Tổng quan đề tài

Đề tài mà em thực hiện là một ứng dụng giúp nhận diện được một chuỗi các nốt nhạc được ký âm trong một file pdf, sau đó mã hóa chuỗi này thành một định dạng sử dụng được trên máy tính (**MML/ABC** hoặc **MIDI**). Mục tiêu là nhận diện được chính xác các nốt nhạc và giữ lại được 2 yếu tố chính: cao độ và trường độ. Để từ đó, em có thể sử dụng những thông tin này để mở rộng thành dự án tiếp theo: tự động nhận diện vị trí phát nhạc trên file pdf đầu vào.

Công nghệ/framework sử dụng dự kiến

* Tensorflow/Keras trên Python để xây dựng mô hình Deep Learning.
* Thư viện Python để chuyển đổi file định dạng pdf thành hình ảnh bitmap (jpg/png/bitmap).
* Thư viện Python để chuyển đổi kết quả nhận dạng (có thể là định dạng MML hoặc ABC) thành định dạng chuẩn MIDI để có thể phát nhạc.

Tài liệu tham khảo

Em đã tìm ra được một số mô hình và kiến trúc mẫu như sau:

* <https://huggingface.co/papers/2402.07596> : Dự án sử dụng mô hình Transformer End-To-End để nhận diện nốt nhạc đa bè trong một trang nhạc.
* <https://huggingface.co/Flova/omr_transformer> : Dự án sử dụng mô hình Transformer để nhận diện nốt nhạc đơn bè trong một hình ảnh chụp/viết tay.
* <https://github.com/afikanyati/cadenCV> : Dự án nhận diện nốt nhạc đơn bè từ file pdf, sử dụng các kỹ thuật/thuật toán/mô hình máy học nhận diện ký tự (OCR), sau đó xuất kết quả thành định dạng MIDI.