Grounding “grounding”: How is grounding used within various AI conferences?

Eric Huang
University of Waterloo
e48huang@uwaterloo.ca

Abstract

Terminology used within linguistics and AI conferences tend to be overused, leading to ambiguous meaning and difficulty navigating new papers. This paper will elucidate the various senses of the word “grounding” through qualitative analysis and deeper quantitative analysis of its various uses. This work will showcase how “grounding” is an overloaded term and guide users to understand how to more easily decipher and understand papers using the term. All code can be found at https://git.uwaterloo.ca/e48huang/cs-784/-/tree/final_project/final_project?ref_type=heads where a University of Waterloo account is required.

1 Introduction

Many conferences centering around Artificial Intelligence have existed for many decades, evolving over time on the types of problems that they tackle. While these problems change over time, so do the terminology, which have a tendency to evolve semantically, leading to overloaded terms. One such term is “grounding”, the idea that one wishes to ensure that there is understanding or a common ground (Nakano et al., 2003). While this term seems simple, it is used in many various contexts, all of which requires different datasets, methods and metrics to evaluate, while being applied in different settings.

To better understand the term “grounding” and its usage, we perform both quantitative analysis and qualitative analysis. This paper explores the “Seed42Lab/AI-paper-crawl” HuggingFace dataset (Forty-Two AI Lab) which collects full-text papers from 11 different conferences spanning from the first year of the conference to 2024. To first select different senses of the word “grounding”, we perform preliminary quantitative analysis to filter for papers to further investigate. From these selected papers, we identify 9 related but distinct meanings of the word “grounding”. We perform some literature review to understand how these different senses are understood, from its various datasets, methods, metrics and applications. Finally, for each of these word senses, we investigate how they have evolved over time.

Conference	Paper Count
AAAI	772
ACL	632
CVPR	862
ECCV	511
EMNLP	575
ICCV	341
ICLR	360
ICML	360
IJCAI	654
NAACL	226
NeurIPS	654

Table 1: Counts of unique papers with “grounding” by conference found in the corpora.

2 Paper Selection

A simple search over the number of papers which have the term “grounding” quickly shows that it is infeasible to cover all possible instances. For example, the Association for Computational Linguistics (ACL) alone has 632 unique papers that have an instance of “grounding” (see Table 1). While not all these instances are due to the paper itself being related to grounding, as they can simply include it within its bibliography, they are indicative that some filtering is necessary.

To filter through these papers, we propose a method which selects the most relevant papers within a conference to the word “grounding”. We take a naive approach where we select the top 10% of papers with the word “grounding”. We determine which papers are more important to "grounding" based on the word frequency, if "grounding" appears more often compared to other words within a paper, then it should be more relevant. This process (while ensuring uniqueness across conferences) resulted in 46 curated papers[1], spanning from the years 2000 to 2024, shown in the following figure. While this selection of papers may not cover the breadth of senses of grounding might entail, as it misses on papers from the 1980's to 2000's, it does cover the most commonly used senses of the word.

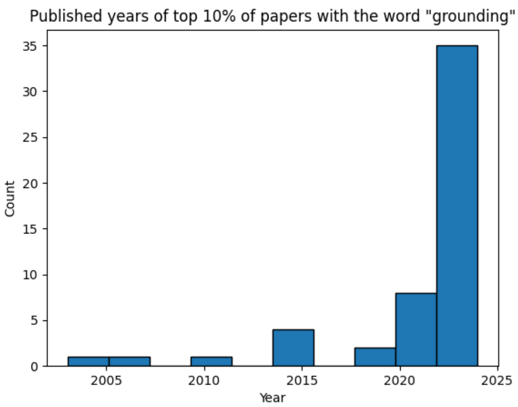

Figure 1: Count of selected conference papers per year

3 Grounding "grounding"

In this section we cover the different senses of the word "grounding" found in the 46 papers covered in the previous section. We will explore the common methods, datasets, applications and metrics revolving around each word. We will continue to elucidate the meanings of the word and its trend within the next section. The following sections also denote its subcategories of "grounding", however these subcategories are loosely defined and are chosen not necessarily based on modality, but the end

