Semantics-aware image understanding

Candidate: Andrea Pasini
Supervisor: Prof. Elena Baralis
Co-supervisor: Prof. Benoit Huet

Politecnico di Torino

Abstract

Deep learning models are characterized by high complexity and low interpretability, which are the payload for obtaining precise results in difficult tasks such as image understanding. Moreover, these models may suffer from an inadequate semantic understanding of the input data, as they are typically focused on a limited task (e.g., classifying images). Conversely, the human brain can learn from different real world activities and derive a more complete semantic knowledge. In this thesis, we design a methodology for inferring semantic knowledge directly from images, with the aim of enhancing image understanding tasks. Our work focuses on the study of object relationships, such as relative position and size, to infer a better semantic understanding of the analyzed pictures.

Our research provides a first application called SAD, a Semantic Anomaly Detection method to identify anomalies in the predictions of semantic segmentation neural networks. Semantic object relationships are exploited to derive an interpretable knowledge base, describing common configurations of normal objects. Our methodology highlights potential classification errors made by a neural network by identifying uncommon object relationships according to the learned knowledge base. The detected anomalies are presented to the user in an interpretable way, facilitating the analysis of the neural network accuracy.

Afterwards, we present SImS (Semantic Image Summarization), a framework designed to summarize big image collections. This task finds applications such as providing previews of personal albums (e.g.,

Google Photos) or suggesting thematic collections based on user interests (e.g., Pinterest). These objectives require a complete semantic understanding of the image collection, as simple visual features and textual tags would not be sufficiently informative. To achieve this goal, we propose a technique based on frequent subgraph mining, which analyzes scene graphs. These data structures are automatically derived from the images by our algorithm and allow a complete representation of the image content. The output summary consists of a set of frequent scene graphs describing the underlying patterns of the collection. These results are more interpretable and provide more interesting descriptions with respect to previous techniques. Moreover, in the experimental results we show that our patterns achieve high summary quality in terms of coverage and diversity.