

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP HCM
KHOA ĐIỆN – ĐIỆN TỬ



HCMUTE

MÔN HỌC: HỌC MÁY VÀ TRÍ TUỆ NHÂN TẠO

COMPUTER VISION ANNOTATION TOOL

GVHD: THẦY NGUYỄN MẠNH HÙNG

Họ và tên:

Trần Thanh Ngọc 19119204

Phan Công Danh 19119160

Võ Minh Hậu 19119174

Nguyễn Hà Nhật Linh 19119190

Phạm Hải Nguyên 19119205

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN.....	3
1.1 GIỚI THIỆU	3
1.2 MỤC TIÊU VÀ GIỚI HẠN ĐỀ TÀI.....	3
1.3 PHƯƠNG PHÁP NGHIÊN CỨU.....	3
1.4 ĐỐI TƯỢNG NGHIÊN CỨU	4
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT	5
2.1 ỨNG DỤNG CỦA CVAT	5
2.1.1 Ứng dụng chính.....	5
2.1.2 Ứng dụng thực tế.....	5
2.2 TÍNH NĂNG	6
2.2.1. Giao diện CVAT	6
2.2.2. Các phím điều hướng cơ bản	7
2.2.3. Cách tạo đối tượng để khảo sát trong CVAT	7
2.2.4. Attribute annotation mode	8
2.2.5. Annotation tool	9
2.3 ƯU VÀ KHUYẾT ĐIỂM	10
2.4 SO SÁNH MANUAL LABELING VÀ SEMI- AUTOMATIC LABELING.....	10
CHƯƠNG 3: DỰ ÁN THỰC TẾ	12
3.1 TẠO LẬP DỰ ÁN	12
3.2 PHÂN TÍCH KẾT QUẢ	13
CHƯƠNG 4: TÀI LIỆU THAM KHẢO	14

CHƯƠNG 1: TỔNG QUAN

1.1 GIỚI THIỆU

Ngày nay, trí tuệ nhân tạo - Artificial intelligence (AI) đang ngày càng phổ biến và góp phần thay đổi nhiều khía cạnh trong cuộc sống hằng ngày. Trong đó thị giác máy tính - Computer Vision là một lĩnh vực cực kỳ quan trọng trong ngành AI bao gồm các phương pháp thu nhận, xử lý ảnh kỹ thuật số, phân tích và nhận dạng hình ảnh. Và công cụ hỗ trợ đắc lực cho việc huấn luyện các mô hình đó chính là những công cụ chú thích hình ảnh.

CVAT (Computer Vision Annotation Tool) là một công cụ chú thích hình ảnh và video mã nguồn mở khá phổ biến được phát triển bởi Intel. CVAT có nhiều tính năng mạnh mẽ: chú thích tự động bằng cách sử dụng mô hình học sâu, phím tắt cho hầu hết các thao tác quan trọng, bảng điều khiển với danh sách các tác vụ chú thích,... Chúng ta có thể sử dụng công cụ này trực tuyến (với một số hạn chế) hoặc cài đặt nó trên máy của mình.

CVAT được sử dụng nhằm đặt nhãn dữ liệu để giải quyết các nhiệm vụ về thị giác máy tính như: Phân loại hình ảnh, phát hiện đối tượng, theo dõi đối tượng, phân đoạn hình ảnh, nhận dạng tư thế. CVAT hỗ trợ nhiều định dạng chú thích như YOLO, Pascal VOC hoặc MS COCO.

1.2 MỤC TIÊU VÀ GIỚI HẠN ĐỀ TÀI

Tìm hiểu cách hoạt động, tính năng, ứng dụng của các công cụ annotation hiện nay là CVAT. Đề tài chỉ thực hiện việc label cho các đối tượng mà không sử dụng các tệp dữ liệu từ đối tượng đó để huấn luyện, do đó khó trong việc đánh giá tệp dữ liệu này có tốt hay không.

