

Ngày nay, kết quả từ lĩnh vực **Thị giác máy tính (Computer Vision)** đã được áp dụng vào trong nhiều lĩnh vực thực tế và đóng một vai trò quan trọng trong các hệ thống giao thông thông minh, hệ thống điểm danh sử dụng khuôn mặt, nhận diện biển báo giao thông, ... Khi xét riêng về bài toán xác định và nhận diện văn bản trong ảnh (text localization – detection and recognition) cũng được áp dụng rộng rãi trong đời sống như nhận diện biển số xe, tự động đọc các giấy tờ tùy thân như CMTNDD, CCCD. Điểm chung của các hệ thống này đó là bước đầu phải xác định được vị trí chứa văn bản (text) trong ảnh – bài toán Text Detection-Localization . Đây cũng là bài toán mà Câu Lạc Bộ AI – Khoa Khoa học Máy tính lấy làm chủ đề cho cuộc thi **AI Tempo Run Mùa Hè** này.

Scene Text Detection (Phát hiện văn bản) là bài toán phát hiện vùng có văn bản trong một bối cảnh phức tạp trong ảnh và bao quanh nó bằng một hình bao (bounding box).



Cuộc thi AI Tempo Run Mùa Hè chọn bài toán này làm chủ đề chính của cuộc thi, với:

- **Đầu vào:** Một tấm ảnh được chụp với bối cảnh bất kỳ.
- **Đầu ra:** Tọa độ của chữ xuất hiện trong bức ảnh.

Lưu ý : Trong quá trình thi, ban tổ chức chỉ cung cấp duy nhất một bộ dữ liệu huấn luyện. Các đội thi được khuyến khích thu thập thêm các dữ liệu bên ngoài, ngoại trừ bộ dữ liệu VinText.

Evaluation Metric

Cuộc thi AI Tempo Run sẽ đánh giá lời giải của các đội thi thông qua thang đo $mAP@[0.5-0.95]$ (mean average precision) theo [chuẩn COCO](#).

Đối với dữ liệu Private test, điểm số sẽ được đánh giá trên cả độ chính xác và tốc độ thực thi. Cụ thể, điểm số cuối cùng được tính theo công thức:

$$\text{score} = \text{mAP} \times \left(80\% + 20\% \times \left(1 - \frac{\text{time}}{7200} \right) \right)$$

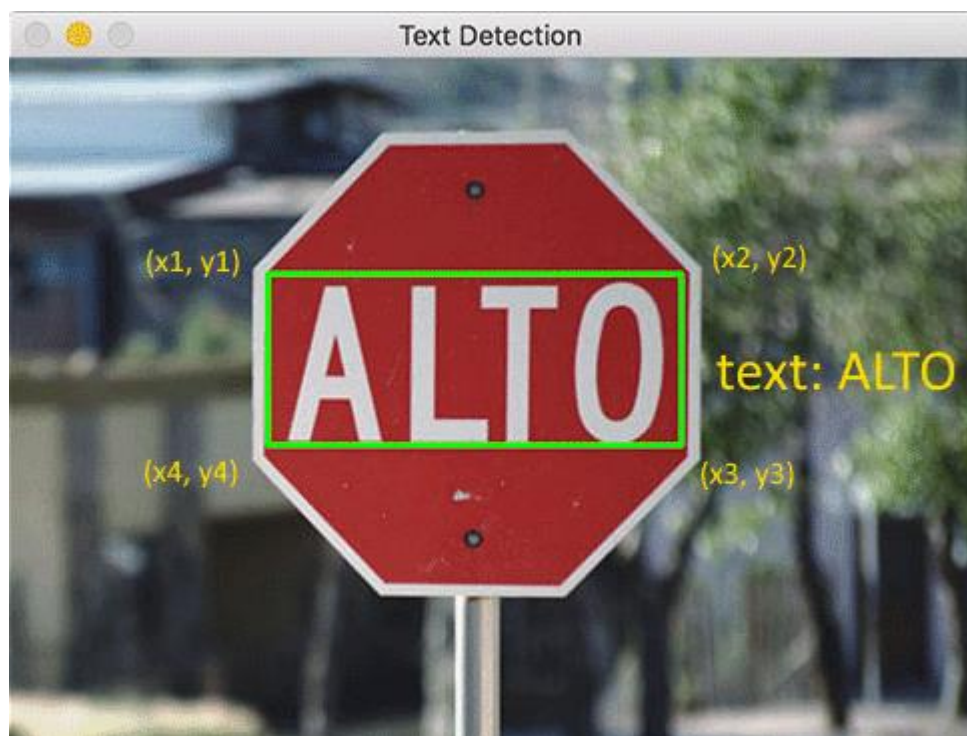
Với "time" là tổng thời gian thực thi của chương trình cho cả bộ dữ liệu, tính bằng giây. Nếu chương trình thực thi quá 02 giờ (7200 giây), lời giải sẽ bị coi là không hợp lệ.

