

# Towards Fairness in Multimodal Scene Graph Generation: Mitigating Biases in Datasets, Knowledge Sources and Models

M. Jaleed Khan<sup>1,\*</sup>, John G. Breslin<sup>1,2</sup> and Edward Curry<sup>1,2</sup>

<sup>1</sup>SFI Centre for Research Training in Artificial Intelligence, Data Science Institute, University of Galway, Ireland.

<sup>3</sup>Insight SFI Research Centre for Data Analytics, Data Science Institute, University of Galway, Ireland.

## Abstract

Fairness is a critical aspect of Artificial Intelligence (AI) techniques. The ability of multimodal AI systems for scene understanding and visual reasoning to make unbiased decisions is crucial for their acceptance and effectiveness in real-world applications. However, inherent biases in data, models, and knowledge sources often lead to unfair outcomes, thereby limiting the potential of these systems. Scene Graph Generation (SGG), a neurosymbolic multimodal method for scene understanding, is no exception to this challenge. SGG, which comprises deep learning-based multi-modal feature learning, symbolic image representation, and structured knowledge infusion, enables a wide range of visual reasoning applications. Despite its potential, various biases associated with the models, datasets, and knowledge sources used in SGG hinder the fairness and effectiveness of these techniques. This paper presents an overview, categorization, and mitigation approaches to these biases. Our aim is to contribute to the development of fairer and more robust SGG techniques, leading to more equitable applications of multimodal scene understanding and visual reasoning in AI.

## Keywords

multimodal scene understanding, scene graph, visual reasoning, bias, fairness

## 1. Introduction

Artificial Intelligence (AI) has become an integral part of various sectors of our society, from government operations to business decisions, demonstrating immense potential in improving the efficiency, accuracy, and scalability of processes. However, as we increasingly rely on AI systems for decision-making, concerns about bias, fairness, and trustworthiness in these systems have come to the forefront [2, 3]. Over 180 human biases have been identified and classified, influencing how we perceive the world and make decisions; these biases, often unconsciously, can be embedded into the AI systems we design, leading to discriminatory decisions and

---

MUWS'23: 2nd International Workshop on Multimodal Human Understanding for the Web and Social Media, October 22, 2023, Birmingham, UK


\*Corresponding author.

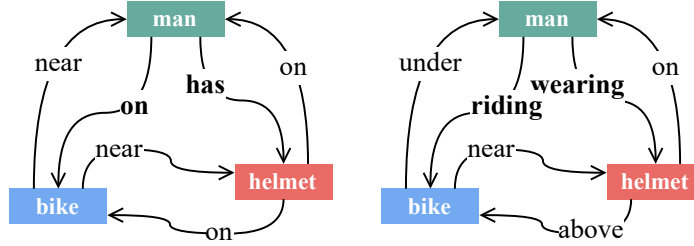
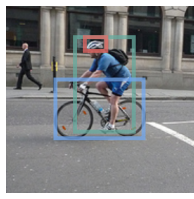
✉ m.khan12@universityofgalway.ie (M. J. Khan); john.breslin@universityofgalway.ie (J. G. Breslin); edward.curry@universityofgalway.ie (E. Curry)

🌐 <https://www.linkedin.com/in/mjaleedkhan/> (M. J. Khan); <http://www.johnbreslin.com/> (J. G. Breslin); <https://edwardcurry.org/> (E. Curry)

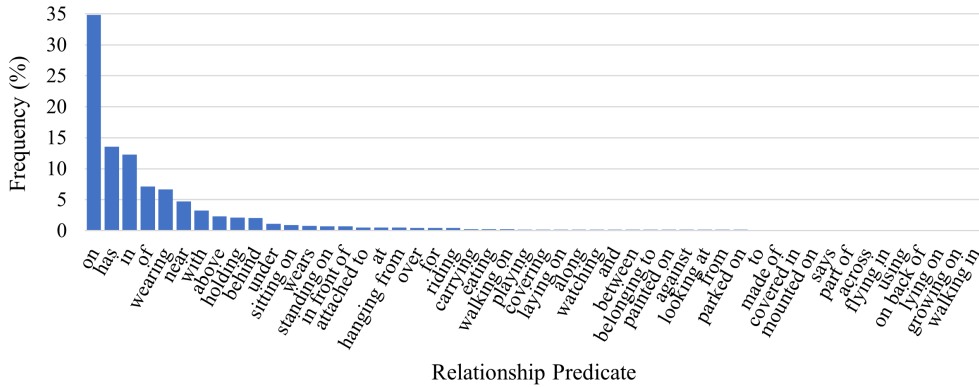
🆔 0000-0003-4727-4722 (M. J. Khan); 0000-0001-5790-050X (J. G. Breslin); 0000-0001-8236-6433 (E. Curry)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)



(a) Biased (left) and unbiased (right) visual relationship prediction in SGG.



