

# Position paper: On the role of abductive reasoning in semantic image segmentation

Andrea Rafanelli<sup>1,2,\*</sup>, Stefania Costantini<sup>2</sup> and Andrea Omicini<sup>3</sup>

<sup>1</sup>*Department of Computer Science, University of Pisa, Italy*

<sup>2</sup>*Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Italy*

<sup>3</sup>*Department of Computer Science and Engineering, ALMA MATER STUDIORUM—Università di Bologna, Italy*

## Abstract

This position paper provides insights aiming at resolving the most pressing needs and issues of computer vision algorithms. Specifically, these problems relate to the scarcity of data, the inability of such algorithms to adapt to never-seen-before conditions, and the challenge of developing explainable and trustworthy algorithms.

This work proposes the incorporation of reasoning systems, and in particular of abductive reasoning, into image segmentation algorithms as a potential solution to the aforementioned issues.

## Keywords

neuro-symbolic, abductive reasoning, semantic segmentation, XAI

## 1. Introduction

Semantic image segmentation is an area of computer vision that has seen significant advancements in recent years. This technique is extremely common today since it facilitates the description, categorization, and visualisation of regions of interest in a picture [1]. The primary goal of semantic image segmentation is to classify distinct portions of an image as separate objects by associating each pixel with a specific class label, and therefore recognising and understanding what is represented in the image at the pixel level.

To date, there is a plethora of applications that make use of these algorithms. Self-driving cars and medical image analysis are some of the most significant applications. In certain scenarios, such as the two reported here, there is an apparent problem of data scarcity. In general, one of the main problems is that Deep Learning (DL) algorithms require annotated data to succeed. This is extremely evident in the domain of computer vision, as it is necessary to annotate every visual element contained in an image. Consequently, there is a remarkable demand of data-sets with a large number of annotations. Furthermore, if one considers the real world, as opposed to a simulated environment, the available data are sequential, subject to continual changes, and frequently unlabelled. In this regard, semantic segmentation algorithms should continuously

---

*Discussion Papers - 22nd International Conference of the Italian Association for Artificial Intelligence (AIXIA 2022)*

\*Corresponding author.

✉ andrea.rafanelli@phd.unipi.it (A. Rafanelli); stefania.costantini@univaq.it (S. Costantini);

andrea.omicini@unibo.it (A. Omicini)

ORCID 0000-0001-8626-2121 (A. Rafanelli); 0000-0002-5686-6124 (S. Costantini); 0000-0002-6655-3869 (A. Omicini)



© 2022 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)

adapt themselves to new contexts and data, asking for a certain degree of *generalisation*. As neural network models are subjected to sparse or low-quality data, it is widely recognised that their performance degrades drastically. Therefore, it is essential to address these issues by proposing alternatives to conventional systems. For instance, via solutions employing both symbolic and sub-symbolic systems (Neuro-Symbolic architectures). Here, the apparent benefit of integrating logical constructs into sub-symbolic models is related to the reduction in learning time, the possibility to improve the prediction accuracy, the decrease in the amount of data required to train the model, and a greater propensity for the model to adapt to previously un-observed situations.

A particular type of reasoning, known as abductive reasoning, is embraced in this proposal. Abductive reasoning was studied and theorised by C.S. Peirce [2]. It is a form of reasoning in which hypotheses are developed in order to explain observations. In general, it is a process of reasoning invoked to explain an observation by inferring causes from observed effects. It can, therefore, be defined as an *explanatory process*. Abduction is a kind of non-monotonic form of reasoning in which the truth of a statement can vary when additional evidence is introduced. This reasoning is also highly adaptable to real-world circumstances, where it is frequently necessary to change some findings in light of fresh information. According to [2], humans use abduction to draw significant conclusions by focusing on a number of observations. On the basis of inadequate evidence, abductive reasoning generates hypotheses. Wherever massive data sets are not available for training a neural network model, the utility of this particular way of reasoning becomes evident.