1.3 PHƯƠNG PHÁP NGHIÊN CỨU

Để thực hiện đề tài này, chúng tôi đã kết hợp các phương pháp nghiên cứu sau:

- Phương pháp phân tích.
- Phương pháp tổng hợp.
- Phương pháp kết luận.

1.4 ĐỐI TƯỢNG NGHIÊN CỨU

Công cụ chú thích hình ảnh và video mã nguồn mở CVAT (Computer Vision Annotation Tool).

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

2.1 ỨNG DỤNG CỦA CVAT

2.1.1 Ứng dụng chính

Ngoài việc chú thích các bounding box để phát hiện đối tượng, CVAT còn cho phép chú thích dữ liệu phục vụ cho việc phân loại hình ảnh, semantic segmentation, instance segmentation và theo dõi đối tượng (object tracking).

Ứng dụng chính của CVAT là công cụ phát hiện vật thể. Công cụ này cho phép vẽ các bounding box, polylines (Polyline là tập hợp các đường thẳng liên nhau tạo thành một khối), polygons (các hình đa giác) và keypoints trên cả dữ liệu hình ảnh và video.

Các tính năng chính của CVAT bao gồm: hỗ trợ các shape, bounding box, polylines, mảng các điểm, polygons; phép nội suy tuyến tính của các bounding box giữa các keyframe trong video; gán thuộc tính đã được cấu hình cho bất kỳ nhãn nào.

Phân lớp hình ảnh cho phép áp dụng các thẻ (tag) với các thuộc tính, bao gồm nhiều định dạng khác nhau như boolean, số hoặc văn bản. Phân đoạn hình ảnh và thao tác với các shape. Mỗi object có z-order để tránh vẽ trùng các đường ranh giới giữa các object. Shape có thể được chỉnh sửa: kéo, thay đổi kích thước, thêm/xóa các điểm. Màu theo label hoặc instance. Hiển thị các vùng không được chú thích.

2.1.2 Ứng dụng thực tế

Chăm sóc sức khỏe: Phân tích ảnh chụp X-ray ngực nhằm phát hiện các bệnh phổi, Covid19; Phát hiện các khối u ác tính bằng việc quét qua da,...

Giúp các nhà bán lẻ cải thiện trải nghiệm của khách hàng, tăng lợi nhuận. Thanh toán không cần thu ngân: sử dụng hàng nghìn camera, cảm biến IoT và hệ thống dựa trên thị giác máy tính (computer vision based system): phát hiện và hiểu các tương tác của khách hàng; theo dõi sự chuyển động của các sản phẩm nhờ các thuật toán phát hiện đối tượng; tự động phát hiện giá.

Ngoài ra còn có ứng dụng lớn trong việc quản lý hàng tồn, computer vision based system được sử dụng để chụp hình ảnh, trước khi sử dụng các thuật toán phân đoạn hình ảnh và phát hiện đối tượng để theo dõi các mặt hàng trên kệ và thực hiện quét toàn bộ hàng tồn kho.