(b) Long-tailed distribution of relationship predicates in VG dataset

**Figure 1:** (a) Biased visual relationship prediction in SGG caused by the (b) long-tailed distribution of visual relationship predicates in crowdsourced datasets [1], indicating imbalance, generalization and evaluation bias

behaviours [4]. This unfair bias is not only ethically problematic but also undermines the utility and acceptance of AI systems. In the context of multimodal scene understanding and visual reasoning, bias, fairness, and trustworthiness are particularly important due to the potential impact of these systems on our perception and understanding of the world. Scene understanding and visual reasoning are fundamental tasks in AI, enabling machines to interpret and interact with the world in a meaningful way [5]. However, biases in these systems can lead to misinterpretations and unfair representations of the world, which can have significant implications for various applications, from autonomous driving to content moderation [6, 7].

Mitigating bias in AI, particularly in multimodal scene understanding and visual reasoning, is critical to ensure the equitable treatment of all individuals and groups. By recognizing and mitigating bias in AI, we can create systems that reflect our values of fairness and justice, and in the process, we may also improve our understanding and awareness of our own biases [8]. Trustworthiness is a fundamental aspect of AI systems, particularly when these systems are used in high-stakes decision-making contexts. Trust in AI extends beyond the accuracy of the system and encompasses aspects such as fairness, transparency, and accountability [9]. Fairness, in particular, is a key pillar of trust, as it is closely intertwined with the concept of justice [10]. Bias can be introduced at any stage in the machine learning pipeline, from problem specification and data engineering to model training and evaluation [11]. Therefore, bias mitigation strategies can be applied at various stages of the pipeline, including pre-processing (applied to training data),

in-processing (applied to a model during its training), or post-processing (applied to predicted labels) [12]. However, these strategies often involve trade-offs between bias and accuracy, as reducing bias may sometimes lower model accuracy [13]. As we continue to develop and deploy multimodal AI systems for scene understanding and visual reasoning, it is crucial that we strive to mitigate bias, promote fairness, and ensure the trustworthiness of these systems. The significance of this endeavor extends beyond the technical realm, as it touches upon our societal values, ethical principles, and the kind of world we want to create with AI [14, 7, 6].

Scene graphs, as symbolic image representations, effectively capture the semantics of visual scenes by modelling objects and their relationships in a structured, semantically grounded manner. Scene Graph Generation (SGG) involves detecting objects, attributes, and relationships in visual scenes and constructing symbolic representations for higher-level downstream visual reasoning. [15]. Bias is introduced in deep learning and computer vision models at various stages of model development and in different ways, which significantly impacts the performance and fairness of these models [16]. The bias in models, datasets, and knowledge sources significantly impacts the robustness of SGG and limits downstream reasoning performance. While efforts have been made to mitigate dataset-related biases, a comprehensive analysis and investigation of biases in SGG are needed to encourage bias mitigation approaches and promote fairer, more robust SGG techniques. Since scene graphs are widely used in several downstream reasoning tasks, fairer SGG will significantly impact scene understanding and visual reasoning applications. This paper presents a detailed overview, categorization and mitigation approaches of biases associated with the models, datasets and knowledge sources used in SGG.

## 2. Scene Graph Generation

Early SGG methods focused on multimodal vision-language feature extraction, while current techniques also utilize common sense knowledge from statistical and language priors and KGs for complementary features [17, 15].

### 2.1. Models

Deep learning models are extensively used in SGG methods. Convolutional Neural Networks (CNN) effectively extract visual features for object detection and pairwise relationship detection [18, 19, 20] but struggle with long-range dependencies and complex relationships. Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks model object dependencies and context by maintaining hidden states to capture long-range dependencies [21, 22, 23, 24, 25] but face challenges with parallelization and scalability. Graph Neural Networks (GNN) excel in SGG due to their graph modelling capabilities. Representing objects as nodes and relationships as edges, GNNs use message passing to iteratively update both, capturing local and global contexts [26, 27]. Attention modules help identify salient regions for prediction and are leveraged in transformers to process object features in parallel, effectively capturing long-range dependencies and complex relationships [19]. Transformers model both local and global contexts by combining multi-head self-attention layers and position-wise feed-forward layers.

