Vademecum for CNNs projects

Preliminary rules

1. Note that independence and autonomy in solving the project assignment is a matter of evaluation. Please avoid to ask questions about implementation issues. This kind of problems must be solved by yourselves, you can ask but such questions are kept into consideration for the final evaluation or even ignored.
2. All the assignment are mandatory, each variation or change must be previously discussed.
3. For the final debate you must provide i) a report (pdf); ii) a set of slides used during discussion; iii) code of the complete pipeline. The required documentation is explained in the following.
4. During oral discussion any kind of questions related to the course topics might be asked.

Required documentation

Must be submitted 3 days before exam date

i. ***Report***: such document must be organized with the following format:
   1) Description of the assignment of the project and the focus of your work;
   2) Theory about CNNs focused on the application outline, providing references;
   3) Detailed description of the method you used (describe most salient aspects of your code too) ; 4) Detailed description of the results (provide graphs, tables, etc.); 5) Results discussion; 6) Future development.
ii. ***Slides***: must be used during oral discussion. Follow the structure of your report.
iii. ***Code***: all the code you used must be provided. Carefully comment it and keep it well organized. Such criteria are matter of evaluation.

Guide for cluster usage

Is MANDATORY to read carefully the documentation at http://hpc.polito.it/download.php in order to fully understand the cluster usage and HOW TO PARALLELIZE. The gain and the usage of such knowledge is an **assignment of the project**, therefore no further information will be provided in addition to what is described in this document. Before running your code on the cluster, carefully check and evaluate it on your local machine (eventually on a very little dataset.) Pay attention to the command you run on the cluster.

A) Instructions for virtual environment creation and management. In brackets a brief explanation for each instruction.

module load intel/python/**3.5** `(python loading)`
virtualenv *virtual_env_name* `(make virtual environment)`
source *virtual_env_name*/bin/activate `(virtual environment activation)`

Install through pip the following packages (pip install *package_name or via -r option*). Please make sure of virtual environment activation before doing *pip*

numpy
matplotlib
pandas
keras
tensorflow*
scikit-learn
Pillow
opencv-python
scikit-image
openslide-python**

N.B. All these modules can be installed via a "*requirements file*" for pip (see pip documentation).

* the default version of tensorflow installed via pip command is the one for CPU computing. On casper cluster you can use only the CPUs (GPUs are not available). CNNs training, if requested in project assignment, will require quite long computational time.

** this library is suited for histological WSIs management (if required in project assignment).

B) Example of .sbatch file that is MANDATORY to submit a job

Please read the whole documentation at
http://hpc.polito.it/download.php

#!/bin/bash
#SBATCH --job-name=*job_name*
#SBATCH --mail-type=ALL
#SBATCH --mail-user=*name.surname*@studenti.polito.it
#SBATCH --partition=global
#SBATCH --time=*hh:mm:ss*
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=32
#SBATCH --output=*job_name*.log
#SBATCH --mem-per-cpu=1024M

module load intel/python/3.5
source */path/to/virtual_env/*bin/activate

python3 -u *script_name.py*

To submit, delete or monitor a job please refer to the above-mentioned online documentation.