There are three main reasons why this particular type of reasoning is proposed. The first reason relates to the explainability of the algorithm. Abduction is defined to provide the formulation of potential explanations. In sensitive domains, such as those discussed previously involving autonomous driving and medical diagnosis, it is crucial to have an explanation for algorithmic decisions. The second reason, on the other hand, concerns the problem of the lack of data and the consequent inability of the algorithm to adapt to new settings, i.e. its lack of generalisation. In fact, providing the system with a suitable number of instances is not simple, firstly because it is difficult to predict in advance which information would be significant for the learning task [3], and secondly because a big amount of data is not always available. In this sense, abduction can be used to generate additional instances for a system and compensate for the scarcity of data in a specific domain. The third reason is related with limiting the search space. Training sub-symbolic models, such as neural networks, takes not only a large amount of data, but also a significant amount of time in order to detect and learn patterns below a particular class label. The greater the number of instances, the greater the algorithm's accuracy and robustness, but the longer it takes to learn. Abduction in this sense could aid in the reduction of the search space. Given a Knowledge Base (KB) and a set of constraints, the minimal number of possible hypotheses  $\theta$  is formed. Let be  $\theta = \{H_1, H_2, H_N\}$ , where  $KB \cup H \models O$ . In practice  $H$  is an hypothesis for the observation  $O$  asserting  $O$  w.r.t the  $KB$ . Consequently, only those feasible inferences formed from  $\theta$  and deduced according to the KB are considered to be viable solutions.

In accordance with the aforementioned considerations, we believe that the employment and integration of abductive reasoning into image segmentation algorithms can lead to improvements in terms of both learning time and the amount of data required for model training. We

also believe that the potential of abduction can lead to a symbolic augmentation resulting in a final model that is more robust, explainable, and trustworthy.

## 1.1. Outline

The remainder of this article is organised as follows: the semantic image segmentation state-of-the-art is detailed in Section 2 through a discussion of the principal methodologies used and the principal limitations inherent in this discipline; the motivations for our proposal are stated in Section 3; Section 4 outlines and explains in depth the primary challenges of this work; finally, two ideas for a framework implementing our proposal are presented in Section 5.

## 2. Background

*Semantic segmentation* refers to the process of identifying which elements occur in an image and where they are located. Convolutional architectures are the most common architectures used for semantic segmentation [4]. Semantic image segmentation employs a multitude of models, the most significant of which are listed below. Fully Connected Layers (FCN) models are one of the best-known techniques [5]. These approaches are innovative since they use skip connections to connect non-adjacent layers. It is important to mention that nearly all the following methodologies used for semantic image segmentation adhere to the concept of skip connections. However, despite their widespread use, one of the major issues with these structures is their fully convolutional nature, which renders them incapable to localise labels within the feature hierarchy, and to process global context knowledge. As a consequence, the approaches involved to solve this issue diverge from the fully-connected models. Object-based techniques, of which R-CNN models are the most well-known, are one example. R-CNNs use a Regional Proposal Network (RPN) to propose unique regions inside a picture, followed by a Convolutional Neural Network (CNN) to locate items within each region. Therefore, the essential structure of this form of architecture consists of a collection of systems for determining the position of bounding boxes and the types of items that can be located within them.

Over the years, other enhancements to the original system have been proposed, including R-CNN [6], Fast R-CNN [7], Faster R-CNN [8], and Mask R-CNN [9]. In this regard, the application of object-based approaches for semantic segmentation is a field that will be considerably advanced in the near future. However, image segmentation is still a young subject with many open problems and challenges. Consequently, new ideas and models are arising. Few-Shot Semantic Segmentation [10] is an example of a technique that attempts to perform segmentation using a small number of annotated instances rather than massive data-sets; similarly, Zero-Shot Semantic Segmentation [11] attempts to build visual features using word embedding vectors when no training data are provided. Semantic segmentation is extended by instance segmentation, which offers distinct labels for independent instances of elements belonging to the same class of objects. According to [12], the problem of instance segmentation is to simultaneously discover a solution for object recognition and semantic segmentation. Although numerous techniques developed over the years have demonstrated good performance in specific situations, a notable downside is that the classifier must be retrained every time new object categories are identified, in addition to the fact that object recognition at different scales is a big

challenge in both semantic and instance segmentation. Other issues, including object occlusion and image degradation, are also highlighted.

### 3. Aims

This research seeks to solve specific difficulties in the field of semantic segmentation by examining the incorporation, with such models, of a reasoning module that might improve classification results and provide reasoning processes to the system. The introduction of a reasoning module to the chain would permit the incorporation of global context information, which would be useful for image segmentation. As also noted by [13], while neural networks are extremely effective at extracting local features and making accurate predictions with a small field of view, they lack the ability to use global context information and, therefore, cannot model interactions between different object instances directly.