## 2.2. Knowledge Infusion

SGG methods leverage common sense knowledge in various forms, such as statistical priors [27, 21, 28], language priors [18, 20], and KGs [22, 19, 23, 26, 25, 24]. Statistical priors model correlations between object pairs and relationships, while language priors use semantic relationships of words. KGs, such as ConceptNet [29], WordNet [30], Wikidata [31] and CSKG [32], are used to extract explicit semantics and common sense knowledge about visual concepts and embedded into the models to improve the performance, expressiveness and interoperability of SGG.

## 2.3. Datasets

The common datasets used to train and evaluate SGG models include Visual Genome (VG) [1], Visual Relationship Detection (VRD) [18], MS COCO [33] and GQA [34]. VG contains over 100,000 images with rich annotations, including 3.3 million objects, 2.8 million attributes, and 5.4 million relationships. VRD contains 5,000 images, 100 object categories, and 70 predicate categories. COCO features 330,000 images with 2.5 million labelled instances, object annotations, segmentation masks, and captions. GQA dataset is a balanced, large-scale dataset comprising over 113,000 images and 22 million questions with functional and spatial relationships.

## 3. Bias in SGG

The following types of biases potentially exist in SGG models, datasets, and knowledge sources, which can limit the performance and fairness of multimodal methods for scene understanding and visual reasoning.

### 3.1. Sociocultural bias

Sociocultural bias refers to the inability of an SGG model to fairly represent and predict visual relationships across diverse populations and contexts. This bias often perpetuates stereotypes or reinforces existing societal biases associated with cultures, languages, genders, and demographics. For example, if a dataset predominantly contains images of male individuals performing a specific action, an SGG model trained on this dataset might perpetuate gender stereotypes by disproportionately associating that action with male individuals [14]. Similarly, if a KG contains biased word associations, such as "nurse" being predominantly linked to "female", the model might perpetuate gender stereotypes by disproportionately associating the "nurse" label with female individuals [35]. Moreover, models themselves can also exhibit sociocultural bias. For instance, if a model has been trained on a dataset with a long-tailed distribution, it might be biased towards predicting more frequent relationships, which could lead to a lack of diversity in the generated scene graphs [6]. This is an example of sociocultural bias in the model used for SGG.

### 3.2. Imbalance bias

Imbalance bias refers to the uneven distribution of object classes and visual relationships in the datasets used for training. This uneven distribution leads to an under-representation of certain relationships or object compositions, which in turn biases SGG models towards frequent classes [36]. Imbalance bias often exhibits a long-tailed distribution of objects, visual relationships, and attributes. This means that a few classes of objects or relationships are very common in the dataset, while many others are rare. This imbalance in the dataset is then reflected in the statistical priors used by the SGG models. For instance, if a dataset contains a high frequency of co-occurrences between the object "person" and the relationship "holding", but only a few instances of "person" and "eating", the model will be biased towards predicting the "holding" relationship, even in contexts where "eating" would be more appropriate [37]. Imbalance bias can be further exacerbated by reporting bias, which refers to the fact that some labels are more likely to be missing from the training data than others. This can lead to an underestimation of the frequency of certain classes, further skewing the distribution of the data [38]. The imbalance bias can also affect the knowledge sources used in SGG. For example, common sense knowledge bases, which are often used to provide additional context for the relationships between objects, may also exhibit a bias towards more common or generic relationships, at the expense of more specific or nuanced ones [39].

### 3.3. Domain bias

Domain bias arises due to the limitations in the scope, diversity, and coverage of the datasets and knowledge sources used in SGG. Domain bias can significantly impact the performance of SGG models when they encounter new domains, situations, or concepts that are not well-represented in the training data. As a result, models may underperform or fail when faced with unfamiliar scenarios. For instance, if a dataset predominantly contains indoor scenes, the model trained on this dataset may not perform well on outdoor scenes due to the lack of exposure to such environments during training. This is a clear example of domain bias, where the model's performance is constrained by the specific characteristics of the training data [25]. Domain bias can also be present in the knowledge sources used for SGG. For example, if a lexical knowledge graph (KG) contains detailed animal taxonomies but lacks comprehensive plant coverage, the model might struggle to recognize and analyze relationships involving plants. This is because the model's understanding of the world is heavily influenced by the knowledge it has been provided, and any gaps or imbalances in this knowledge can lead to biased predictions [40]. Moreover, domain bias can also manifest in the form of an imbalance in the representation of different types of relationships in the training data. For example, if a dataset contains a disproportionate number of certain types of relationships, the model may become biased towards predicting these relationships, even when they may not be the most appropriate or accurate in a given context [6]. In addition, domain bias can also be introduced through the process of data collection. For instance, if the data collection process is biased towards certain types of scenes or objects, this can lead to a skewed representation of the world in the dataset, which in turn can lead to biased predictions by the model [41, 42].

