## COMPUTER AND CONTROL ENGINEERING

### Neural-symbolic scene integration & generation

| Funded By | Dipartimento DAUIN  
Centro Interdipartimentale SmartData@PolI.TO |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisor</td>
<td>LAMBERTI FABRIZIO - <a href="mailto:fabrizio.lamberti@polito.it">fabrizio.lamberti@polito.it</a></td>
</tr>
</tbody>
</table>
| Contact | LAMBERTI FABRIZIO - fabrizio.lamberti@polito.it  
MORRA LIA - lia.morra@polito.it |

### Context of the research activity

Deep neural networks (DNNs), despite achieving unprecedented success in most computer vision tasks, lack desirable properties including top-down control, transparency, and generalization to unseen data. Neural-symbolic techniques combine DNNs for (sub-symbolic) representation learning with (symbolic) Knowledge Representation and Reasoning techniques. To achieve this, Knowledge Base (KB) is grounded (i.e., mapped) in a tensor space. These techniques are especially suited to visual tasks that combine efficient pattern recognition with higher cognitive capabilities, such as scene graph generation (SGG), event detection in videos, and procedural generation of 3D scenes. Besides the ability to explicitly represent and manipulate abstract concepts in the KB, neural-symbolic architectures can encode prior knowledge, e.g., logical axioms, which improves learning from small-scale, biased, incomplete, or conflicting reference standards, as is often the case in SGG. Despite their potential, neural-symbolic architectures have reached limited diffusion so far, due to issues related to scalability, and lack of demonstrated applications. The objective of this Ph.D. project is thus to improve the scalability and ease of train of neural-symbolic architectures and investigate their application on large-scale benchmarks and realistic tasks, including image classification or scene graph generation, and procedural generation of 3D and image content.

### Objectives

Neuro-symbolic techniques have already been proposed for semantic image interpretation. The neural-symbolic component is typically placed on top of representations learnt by DNNs: for instance, SGG architectures include an object detector, followed by a visual relationship detection module. Preliminary work has proven the feasibility of jointly training neural-symbolic and representation learning components. Yet, research remains limited on how to effectively unify the two components for various tasks and find the best trade-off between modularity and end-to-end training (in terms of efficiency and accuracy). Additionally, scaling to large datasets remains challenging. For instance, many methods rely on a predefined KB which includes declarative and prior knowledge in the form of logical axioms. As only groundings are learnt from data, a new KB must be manually defined for each dataset. To progress towards more scalable and largely applicable techniques, the following objectives are identified:
- Develop techniques and tools to design general purpose KBs that can be shared among multiple tasks or datasets.
- Design and validate neural-symbolic architectures for semantic image interpretation.
- Design and validate generative models to generate an image from a structured latent model (e.g., scene graphs): solving the inverse problem of mapping the scene graph back to the image domain is an important stepping stone towards allowing semi-supervised. Besides, it has several important practical applications, e.g., for the automatic generation of computer graphics content.
- Investigate the feasibility of learning the KB base in an unsupervised or semi-supervised fashion from raw data.

| Skills and competencies for the development of the activity | Deep learning e computer vision. |