The objective is therefore to introduce a form of reasoning, the abductive one, that provides the system with the ability to abstract from its predictions, thereby guaranteeing an higher degree of generalisation of the model. In fact, the reasoning module could circumvent the problem of limited data by generating abductive hypotheses to be utilised as new examples in the training phase and by providing the system with some form of commonsense reasoning that can produce responses in the presence of unexplored new scenarios. In addition, this strategy could enable the model to obtain some explainability and interpretability characteristics.

According to [2], abductive reasoning is motivated by the fact that, if the hypothesis were true, observable events would have been inevitable consequences of them. Abductive reasoning is considered as reliable when applied to real-world problems since it starts by providing initial explanations. Obviously, these explanations will require further investigation in the future to express their veracity. This form of reasoning can enable the neural network model (which is a black-box model) to provide external users with explanations of the outputs generated by its internal complex processes. Consequently, the main aim of this study is to propose a general hybrid framework for semantic image segmentation that can overcome some of the issues highlighted above and that, in the future: (i) can be applied to a variety of scenarios; (ii) can be used with a small amount of data; and (iii) is ultimately explicable, interpretable, and reliable.

### 4. Challenges

The main idea here is to integrate abductive reasoning capabilities with CNNs—specifically segmentation-specific CNNs. The main challenges we aim to solve are outlined below:

**Data scarcity** — One of the primary advantages of employing Neuro-Symbolic systems, i.e., systems incorporating the integration of symbolic and sub-symbolic systems, is the ability to use and combine the strengths of the two types of systems. Among them, it is argued that these hybrid models can be trained end-to-end with limited data. Evidently, in the case of computer vision, this strategy proves effective. In particular, one of the reasons why the combination of convolutional neural networks with a symbolic approach that

exploits abductive reasoning is recommended is the needed to attempt to solve the data scarcity issue that frequently arises in this field.

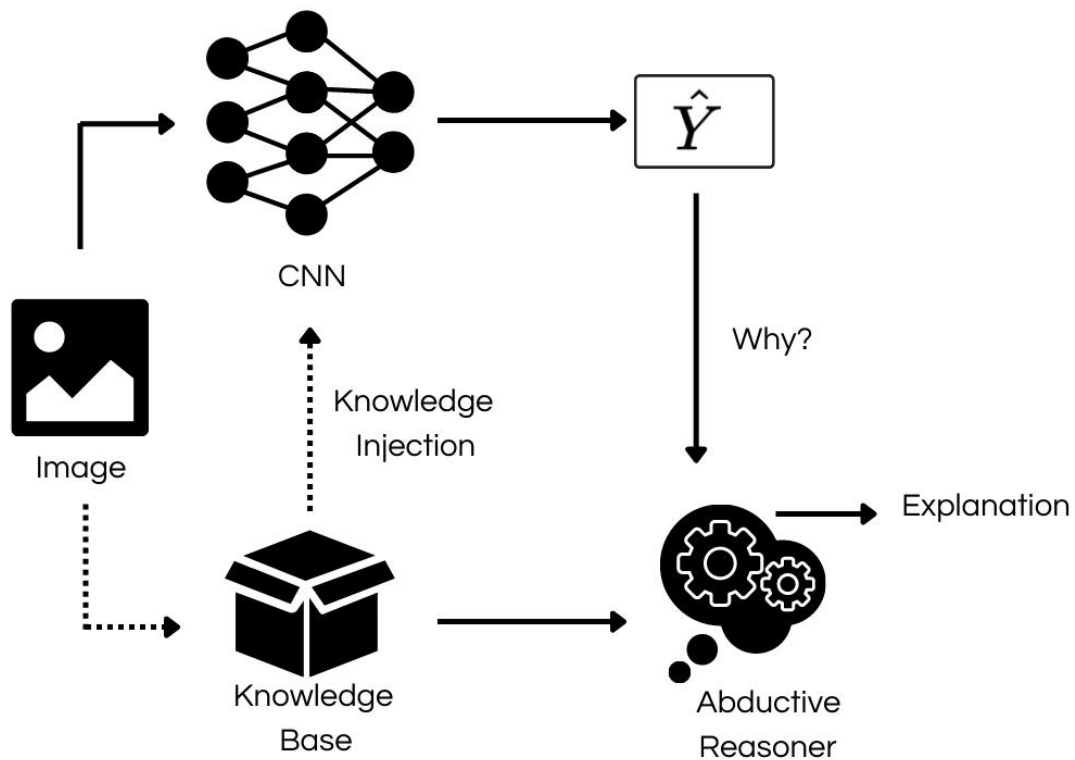
**Robustness** — In the area of computer vision, there are typically issues with image distortion or, as discussed in Section 2, issues with image occlusion, degradation, etc. If the model is not robust enough, these occurrences can result in a substantial loss of performance. Model robustness relates to system's resistance to data disturbances and specifies the extent to which findings from a sample may be applied to other samples from the population. In a way, it is desired to believe that symbolic knowledge can be of support in this regard. In fact, the use of logical terms enables the neural network to learn and reason on the image without requiring millions of samples to comprehend what lies behind a blurred or obstructed image. It has been proven that using symbolic knowledge within a DNN can bring robustness to the learning system [14]. In the specifics of this proposed integration of abductive reasoning with CNN models, this can be seen as a ways of improving the generalisability of the model by generating and introducing additional hypotheses. Such hypotheses, in practice, are used as new possible answers to be considered and explored, hence enhancing the model's ability to reason and drop into previously untested and unobserved scenarios.