3.1 Visual Grounding

Visual grounding, also known as image, phrase or referring expression grounding (Xiao et al., 2024; Li et al., 2024; Ma et al., 2020; Islam et al., 2023; Jiang et al., 2019; Lu et al., 2022; Dou and Peng, 2021; Surikuchi et al., 2023) refers to the challenge of trying to localize specific regions within an image based on some textual description[2]. Traditionally, this involved finding the phrase's referring region by predicting a bounding box around said region. As time has gone on however, there has been more and more types of challenges that one could tackle within visual grounding (Xiao et al., 2024). From the 46 papers we filtered for, we have discovered the following subcategories involved in visual grounding:

3.1.1 Classical Visual Grounding

This entails the traditional problem of trying to predict a bounding box around a region (Li et al., 2024; Huang et al., 2021; Peng et al., 2023). Recent papers have found different methods in an attempt to improve performance. Zeng et al. (2024) improves compositional understanding through providing a harder dataset that relies on introducing different compositions of objects. Zhang et al. (2020) improves performance in weakly supervised (no bounding box annotations) settings through contrastive learning. Ma et al. (2024) provides a new dataset for higher resolution images at different granularity and bounding box sizes. Lee and Sung (2024) has shown improvements in image generation as well.

This visual grounding task can be understood as the inverse problem to image captioning, where one is given an image and need to provide the text portion. In fact, this inverse paradigm has led to better models (Wang et al., 2023a) involving cyclic updates.

3.1.2 Answer Grounding

Rather than fit a bounding box to various objects, Visual Question Answering (VQA) grounding attempts to find specific parts of an image that corresponds with inputted questions rather than prompts (Chen et al., 2022, 2023).

3.2 Action Grounding

Relying on other types of grounding such as image grounding, action grounding is a term that refers to building a model that is able to take some grounding and relate it to a set of actions. Recent works utilize LLMs in the fields of chat agents, web agents and robotics (Zhang et al., 2023; Cheng et al., 2024; Zheng et al., 2024; Tellex et al., 2011; Wang et al., 2023b) to motivate better actions that are aligned with people's understanding of the world.

[1] https://git.uwaterloo.ca/e48huang/cs-784/-/blob/e09a1c22c0de7e331ca16109a5f32b226dc6d9c5/final_project/grounding_top_p.txt

[2] other terms include natural language object retrieval or phrase localization (Ma et al., 2024)

Contrary to using image grounding, Kameko et al. (2015) matches certain states of games to commentary in an attempt to understand how various actions are grounded in language or its symbols. They refer to this type of grounding as "symbol grounding" but essentially attempts to relate some action to some other observation.

3.3 Audio Grounding

Audio grounding is the task of taking static images and sounds and attempting to identify which parts of the image are correlated with certain parts of audio. For example, Tian et al. (2021) attempts to separate images of bands into which instruments produce what kinds of audio.

3.4 Video Grounding

Another related grounding task to visual grounding is the idea of video grounding or spatio-temporal grounding. This task is to identify various portions of a video or the entities within them to provide an understanding for a certain prompt (Jiang et al., 2024). These different grounding tasks can be split into its own categories defined in the next sections.

3.4.1 Object Tracking

Object tracking relies on the idea that given some natural language prompt, to both identify the specific object within the video but also to continuously track it throughout the video or still frames (Zhou et al., 2023).

3.4.2 Natural Language Spatial Video Grounding

This video grounding task is an extension of classic visual grounding, where the model attempts to set a bounding box for each frame of a video (Li et al., 2022; Ma et al., 2020).

3.4.3 Temporal Video Grounding

This video grounding task is to identify the timestamps in which a prompt holds true for a video (Li et al., 2024; Afouras et al., 2023; Bao et al., 2021; Chen et al., 2018).

3.4.4 Spatio-temporal Video Grounding

This video grounding task combines the last two tasks and attempts to identify both the bounding boxes and the timestamps in which a prompt holds true for a video (Wasim et al., 2024; Chen et al., 2024; Jin et al., 2022). It can be used within various settings including video entailment which determines whether a prompt holds true for some video (Chen and Kong, 2021). Similar to image grounding, video grounding can also be used within video generation tools (Jeong and Ye, 2024).

3.5 3D Grounding

Similar to image grounding, 3D grounding adds a dimension and attempts to put bounding boxes around 3D models which are often represented as point clouds. These 3D grounding tasks share similar strategies to image grounding, using captioning tasks to improve performance (Cai et al., 2022; Yang et al., 2023; Miyanishi et al., 2023; Wang et al., 2023c). Some papers have even used 2D object representations to improve 3D grounding (Yang et al., 2021), while others have improved 3D visual grounding with reasoning (Zhu et al., 2024).