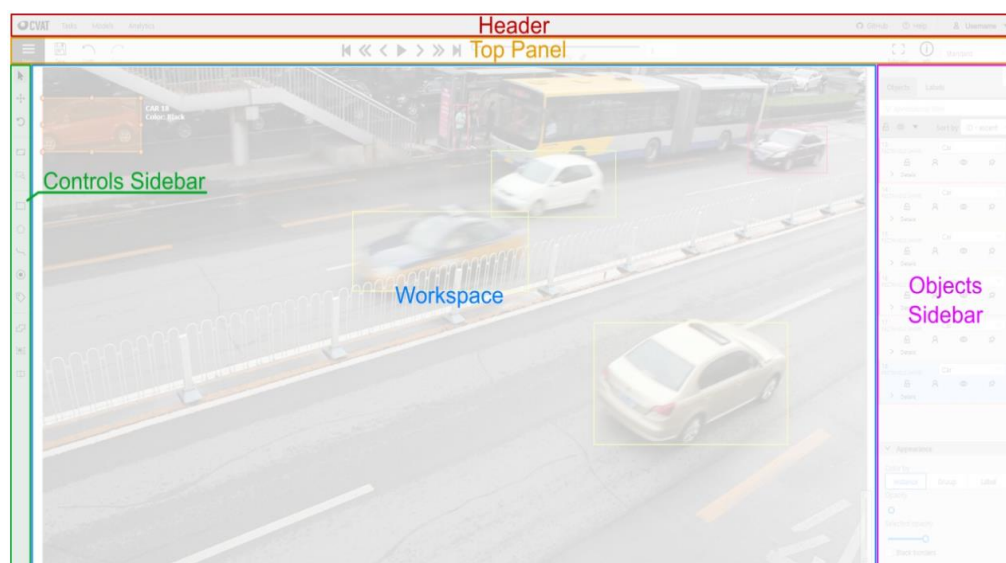
Đối với cộng đồng, CVAT hỗ trợ trong việc phân tích đối tượng trong phạm vi nhất định. Phát hiện khách hàng vào và ra khỏi cửa hàng, phân tích các số liệu cho phép nhà bán lẻ hiểu lý do tại sao mọi người mua sản phẩm của họ, phản ứng trong cửa hàng đối với các sản phẩm, cũng như sở thích của khách hàng.

Đối với việc phát triển nông nghiệp. CVAT hỗ trợ giám sát cây trồng và đất. Thu thập các hình ảnh về đất và cây trồng, train computer vision sử dụng các hình ảnh đó để theo dõi sự phát triển của cây trồng; dự đoán năng suất chính xác; phát hiện các vấn đề về của cây trồng nhanh hơn nhiều so với con người; Hệ thống tưới thông minh, phát hiện côn trùng gây hại.

2.2 TÍNH NĂNG

2.2.1. Giao diện CVAT

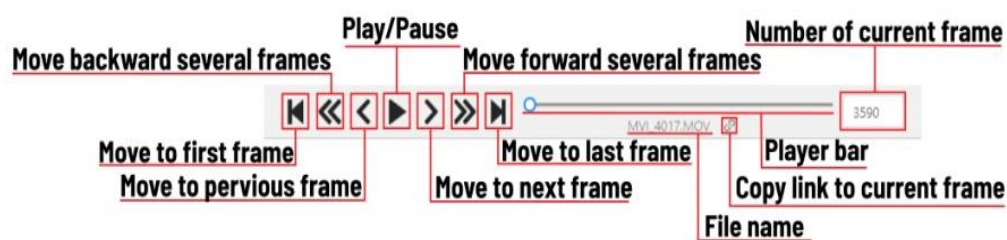
Giao diện CVAT bao gồm 4 khu vực: Header, Top Panel, Workspace và Control sidebars. Với header mang ý nghĩa như thanh điều hướng và các cài đặt tài khoản, top panel chứa các nút điều hướng, các chức năng chính và hỗ trợ người dùng truy cập menu. Khu vực lớn nhất cũng như cần xử lý nhiều nhất đó là workspace, với ý nghĩa mang hình ảnh để người dùng xử lý, khảo sát. Control sidebars chứa các công cụ để điều hướng hình ảnh, thu phóng, tạo hình và chỉnh sửa cấu trúc của tệp hình ảnh (hợp nhất, chia, nhóm). Và cuối cùng là objects sidebar: Chứa bộ lọc nhãn gồm hai danh sách: các đối tượng (trên frame) và nhãn (của các đối tượng trên frame).