### 3.4. Contextual bias

Contextual bias arises from ambiguous or unclear relationships or situations in the data. Contextual bias can affect the ability of a model to understand the context and predict accurate relationships accordingly, leading to misinterpretations and errors in predictions. For instance, language priors biased towards certain word combinations can cause the model to incorrectly predict relationships based on co-occurrence frequencies rather than the actual visual context [43]. In the process of generating a scene graph, a model can incorrectly associate certain objects or relationships based on the frequency of their occurrence in the training data, rather than their actual presence or relationship in the image. This is particularly problematic in cases where the training data has a long-tailed distribution, with certain objects or relationships being significantly more common than others. This can lead to the model overemphasizing these common relationships and underrepresenting or misinterpreting less common ones [6]. Contextual bias can also be present in the datasets used for training SGG models. For instance, if a dataset primarily contains images of certain types of scenes or objects, the model may struggle to accurately interpret and represent scenes or objects that are underrepresented in the dataset. This can lead to the model developing a bias towards the types of scenes or objects that are more common in the dataset, and potentially misinterpreting or misrepresenting those that are less common [44]. Similarly, knowledge sources used in SGG can also exhibit contextual bias. For example, if a knowledge graph used for infusing common sense knowledge into the model contains more detailed or comprehensive information about certain types of objects or relationships, the model may develop a bias towards these objects or relationships. This can lead to the model overemphasizing these aspects and potentially misinterpreting or underrepresenting others [43].

### 3.5. Generalization bias

Generalization bias arises due to various factors such as data pre-processing, model training, and architectural choices in SGG. Generalization bias can lead to sub-optimal performance, overfitting, and poor performance on unseen data. For instance, a model trained on the VRD dataset [18], which contains a high number of images with people "holding" objects, may struggle to recognize other relationships, such as people "standing next to" objects, because it has overfitted to the "holding" relationship [6]. Similarly, if a Knowledge Graph (KG) contains extensive information on European history but lacks comprehensive data on African history, the model will struggle to accurately predict relationships involving African historical figures or events. This is an example of how bias in knowledge sources can limit the generalizability of SGG models. Moreover, the long-tailed distribution of training data in datasets can also introduce generalization bias. For example, SGG methods often suffer from sub-optimal scene graph generation due to the long-tailed distribution of training data, which can lead to most frequent relation predictions caused by capricious visual features and trivial dataset annotations [45].

### 3.6. Evaluation bias

Evaluation bias is closely tied to the metrics and benchmarks used for evaluating the performance of SGG models. These metrics and benchmarks may not accurately reflect the model's performance in real-world scenarios, potentially leading to overfitting or biased model development. This can result in misleading or unrepresentative evaluation metrics, which do not truly reflect the model's ability to generalize to diverse images or scenarios. For instance, consider a scenario where only the ten most frequent relationships in the VG dataset [1] are correctly classified in a test. In this case, the accuracy could reach 90%, even if the rest of the forty relationships are all wrong. This means that the model might perform well on the evaluation set but fail to generalize to more diverse images or scenarios. This is a clear example of evaluation bias, where the evaluation metric (accuracy in this case) does not accurately reflect the model's performance in real-world scenarios. In addition to this, recent studies have pointed out that conventional evaluation metrics such as Recall@K, which measures the ratio of correctly predicted triplets that appear in the ground truth, cannot capture the global semantic information of scene graphs and measure the similarity between images and generated scene graphs [46]. This further limits the usability of scene graphs in downstream tasks and contributes to the evaluation bias. Furthermore, the long-tailed distribution of training data also contributes to evaluation bias in SGG. For instance, models trained on datasets with a long-tailed distribution often perform well on frequent categories but struggle with infrequent ones [6]. This discrepancy in performance is often not captured by conventional evaluation metrics, leading to a biased evaluation of the model's performance.

