Assignment #2: Action Recognition From Still Images Using Deep Learning Networks

- The deadline for the delivery of this exercise is **Jan 03th, 2025**

- A zip file with the source codes in Python and a PDF file must be sent.

- The code files and the results should be exported from Visual studio code, Google Colab or Jupyter. If not, include the necessary instructions files to understand the program.

- A detailed documentation in PDF is required

- The submission of the source and documentation must be done using Moodle.

Description:

Action recognition, the ability to identify and categorize human actions from visual data, has been a long-standing challenge in the field of computer vision. Traditionally, this task has been tackled using video footage, where the temporal information provided by consecutive frames allows for a more robust understanding of the action's dynamics. Recent advances in deep learning have enabled action recognition to be achieved with impressive accuracy using still images, even in challenging conditions.

Indeed, everyday human actions like "climbing," "fishing," or "phoning" can also be effectively described in still images. Furthermore, certain actions captured in videos, such as "taking photos," are inherently static and may require recognition methods solely based on static cues. Driven by the potential implications of recognizing actions in still images and the comparative neglect of this problem in computer vision, this assignment delves into the recognition of human actions utilizing a single photograph.

Walking the dog   Watching TV   Riding horse   Using a computer   Washing dishes   Rowing boat   Running

For this project, the accompanying dataset encompasses a training set and a test set, encompassing actions across 40 distinct categories. The Stanford 40 Action Dataset comprises images depicting individuals executing 40 different actions. For each image, we provide a bounding box surrounding the person performing the action, as indicated by the image's filename. The dataset comprises 9532 images in total, with 180-300 images per action category. The dataset is attached to this file for your convenience. You can also download the dataset from:

http://vision.stanford.edu/Datasets/40actions.html

We aim to develop automatic systems to predict actions from still images containing:

- A data loader to read the training and testing sets from the Stanford 40 dataset.
- An end-to-end model built from scratch (customized CNN network). You have to find empirically the number of the required layers and the structure of each layer.
- An end-to-end model built on pre-trained deep learning networks (e.g. ResNet, Googlenet, VGG, MobileNet) to classify the 40 human actions.

Requirements:

- Implement the action recognition algorithm from still images in Python.
- Train an end-to-end model with the train set of the dataset.
- Validate the implemented algorithms with the test set of the dataset.
- Write a report in which you describe and discuss the proposed algorithms' results indicating the best model that can be used to solve the human action from a still image.

Deliverables [Due 03/01/2025 - 23:59:00]:

- A comprehensive report documenting the methodology, algorithms, and techniques employed in the automatic pieces detection system.

- Codebase in Python (using OpenCV, matplotlib, etc..) with clear comments and documentation to facilitate understanding the flow chart of the code.

- Visual outputs demonstrating the effectiveness of the detection system on sample thermographic images in the shared zip file (i.e., including 7 images).