Introduction

To Cinzia, Danilo and Federico,
My family, My safe harbour.
And to Irene. Of course to Irene.

In this manuscript, I resume the most important works that I conducted with my supervisor and other colleagues during the PhD studies. The leitmotif connecting these works is the idea of merging symbolic reasoning, generally represented by logic rules, with machine perception, usually a neural network. The creation of a hybrid model allows tackling some of the most critical issues of deep learning. Among several possible choices, in these works we followed the approach of representing First-order-Logic (FOL) rules as logic constraints. In this way, we can exploit the available domain knowledge to enhance neural networks (both at training and at evaluation time), within the paradigm of Learning From Constraints. On the other side, when no knowledge is available, we can extract it from a deep neural network by analysing which logic constraints are satisfied, following a Learning Of Constraints approach. This also allows explaining the predictions of a neural network, otherwise commonly considered black-box.

This thesis is organized as follows. In Chapter 1, we briefly summarize what Deep Learning is with its strengths and, most importantly, its issues. A formal background on how to combine it with logic constraints is given in Chapter 2. The following chapters focus on the Learning From Constraints approach and how it can be exploited to tackle two different deep learning issues. Chapter 3 presents a way to minimize the number of supervisions required to train a model, in the context of active learning. Chapter 4 describes how we can defend a neural network from adversarial attacks by enforcing logic constraints. In the second part of the thesis, we show

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

two possible approaches to Learning Of Constraints. The first focuses on explaining a given classifier by means of another network (see Chapter 5). The latter introduces a novel neural layer which allows building explainable-by-design networks (see Chapter 6). and checking their violation. A resume of the experimental chapters is provided in Chapter 7, together with possible future work. The works resumed in this thesis have been published (or submitted) in journals like Neural Networks, IEEE TNNLS, Artificial Intelligence, Neurocomputing and IEEE TPAMI, or international conferences such as IJCAI, AAAI, IJCNN and ECML.