Giao diện CVAT

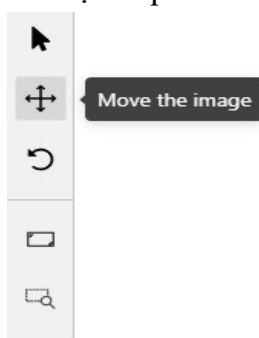
2.2.2. Các phím điều hướng cơ bản

Sử dụng các phím mũi tên bên dưới để chuyển sang frame tiếp theo/trước. Sử dụng thanh trượt/thanh cuộn để cuộn qua các frame. Hầu như mọi nút đều có một phím tắt. Để có được một gợi ý về một phím tắt, chỉ cần di chuyển con trỏ chuột lên các phím mũi tên này.



Phím điều hướng cơ bản

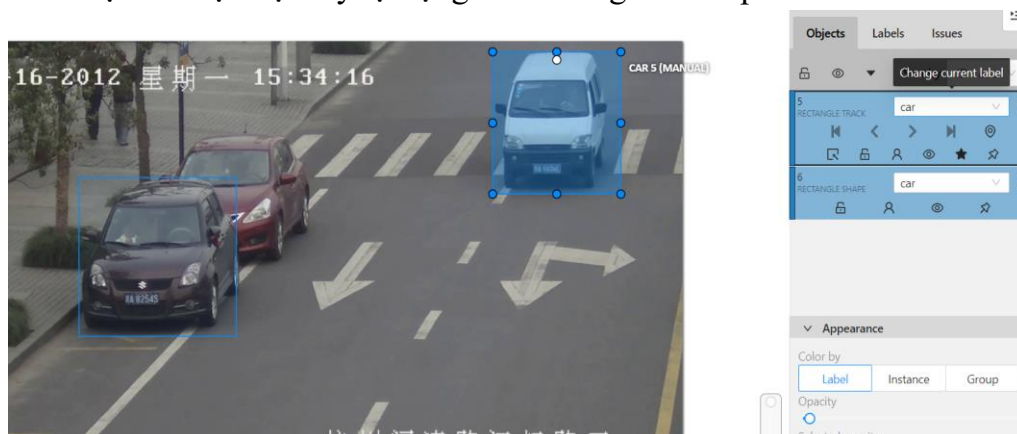
Để di chuyển hình ảnh, sử dụng chức năng trên controls sidebar. Một cách khác mà một hình ảnh có thể được di chuyển/dịch chuyển là bằng cách giữ nút chuột trái và di chuyển. Nếu lăn bánh xe chuột, thì tất cả các đối tượng được chú thích đều bị bỏ qua.



Phím tắt di chuyển hình ảnh

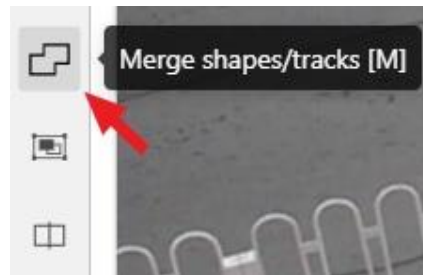
2.2.3. Cách tạo đối tượng để khảo sát trong CVAT

Ở thanh controls sidebar, với chế độ shaping, một ô lưới mới trong khung ảnh có thể được người dùng tạo ra để xử lý. Ở chế độ tracking, hình chữ nhật sẽ được nội suy tự động trên khung hình tiếp theo.



Ví dụ về 2 đối tượng shaping và tracking

Nếu đối tượng không hiển thị trên một vài khung hình và sau đó xuất hiện lại, có thể sử dụng tính năng merge để hợp nhất một số tracks riêng lẻ thành một.



Merge shape/tracks

Để tracking không hiện nữa khi vật thể bị che hoặc biến mất, tiến hành làm như sau: Bằng cách lựa chọn vật thể qua Merge (click M) ở từng thời điểm mà vật thể xuất hiện ở từng thời điểm vật thể xuất hiện 1 cách xen kẽ. Sau đó ta có thành quả như sau.