### 3.7. Integration bias

Integration bias arises due to the integration of multi-modal data features or multiple knowledge sources in SGG, which can affect the reasoning capability of the system and lead to an inconsistent and conflicting understanding of relationships. For instance, Chen et al. [6] discussed how the integration of different types of data can lead to biases in the model's understanding of relationships. In this case, the model might overemphasize certain relationships due to the uneven distribution of data, leading to a biased understanding of the scene. Gu et al. [25] discussed how the integration of external knowledge can lead to biases in SGG. They argue that the use of external knowledge bases can help improve the generalizability of SGG, but it can also introduce biases if the knowledge base itself is biased or incomplete. Similarly, Zhang et al. [47] discussed how the integration of different types of data can lead to biases in the understanding of 3D scenes. They argue that the use of edge-oriented reasoning can help improve the accuracy of scene graph generation, but it can also introduce biases if the data used for edge reasoning is biased or incomplete. Chang et al. [35] discussed how cognitive biases can be introduced in SGG through the integration of linguistic features and visual representations. They argue that while these biases can help improve the model's understanding of the scene, they can also lead to biases if the linguistic features or visual representations are biased or incomplete. Lastly, Luo et al. [48] discussed how integrating different types of data can lead to biases in optimising scene layouts. They argue that the use of end-to-end optimization can help improve the accuracy of scene layout generation, but it can also introduce biases if the data used for

optimization is biased or incomplete.

## 4. Bias Mitigation Approaches

Mitigating bias in SGG is crucial to ensure fair and accurate scene understanding and downstream visual reasoning. Efforts to mitigate bias in SGG due to long-tailed relationship distribution in datasets include incorporating contextual information with gated recurrent units [7] and introducing the total direct effect loss function to distinguish between good and bad context biases [14]. Guo et al. [38] developed a domain transfer framework to target bad bias and predicate imbalance. The explicit ontological adjustment method [49] adjusts predicate logits using KG adjacency matrices to prioritize informative predicates. The Mean Recall (mR@K) metric [28, 50] balances evaluation by emphasizing infrequent but valuable predicates. Malawade et al. [51] introduced *roadscene2vec*, an open-source tool for extracting and embedding road scene graphs, which helps mitigate bias in the generation of road scene graphs. Chen et al. [6] proposed Resistance Training using Prior Bias (RTPB) for SGG, which uses a distributed-based prior bias to improve the detecting ability of models on less frequent relationships during training, thus improving the model generalizability on tail categories. Gu et al. [25] proposed a novel scene graph generation algorithm with external knowledge and image reconstruction loss to overcome dataset issues such as bias, noise, and missing annotations.

The primary focus so far has been on mitigating imbalance and contextual bias related to datasets, while evaluation and domain biases have also been investigated to some extent. Other biases, such as sociocultural, generalization and integration biases, remain unexplored. Due to its significant impact on the performance and fairness of SGG methods, investigating biases is a crucial future direction for scene understanding and visual reasoning research. This will involve developing novel bias mitigation techniques, refining existing techniques, and the exploration of new evaluation metrics and benchmarks that can more accurately reflect the performance of SGG models in diverse and real-world scenarios.

The bias mitigation techniques developed for machine learning can be adapted and applied to reduce biases in the datasets, models, and knowledge sources used in SGG. These techniques work by altering the training data, the learning algorithm, or the output predictions to lessen the impact of biases [52]. There are several bias mitigation algorithms that focus on modifying the training data. For instance, the Reweighting method [53] assigns different weights to the training examples in each (group, label) combination to ensure fairness before the classification process. Optimized pre-processing [54] is another method that learns a probabilistic transformation to edit the features and labels in the data. This method is designed to meet group fairness, individual distortion, and data fidelity constraints and objectives. Learning fair representations [55] is a technique that aims to find a latent representation that encodes the data well but obscures information about protected attributes. Disparate impact remover [56] edits feature values to increase group fairness while preserving rank ordering within groups. These bias mitigation techniques can be used to balance the representation of different groups in the datasets used for SGG. These techniques can also ensure that the knowledge sources used in SGG do not disproportionately favour or disadvantage any particular group.

