1. Cách run inference cho vùng quan tâm.

   File chính: roi_inference/run_pipeline.py
   Các files liên quan:
   - *roi_inference/download_data.py hoặc roi_inference/download_data_fixed.py*
   - *roi_inference/extract_data.py*
   - *roi_inference/merge_and_process.py*
   - *inference_emsemble_roi.py* (file này nằm ngoài folder *roi_inference* để có thể thuận tiện gọi lớp định nghĩa model DenseSM trong file *DenseWideNet.py*).
   - *roi_inference/prediction_visualize.py*

   Câu lệnh chạy:
   Ví dụ:
   python roi_inference/run_pipeline.py --region ngocnhat2 --start_date 2024-12-25 --end_date 2025-02-05 --download --extract --process --inference --visualize.

   Đầu vào:
   region: Tên vùng (folder) lấy dữ liệu,
   Nếu đang dùng file grid thì đưa trên phiên hiệu vào chương trình sẽ tự động lấy thông tin tọa độ từ file roi_inference/Grid_50K_MatchedDates.geojson.
   Hoặc có thể đưa đường dẫn shapefile vào để xác định vùng.

   start_date : Ngày bắt đầu lấy dữ liệu dựa trên Sentinel-1.

   end_date : Ngày cuối cùng lấy dữ liệu dựa trên Sentinel-1.

   Các bước trong pipeline, các bước này có thể chạy riêng lẻ. Ví dụ đã download và extract dữ liệu rồi thì có thể chạy 3 bước còn lại:
   python roi_inference/run_pipeline.py --region ngocnhat2 --start_date 2024-12-25 --end_date 2025-02-05 --process --inference --visualize.

   download : Gọi file *roi_inference/download_data.py*, thực hiện tải dữ liệu.

   extract : Gọi file *roi_inference/extract_data.py*, thực hiện trích xuất dữ liệu thành csv từ ảnh tải về.

   process : Gọi file *roi_inference/merge_and_process.py*, thực hiện tiền xử lý và ghép các nguồn dữ liệu đầu vào lại với nhau (ndvi, dem,...), lưu file theo từng ngày.

   inference : Gọi file *inference_emsemble_roi.py*, thực hiện run inference.

   visualize : Gọi file *roi_inference/prediction_visualize.py*, thực hiện biến đổi chuỗi giá trị dự đoán thành ảnh tiff.

2. Huấn luyện mô hình
   Dữ liệu huấn luyện sau khi được tải về, xử lý và kết hợp sẽ được lưu trong thư mục *traing_data/fusion*, tùy vào cách kết hợp sẽ có các file fusion khác nhau.
   Để thực hiện huấn luyện cần đầu vào:
   + File tổng hợp dữ liệu trong thư mục training_data/*fusion*.
   + Đường dẫn đến thư mục chứa các pre-trained models được cung cấp sẵn: *pretrained_models/DenseSM_9km*

   Chúng ta sẽ huấn luyện cả 25 biến thể của mô hình của DenseSM. Và thực hiện huấn luyện nhiều lần tùy thuộc và số lượng vòng lặp được chọn trong vòng for.
   Ví dụ: `for r in range(3):` -> Sẽ train 25 models 3 lần, tạo ra 3 tập hợp trained model được chứa trong các thư mục ký hiệu _r0, _r1, _r2.

   Các mô hình trained được lưu trữ trong thư mục: *Demo/ft12_model*.

3. Chương trình thu thập dữ liệu đầu vào cho huấn luyện.
   Sau khi xác định các vị trí điểm (site) cần lấy dữ liệu cùng với giá trị sm tại các thời điểm tại vị trí đó, ta cần lấy các thông tin đầu vào cho quá trình huấn luyện. Chương trình *data_pre/Prepare_samples.ipynb* thực hiện điều này:
   Đầu vào:

- File lưu thông tin các vị trí cần lấy dữ liệu, thương có hậu tố trên là *_site_info.csv* , bao gồm các thông tin:
  + network: Tên vùng lấy dữ liệu, phân biệt các bộ dữ liệu đầu vào khác nhau, gồm có: CHINA, INDIA, CHINA_1km, INDIA_1km, VN.
  + station: Tên điểm trong network, thương là id từ 1-> n, n là số lượng điểm lấy trong vùng.
  + lat: Latitude
  + lon: Longitude
  + s_depth: độ sâu bắt đầu, cái này không quan trọng, thêm vào để giống mẫu có sẵn. Vì đo sm mặt đất nên s_depth = 0.
  + e_depth: độ sâu cuối, cái này không quan trọng, thêm vào để giống mẫu có sẵn. Vì đo sm mặt đất nên e_depth = 5.
- Các file lưu thông tin độ ẩm đất và tọa độ cho từng điểm trong vùng cần lấy dữ liệu, mỗi file tương ứng với một điểm (site). Trong file bao gồm các thông tin sau:
  + id: tương tự như station, tên điểm bắt đầu từ 1->n.
  + date: là ngày có dữ liệu sm cần được lấy thông tin đầu vào tương ứng.
  + latitude
  + longitude
  + pixel_value: không quan trọng, không có cũng được.
  + sm: giá trị độ ẩm đất của điểm i tại thơi điểm t.
  + time: giống như date.
  + DoY: Date of Year biến đổi từ cột time.
  + station: tương tự id
  + sm count: không quan trọng

Cần lưu thay đổi giá trị grid_size khi chạy chương trình, nếu lấy dữ liệu 100m -> grid_size = 0.1; nếu lấy dữ liệu 1km -> grid_size = 1.0.

Các dữ liệu đầu vào này được tạo ra từ các chương trình trong thư mục: *100m_data, 1km_vn_data, 1km_global_data.*

4. Các chương trình chọn điểm và lấy dữ liệu sm.
   Với mỗi tập dữ liệu sẽ có một folder riêng chứa các chương trình để xử lý để tránh code rõ ràng.
   Nhưng quy trình chung để xử lý để tạo ra các file csv làm đầu vào chương trình thu thập như sau:
   - Chia vùng lấy thành các grid (kích thước sẽ khác nhau tùy vùng và loại dữ liệu 100m hay 1km).
   - Từ grid kết hợp với dữ liệu land cover để lọc grid cell và chọn các điểm ngẫu nhiên trong mỗi grid cell được chọn.
   - Từ các điểm được chọn, trích xuất thông tin sm từ các ảnh tif được tải về từ Planet hoặc NSIDC.
   - Thu thập thêm dữ liệu về ngày có Sentinel-1 ở các vùng, dùng dữ liệu này để lọc và chỉ dữ lại các giá trị sm có ngày trùng với Sentinel-1 hoặc sau ngày Sentinel-1 1 ngày.
   - Tạo lập các file csv đầu vào chứa các thông tin cần thiết để tải các dữ liệu đầu vào (NDVI, DEM, soil texture, sentinel-1,...) tại chương trình: *data_pre/Prepare_samples.ipynb*