Yêu cầu bài toán:

- Xác định được vùng chữ có trong ảnh chụp.

- Xác định được chữ trong ảnh.

- Dịch được nó sang Tiếng Việt.

Kiến thức và dữ liệu đã có:

- Bộ dữ liệu MSRA-TD500: gồm 500 ảnh có tọa độ của vùng chữ trong mỗi ảnh, chủ yếu là Tiếng anh và Tiếng Trung.

- Bộ dữ liệu crawl từ kết quả tìm kiếm của google và flick: chứa hình ảnh của các đền chùa cùng chữ Hán xuất hiện trong đó.

- Bộ dữ liệu SynthText chứa hơn 800.000 ảnh cùng với text trong mỗi ảnh. Bộ này được sử dụng để train cho mô hình: "Multi-Oriented Scene Text Detection". (<https://arxiv.org/pdf/1802.08948.pdf>)

- Bộ chữ viết tay Tiếng Trung: CASIA-OLHWDB, được dùng trong paper nhận dạng chữ viết tay "Learning Spatial-Semantic Context " (<http://discovery.ucl.ac.uk/1569458/1/TPAMI-2016-08-0656-R2.pdf>)

- <https://github.com/tesseract-ocr/tesseract>

Công việc cần làm ngay sau khi về nhà:

- Download bộ chữ viết tay tiếng Trung

- Download & install matlab để đọc file mat trong dataset SynthText

Nhận xét của GV:

- Cần viết 1 bản khảo sát (survey) về các phương pháp người ta đã sử dụng bao gồm cả Text Detection và Text Regconition:

- Phương pháp tiếp cận, kể cả dùng và không dùng deeplearning.

- Độ hiệu quả, độ chính xác của mỗi phương pháp.

- Mức độ phức tạp của từng phương pháp, có thể áp dụng mô hình vào app?

- Ưu nhược điểm của mỗi phương pháp.

- Về vấn đề translate có thể dùng của bên thứ 3 như google mà không phải build lại mô hình nhưng cũng cần phải đánh giá độ chính xác của nó trong trường hợp mình đang áp dụng.

- Đánh giá phương pháp tiếp cận trước khi tiếp cận, tìm hiểu sâu về phương pháp giải bài toán.