Adversarial debiasing [57] and prejudice remover [58] are two bias mitigation techniques



that operate by modifying the learning algorithm itself, and they can be effectively applied to mitigate biases in the deep learning and knowledge enrichment pipelines for SGG. Adversarial debiasing [57] is a technique that trains a classifier to not only maximize prediction accuracy but also minimize the ability of an adversary to determine the protected attribute from the predictions. In the context of SGG, this could mean training the model to generate scene graphs that accurately represent the visual scene while making it difficult for an adversary to determine protected attributes (like gender or race) from the generated scene graphs. This approach can help reduce sociocultural bias by ensuring that the model predictions do not unfairly favour or discriminate against certain groups based on protected attributes. On the other hand, the prejudice remover [58] adds a discrimination-aware regularization term to the learning objective. This means that during the training process, the model is trying to minimize the prediction error and the discriminatory bias. In the context of SGG, this could mean training the model to generate scene graphs that are accurate and fair, in the sense that they do not disproportionately favour certain objects or relationships based on biased training data. This approach can help reduce imbalance bias by ensuring that the model predictions are not skewed towards overrepresented classes in the training data.

Bias mitigation techniques that modify the predictions of a model can also be employed to reduce biases in SGG. These techniques adjust the output labels of a model to ensure fairness. One such technique is equalized odds post-processing [59], which uses a linear program to determine the probabilities for changing output labels, with the goal of optimizing equalized odds. This ensures that the predictions should be equally accurate for all groups. Calibrated equalized odds post-processing [60] is a similar technique, but it optimizes over the calibrated classifier score outputs instead of the output labels directly. This method finds the probabilities for changing output labels with an equalized odds objective, ensuring that the false positive and false negative rates are similar across all groups. This can help reduce evaluation bias by ensuring that the performance is evaluated fairly across different groups. Reject option classification [61] is a technique that provides favourable outcomes to underprivileged groups and unfavourable outcomes to privileged groups in a confidence band around the decision boundary with the highest uncertainty. This can help reduce imbalance bias by ensuring that underrepresented groups are not disproportionately disadvantaged by the predictions.

## 5. Conclusion

Bias in SGG significantly impacts the fairness and effectiveness of multimodal scene understanding and visual reasoning techniques. Mitigating these biases is a complex, ongoing process that requires a comprehensive understanding of the sources and impacts of bias, and the application of effective bias mitigation strategies. The investigation of biases and fairness in SGG is a crucial future direction for multimodal scene understanding and visual reasoning research. The development of fairer and more robust SGG techniques will lead to more equitable applications of multimodal scene understanding and visual reasoning.

## Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6223 and 12/RC/2289\_P2. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## References

- [1] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, et al., Visual genome: Connecting language and vision using crowdsourced dense image annotations, *International journal of computer vision* 123 (2017) 32–73.
- [2] S. Barocas, A. D. Selbst, Big data’s disparate impact, *California law review* (2016) 671–732.
- [3] S. Verma, Weapons of math destruction: how big data increases inequality and threatens democracy, *Vikalpa* 44 (2019) 97–98.
- [4] K. Crawford, The trouble with bias, *NIPS 2017 Keynote* (2017).
- [5] J. Johnson, R. Krishna, M. Stark, L.-J. Li, D. Shamma, M. Bernstein, L. Fei-Fei, Image retrieval using scene graphs, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3668–3678.
- [6] C. Chen, Y. Zhan, B. Yu, L. Liu, Y. Luo, B. Du, Resistance training using prior bias: toward unbiased scene graph generation, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2022, pp. 212–220.
- [7] D. Xu, Y. Zhu, C. B. Choy, L. Fei-Fei, Scene graph generation by iterative message passing, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5410–5419.
- [8] A. G. Greenwald, L. H. Krieger, Implicit bias: Scientific foundations, *California Law Review* 94 (2006) 945–967.
- [9] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, et al., Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations, *Minds and Machines* 28 (2018) 689–707.
- [10] R. Binns, Fairness in machine learning: Lessons from political philosophy, *Proceedings of Machine Learning Research* 81 (2018) 1–11.
- [11] S. A. Friedler, C. Scheidegger, S. Venkatasubramanian, S. Choudhary, E. P. Hamilton, D. Roth, A comparative study of fairness-enhancing interventions in machine learning, in: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019, pp. 329–338.
- [12] R. K. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, et al., Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias, *arXiv preprint arXiv:1810.01943* (2018).
- [13] S. Corbett-Davies, S. Goel, The measure and mismeasure of fairness: A critical review of fair machine learning, *arXiv preprint arXiv:1808.00023* (2018).