3.6 Dialogue Grounding

This term of "dialogue grounding" is loosely defined, usually seen in literature simply as "grounding". Within these papers, "grounding" refers to the idea of trying to build a common ground of understanding between two or more actors within a conversation. It includes attempting to analyze nonverbal behaviours (Nakano et al., 2003; Roque, 2007; Liu et al., 2012; Shaikh et al., 2024).

3.7 Markov Logic Networks Grounding

"Grounding" in Markov Logic Networks (MLNs) differs significantly from the other senses of the word (Venugopal and Gogate, 2014). MLNs refer to a statistical model for probabilistic logic reasoning, where by developing a set of first-order logic rules known as "grounds" one is able to form a weighted satisfiability problem with an optimized solution. In particular, grounding within Markov Logic Networks refers to the process of forming the weighted graph (Fang et al., 2023).

3.8 Physical Dynamics Grounding

Attempting to model physical dynamics purely from states and its transitions tend to be difficult, requiring a ton of resources to supervise consecutive particle properties. Instead of requiring this supervision, a new field has emerged to attempt to understand these physical dynamics from visual observations (Cao et al., 2024). One such application is in fluid dynamics grounding; which attempts to build an understanding of fluid particle systems from sequential visual observations (Guan et al., 2022).

4 Analyzing “grounding”’s usage

In this section, we will build a quantitative understanding of “grounding” and its senses over time. We will explore how the word has been used throughout the years, and dive deeper into a few senses of the word. We will accomplish this through observing the co-occurrence trends over time with other key words for each specific sense.

4.1 How has “grounding” evolved over time?

In this section, we explore how the term “grounding” has evolved over time through analyzing how many papers have included the term “grounding”. We aggregate over all the data splits while showcasing a more fine-grained example for a specific conference to avoid any patterns lost through aggregation.

In particular, we observe that the number of instances of “grounding” has increased both in terms of pure count and frequency over time (see Fig 2 and Fig 3). Due to the number of papers being published in general, we need to normalize for this factor and measure the proportion of papers which include “grounding”. Because the pure count and frequency have both increased over time, we conclude that “grounding” has been a terminology that is becoming more and more utilized, likely due to it becoming more relevant with the uprise of LLMs and a need to interpret and improve these models. We provide more analysis on this co-occurrence within the next section.

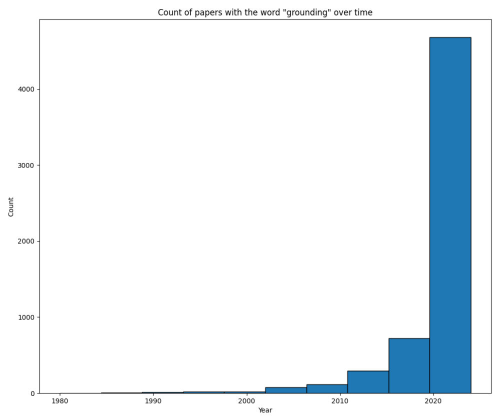

Figure 2: Count of “grounding” for all the conferences

To see how each individual conference’s count and frequency changes over time, see Appendix A.

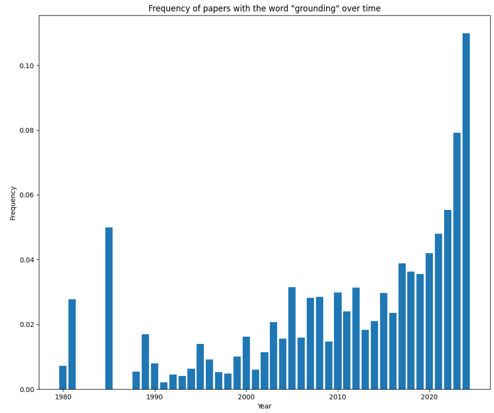

Figure 3: Frequency of papers containing the word “grounding” for all conferences over time

4.2

4.2.1 Visual Grounding

TODO: Also explore why there are so many different ways to mean visual grounding. Also explore which datasets are most popular within these papers. Also, explore with removing IJCAI, as empirically it seems to have a lot of overlap with other conferences...

5 Datasets and Methods

6 Discussion

7 Conclusion

Limitations

References

Triantafyllos Afouras, Effrosyni Mavroudi, Tushar Nagarajan, Huiyu Wang, and Lorenzo Torresani. 2023. Ht-step: Aligning instructional articles with how-to videos. In *Advances in Neural Information Processing Systems*, volume 36, pages 50310–50326. Curran Associates, Inc.