Hiểu theo một cách khác, điểm mAP chỉ đóng góp 80% tỉ trọng vào số điểm cuối cùng, phần còn lại được tính dựa trên thời gian thực thi.

Format nộp bài:

Các đội thi nộp bài lên hệ thống theo định dạng sau:

File mẫu: [sample_submission.txt](#)



Cụ thể, mỗi dòng trong file nộp bài tương ứng với một đơn vị văn bản xuất hiện trong tập dữ liệu test, được thể hiện bởi tọa độ 4 đỉnh đường bao (bounding box) của chữ, kèm theo tên file chứa văn bản:

file_name, x1, y1, x2, y2, x3, y3, x4, y4

Trong đó:

- **file_name:** Tên file ảnh chứa đơn vị văn bản này (VD: test_0001.jpg, test_0923.jpg, v.v.)

- (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , (x_4, y_4) lần lượt là tọa độ bốn đỉnh đường bao chứa đơn vị văn bản. Bốn đỉnh cần tạo thành một tứ giác lồi để được xem là đường bao hợp lệ.

Nội quy cuộc thi:

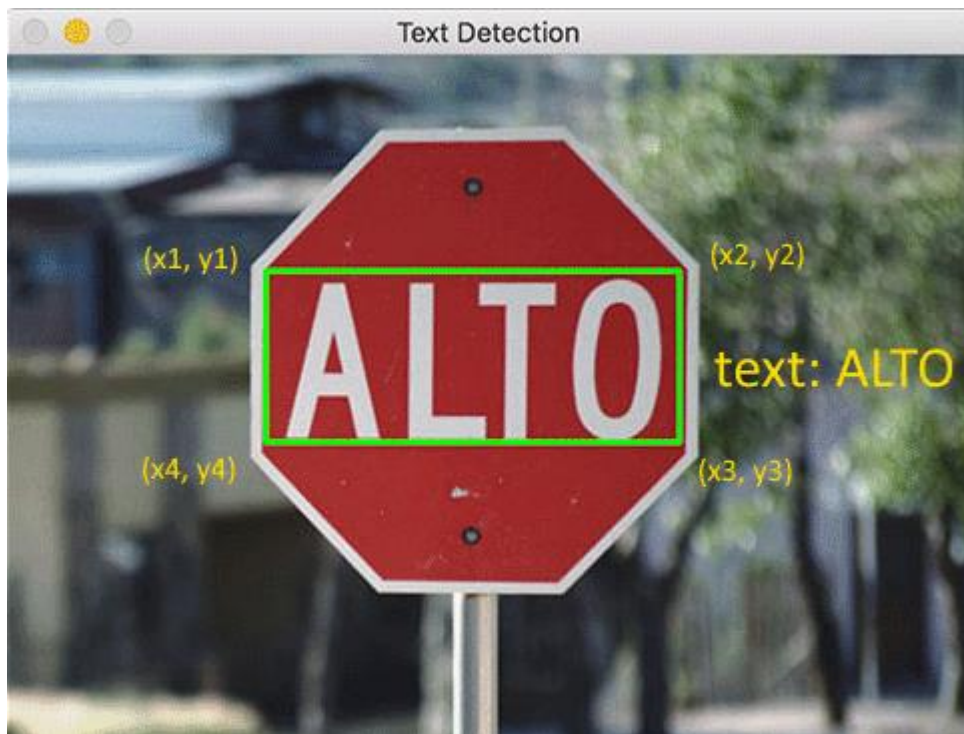
1. Thí sinh tham gia cuộc thi theo hình thức cá nhân hoặc theo tổ đội (tối đa 3 thành viên). Mỗi thí sinh chỉ tham gia vào tối đa một đội và chỉ được đăng ký một lần.
2. Khi đặt tên đội thi cần tuân thủ thuần phong mỹ tục. Không xúc phạm đến danh dự và nhân phẩm của người khác.
3. Chỉ các đội đã đăng ký hợp lệ mới được xem là tham gia cuộc thi. Đăng ký tham gia [ngay tại đây](#).
4. Các đội thi được khuyến khích sử dụng các bộ dữ liệu bên ngoài hoặc các mô hình Học sâu được huấn luyện sẵn.
5. Nghiêm cấm mọi hành vi gian lận. Không chia sẻ và trao đổi lời giải giữa các đội trong quá trình thi.
6. **Nhằm đề cao tính học thuật và công bằng trong cuộc thi, các đội vui lòng không sử dụng bộ dữ liệu VinText để huấn luyện mô hình.**
7. Tập dữ liệu **Public test** được sử dụng để các đội thi tự đánh giá được phương pháp và thuật toán đội mình phát triển. Kết quả cuối cùng của cuộc thi sẽ được quyết định dựa trên kết quả của các đội trên tập dữ liệu **Private test**.
8. Các đội thi không bị giới hạn số lần nộp bài trên tập dữ liệu **Public test**. Tuy nhiên mỗi lần nộp bài cần cách nhau ít nhất 01 giờ nhằm giảm tải cho máy chủ.
9. Ban tổ chức (BTC) sẽ mở form nộp bài cho bộ dữ liệu **Private test** vào **17h00 thứ Năm ngày 05/08/2021**. Các đội thi cần nộp source code cho lời giải của mình để BTC thực hiện chấm bài và ra quyết định cho kết quả cuối cùng. Source code được nộp dưới dạng tệp tin .ipynb hoặc đường dẫn đến Github/Gitlab Repository.
10. Giải thưởng "**Ý tưởng sáng tạo**" sẽ dành cho đội có lời giải sáng tạo nhất kèm theo chú thích (comment) trong source code để thuyết minh cho lời giải của mình. Lời giải này sẽ được BTC chia sẻ rộng rãi đến các sinh viên và đội thi khác nhằm khuyến khích tinh thần học hỏi và chia sẻ lẫn nhau.
11. Các đội vi phạm nội quy cuộc thi sẽ bị tước quyền tham gia cuộc thi này và tất cả các cuộc thi khác do Câu lạc bộ AI tổ chức.

