Lý do chọn đề tài:

* 7650 bộ hồ sơ gồm nhiều loại giấy tờ khác nhau.

người làm : lấy dữ liệu đọc, ghi ra phân tích, … → 15’ xử lý xong 1 bộ → 60’ sẽ xử lý xong 4 bộ → 8h/32 bộ → 7650 bộ sẽ mất → 239 ngày (8h / ngày) → 5 người cùng làm sẽ mất 47 ngày (1 tháng rưỡi)

* 4’ xử lý 1 bộ hồ sơ (hơn 100 trang) → 60’ xử lý được 15 bộ → 24h/360 bộ → 7650 bộ sẽ mất 21 ngày làm (đây chỉ là 1 sevice chỉ chiếm 3gb vram) → x 5 lần lên (5 con sevice) -> chỉ mất 4 ngày + cộng thêm thời gian kiểm tra dữ liệu (đã phân ra được các cụm loại hồ sơ) 14 ngày = 18 ngày | x2 x3 lần rồi. Ngoài ra còn lưu trữ được data ở dạng pdf, ảnh, ko lo bị hủy dữ liệu.

Text Detection:

* Apdaptive Scale Fusion module: Giúp tổng hợp đặc các đặc trưng ở nhiều tỷ lệ giúp tăng cường khả năng biểu diễn của mô hình.

Text Recignition:

* Patch Embedding: Lấy ý tưởng từ patch image trong mô hình vision transformer → giúp tách các ảnh đầu vào thành các phần tử, mỗi phần tử đóng vai trò như các ký tự trong các bài toán xử lý ngôn ngữ.
* Mixing Blocks: Gồm 2 loại Local mixing / Global mixing block
  + Global Mixing : Dùng để biểu diễn giữa các phần tử non-text và text qua đó biểu diễn phụ thuộc xa giữa các ký tự với nhau.
  + Local mixing : Dùng để biểu diễn mối quan hệ giữa các nét trong cùng một ký tự.