stop_sign_annotation_protocol - Dan Szwec

The document first presents the assumptions and trade-offs behind the labeling protocol, followed by the detailed annotation instructions. It concludes with a brief explanation of how the labeled dataset supports effective training of stop-sign behavior models.

Assumptions:

1. The detection algorithm does not have 100% accuracy, and in some cases we may receive videos without any sign.

2. There is a technical guide for annotators explaining how to review the videos (specifically using a video player that allows frame-level navigation), as well as a system that cuts the videos and uploads them into the annotation tool (CVAT).

3. No additional information is available besides the video frames themselves. Annotators cannot access metadata such as speed, GPS, or sensor readings; all annotations must be based solely on what is visible in the frames

4. **I do not have enough examples to generalize and capture all edge cases. To address this, I would require more data and provide better instructions and visual examples for the taggers to ensure consistent and accurate annotations.**

5. **The annotation instructions include a section on Examples and Edge Scenarios, provided to the annotators by the team, offering visual examples and full-video annotation cases to help ensure consistent labeling.**

6. **Finally, in real life I will also test the system on some of my own videos and share it with the development team for hands-on experimentation.**

Tradeoffs:

Our annotation guide demands significant time and concentration from annotators. While this lowers throughput, it ensures higher-quality, richer labels, reducing the need for re-annotation and providing data that may benefit future models across the company (maybe including advanced systems like Nexar's CityStream and Live HD Maps).

Near Stop Sign Behavior - Annotation instructions

This document defines a concise, objective per-video labeling protocol for stop-sign behavior in dashcam clips, optimized for high inter-annotator consistency and strong training signal.

Annotations are divided into two types: per-frame (bounding boxes for objects) and per-video (behavioral outcomes). All decisions are made based on visual analysis of the video content only, without access to speed, GPS, or sensor data.

Key Requirement: Consistency across annotators and objective analysis based solely on what is visually observable are essential. Do not make assumptions beyond what you can clearly see in the video. When in doubt, use "uncertain" rather than guessing.

Document Structure:

1. Per-Video Annotation Instructions

Scope

Watch the entire clip **once** before starting any annotations. Use this initial viewing to understand the scene, the vehicles, and the stop sign context.

For all fields, “U” represents uncertainty. Use it whenever you cannot clearly determine the correct value from the video.

If the video is corrupted, write **“corrupted”** in the ‘video_id’ field and skip to the next video.

Required Fields:

- **video_id**: unique identifier for the clip

- **more_than_one_stop_sign:** yes / no
  mark **yes** if there are two or more stop signs in the video; otherwise, mark **no**.

  If **yes**, you finish with this video. Cut the video according to the technical guide, rename it, and process the resulting cut videos

- **stop_sign_present:** yes / no / U
  Mark **yes** only if the **entire stop sign** is clearly visible; otherwise, mark **no** if absent or **U**

- **sign_first_seen_frame:** digit (e.g., 1, 2, 12) / None
  Frame number where the stop sign is first clearly visible; use None if the stop_sign_present is 'no' or U.

- **sign_last_seen_frame:** digit (e.g., 1, 2, 12) / None
  mark the frame number where the stop sign is last clearly visible; use None if the stop_sign_present is 'no' or U.

- **sign_condition:** good / poor / U
  mark good if the stop sign is intact and clearly readable, poor if damaged, faded, or partially destroyed, U if unclear.

- **multi_way_stop (ALL WAY/4-WAY plate visible):** ALL WAY / 4-WAY / no / U
  mark ALL WAY or 4-WAY if the corresponding plate is clearly visible; mark no if absent, blocked, or corrupted; mark U if unclear.

- **stop line present:** yes / no / U
  Mark yes only if the entire stop sign is clearly visible; otherwise, mark no if absent or U

- **line_condition:** good / poor / U
  mark good if the stop line is intact and clearly visible, poor if faded, worn, or damaged, U if unclear.