**Explainability** — In recent years, the demand for explainable algorithms has increased, in part because of the regulations and guidelines provided by the EU in the GDPR. Nonetheless, there is no unanimity on how to build these explanations. The integration we propose here may be an option. The employment of a symbolic basis with a neural network can produce explanations that closely resemble how human thinking works while preserving cutting-edge performance [15]. In addition, abductive reasoning can be viewed as a means of fortifying and enhancing the explainability of a model. Abduction is frequently defined as a method for *inferring the best explanation* since it tries to make inexplicable facts understandable. In a sense, it is anticipated that the ability to employ this form of reasoning will enable human-comprehensible explanations of the algorithm's findings.

**Trustworthiness** — Multiple features contribute to the trustworthiness of artificial intelligent systems, such as explainability, robustness, transparency, and accuracy [16]. For the same reasons discussed above, it can be argued that combining abductive reasoning into neural network can give rise to reliable systems. Specifically, a system is considered trustworthy when humans have confidence in the algorithm's decisions, i.e., when they comprehend how and why certain decisions were made [17]. In this sense, system's trustworthiness depends largely on its explicability and interpretability. Abductive reasoning, as stated in the previous point, is particularly adept at finding potential solutions in the presence of incomplete or misleading information, making it ideal for real-world situations. In this sense, this type of reasoning should be incorporated into artificial intelligence systems prior to achieving trusted autonomy, which is, according to a Forbes article<sup>1</sup>, the point at which human trust artificial intelligence systems to perform complex tasks requiring

---

<sup>1</sup><https://www.forbes.com/sites/forbestechcouncil/2017/08/30/why-human-like-reasoning-is-the-key-to-trusted-ai/?sh=2dcf1ff54edf>



**Figure 1:** Knowledge injection

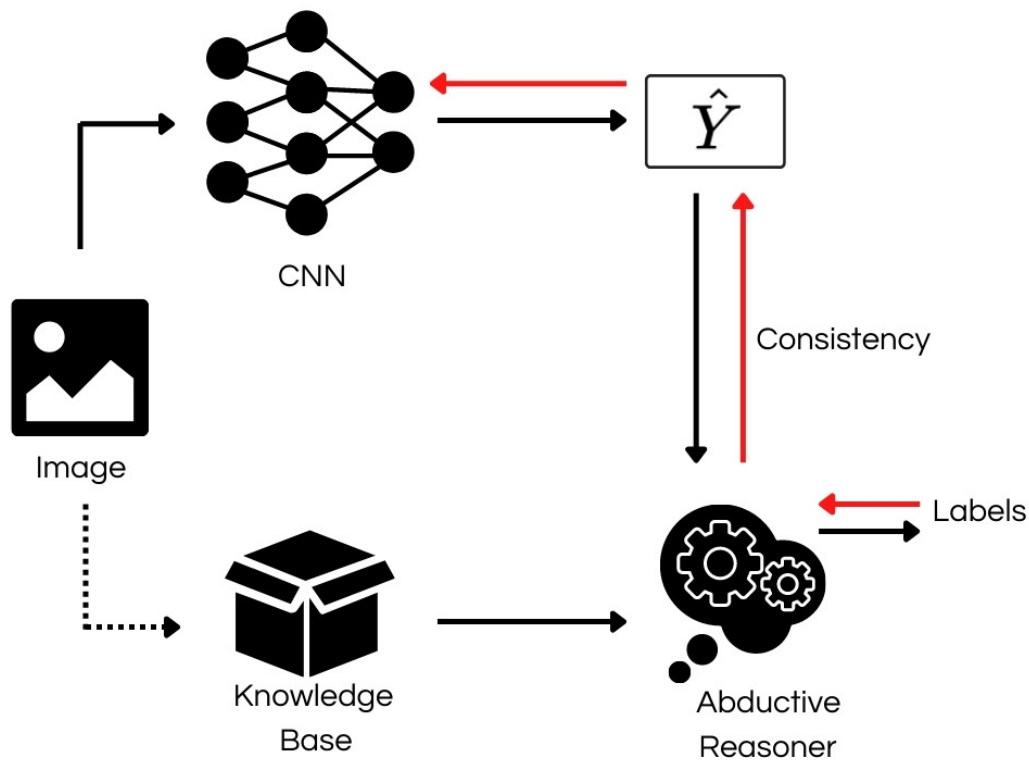
flexibility and rapid decision-making, even in the presence of incomplete or inaccurate data.

