

# **HƯỚNG TIẾP CẬN SỬ DỤNG THƯ VIỆN HỖ TRỢ CHO BÀI TOÁN NHẬN DẠNG VĂN BẢN TRONG HÌNH ẢNH VÀ XÂY DỰNG BỘ DỮ LIỆU VỀ NHẬN DẠNG VĂN BẢN TRONG HÌNH ẢNH CHO NGÔN NGỮ TIẾNG VIỆT**

<b>Thái Trần Khánh Nguyên</b>	<b>- 19520188</b>
<b>Nguyễn Khánh Như</b>	<b>- 19520209</b>
<b>Lê Đoàn Thiện Nhân</b>	<b>- 19520197</b>

# Tóm tắt

- Lớp: CS519.M11.KHCL
- Link Github của nhóm: <https://github.com/ttknguyen/CS519.M11.KHCL>
- Link YouTube video:
- Tổng số slides không vượt quá 10



Thái Trần Khánh Nguyễn  
19520188



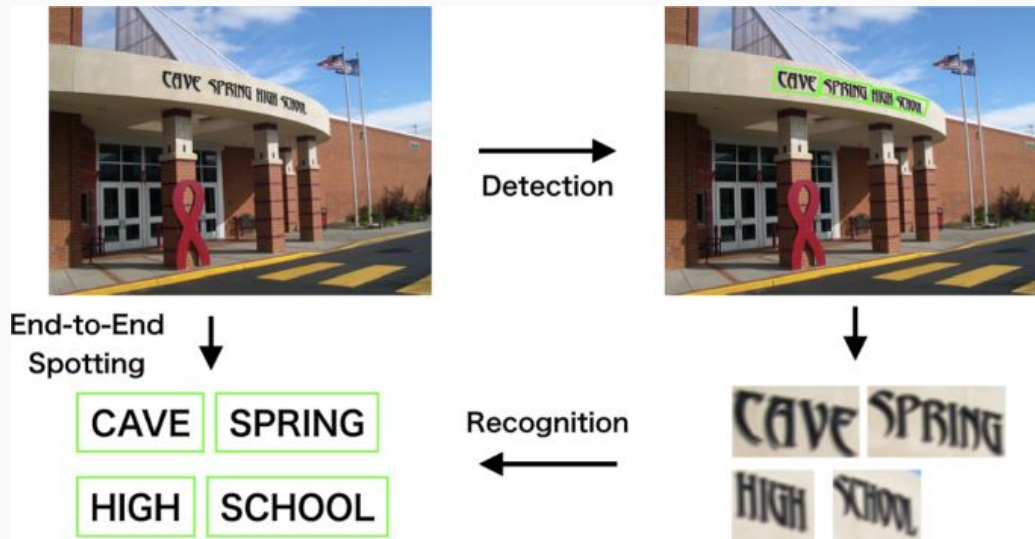
Nguyễn Khánh Như  
19520209



Lê Đoàn Thiện Nhân  
19520197

# Giới thiệu

- Nhận dạng văn bản trong hình ảnh

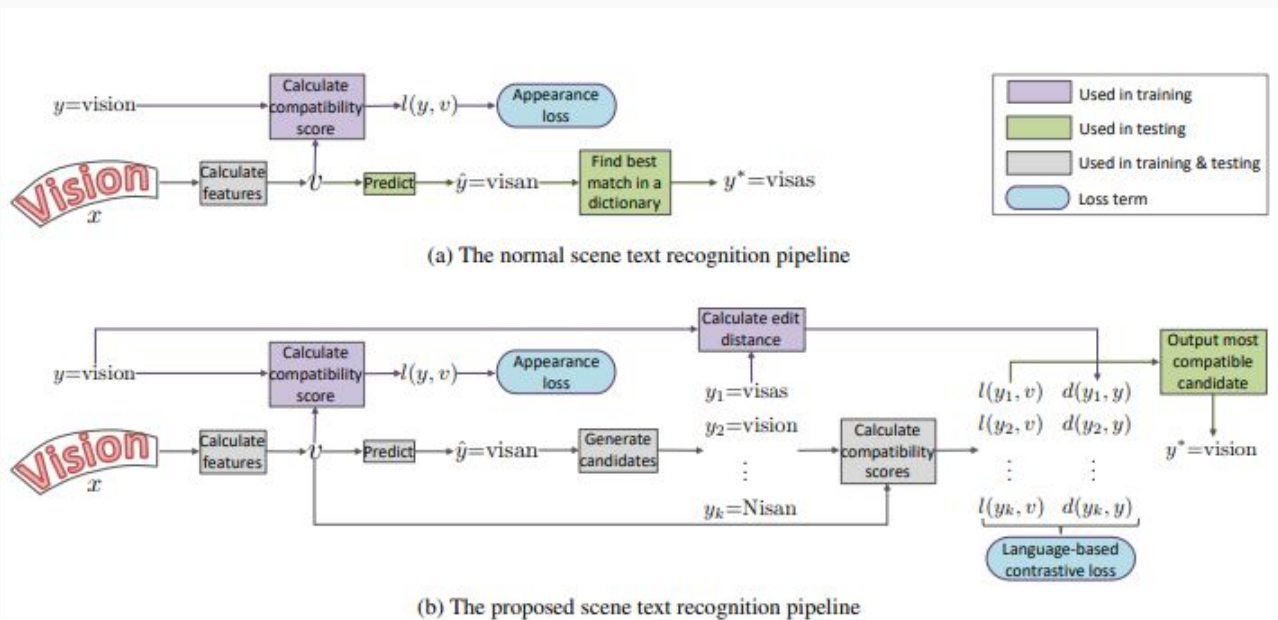


# Giới thiệu

- Sử dụng từ điển để cải thiện hiệu quả mô hình:
  - Hướng tiếp cận hiện tại còn đơn giản
  - Một vài bất cập khi sử dụng phương pháp
    - Nhiều trường hợp văn bản là các từ nước ngoài hoặc được tạo ra không có trong từ điển
    - Không có sự phản hồi trong quy trình suy luận khi huấn luyện mô hình
    - Khoảng cách dùng để xác định từ được xuất ra còn nhiều hạn chế

# Giới thiệu

- Đề xuất hướng tiếp cận mới
- Xây dựng bộ dữ liệu dành cho Tiếng Việt

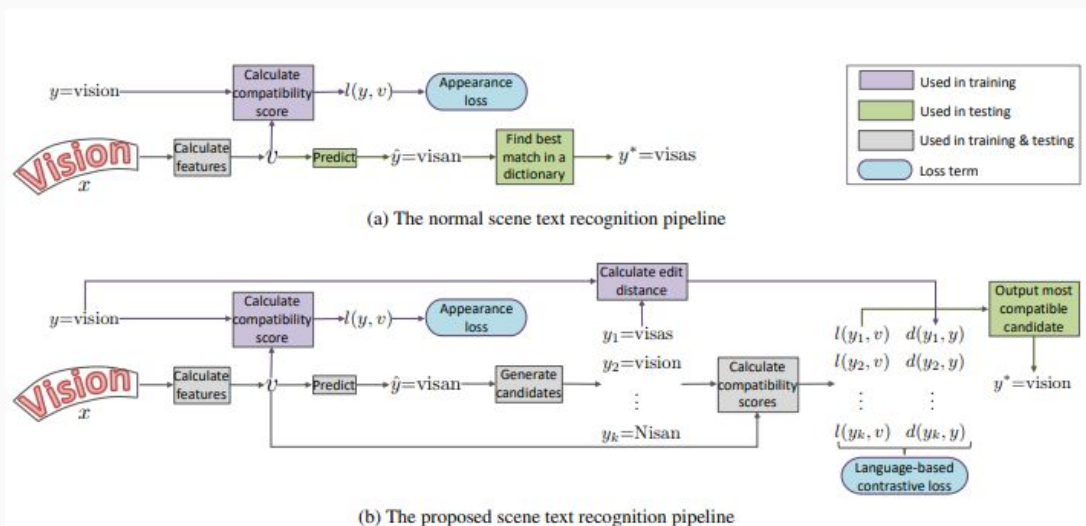


# Mục tiêu

- Đề xuất cách tiếp cận mới, thay vì buộc mô hình phải dự đoán đầu ra là một từ thuộc từ điển thì tích hợp từ điển vào mô hình. Từ đó, tạo ra danh sách các từ ứng cử viên phù hợp, sau đó, đưa vào module tính điểm để tìm ra từ tương thích nhất.
- So sánh kết quả của mô hình Scene Text Recognition cơ bản và phiên bản cải tiến (tích hợp từ điển vào mô hình ban đầu) bằng cách sử dụng các bộ dataset TotalText, ICDAR2013, ICDAR2015 và VinText để đánh giá.