Peijun Bao, Qian Zheng, and Yadong Mu. 2021. Dense events grounding in video. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(2):920–928.

Daigang Cai, Lichen Zhao, Jing Zhang, Lu Sheng, and Dong Xu. 2022. 3djcg: A unified framework for joint dense captioning and visual grounding on 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16464–16473.

Junyi Cao, Shanyan Guan, Yanhao Ge, Wei Li, Xiaokang Yang, and Chao Ma. 2024. Neuma: Neural material adaptor for visual grounding of intrinsic dynamics. In *Advances in Neural Information Processing Systems*, volume 37, pages 65643–65669. Curran Associates, Inc.

Brian Chen, Nina Shvetsova, Andrew Rouditchenko, Daniel Kondermann, Samuel Thomas, Shih-Fu Chang, Rogerio Feris, James Glass, and Hilde Kuehne. 2024. What when and where? self-supervised spatio-temporal grounding in untrimmed multi-action videos from narrated instructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18419–18429.

Chongyan Chen, Samreen Anjum, and Danna Gurari. 2022. Grounding answers for visual questions asked by visually impaired people. *Preprint*, arXiv:2202.01993.

Chongyan Chen, Samreen Anjum, and Danna Gurari. 2023. Vqa therapy: Exploring answer differences by visually grounding answers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15315–15325.

Jingyuan Chen, Xinpeng Chen, Lin Ma, Zequn Jie, and Tat-Seng Chua. 2018. Temporally grounding natural sentence in video. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 162–171, Brussels, Belgium. Association for Computational Linguistics.

Junwen Chen and Yu Kong. 2021. Explainable video entailment with grounded visual evidence. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2021–2030.

Kanzhi Cheng, Qiushi Sun, Yougang Chu, Fangzhi Xu, Li YanTao, Jianbing Zhang, and Zhiyong Wu. 2024. SeeClick: Harnessing GUI grounding for advanced visual GUI agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9313–9332, Bangkok, Thailand. Association for Computational Linguistics.

Zi-Yi Dou and Nanyun Peng. 2021. Improving pre-trained vision-and-language embeddings for phrase grounding. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6362–6371, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Huang Fang, Yang Liu, Yunfeng Cai, and Mingming Sun. 2023. Mln4kb: an efficient markov logic network engine for large-scale knowledge bases and structured logic rules. In *Proceedings of the ACM Web Conference 2023*, WWW '23, page 2423–2432, New York, NY, USA. Association for Computing Machinery.

Forty-Two AI Lab. Seed42lab/ai-paper-crawl. `https://huggingface.co/datasets/Seed42Lab/AI-paper-crawl`.

Shanyan Guan, Huayu Deng, Yunbo Wang, and Xiaokang Yang. 2022. Neurofluid: Fluid dynamics grounding with particle-driven neural radiance fields. *Preprint*, arXiv:2203.01762.

Jianqiang Huang, Yu Qin, Jiaxin Qi, Qianru Sun, and Hanwang Zhang. 2021. Deconfounded visual grounding. *Preprint*, arXiv:2112.15324.

Md Mofijul Islam, Alexi Gladstone, and Tariq Iqbal. 2023. Patron: Perspective-aware multitask model for referring expression grounding using embodied multimodal cues. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1):971–979.

Hyeonho Jeong and Jong Chul Ye. 2024. Ground-a-video: Zero-shot grounded video editing using text-to-image diffusion models. *Preprint*, arXiv:2310.01107.

Ming Jiang, Qiuyuan Huang, Lei Zhang, Xin Wang, Pengchuan Zhang, Zhe Gan, Jana Diesner, and Jianfeng Gao. 2019. TIGEr: Text-to-image grounding for image caption evaluation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2141–2152, Hong Kong, China. Association for Computational Linguistics.

Wenhui Jiang, Yibo Cheng, Linxin Liu, Yuming Fang, Yuxin Peng, and Yang Liu. 2024. Comprehensive visual grounding for video description. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(3):2552–2560.

Yang Jin, Yongzhi Li, Zehuan Yuan, and Yadong Mu. 2022. Embracing consistency: A one-stage approach for spatio-temporal video grounding. *Preprint*, arXiv:2209.13306.