- [14] K. Tang, Y. Niu, J. Huang, J. Shi, H. Zhang, Unbiased scene graph generation from biased training, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3716–3725.
- [15] X. Chang, P. Ren, P. Xu, Z. Li, X. Chen, A. Hauptmann, A comprehensive survey of scene graphs: Generation and application, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45 (2023) 1–26. doi:10.1109/TPAMI.2021.3137605.
- [16] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, R. Socher, Deep learning-enabled medical computer vision, *NPJ digital medicine* 4 (2021) 5.
- [17] M. J. Khan, J. G. Breslin, E. Curry, Common sense knowledge infusion for visual understanding and reasoning: Approaches, challenges, and applications, *IEEE Internet Computing* 26 (2022) 21–27.
- [18] C. Lu, R. Krishna, M. Bernstein, L. Fei-Fei, Visual relationship detection with language priors, in: *European Conference on Computer Vision*, Springer, 2016, pp. 852–869.
- [19] Y. Guo, J. Song, L. Gao, H. T. Shen, One-shot scene graph generation, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 3090–3098.
- [20] X. Liang, L. Lee, E. P. Xing, Deep variation-structured reinforcement learning for visual relationship and attribute detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 848–857.
- [21] R. Zellers, M. Yatskar, S. Thomson, Y. Choi, Neural motifs: Scene graph parsing with global context, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5831–5840.
- [22] M. J. Khan, J. G. Breslin, E. Curry, Expressive scene graph generation using commonsense knowledge infusion for visual understanding and reasoning, in: *The Semantic Web: 19th International Conference, ESWC 2022, Hersonissos, Crete, Greece, May 29–June 2, 2022*, Proceedings, Springer, 2022, pp. 93–112.
- [23] X. Kan, H. Cui, C. Yang, Zero-shot scene graph relation prediction through commonsense knowledge integration, in: *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021*, Proceedings, Part II 21, Springer, 2021, pp. 466–482.
- [24] M. J. Khan, J. Breslin, E. Curry, Neusyre: Neuro-symbolic visual understanding and reasoning framework based on scene graph enrichment, *Semantic Web* (2023).
- [25] J. Gu, H. Zhao, Z. Lin, S. Li, J. Cai, M. Ling, Scene graph generation with external knowledge and image reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1969–1978.
- [26] A. Zareian, S. Karaman, S.-F. Chang, Bridging knowledge graphs to generate scene graphs, in: *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020*, Proceedings, Part XXIII 16, Springer, 2020, pp. 606–623.
- [27] H. Zhou, Y. Yang, T. Luo, J. Zhang, S. Li, A unified deep sparse graph attention network for scene graph generation, *Pattern Recognition* 123 (2022) 108367.
- [28] T. Chen, W. Yu, R. Chen, L. Lin, Knowledge-embedded routing network for scene graph generation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6163–6171.
- [29] R. Speer, J. Chin, C. Havasi, Conceptnet 5.5: An open multilingual graph of general

