**UIT - CS332 - Vietnam’s Traffic Signs Dataset**

**[BÁO CÁO] Thực hành thu thập dataset**

* **Thành viên nhóm:**

14520156 - Trần Quang Đạt

14520247 - Trần Xuân Hải

14520956 - Hoàng Hữu Tín

* **Mã Lớp: CS332.I11.KHTN**
* **Thống kê dữ liệu hiện tại:**
* Tổng thời gian videos: **01:24:44**
  + Tổng số Videos: 40 videos
  + Thời gian thu thập dữ liệu: 4 ngày
  + Thiết bị sử dụng: Canon EOS 700D, iPhone 6s
  + Địa điểm: Tp. Hồ Chí Minh
* Tổng số lượng ảnh đã tách: **1559 ảnh**
  + Thời gian tách ảnh và đánh Annotations: 2 ngày
  + Tổng số loại biển báo đường bộ VN được lấy: 195
  + Sổ loại biển báo đã có ảnh trong Dataset: 69

**Chi tiết thống kê cho từng loại biển báo xem ở file:**

[DatasetStatistics\_Excel.xlsx](https://drive.google.com/open?id=0B9hMAZTpHpyCN0l1WU9mQ2RxT1U)

**Chi tiết cách đánh Annotations cho từng ảnh xem ở file:**

[Annotation Format Guide](https://drive.google.com/open?id=16Fe2-b_eLCv-YRgcm3exkJzLIDgldua4pcFKYUe4rs8)

**Thư mục chính chứa toàn bộ dataset:**

[**https://drive.google.com/open?id=0B9hMAZTpHpyCclFwT2NFWTRYSjg**](https://drive.google.com/open?id=0B9hMAZTpHpyCclFwT2NFWTRYSjg)

* **Các thành phần có trong thư mục chính:**
  + Folder **TrafficSigns\_ForDetection**: thư mục chứa các ảnh sau khi đã đánh Annotations
    - Gồm 195 thư mục con tương ứng với 195 loại biển báo giao thông đường bộ Việt Nam được nhóm chọn lọc.
    - Mỗi thư mục con này luôn chứa 2 file:
      * File \_Annotations.txt: Thông tin về loại biển báo
      * File ảnh \_ORG\_0.jpg: Icon cho biển báo
    - Trong đó có 69 thư mục con đã có ảnh mà nhóm đã gặp được trong lúc quay videos.
    - Còn 126 thư mục chưa có ảnh vì các biển báo này khá hiếm và nhóm không đủ thời gian để đi tìm hết các loại biển báo.
  + Folder **AnnotationGenerator\_Program:** Phần mềm được nhóm tạo ra dùng để thuận tiện tách ảnh và đánh Annotations. Sơ lược cách sử dụng phần mềm này xem tại:

<https://www.youtube.com/watch?v=EjKDYGtHm9A>

* + Folder **Raw\_Videos**: Chứa các videos mà nhóm đã quay được.
  + Folder **Utilities\_Codes:** Một số đoạn code thực hiện tự động:
    - Crawler icon và tạo thư mục từ các link trên mạng.
    - Merge các phần đánh Annotations của các thành viên thành một folder.
    - Tính statistics của Dataset.
  + File [Báo cáo], Dataset\_Statistics và file Annotation Format Guide
  + File TrafficSigns\_ORG.rar chứa thư mục chỉ gồm 195 thư mục con được crawler từ website về, chưa gồm các ảnh được tách.

**Các task có thể sử dụng Dataset này**

* **Traffic signs Detection:**
  + Thực hiện detect những biển báo có trong ảnh.
  + Dựa trên tọa độ đã đánh trong Annotations có thể làm groundtruth để tính độ chính xác của thuật toán Detection
  + Có thể dựa trên các cờ trạng thái để đo độ chính xác của thuật toán khi chạy trên các trường hợp khó như ảnh ban đêm, ảnh Blurred,...
* **Traffic signs Classification:**
  + Nhận vào một ảnh đã crop của một biển báo cho biết ảnh đó thuộc biển báo nào.
  + Thực hiện crop ảnh dựa trên các Bounding Box trong Annotation của mỗi ảnh để có được training dataset cho Classifier.
  + Các flag trạng thái của biển báo có thể được dùng để chỉ tách các biển báo tốt Visible hay lấy luôn cả các biển báo có flag khác dùng làm training dataset.