Thành quả sau khi Merge/Tracks

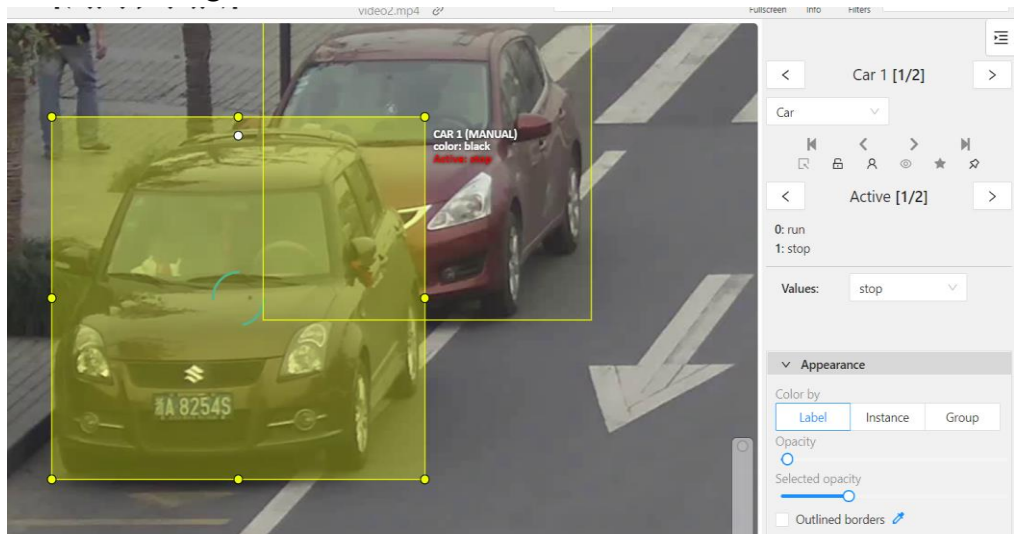
2.2.4. Attribute annotation mode

Đây là chế độ chú thích các thuộc tính cơ bản của vật thể trong CVAT



Hình ảnh minh họa

Với xe trên hình quan sát được các hiển thị thuộc tính hiện ra ở bên cạnh, như màu và trạng thái, ở đây là màu đen và trạng thái dừng. Các thuộc tính đang active để chỉnh sửa sẽ hiện màu đỏ, như:



Hình ảnh minh họa

Quan sát trên hình thấy được đang trở tới xe thứ nhất, trạng thái active với 2 sự lựa chọn là run và stop.

2.2.5. Annotation tool

Để annotation tool là một công cụ tốt, tuân theo các tiêu chí như label dễ dàng và kết quả tối ưu nhất có thể. Đặc biệt là trong các tình huống khác nhau thì các classes phải mang tính chất linh động, việc import và export data cũng khác nhau.

2.3 ƯU VÀ KHUYẾT ĐIỂM

CVAT là mã nguồn mở, sử dụng nền tảng Web giúp mọi người dễ dàng tiếp cận sử dụng. Đối với các công ty, người quản lý chỉ cần tạo một project và chia nhỏ công việc cho các nhân viên một cách dễ dàng. CVAT hỗ trợ tính năng dán nhãn bán tự động (Semi-automatic Labeling) tăng tốc độ dán nhãn các đối tượng cung như hỗ trợ nhập xuất nhiều loại file phù hợp cho nhiều đối tượng và nhiều mô hình cần những định dạng file khác nhau. Với CVAT, người sử dụng không cần phải label song song với huấn luyện tập dữ liệu hiện có. Khi huấn luyện mô hình, ta có thể sử dụng chính mô hình đó để label dữ liệu. Dữ liệu được label từ mô hình train ban đầu có thể sai số nhiều, nhưng trong quá trình ấy, nếu ta liên tục định hướng và chỉnh sửa cho mô hình, nó sẽ hoạt động tốt hơn và dần thay thế cho con người. Tuy nhiên, CVAT thường xuyên có vấn đề về hiệu suất làm việc và thiếu đi các tính năng chọn lọc và sắp xếp nâng cao.