- knowledge, in: Thirty-first AAAI conference on artificial intelligence, 2017, pp. 4444–4451.
- [30] G. A. Miller, Wordnet: a lexical database for english, *Communications of the ACM* 38 (1995) 39–41.
- [31] F. Ilievski, P. Szekely, D. Schwabe, Commonsense knowledge in wikidata, *arXiv preprint arXiv:2008.08114* (2020).
- [32] F. Ilievski, P. Szekely, B. Zhang, Cskg: The commonsense knowledge graph, in: *The Semantic Web: 18th International Conference, ESWC 2021, Virtual Event, June 6–10, 2021, Proceedings* 18, Springer, 2021, pp. 680–696.
- [33] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V* 13, Springer, 2014, pp. 740–755.
- [34] D. A. Hudson, C. D. Manning, Gqa: A new dataset for real-world visual reasoning and compositional question answering, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6700–6709.
- [35] X. Chang, T. Wang, C. Sun, W. Cai, Biasing like human: A cognitive bias framework for scene graph generation, *arXiv preprint arXiv:2203.09160* (2022).
- [36] N. Shahbazi, Y. Lin, A. Asudeh, H. Jagadish, Representation bias in data: A survey on identification and resolution techniques, *ACM Computing Surveys* (2023).
- [37] M.-J. Chiou, et al., Recovering the unbiased scene graphs from the biased ones (2021).
- [38] Y. Guo, L. Gao, X. Wang, Y. Hu, X. Xu, X. Lu, H. T. Shen, J. Song, From general to specific: Informative scene graph generation via balance adjustment, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 16383–16392.
- [39] J. Zhang, et al., Scene graph generation with external knowledge and image reconstruction (2019).
- [40] X. Yang, K. Tang, H. Zhang, J. Cai, Auto-encoding scene graphs for image captioning, *arXiv preprint arXiv:1812.02378* (2018).
- [41] J. Wald, H. Dharmo, N. Navab, F. Tombari, Learning 3d semantic scene graphs from 3d indoor reconstructions, *arXiv preprint arXiv:2004.03967* (2020).
- [42] S. Garg, H. Dharmo, A. Farshad, S. Musatian, N. Navab, F. Tombari, Unconditional scene graph generation, *arXiv preprint arXiv:2108.05884* (2021).
- [43] J. Yang, J. Lu, S. Lee, D. Batra, D. Parikh, Graph r-cnn for scene graph generation, in: *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 670–685.
- [44] L. Xu, H. Qu, J. Kuen, J. Gu, J. Liu, Meta spatio-temporal debiasing for video scene graph generation, in: *European Conference on Computer Vision*, Springer, 2022, pp. 374–390.
- [45] B. Xu, R. Liao, L. Sigal, Self-supervised relation alignment for scene graph generation, *arXiv preprint arXiv:2302.01403* (2023).
- [46] Y. Cong, W. Liao, B. Rosenhahn, M. Y. Yang, Learning similarity between scene graphs and images with transformers, *arXiv preprint arXiv:2304.00590* (2023).
- [47] C. Zhang, J. Yu, Y. Song, W. Cai, Exploiting edge-oriented reasoning for 3d point-based scene graph analysis, *arXiv preprint arXiv:2103.05558* (2021).
- [48] A. Luo, Z. Zhang, J. Wu, J. B. Tenenbaum, End-to-end optimization of scene layout, *arXiv preprint arXiv:2007.11744* (2020).
- [49] Z. Chen, S. Rezayi, S. Li, More knowledge, less bias: Unbiasing scene graph generation

- with explicit ontological adjustment, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 4023–4032.
- [50] K. Tang, H. Zhang, B. Wu, W. Luo, W. Liu, Learning to compose dynamic tree structures for visual contexts, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 6619–6628.
- [51] A. V. Malawade, S.-Y. Yu, B. Hsu, H. Kaeley, A. Karra, M. A. Al Faruque, Roadscene2vec: A tool for extracting and embedding road scene-graphs, Knowledge-Based Systems 242 (2022) 108245.
- [52] B. d’Alessandro, C. O’Neil, T. LaGatta, Conscientious classification: A data scientist’s guide to discrimination-aware classification, Big data 5 (2017) 120–134.
- [53] F. Kamiran, T. Calders, Data preprocessing techniques for classification without discrimination, Knowledge and information systems 33 (2012) 1–33.
- [54] F. Calmon, D. Wei, B. Vinzamuri, K. Natesan Ramamurthy, K. R. Varshney, Optimized pre-processing for discrimination prevention, Advances in neural information processing systems 30 (2017).
- [55] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, C. Dwork, Learning fair representations, in: International conference on machine learning, PMLR, 2013, pp. 325–333.
- [56] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, S. Venkatasubramanian, Certifying and removing disparate impact, in: proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, 2015, pp. 259–268.
- [57] B. H. Zhang, B. Lemoine, M. Mitchell, Mitigating unwanted biases with adversarial learning, in: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 2018, pp. 335–340.
- [58] T. Kamishima, S. Akaho, H. Asoh, J. Sakuma, Fairness-aware classifier with prejudice remover regularizer, in: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2012, Bristol, UK, September 24–28, 2012. Proceedings, Part II 23, Springer, 2012, pp. 35–50.
- [59] M. Hardt, E. Price, N. Srebro, et al., Equality of opportunity in supervised learning, in ‘advances in neural information processing systems’ (2016).
- [60] G. Pleiss, M. Raghavan, F. Wu, J. Kleinberg, K. Q. Weinberger, On fairness and calibration, Advances in neural information processing systems 30 (2017).
- [61] F. Kamiran, A. Karim, X. Zhang, Decision theory for discrimination-aware classification, in: 2012 IEEE 12th international conference on data mining, IEEE, 2012, pp. 924–929.