Thông tin về bộ dữ liệu:

Dataset sử dụng trong cuộc thi được cung cấp dưới dạng tệp tin ảnh, tương ứng với nhãn dữ liệu theo định dạng ICDAR15.

Dữ liệu cuộc thi gồm 3338 ảnh chụp ở các bối cảnh ngẫu nhiên được chia thành 3 tập **Train set**, **Public test**, **Private test** với tỉ lệ như sau:

- **Train set:** 1240 ảnh đã được gán nhãn tọa độ và kích thước của văn bản.
- **Public test:** 987 ảnh chưa được gán nhãn và thí sinh cần dự đoán nhãn dữ liệu dưới dạng tọa độ xuất hiện của các đối tượng văn bản để hệ thống chấm điểm.
- **Private test:** Tương tự tập dữ liệu Public test với 1111 ảnh. Tập dữ liệu này không được cung cấp trực tiếp cho thí sinh mà chỉ được dùng để chấm điểm ở vòng chung cuộc.



Cụ thể, mỗi ảnh trong tập Train set sẽ có một tệp tin nhãn, trong đó mỗi dòng tương ứng với một đơn vị văn bản, được thể hiện bởi tọa độ 4 đỉnh đường bao (bounding box) của chữ, kèm theo nội dung bên trong:

$x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4, \text{text}$

Lưu ý: BTC không đảm bảo đỉnh (x_1, y_1) sẽ ở góc trái trên, và không đảm bảo thứ tự các đỉnh được sắp xếp theo chiều kim đồng hồ.

Acknowledgements: Bộ dữ liệu trong cuộc thi được xây dựng dựa theo bộ dữ liệu VinText giới thiệu trong bài báo [\[1\]x](#). Chúng tôi cam kết rằng bộ dữ liệu này chỉ được sử dụng với mục đích phi lợi nhuận. Nếu bộ dữ liệu này được sử dụng trong các công trình nghiên cứu, xin vui lòng dẫn nguồn đến tác giả gốc.