2.4 SO SÁNH MANUAL LABELING VÀ SEMI- AUTOMATIC LABELING

Dán nhãn đối tượng là một vùng bao xung quanh đối tượng bằng các điểm được liên kết với nhau. CVAT có 2 cách để dán nhãn đối tượng là Dán nhãn thủ công (Manual Labeling) và Dán nhãn bán tự động (Semi-Automatic Labeling). Hai cách trên đều có những ưu điểm và khuyết điểm của mình nhưng trong bảng dưới đây.

Dán nhãn thủ công	Dán nhãn bán tự động
Sử dụng nhiều thao tác chuột để tạo các điểm bao quanh đối tượng Mất nhiều thời gian để dán nhãn đối tượng, không thích hợp với các mô hình có nhiều đối tượng	Cần ít thao tác chuột để hệ thống tự dán nhãn đối tượng Thời gian dán nhãn đối tượng nhanh và tiện lợi


Độ chính xác cao do người dùng phải tự dán nhãn đối tượng	Độ chính xác không cao do có thể đối tượng có một bộ phận không giống như mô tả của hệ thống (xe hơi bị bóp méo)
---	--

So sánh Dán nhãn thủ công và Dán nhãn bán tự động

CHƯƠNG 3: DỰ ÁN THỰC TẾ

3.1 TẠO LẬP DỰ ÁN

Trước khi tạo lập dự án, bước đầu tiên cần phải xác định đối tượng mình đang hướng đến. Ở đây, đối tượng cụ thể nghiên cứu sẽ là người và phương tiện (ô tô). Sau khi đã xác định được đối tượng, công việc tiếp theo sẽ là tạo lập label cho đối tượng đó, mỗi đối tượng sẽ có những đặc điểm khác nhau, vì thế ngoài việc tạo lập label, thì còn phải tạo lập thuộc tính cho label đó. Với phương tiện (ô tô) sẽ có các thuộc tính là trạng thái và màu sắc. Với người sẽ có thuộc tính là trạng thái.



Task #256356 Created by ngocyt001 on May 9th 2022

Issue Tracker
Not specified [🔗](#)

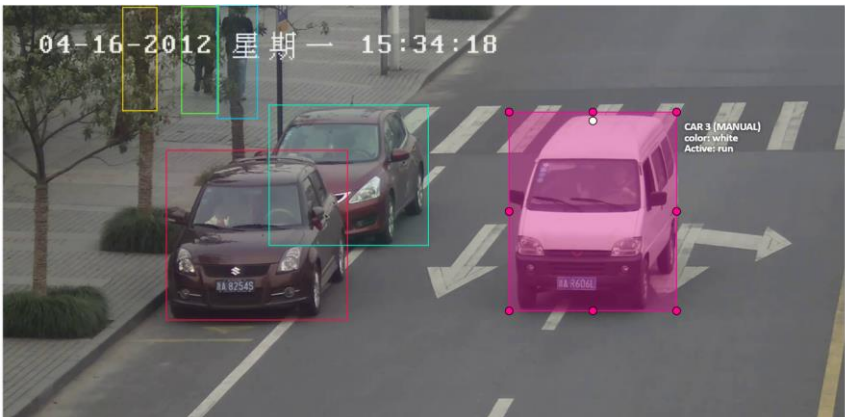
[Raw](#) [Constructor](#)

[Add label +](#) [people](#) [Car](#)

Overlap size	Segment size	Image quality
5	749	70

Thiết lập label và thuộc tính

Sau khi đã thiết lập các thông số cơ bản, tiến hành gắn nhãn cho đối tượng. Ở hai đối tượng được xét đến, các đối tượng này liên tục di chuyển trong các frame liên tục do đó công cụ tracking là cần thiết.