Hirotaka Kameko, Shinsuke Mori, and Yoshimasa Tsuruoka. 2015. Can symbol grounding improve low-level NLP? word segmentation as a case study. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2298–2303, Lisbon, Portugal. Association for Computational Linguistics.

Phillip Y. Lee and Minhyuk Sung. 2024. Reground: Improving textual and spatial grounding at no cost. *Preprint*, arXiv:2403.13589.

Mengze Li, Tianbao Wang, Haoyu Zhang, Shengyu Zhang, Zhou Zhao, Jiaxu Miao, Wenqiao Zhang, Wenming Tan, Jin Wang, Peng Wang, Shiliang Pu, and Fei Wu. 2022. End-to-end modeling via information tree for one-shot natural language spatial video grounding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8707–8717, Dublin, Ireland. Association for Computational Linguistics.

Zhaowei Li, Qi Xu, Dong Zhang, Hang Song, YiQing Cai, Qi Qi, Ran Zhou, Junting Pan, Zefeng Li, Vu Tu, Zhida Huang, and Tao Wang. 2024. GroundingGPT:

Language enhanced multi-modal grounding model. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6657–6678, Bangkok, Thailand. Association for Computational Linguistics.

Changsong Liu, Rui Fang, and Joyce Chai. 2012. Towards mediating shared perceptual basis in situated dialogue. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 140–149, Seoul, South Korea. Association for Computational Linguistics.

Panzhong Lu, Xin Zhang, Meishan Zhang, and Min Zhang. 2022. Extending phrase grounding with pronouns in visual dialogues. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7614–7625, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Chih-Yao Ma, Yannis Kalantidis, Ghassan AlRegib, Peter Vajda, Marcus Rohrbach, and Zsolt Kira. 2020. Learning to generate grounded visual captions without localization supervision. *Preprint*, arXiv:1906.00283.

Tao Ma, Bing Bai, Haozhe Lin, Heyuan Wang, Yu Wang, Lin Luo, and Lu Fang. 2024. When visual grounding meets gigapixel-level large-scale scenes: Benchmark and approach. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22119–22128.

Taiki Miyanishi, Fumiya Kitamori, Shuhei Kurita, Jungdae Lee, Motoaki Kawanabe, and Nakamasa Inoue. 2023. Cityrefer: Geography-aware 3d visual grounding dataset on city-scale point cloud data. In *Advances in Neural Information Processing Systems*, volume 36, pages 77758–77770. Curran Associates, Inc.

Yukiko Nakano, Gabe Reinstein, Tom Stocky, and Justine Cassell. 2003. Towards a model of face-to-face grounding. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 553–561, Sapporo, Japan. Association for Computational Linguistics.

Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, and Furu Wei. 2023. Kosmos-2: Grounding multimodal large language models to the world. *Preprint*, arXiv:2306.14824.

Antonio Roque. 2007. Reacting to agreement and error in spoken dialogue systems using degrees of groundedness. In *AAAI Conference on Artificial Intelligence*.

Omar Shaikh, Kristina Gligorić, Ashna Khetan, Matthias Gerstgrasser, Diyi Yang, and Dan Jurafsky. 2024. Grounding gaps in language model generations. *Preprint*, arXiv:2311.09144.

Aditya K Surikuchi, Sandro Pezzelle, and Raquel Fernández. 2023. GROOViST: A metric for grounding objects in visual storytelling. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3331–3339, Singapore. Association for Computational Linguistics.

Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew Walter, Ashis Banerjee, Seth Teller, and Nicholas Roy. 2011. Understanding natural language commands for robotic navigation and mobile manipulation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 25(1):1507–1514.

Yapeng Tian, Di Hu, and Chenliang Xu. 2021. Cyclic co-learning of sounding object visual grounding and sound separation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2745–2754.

Deepak Venugopal and Vibhav Gogate. 2014. Scaling-up importance sampling for markov logic networks. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Ning Wang, Jiajun Deng, and Mingbo Jia. 2023a. Cycle-consistency learning for captioning and grounding. *Preprint*, arXiv:2312.15162.

Renhao Wang, Jiayuan Mao, Joy Hsu, Hang Zhao, Jiajun Wu, and Yang Gao. 2023b. Programmatically grounded, compositionally generalizable robotic manipulation. *Preprint*, arXiv:2304.13826.

Zehan Wang, Haifeng Huang, Yang Zhao, Linjun Li, Xize Cheng, Yichen Zhu, Aoxiong Yin, and Zhou Zhao. 2023c. 3drp-net: 3d relative position-aware network for 3d visual grounding. *Preprint*, arXiv:2307.13363.

