

Leveraging Neurosymbolic AI for Slice Discovery

Michele Collevati^[0000-0001-7958-7841], Thomas Eiter^[0000-0001-6003-6345], and
Nelson Higuera^[0000-0003-3172-723X]

Institute of Logic and Computation, Technische Universität Wien,
Favoritenstraße 9–11, 1040 Vienna, Austria
{michele.collevati, thomas.eiter, nelson.ruiz}@tuwien.ac.at

Abstract. Computer Vision (CV) has made significant progress thanks to the remarkable recent developments in deep neural networks. However, empirical studies show that CV models often make systematic errors on relevant subsets of data, called *slices*, which are groups of data sharing a set of attributes. Consequently, the *slice discovery problem* consists of discovering semantically meaningful slices on which the model achieves poor performance, called *rare* slices. To address this problem and foster the explainability of CV models, we propose a modular Neurosymbolic (NeSy) AI approach to extract logical rules that describe rare slices expressed in the Answer Set Programming (ASP) language. For experimental evaluations, we also present a methodology for inducing the occurrence of rare slices in a model by producing controlled datasets using our image generator leveraging on Super-CLEVR. Experimental results show that our approach succeeds in correctly identifying rare slices via ASP rules. To mend the model’s behaviour, the extracted rules can be directly integrated into it or exploited to generate new training data so as to improve the model’s inference capabilities.¹

Keywords: Neurosymbolic AI · Slice Discovery · Inductive Logic Programming

1 Neurosymbolic Framework for Slice Discovery

A slice is a group of data sharing a set of attributes, and the *slice discovery problem* [4, 10] has been described as mining unstructured input data for semantically meaningful slices on which the model performs poorly. To address this problem, we have contributed the following in our work:

1. We propose our *Slice Discovery Method (SDM)* consisting of a modular NeSy AI framework [5] as shown in Fig. 1, in which the generation of training data, the classification of images, the generation of scene graphs describing the semantic contents of images, the learning of rules to detect rare slices, and the mending of the neural network model form a closed loop. To achieve these tasks, we provide an image generator for datasets with rare slices that leverages on Super-CLEVR, which we use to train YOLOv5 [11]. We then translated YOLOv5’s classifications into scene graphs in the language of Inductive Logic Programming (ILP) [2, 9], which,

¹The code for reproducing our experiments is available as an online repository: <https://gitlab.tuwien.ac.at/kbs/nesy-ai/ilp4sd>.

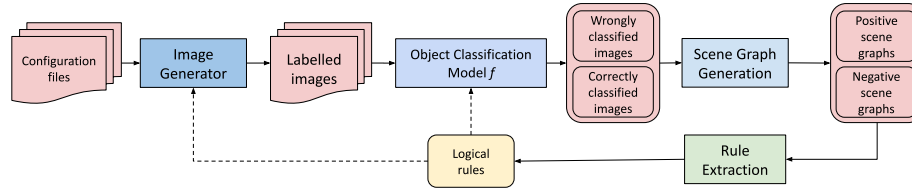


Fig. 1: Overview of the proposed neurosymbolic SDM framework. The solid arrows show the data flow, while the dashed arrows denote the exploitation of the extracted rules to improve the performance of the model, either by incorporating them into it or by using them to generate specific training data.

depending on the ground truth, constitute positive and negative examples, i.e., where the neural network incorrectly resp. correctly classified the image. We then use an ILP system, ILASP [6, 7], to obtain succinct ASP rules that reveal which images are hard for the model to classify. Finally, the neural model is trained on its checkpoint with data generated using these rules.

2. While the detection of rare slices has been widely studied [1, 4, 3], the generation of datasets with rare slices has received less attention. Therefore, we pursue a taxonomy-based approach and present a methodology for building datasets with rare slices. To this end, we leverage on Super-CLEVR [8], which is a well-known synthetic dataset that comes with a data generator for images with objects organised in hierarchical classes, see Fig. 2 for an example.
3. We provide an implementation and experimental results for datasets that we generate to test for the efficacy of the rare slice generation, the rule extraction on the neural network’s classification results, and the mending of the network model. The results show that our approach could reliably generate rare slices, and that rule learning delivered meaningful rules describing rare slices. Furthermore, feeding training data generated by using such rules to the network resulted in significant performance improvement, as misclassifications were almost eliminated.