## 5. Proposal

The potential of incorporating a reasoning module – specifically, one that performs abductive reasoning – into convolutional analysis is likely to result as beneficial in light of the aforementioned limitations and requirements in the field of image segmentation. This section covers potential implementation frameworks for the development of this hybrid system. Obviously, the frameworks outlined at this time are primarily hypothetical and will require further research, implementation, and extensive experiments.

The initial phase of both frameworks is identical: the neural network receives the image as input, while concepts are taken from the image to construct a Knowledge Base (KB) to symbolically represent what the black-box has learned. Next, two techniques should be investigated:

**Knowledge injection** – The KB constrains the neural network in order to enhance its performances. The reasoner provides an explanation of the ultimate decision at the conclusion



**Figure 2:** Knowledge learning

of the procedure, based on the abductive hypotheses that were formed. In addition to providing explainability to the model, the advantage of this method is to supply symbolic information to the neural network during the training phase, hence reducing the learning time and the amount of input data (Figure 1);

**Knowledge learning** – An abductive reasoner is utilised to adjust and correct the network’s output. In this sense, revisions are generated by abduction based on the knowledge base’s contents (Figure 2). Specifically, this concept is derived from the ABL framework [18] [19], which consists of three steps: i) initially, the model  $f$  is utilised to get the symbolic predictions  $z = f(x)$  as pseudo-labels, ii) ABL then modifies the pseudo-labels  $z$  to  $\hat{z}$  via abductive reasoning. There are typically multiple candidates  $\hat{z}$  that are consistent with KB. The reasoning model deduces the most likely right pseudo-labels  $\hat{z}$  based on the idea of minimal inconsistency. iii) ABL treats the  $\hat{z}$  labels as ground-truth labels in order to modify the model. This procedure is repeated iteratively.

Given the two frameworks above, one might also consider a fusion of the two, potentially resulting in a system where both symbolic knowledge injection and a review of the final results with eventual system adjustments are possible. In truth, there are several paths worth to

investigate, and it is currently unclear which ones should be actually pursued.

## 6. Conclusion

In conclusion, this proposal highlights the potentials of combining a symbolic system that performs abductive reasoning with a sub-symbolic one in the context of semantic image segmentation. This integration may enable model training and adaptation without the need for a large number of samples. In this way, it would be possible to design a system that is very generalisable across domains by using knowledge to make up for the small number of training instances. Due to the difficulty of studying the core representations and reasoning operations of neural networks, this could serve as the foundation for support in a scenario involving explainable artificial intelligence (eXplainable AI). The resulting integrated system will produce an explainable framework intended to incorporate knowledge representation and reasoning into black-box technologies. Even in this regard, as discussed above, the use of abductive reasoning is not arbitrary in our proposal; in fact, it is one of the most well-studied forms of logic to be explicable. This is because the abductive approach identifies plausible explanations for each alternative and then selects the most probable candidate. In perspective, the conclusions that might be reached from this line of reasoning are consequently open to interpretation and explanation.

## Acknowledgments

This work has been partially supported by the European Community Horizon 2020 programme under the funding schema ERC-2018-ADG G.A. 834756 *XAI: Science and technology for the eXplanation of AI decision making*, and by the CHIST-ERA IV project “EXPECTATION” – CHIST-ERA-19-XAI-005 –, co-funded by EU and the Italian MUR (Ministry for University and Research).