Syed Talal Wasim, Muzammal Naseer, Salman Khan, Ming-Hsuan Yang, and Fahad Shahbaz Khan. 2024. Videogrounding-dino: Towards open-vocabulary spatio-temporal video grounding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18909–18918.

Linhui Xiao, Xiaoshan Yang, Xiangyuan Lan, Yaowei Wang, and Changsheng Xu. 2024. Towards visual grounding: A survey. *Preprint*, arXiv:2412.20206.

Li Yang, chunfeng yuan, Ziqi Zhang, Zhongang Qi, Yan Xu, Wei Liu, Ying Shan, Bing Li, Weiping Yang, Peng Li, Yan Wang, and Weiming Hu. 2023. Exploiting contextual objects and relations for 3d visual grounding. In *Advances in Neural Information Processing Systems*, volume 36, pages 49542–49554. Curran Associates, Inc.

Zhengyuan Yang, Songyang Zhang, Liwei Wang, and Jiebo Luo. 2021. Sat: 2d semantics assisted training for 3d visual grounding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1856–1866.

Yunan Zeng, Yan Huang, Jinjin Zhang, Zequn Jie, Zhenhua Chai, and Liang Wang. 2024. Investigating compositional challenges in vision-language models for visual grounding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14141–14151.

Hao Zhang, Hongyang Li, Feng Li, Tianhe Ren, Xueyan Zou, Shilong Liu, Shijia Huang, Jianfeng Gao, Lei Zhang, Chunyuan Li, and Jianwei Yang. 2023. Llava-grounding: Grounded visual chat with large multimodal models. *Preprint*, arXiv:2312.02949.

Zhu Zhang, Zhou Zhao, Zhijie Lin, jieming zhu, and Xiuqiang He. 2020. Counterfactual contrastive learning for weakly-supervised vision-language grounding. In *Advances in Neural Information Processing Systems*, volume 33, pages 18123–18134. Curran Associates, Inc.

Boyuan Zheng, Boyu Gou, Jihyung Kil, Huan Sun, and Yu Su. 2024. Gpt-4v(ision) is a generalist web agent, if grounded. *Preprint*, arXiv:2401.01614.

Li Zhou, Zikun Zhou, Kaige Mao, and Zhenyu He. 2023. Joint visual grounding and tracking with natural language specification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23151–23160.

Chenming Zhu, Tai Wang, Wenwei Zhang, Kai Chen, and Xihui Liu. 2024. Scanreason: Empowering 3d visual grounding with reasoning capabilities. *Preprint*, arXiv:2407.01525.

A Distribution of years per conference

In Fig 2 and Fig 3, we only showed what the aggregated counts and frequency of “grounding” over time were, possibly hiding some trends. The following figures showcase that the individual conference trends follow the overall trend of increasing in both count and frequency over time.