04-16-2012 星期一 15:34:18

CAR 3 (MANUAL)
color: white
Active: run

CAR 1 (MANUAL)
color: red
Active: run

CAR 2 (MANUAL)
color: blue
Active: run

Objects Labels Issues

Sort by ID - as...

1 RECTANGLE TRACK Car

2 RECTANGLE TRACK Car

3 RECTANGLE TRACK Car

Appearance

Color by Label Instance Group

Opacity

Selected opacity

Tracking label

3.2 PHÂN TÍCH KẾT QUẢ

Kết quả của việc gắn nhãn sẽ là tập dữ liệu huấn luyện. Tương ứng mỗi mô hình AI sẽ là mỗi tập dữ liệu khác nhau. Ở đây, chúng ta sẽ phân tích hai dạng dữ liệu phổ biến là dạng file XML và file text.

```
▼<track id="0" label="Car" source="manual">
  ▼<box frame="132" outside="0" occluded="0" keyframe="1" xtl="1431.70" ytl="35.13" xbr="1719.93" ybr="394.90" z_order="0">
    <attribute name="color">white</attribute>
    <attribute name="Active">run</attribute>
  </box>
  ▼<box frame="133" outside="0" occluded="0" keyframe="0" xtl="1427.44" ytl="38.96" xbr="1717.46" ybr="399.84" z_order="0">
    <attribute name="color">white</attribute>
    <attribute name="Active">run</attribute>
  </box>
  ▼<box frame="134" outside="0" occluded="0" keyframe="0" xtl="1423.18" ytl="42.79" xbr="1714.99" ybr="404.78" z_order="0">
    <attribute name="color">white</attribute>
    <attribute name="Active">run</attribute>
  </box>
  ▼<box frame="135" outside="0" occluded="0" keyframe="0" xtl="1418.93" ytl="46.63" xbr="1712.51" ybr="409.72" z_order="0">
    <attribute name="color">white</attribute>
    <attribute name="Active">run</attribute>
  </box>
▼<track id="10" label="Car" source="manual">
  ▼<box frame="0" outside="0" occluded="0" keyframe="1" xtl="606.81" ytl="226.90" xbr="970.08" ybr="546.34" z_order="0">
    <attribute name="color">red</attribute>
    <attribute name="Active">stop</attribute>
  </box>
  ▼<box frame="1" outside="0" occluded="0" keyframe="0" xtl="606.81" ytl="226.90" xbr="970.08" ybr="546.34" z_order="0">
    <attribute name="color">red</attribute>
    <attribute name="Active">stop</attribute>
  </box>
  ▼<box frame="2" outside="0" occluded="0" keyframe="0" xtl="606.81" ytl="226.90" xbr="970.08" ybr="546.34" z_order="0">
    <attribute name="color">red</attribute>
    <attribute name="Active">stop</attribute>
  </box>
```

Export XML file

Các dữ liệu trả về thể hiện đặc tính của label đó trong frame đang xét. Các đặc điểm điểm bao gồm vị trí của đối tượng trong frame và các đặc tính của đối tượng

CHƯƠNG 4: TÀI LIỆU THAM KHẢO

1. TaSeeMba/Github , CVAT user guide,
<https://github.com/TaSeeMba/cvat/blob/master/cvat/apps/documentation/user_guide.md>, 10.5.2022.
2. Alberto Rizzoli (2022), The complete Guide to CVAT - Pros & Cons [20220,
<<https://www.v7labs.com/blog/cvat-guide>>, 15.5.2022.
3. Alberto Rizzoli (2022), 6 AI Applications Shaping the Future of Retail,
<<https://www.v7labs.com/blog/ai-in-retail>>, 15.5.2022.
4. Tomas Laurinavicius (2022), 7 Life-Saving AI use cases in Healthcare,
<<https://www.v7labs.com/blog/ai-in-healthcare>>, 15.5.2022.