks (I don't expect the survey to be fully exhaustive, of course). These are also inherently NeSy. It could in addition make challenges section interesting. 2) Hyperbolic embedding approaches (which takes either KG or taxonomy into account) relevant to common-sense reasoning e.g., Xiong et al. 2022, Hyperbolic Embedding Inference for Structured Multi-Label Prediction. Relevant to their section 2.3.2. the hierarchical semantic segmentation. Hyperbolic Image Segmentation Ghadimi et al. 2022. There is actually a great survey: by Mettes et al. 2022 "Hyperbolic Deep Learning in Computer Vision: A Survey".

Thank you for pointing out this gap and suggesting the related literature. We have included the suggested causality-based methods in Section 2.2.1 (page 8, lines 41-46) and hyperbolic methods in Section 2.3.2 (page 12, lines 1-7), referencing the suggested literature. We would also like to clarify that this survey paper focuses on methods involving both scene graphs and common sense knowledge, which we have clarified in the revised manuscript (page 2, lines 42-44).

- ML architecture classification reads exhaustive: I am not sure I would put the deep learning architecture as a exhaustive (also in figures) because there can be new architectures to be used, and this would make your survey more obsolete than it should be when time passes. There could be a subsection "Other" which could explain this fact. I leave it to authors' judgements.

We acknowledge the potential for new architectures, and have included a subsection titled "Other" (Section 2.1.4) as suggested to ensure the survey remains relevant as the field evolves. We have included DQN- and transformer-based methods in this section (page 7, lines 41-49).

Minor issues:

- It is not clear whether the performance evaluation for instance Table 4, has the authors themselves did or transferred from the papers. Should be clarified.

- lots of top K -> top-\$K\$

- MNM -> MMN

- Section 1.2 the lines 46 to 51 reads redundant: "deep learning, common sense knowledge and NeSy integration for scene representation and visual reasoning." twice.
- I would expect to see Figure 2, from left to right. (But I guess, authors want us to compare the left bottom to the right bottom.) Still something to reconsider. (optional).

We have clarified in Section 4.1 that the performance reported in Table 4 was sourced from the referenced papers (page 14, lines 38-39). We have corrected the formatting issues and typographical errors. We have also removed the redundancy in Section 1.2 for a more concise presentation. In Figure 2, we have revised the caption to state that the SGG process begins at the bottom left and concludes at the bottom right for clarity.