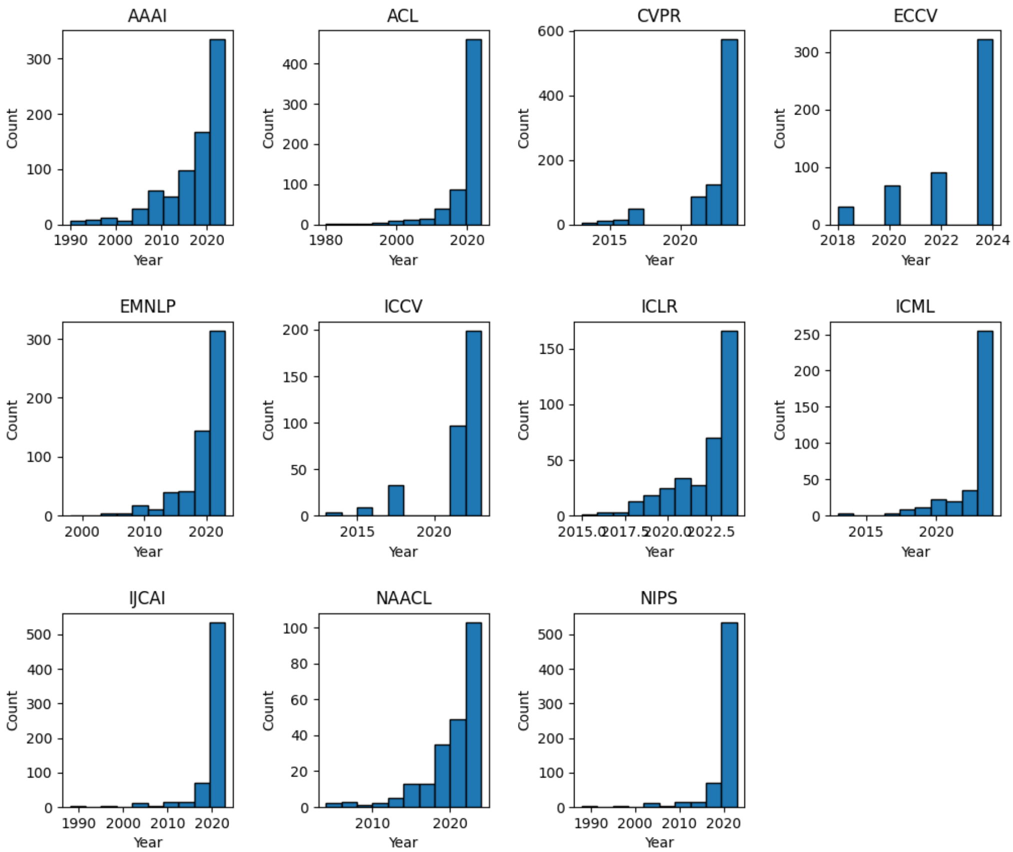

Figure 4: Count of “grounding” for all conferences over time

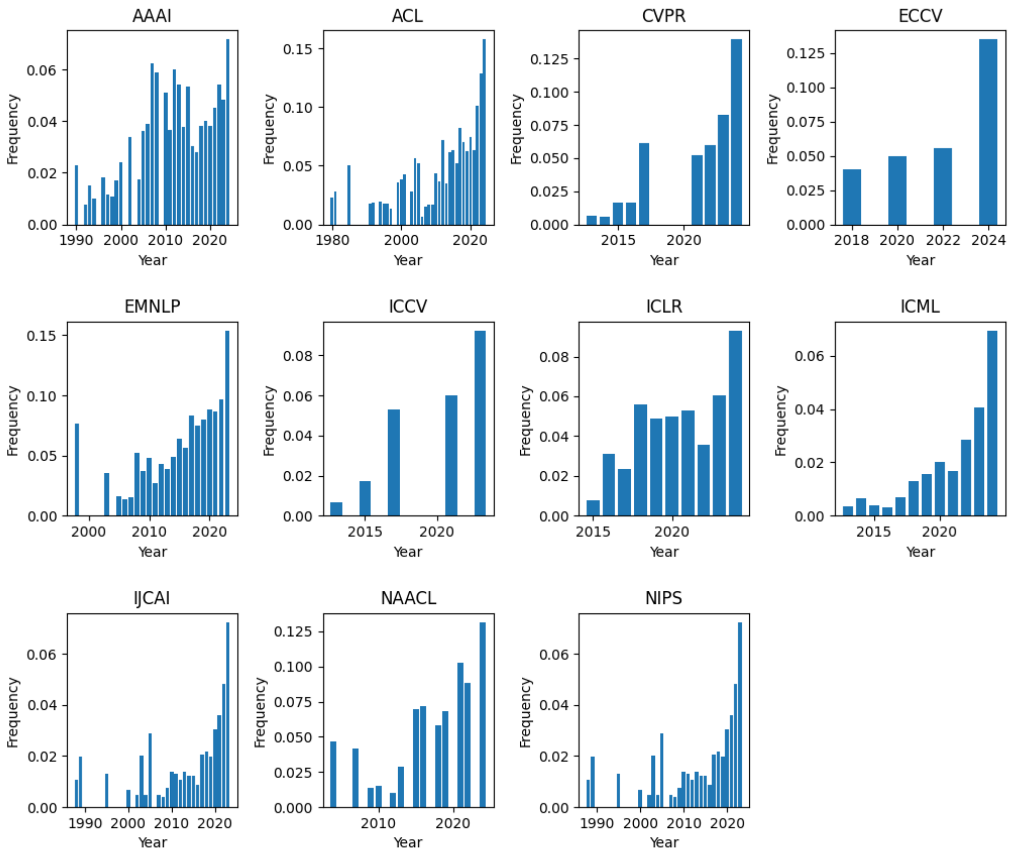

Figure 5: Frequency of papers containing the word "grounding" for all conferences over time