

Borsa di studio attivata ai sensi di quanto disposto dal D.M. n. 1061 del 10/08/2021

Titolo del progetto: Estrazione di informazioni per il monitoraggio ambientale da serie temporali di immagini satellitari con metodi matematici avanzati di deep learning e di teoria dell'informazione.

La borsa sarà attivata sul seguente corso di dottorato accreditato per il XXXVII ciclo:
MATEMATICA

Responsabile scientifico: Emanuele Caglioti
Area per la quale si presenta la richiesta: GREEN
Numero di mensilità da svolgere in azienda: 12
Azienda: e-Geos S.p.A. (an ASI/Telespazio company)

Progetto di ricerca:

Il numero sempre maggiore di missioni satellitari di osservazione della Terra, con sempre migliori risoluzioni temporali, spaziali, e spettrali, insieme al continuo incremento della potenza dei computer, rendono necessario e possibile l'utilizzo di algoritmi sempre più efficaci per l'estrazione dell'informazione dalla grande mole di dati disponibili per il monitoraggio dell'ambiente. In particolare, nuove tecnologie di elaborazione dati sono necessarie per l'utilizzo congiunto e ottimale, dell'informazione spaziale, spettrale e temporale contenuta nei dati allo scopo di migliorare la ricostruzione dei parametri di interesse.

Tecniche di intelligenza artificiale come il deep learning si sono dimostrate molto efficaci per diverse applicazioni ma normalmente richiedono molti dati con verità (labels) per la fase di addestramento (training). Per superare questo problema, sono state sviluppate tecnologie denominate unsupervised o self-supervised learning, che permettono di addestrare reti neurali profonde su dei casi "giocattolo" (pretext tasks) dove il risultato da ottenere è noto in quanto il pretext task è ricavato da opportune manipolazioni dei dati stessi. Uno degli pretext task più efficaci per l'addestramento si basa su compressione e decompressione dei dati (architetture autoencoder).

In questo contesto, ci proponiamo di esplorare l'utilizzo di tecniche e di concetti della teoria dell'informazione per migliorare gli approcci di self-supervised learning. La teoria dell'informazione offre strumenti molto importanti al dominio del machine learning e deep learning. Esempi importanti, a seconda delle varie architetture, sono le funzioni di costo che vengono minimizzate nel processo di addestramento come la mutua informazione (denoising autoencoder) o l'entropia relativa o Kullback-Leibler divergence (variational autoencoder).

Inoltre, la teoria dell'informazione, e in particolare le tecniche di compressione, forniscono anche gli strumenti per un approccio efficace allo sfruttamento congiunto dell'informazione temporale, spaziale e temporale contenuta nei dati. Le tecniche sviluppate quindi verranno sperimentate su dati e casi d'uso reali per il monitoraggio ambientale. Possibili esempi sono:

- serie temporali di dati ottici (Sentinel-5) per il monitoraggio di aerosol e inquinanti nell'atmosfera
- serie temporali di dati SAR e ottici multispettrali (Sentinel-1 e Sentinel-2) per analisi delle variazioni della copertura del suolo (permeabilità, deforestazione, erosione coste, etc.)
- serie temporali di dati SAR ad alta ed altissima risoluzione (Sentinel-1, COSMO SkyMed, Cosmo Second Generation) per analisi stabilità di infrastrutture e suolo.
- immagini iperspettrali (missione PRISMA) per la rilevazione di materiali inquinanti (eternit in aree urbane, plastiche in acque, anomalie in aree ambientali protette, etc.)

Titolo del progetto (inglese): Extraction of information for environmental monitoring from time series of satellite images with advanced mathematical methods of deep learning and information theory.

Progetto di ricerca (inglese):

The increasing number of Earth observation satellite missions, with ever better temporal, spatial, and spectral resolutions, together with the continuous increase in computer power, make it necessary and possible to use increasingly effective algorithms for the extraction of information from the large amount of data available for monitoring the environment.

In particular, new data processing technologies are necessary for the joint and optimal use of the spatial, spectral and temporal information contained in the data in order to improve the reconstruction of the parameters of interest. Artificial intelligence techniques such as deep learning have proved very effective for different applications but normally require a lot of data with truth (labels) for the training phase. To overcome this problem, technologies called unsupervised or self-supervised learning have been developed. which allow to train deep neural networks on "toy" cases (pretext tasks) where the result to be obtained is known as the pretext task is obtained from appropriate manipulations of the data themselves. One of the most effective pretext tasks for training is based on data compression and decompression (autoencoder architectures). In this context, we aim to explore the use of information theory techniques and concepts to improve self-supervised learning approaches.

Information theory offers very important tools to the domain of machine learning and deep learning. Important examples, depending on the various architectures, are the cost functions that are minimized in the training process such as mutual information (denoising autoencoder) or relative entropy or Kullback-Leibler divergence (variational autoencoder). In addition, information theory, and in particular compression techniques, also provide the tools for an effective approach to the joint exploitation of the temporal, spatial and temporal information contained in the data. The techniques developed will then be tested on real data and use cases for environmental monitoring. Possible examples are: - time series of optical data (Sentinel-5) for monitoring aerosols and pollutants in the atmosphere - time series of SAR and optical multispectral data (Sentinel-1 and Sentinel-2) for analysis of land cover variations (permeability, deforestation, coastal erosion, etc.) - time series of high and very high resolution SAR data (Sentinel-1, COSMO SkyMed, Cosmo Second Generation) for stability analysis of infrastructures and soil. - hyperspectral images (PRISMA mission) for the detection of polluting materials (eternit in urban areas, plastics in water, anomalies in protected environmental areas, etc.)