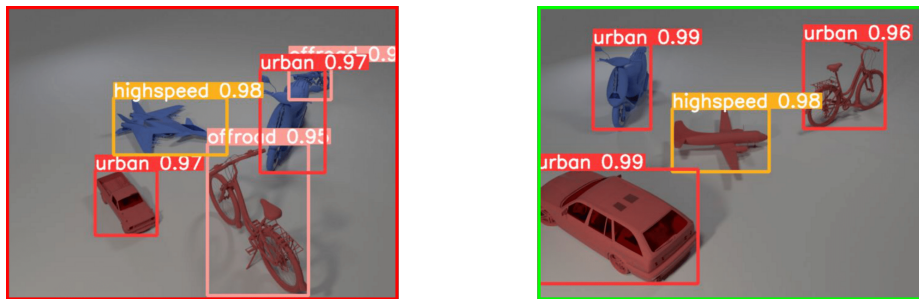



Fig. 2: The left figure shows a Super-CLEVR scene in which vehicles corresponding to the “utility bicycle” and “pickup” rare slices are misclassified by YOLOv5 into the “offroad” and “urban” classes, respectively. In contrast, the right figure shows a different scene in which the “utility bicycle” is correctly classified into its “urban” class.

Acknowledgments. The project leading to this research has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101034440. Additionally, this work was supported by funding from the Bosch Center for AI (BCAI) in Renningen, Germany. 

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chung, Y., Kraska, T., Polyzotis, N., Tae, K.H., Whang, S.E.: Slice Finder: Automated Data Slicing for Model Validation. In: 35th IEEE International Conference on Data Engineering, ICDE 2019, Macao, China, April 8-11, 2019. pp. 1550–1553. IEEE (2019). <https://doi.org/10.1109/ICDE.2019.00139>, <https://doi.org/10.1109/ICDE.2019.00139>
2. Cropper, A., Dumancic, S.: Inductive Logic Programming At 30: A New Introduction. *J. Artif. Intell. Res.* **74**, 765–850 (2022). <https://doi.org/10.1613/JAIR.1.13507>, <https://doi.org/10.1613/jair.1.13507>
3. d’Eon, G., d’Eon, J., Wright, J.R., Leyton-Brown, K.: The Spotlight: A General Method for Discovering Systematic Errors in Deep Learning Models. In: FAccT ’22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, June 21 - 24, 2022. pp. 1962–1981. ACM (2022). <https://doi.org/10.1145/3531146.3533240>, <https://doi.org/10.1145/3531146.3533240>
4. Eyuboglu, S., Varma, M., Saab, K.K., Delbrouck, J., Lee-Messer, C., Dunnmon, J., Zou, J., Ré, C.: Domino: Discovering Systematic Errors with Cross Modal Embeddings. In: The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022. OpenReview.net (2022), <https://openreview.net/forum?id=FPCMqjI0jXN>
5. Hitzler, P., Sarker, M.K. (eds.): Neuro-Symbolic Artificial Intelligence: The State of the Art, Frontiers in Artificial Intelligence and Applications, vol. 342. IOS Press (2021). <https://doi.org/10.3233/FAIA342>, <https://doi.org/10.3233/FAIA342>
6. Law, M., Russo, A., Broda, K.: The ILASP system for learning Answer Set Programs. www.ilasp.com (2015)
7. Law, M., Russo, A., Broda, K.: The ILASP system for inductive learning of answer set programs. *CoRR* **abs/2005.00904** (2020), <https://arxiv.org/abs/2005.00904>
8. Li, Z., Wang, X., Stengel-Eskin, E., Kortylewski, A., Ma, W., Durme, B.V., Yuille, A.L.: Super-CLEVR: A Virtual Benchmark to Diagnose Domain Robustness in Visual Reasoning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023. pp. 14963–14973. IEEE (2023). <https://doi.org/10.1109/CVPR52729.2023.01437>, <https://doi.org/10.1109/CVPR52729.2023.01437>
9. Muggleton, S.H.: Inductive logic programming. *New Gener. Comput.* **8**(4), 295–318 (1991). <https://doi.org/10.1007/BF03037089>, <https://doi.org/10.1007/BF03037089>
10. Oakden-Rayner, L., Dunnmon, J., Carneiro, G., Ré, C.: Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. In: Ghassemi, M. (ed.) ACM CHIL ’20: ACM Conference on Health, Inference, and Learning, Toronto, Ontario, Canada, April 2-4, 2020 [delayed]. pp. 151–159. ACM (2020). <https://doi.org/10.1145/3368555.3384468>, <https://doi.org/10.1145/3368555.3384468>
11. Redmon, J., Divvala, S.K., Girshick, R.B., Farhadi, A.: You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. pp. 779–788. IEEE Computer Society (2016). <https://doi.org/10.1109/CVPR.2016.91>, <https://doi.org/10.1109/CVPR.2016.91>