
**BÁO CÁO CUỐI KỲ
HỌC SÂU**

Câu hỏi 1. Thư mục “BT_01” chứa 2 thư mục con chứa hình ảnh của vùng da bị ung thư (thư mục data) và nhãn phân vùng da bị ung thư (thư mục mask). Sinh viên hãy thực hiện các yêu cầu sau:

- Xây dựng một mô hình học sâu để phân vùng ảnh da bị ung thư. Lưu ý, mô hình sinh viên chọn có thể là mô hình có kiến trúc đã có sẵn từ các bài báo khoa học.
- Sử dụng bảng để liệt kê và mô tả thông tin của tất cả các lớp trong mô hình và giải thích lý do lựa chọn kiến trúc này.
- Chọn và mô tả hàm mất mát (loss function) và phương pháp tối ưu (optimization method).
- Huấn luyện mô hình sử dụng 80% dữ liệu trong thư mục “data” và “mask” tương ứng cho quá trình huấn luyện và 20% cho quá trình kiểm tra (validation). Báo cáo độ chính xác (accuracy) và hàm mất mát (loss) trên tập huấn luyện và tập kiểm tra trong ít nhất 20 epoch.
- Chọn 10 hình ảnh từ tập kiểm tra (đã chọn ở câu d.), sau đó báo cáo kết quả dự đoán cho các hình ảnh này (Sinh viên phải đính kèm hình ảnh đầu vào, kết quả đầu ra và nhãn phân vùng tương ứng trong bài báo cáo).
- Đề xuất và thực hiện giải pháp cải thiện kết quả đã đạt được ở câu d.

Câu hỏi 2. Các tệp “train_02.csv” và “test_02.csv” được cung cấp chứa dữ liệu văn bản dùng cho bài toán phân loại chủ đề Tin tức. Mỗi hàng trong mỗi tệp gồm hai cột: chỉ số lớp (class index) và phần mô tả (description). Các chỉ số lớp bao gồm: : 1-World, 2-Sports, 3-Business, 4-Science and Technology. Dựa trên phần mô tả, sinh viên cần thực hiện các yêu cầu sau:

- Xây dựng một mô hình phân loại sử dụng nhiều hơn 2 lớp LSTM cùng với các lớp khác để phân loại 4 chủ đề Tin tức. Liệt kê và mô tả tất cả các lớp trong mô hình phân loại, bao gồm: tên lớp, hàm kích hoạt, kích thước đầu vào/đầu ra, số bước thời gian (timesteps) và kích thước vec-tơ đầu vào (đối với các lớp LSTM). Giải thích lý do lựa chọn kiến trúc mô hình này.
- Trong báo cáo, mô tả chi tiết các bước xử lý dữ liệu (đầu vào và nhãn). Trình bày lý do thực hiện các bước.
- Trong mã nguồn, hiển thị số lượng tham số của từng lớp trong mô hình và trình bày trong báo cáo kết quả đó.
- Trình bày hàm mất mát, thuật toán tối ưu và tốc độ học (learning rate) do sinh viên lựa chọn. Giải thích lý do cho các lựa chọn này. Huấn luyện mô hình trong ít nhất 10 lần lặp (iterations). Sau khi huấn luyện, trình bày và nhận xét kết quả huấn luyện và kiểm thử mà sinh viên đạt được.
- Đề xuất và thực hiện giải pháp cải thiện kết quả đã đạt được ở câu d.