

PyTorch/Matlab Project

- Topics eligible for the projects are:
- Protein simulation, both in oncology and <https://en.wikipedia.org/wiki/Bacteriophage>;
- Methods for representation of DNA / RNA string for feeding a neural network (either CNN or 1D input networks);
- Creation of artificial images (i.e. data augmentation step) for Lymphoma diagnosis;
- Image filtering to improve Plankton classification;
- Create artificial examples (i.e. data augmentation step) to train a bioacoustics sound classification system (bird classification);
- Development of an activation function for deep network;
- 1D descriptor for representing audio file classification;
- Develop a loss function to segment polyps in colonoscopies;
- Develop a method to encode a RGB image starting from a multi-channel image for the identification of planktic foraminifera;
- Develop multilabel neural network topology to predict drug side effects.

Python lecture notes: https://mega.nz/file/xBQhyJjJ#xrNlMjss16l3zY9l8QuKtSCDML7DEguv_H1i5JhaWs

Numpy/PyTorch materials are available at <https://mega.nz/file/IdZTSaaL#nPogI0TAv1mTO-3DtmNL6OuVe5ObdsXiTIojvxZ1Ans>

- Each project can be done by one->three people (no more), if multiple projects are based on the same idea and the code is almost identical they will be invalidated until proven otherwise. If in discussion among you, you have a similar idea and want to test it in different groups, each one will have to proceed with the implementation/test independently.
 - Ideas can obviously be taken from the scientific literature, it is important that they are NOT those of the reference papers / tools (see next lecture notes).
- In the next pages each project is detailed, for them baseline Matlab code is already available, for PyTorch project you can read it as baseline

For any question related to the PyTorch project, please contact the tutor daniel.fusaro@phd.unipd.it.

For any question related to the matlab project, write to me loris.nanni@unipd.it.

For both projects, please send me a single .rar/zip file containing the final paper, the code and the README to use the code.

Further notes

- **PyTorch** project is MANDATORY:

PyTorch, it must be in before exam date

- **Matlab** project is NOT mandatory

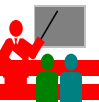
The Matlab work can be delivered when you want, even after exam date. In this case, immediately before the exam, you must notify me not to record the grade, but to wait for the delivery of the project (then you will have to register for the next session so that I can record the grade).

You have to submit both code and short paper that describe your work, example of short paper:

https://mega.nz/file/gBxGUJhI#-62kPJFJveZnkBUdD1BWI6vu-lWNFUdQK_J1IfIIPV8

- BOTH the projects are evaluated [1-3], this evaluation will be added to the final mark, only if the exam will be ≥ 18 . Also the ease of reading the developed code will be part of the vote (e.g. well commented code) and the formalization of the method in a rigorous way.
- If you submit **both PyTorch and Matlab** projects and the **topics** of these two projects are **different** then you can write two **short papers**. If you want to work in the **same topic** using both **PyTorch and Matlab** you must implement **different ideas** in matlab and PyTorch, moreover you have to **submit a standard scientific paper and not a short one**.
- If you were one of **my students in the AI course**, and you had done the optional project, you can also choose the same topic, but you have to test a different approach than that proposed in the project submitted for the AI course.
- Many **datasets are available as .mat file**, you can read them in Matlab, using python there are several tools for read .mat files, e.g. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.io.loadmat.html>
- Dataset in the .mat file are often stored in **cell array**, for **python** project you have to **manage them** in your PyTorch environment:

<https://stackoverflow.com/questions/40363277/import-matlab-cell-array-into-python-for-scikit-learn>



Loss Function - Segmentation

- To implement a new loss function for a segmentation network
- Application: Polyp detection in colonoscopy <https://arxiv.org/abs/2112.12955>
- Training set: <https://zenodo.org/record/5834916#.YfcOp-rMI2w> polyp dataset-> only the training set without data augmentation
- Test sets: <https://zenodo.org/record/5579392#.YdzC1f7MKUk>
- Matlab baseline code:
https://mega.nz/file/xUp3TCZC#yqo9UxuK7gt8hY3zaVp1gct3iV_T_X8myWNEIsiUIW0
trainSegmentationPreTrained.m main file to run the tests
designDeepLabV3plus.m file to update for replacing the loss, see rows 22-40
the folder \NewLoss stores some examples of loss functions (<https://arxiv.org/abs/2112.12955>)
Be careful, the data have to be managed as darray (inside loss function), in this way matlab calculates automatically the gradient

PyTorch: modify loss of the following network <https://github.com/james128333/HarDNet-MSEG>

Moreover, see

<https://mega.nz/file/8EhU1LYR#jyhfwQPuJHEpRN84KpcE5qaEwHrwspYJfPkUXDngOYA>
for a further help for modifyng the loss of HardNet