- Xây dựng bộ dataset cho bài toán Scene Text Recognition trong ngữ cảnh tiếng Việt.

# Nội dung và Phương pháp

- Nghiên cứu về cách thức sinh danh sách ứng cử viên trong quá trình suy luận và huấn luyện
- Nghiên cứu về các khoảng cách để có thể tìm ra cách điều chỉnh khoảng cách thích hợp với hướng tiếp cận mới này



# Nội dung và Phương pháp

- Thực hiện việc thu thập dữ liệu cho bộ dữ liệu VinText thông qua các cách như thu thập trên internet và chụp hình thực tế.





# Kết quả dự kiến

- Hướng tiếp cận mới sẽ đạt được state-of-the-art khi so sánh (h-means) với các giải pháp trước đây
- Chứng minh rằng bộ dataset cho vấn đề Scene Text Recognition đến từ team Vin (VinText) là một bộ dataset có độ tin cậy cao và đáng để được đưa vào sử dụng rộng rãi.

# Tài liệu tham khảo

- [1]. Chee Kheng Chng and Chee Seng Chan. Total-text: A comprehensive dataset for scene text detection and recognition. In IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017
- [2]. Y. Liu, Hao Chen, Chunhua Shen, Tong He, Lian-Wen Jin, and L. Wang: ABCNet: Real-time scene text spotting with adaptive bezier-curve network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020
- [3]. Minghui Liao, Guan Pang, J. Huang, Tal Hassner, and X. Bai. Mask textspotter v3: Segmentation proposal network for robust scene text spotting. ArXiv, 2020
- [4]. Wen-Yang Hu, Xiaocong Cai, J. Hou, Shuai Yi, and Zhiping Lin. Gtc: Guided training of ctc towards efficient and accurate scene text recognition. ArXiv, 2020.