- **stop_frame_start:** digit(e.g., 5,6,7) / None
  Frame number where the vehicle comes to a full stop; enter a digit for the frame or None if 'stop_behivor' is not "full_stop".

- **stop_frame_end:** digit (e.g., 10, 11, 12) / None
  Frame number where the vehicle starts moving again after a full stop; enter a digit if 'stop_behivor' is full_stop, or None if the vehicle does not fully stop.

- **stop_behavior:** full_stop / slow_down / no_stop / U / None
  full_stop: if the vehicle fully stops
  slow_down if it decreases speed but does not stop.
  no_stop if it keeps the same speed;
  use None if stop_sign_present ≠ yes, or U if uncertain.

- **stop_location:** before_line / after_line / None / U
  mark based on where the vehicle stops; use None if stop_sign_present = no, or **U** if unclear.

- **maneuver:** straight / left / right / backward / U
  mark straight if the vehicle does not turn during the video, left or right for corresponding turns, backward if reversing, U if unclear. Examples provided.

- **pedestrian_or_cyclist_crossing:** yes / no / U
  mark yes if a pedestrian or cyclist crosses the road anywhere in the video, no if absent, U if unclear.

- **cross_traffic_present:** yes / no / U
  mark yes if crossing traffic is clearly visible, no if absent, U if unclear.

- **weather:** clear / rain / fog / other / U
  mark based on visible weather conditions; U if unclear.

- **ego_collision:** frame number (e.g., 1,2,3) / None / U
  mark the frame where the ego vehicle collides, None if no collision occurs, U if unclear.

- **other_collision:** frame number (e.g., 1,2,3) / None / U
  mark the frame where another vehicle or object collides, None if no collision occurs, U if unclear.

- **construction_present:** yes / no / U
  mark yes if construction is clearly visible, no if absent, U if unclear.

- **road_condition:** good / poor / U
  mark good if the road surface is intact, poor if damaged, faded, or uneven, U if unclear.

2. Per-Frame Annotation Instructions

Scope:

If **stop_sign_present = yes**, load the video into the “by_frame_annotation_tool” for annotation; otherwise, skip to the next video.

Required Fields:

1. **Stop signs**: annotate every stop sign that is at least **50% visible** in the frame.

2. **Stop lines**: annotate every stop line that is visible; see examples for reference.

3. **Pedestrians**: annotate only if the **upper body** is visible.

4. **Cyclists**: annotate only if the **entire cyclist** is visible; create **one box for the cyclist** and **one for the person riding it** if the rider stands at the lower half of the upper body.

5. **Other traffic signs**: annotate the same way as stop signs, only if at least **50% visible.**

6. **Vehicles**: annotate all visible vehicles; include the **entire visible part**; if less than 50% is visible, do not annotate.

7. **Obstacles**: annotate cones, barriers, debris, or other visible objects on or near the road.

8. **Speed cameras**: annotate any visible speed camera.

General Rules for All Classes:

1. Use **one bounding box per instance**.

2. If an instance is **partially occluded**, annotate only the visible portion.

3. Do **not** annotate objects outside the frame, too far to identify, or not clearly visible.

4. Maintain **tight and consistent bounding boxes** (centered around the visible object, including all clearly visible parts).

5. Examples and edge cases are provided in the **Examples and Edge Scenarios** section.

short explanation of how your labeling scheme supports training an effective model for this task:

The labeling protocol ensures high-quality, consistent annotations for stop-sign behavior by clearly defining each field. By providing both per-frame and per-video labels, along with features such as road conditions, cross traffic, multi-way stops, vehicles, pedestrians, and other objects, the dataset becomes very rich and detailed.
Consistency across annotators reduces noise in the training data, and these comprehensive labels allow models to learn precise stop-sign recognition and driver behavior patterns.
This structured approach balances efficiency with coverage, making the dataset suitable for training robust models and supporting potential future tasks.