## References

- [1] M. K. Kar, M. K. Nath, D. R. Neog, A review on progress in semantic image segmentation and its application to medical images, *SN Computer Science* 2 (2021) 397. doi:10.1007/s42979-021-00784-5.
- [2] F. Bellucci, A. Pietarinen, Peirce on the justification of abduction, *Studies in History and Philosophy of Science Part A* 84 (2020) 12–19. doi:10.1016/j.shpsa.2020.04.003.
- [3] F. Bergadano, V. Cutello, D. Gunetti, Abduction in machine learning, in: D. M. Gabbay, R. Kruse (Eds.), *Abductive Reasoning and Learning*, Springer Netherlands, Dordrecht, 2000, pp. 197–229. doi:10.1007/978-94-017-1733-5\_5.
- [4] F. Lateef, Y. Ruichek, Survey on semantic segmentation using deep learning techniques, *Neurocomputing* 338 (2019) 321–348. doi:10.1016/j.neucom.2019.02.003.
- [5] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (2017) 640–651. doi:10.1109/TPAMI.2016.2572683.



- [6] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587. doi:10.1109/CVPR.2014.81.
- [7] R. Girshick, Fast R-CNN, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448. doi:10.1109/ICCV.2015.169.
- [8] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in: C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 28, Curran Associates, Inc., 2015. URL: <http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks>.
- [9] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988. doi:10.1109/ICCV.2017.322.
- [10] K. Wang, J. H. Liew, Y. Zou, D. Zhou, J. Feng, PANet: Few-shot image semantic segmentation with prototype alignment, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. doi:10.1109/ICCV.2019.00929.
- [11] M. Bucher, T.-H. VU, M. Cord, P. Pérez, Zero-shot semantic segmentation, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 32, Curran Associates, Inc., 2019. URL: <http://papers.nips.cc/paper/8338-zero-shot-semantic-segmentation>.
- [12] F. Fan, X. Zeng, S. Wei, H. Zhang, D. Tang, J. Shi, X. Zhang, Efficient instance segmentation paradigm for interpreting SAR and optical images, *Remote Sensing 14* (2022). doi:10.3390/rs14030531.
- [13] M. Teichmann, R. Cipolla, Convolutional CRFs for semantic segmentation, in: 30th British Machine Vision Conference 2019 (BMVC 2019), BMVA Press, 2019, p. 142. URL: <https://bmvc2019.org/wp-content/uploads/papers/0865-paper.pdf>.
- [14] A. S. d'Avila Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, S. N. Tran, Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning, *Journal of Applied Logics* 6 (2019) 611–632. URL: <https://collegepublications.co.uk/ifcolog/?00033>.
- [15] R. R. Hoffman, W. J. Clancey, S. T. Mueller, Explaining AI as an exploratory process: The peircean abduction model, *CoRR abs/2009.14795* (2020). URL: <https://arxiv.org/abs/2009.14795>. arXiv:2009.14795.
- [16] M. Harbers, R. Verbrugge, C. Sierra, J. Debenham, The examination of an information-based approach to trust, in: J. S. Sichman, J. Padget, S. Ossowski, P. Noriega (Eds.), *Coordination, Organizations, Institutions, and Norms in Agent Systems III*, volume 4870 of *Lecture Notes in Computer Science*, Springer, 2008, pp. 71–82. doi:10.1007/978-3-540-79003-7\_6.
- [17] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nature Machine Intelligence* 1 (2019) 206–215. doi:10.1038/s42256-019-0048-x.
- [18] W. Dai, Q. Xu, Y. Yu, Z. Zhou, Bridging machine learning and logical reasoning by abductive learning, in: H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019*

(NeurIPS 2019), 2019, pp. 2811–2822. URL: <https://proceedings.neurips.cc/paper/2019/hash/9c19a2aa1d84e04b0bd4bc888792bd1e-Abstract.html>.

- [19] Y.-X. Huang, W.-Z. Dai, L.-W. Cai, S. H. Muggleton, Y. Jiang, Fast abductive learning by similarity-based consistency optimization, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan (Eds.), *Advances in Neural Information Processing Systems*, volume 34, Curran Associates, Inc., 2021, pp. 26574–26584. URL: <https://proceedings.neurips.cc/paper/2021/hash/df7e148cabfd9b608090fa5ee3348bfe-Abstract.html>.