Politecnico di Torino
Dipartimento di Automatica e Informatica

PhD in Computer and Control Engineering
XXXV cycle

Supervisor
Fabrizio Lamberti

Human-Centered Machine Learning Applications

PhD Candidate: *Fabio Garcea*

1. Introduction

Deep Learning (DL) models called Convolutional Neural Networks (CNNs) have become the de-facto standard approach to tackle Computer Vision problems. Unfortunately, DL models are black-boxes and it is hard for humans to understand the rationale behind the model decision process. Many research areas try to provide more insights on neural network decisional processes, spanning from the analysis of these network with Explainable-AI (XAI) techniques to an analysis of their behavior from a more data-oriented perspective.

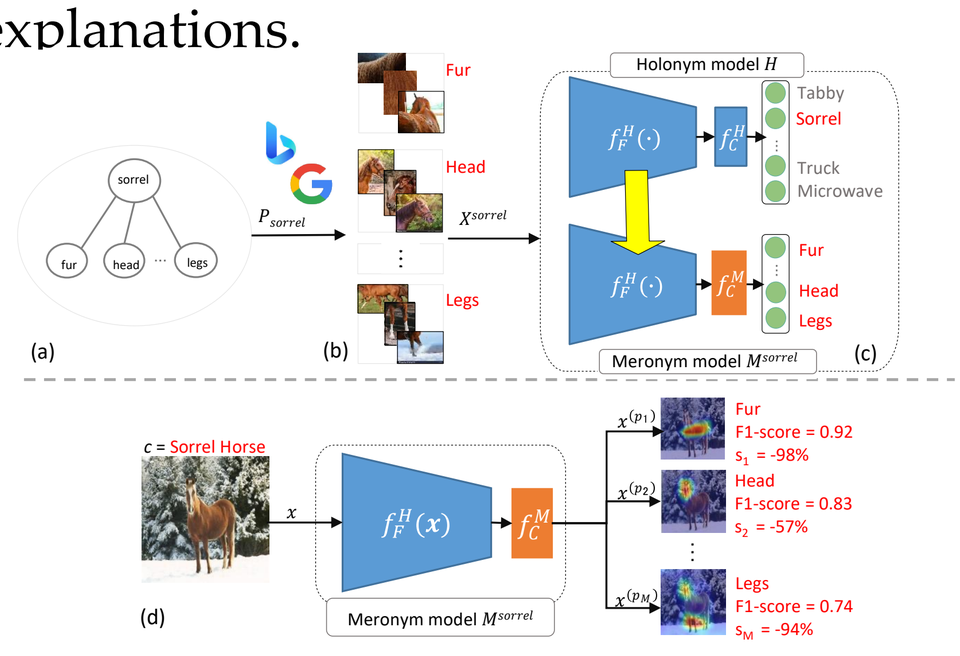
2. Objectives

The objectives of this research focus on two main pillars: (1) studying methods that provide new insights on DNNs, (2) analysing network behaviour in presence on data related phenomena such as concept drift.

3. Methodologies

3.1 Innovative XAI techniques

An innovative methodology leveraging ontologies and the holonym-meronym relationship has been proposed to provide post-hoc model explanation in the form of part-based attention maps, thus taking a step further with respect to standard label-level explanations.

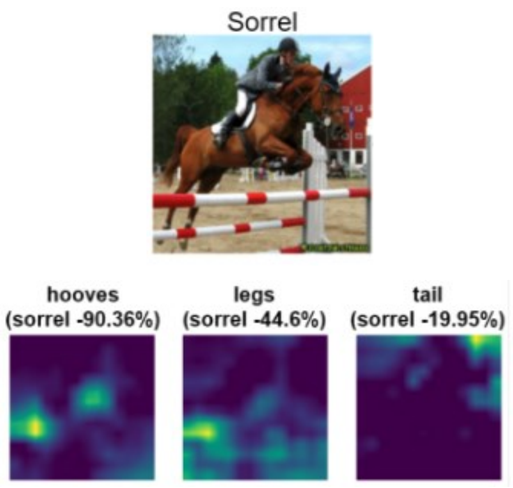


3.1 Concept Drift Detection

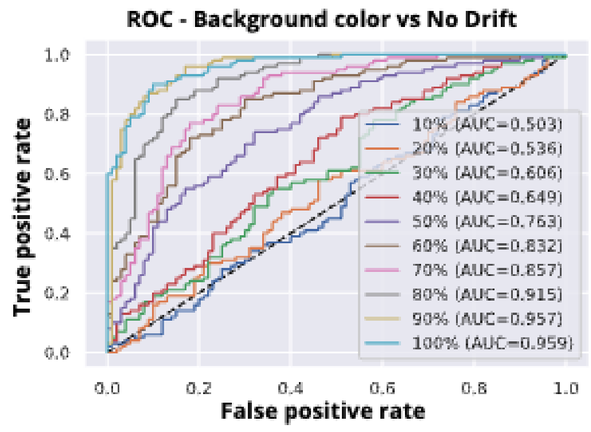
Concept drift occurs whenever the statistical properties of the output variable(s) that the ML model is trying to predict evolve over time. In this research a novel methodology for unsupervised drift detection was proposed to detect and monitor the occurrence of drift in the context of synthetically generated documents. The proposed methodology can identify drift and correlates well with the performance degradation of the model.

4. Results

The proposed methods allow to glimpse richer explanations in one case, and to efficiently detect drift in the data in the other.



Per-part scrore drop evaluation made by HOLMES for an image of a sorrel described in Section 3.1.



AUC values at different levels of drift injection. Predictions determined through the Hellinger distance as described in Section 3.2.

5. References

1. Piano, Luca; Garcea, Fabio; Gatteschi, Valentina; Lamberti, Fabrizio; Morra, Lia Detecting drift in deep learning: A methodology primer. In: IT PROFESSIONAL. ISSN 1520-9202
2. Garcea, Fabio; Famouri, Sina; Valentino, Davide; Morra, Lia; Lamberti, Fabrizio iNNvestigate-GUI - Explaining neural networks through an interactive visualization tool /. 294(2020), pp. 291-3Workshop on Artificial Neural Networks in Pattern Recognition (ANNPR 2020), Winterthur, Switzerland