SF Lizard Project

This project was done for the final thesis of the Master MSE in data science of Léonard Favre.

The goal of this project is to improve classification of colonic nuclear instance using graph neural networks.

Results

The proposed pipeline first segments the image with the Stardist model, then improves the classification with a custom graph neural network using graphSAGE layers.

The result of the pipeline on the test dataset is visible here: metric_report.

Installation

Before installing this project, gcc compiler is mandatory.

To install it, run the following command:

```
sudo apt install build-essential
```

Git-LFS is also required in order to properly downloads large files from the repository. To install it, run the following commands:

```
curl -s https://packagecloud.io/install/repositories/github/git-
lfs/script.deb.sh | sudo bash
sudo apt install git-lfs
```

More info on Git-LFS: https://git-lfs.com/

It is recommended tu use a python environement to install and run the project.

To do this using miniconda, run the following command:

```
conda create -n sflizard python=3.9
```

The python version used during this project was 3.9.16

To install this project, run the following command:

```
pip install -e .
```

After this initial installation, it is mandatory to install PytorchGeometric. This installation depends on the CUDA version installed on your system. This code was develloped and tested with the version 11.7 of CUDA. To install with the 11.7 version of CUDA, run the following command:

```
pip install pyg-lib torch-scatter torch-sparse torch-cluster torch-spline-conv torch-geometric -f https://data.pyg.org/whl/torch-1.13.0+cu117.html
```

For other configuration, please visit the dedicated page of the library: https://pytorch-geometric.readthedocs.io/en/latest/install/installation.html

Wandb logging

The Wandb tool was used to monitor the trainings of this project. The code responsible for logging with wandb has been commented to improve code portability. To enable again the wandb logging, uncomment the code in the following files:

- sflizard.training.py
- sflizard.stardist_model.stardist_model.py
- sflizard.graph_model.graph_model.py

Install wandb library with the following command:

```
pip install wandb
```

For more informations about wandb, and to configure an account, please visit https://wandb.ai

Usage / Reproduction

Reproduce test results with Stardist & graph:

1. Download the data
    - Lizard dataset:
2. Extract the data with sflizard.data_utils.data_extraction.py (PKL mode)
3. Run test pipeline with sflizard.run_test_pipeline.py

Reproduce test results on Hovernet:

1. Download the data
    - Lizard dataset:
2. Extract the data with sflizard.data_utils.data_extraction.py (PKL and FILE mode)
3. Compute output data for HoverNet
    - Compute data like this :

```
python external_models/hover_net/run_infer.py --gpu='1' --
nr_types=7 --model_path=weights/hover_net_trained.tar --
model_mode=original --type_info_path=type_info_lizard.json tile -
-input_dir data/Lizard_dataset_split/patches/Lizard_Images_test -
-output_dir external_models/output/Lizard_test_out
```

- Run the segmentation metric tool on hovernet data:

```
python sflizard/pipeline/run_segm_metric_on_hovernet.py
```

Run training:

1. Download the data
   - Lizard dataset:
2. Extract the data with sflizard.data_utils.data_extraction.py (PKL mode)
3. Run train with sflizard.training.py

Full documentation

The full documentation is available here.