

Slide 1/10 · TV1 · 0:30 · bắt đầu 0:00

Mở đầu: Camera Fisheye 360° & SVM

Mọi người đã từng lùi xe và nhìn màn hình 360 độ chưa?

Bình thường chúng ta không thể nhìn hết mọi thứ quanh xe. Vì vậy, người ta lắp bốn camera ở phía trước, phía sau và hai bên gương.

Câu nhớ nhanh: SVM: Bốn mắt camera giúp xe nhìn quanh mình.



Slide 2/10 · TV1 · 0:30 · bắt đầu 0:30

SVM là gì và dùng khi nào?

Những camera này giúp xe nhìn được các vật cản sát bên, đặc biệt khi đỗ hoặc lùi xe.

Nhưng có một vấn đề: camera nhìn càng rộng thì hình càng méo. Và đó chính là thử thách khi chúng ta gán nhãn dữ liệu.

Câu nhớ nhanh: Fisheye: Nhìn càng rộng, mép càng méo.



Slide 3/10 · TV2 · 1:30 · bắt đầu 1:00

Vì sao fisheye nhìn rộng nhưng ảnh bị cong?

Mọi người tưởng tượng đang nhìn qua mắt thần ở cửa nhà.

Nó giúp mình thấy một vùng rất rộng, nhưng cửa, tường và những đường thẳng lại bị cong.

Camera mắt cá cũng vậy. Nó giống như cố nhét cả một khung cảnh rất rộng vào một bức ảnh nhỏ.

Càng ra mép ảnh, vật thể càng dễ bị biến dạng. Một chiếc xe có thể trông nghiêng đi, còn vạch đường thẳng có thể thành một đường cong.

Câu hỏi là: Khi gán nhãn, mình vẽ đường thẳng như ngoài đời hay vẽ cong như trong ảnh?

Tương tác nhanh

- Cho cả lớp chọn A – đường thẳng, B – đường cong.
- Chưa cần giải thích dài, chuyển ngay sang slide 4.



Slide 4/10 · TV3 · 1:15 · bắt đầu 2:30

Thấy cong thì vẽ theo cong

Ví dụ mọi người muốn vẽ một khúc cua.

Nếu lấy cây thước kẻ một đường thẳng từ đầu tới cuối thì chắc chắn sẽ lệch khỏi đường.

Nhưng nếu nối nhiều đoạn thẳng ngắn, chúng ta sẽ bám được khúc cua.

Gán nhãn cũng y như vậy. Mình dùng công cụ Polyline, chấm nhiều điểm theo đúng vạch nhìn thấy.

Chỗ nào cong nhiều thì thêm nhiều điểm. Chỗ nào gần thẳng thì ít điểm hơn.

Bài này yêu cầu vẽ lệch không quá 5 pixel. **Không cần làm cho đường thẳng lại, chỉ cần bám đúng ảnh gốc.**

Câu nhớ nhanh: Polyline: Cong chỗ nào, thêm điểm chỗ đó.

Lưu ý: Ngưỡng 5 px là quy định của bài học, không phải chuẩn chung cho mọi hệ thống.



Slide 5/10 · TV3 · 1:00 · bắt đầu 3:45

Demo A: Vẽ polyline thế nào cho đúng?

Đây là vạch đường thật trên ảnh. Nếu chỉ chấm hai đầu, đoạn giữa sẽ bị lệch.

Bây giờ mình thêm vài điểm dọc đường cong. Mọi người thấy đường vẽ đã bám sát vạch thực tế hơn.

Chúng ta không cần thật nhiều điểm. Chỉ cần đủ điểm để vẽ chính xác.

Thao tác trên CVAT

1. Chọn Polyline và vạch đường cong.
2. Thử nối 2 điểm để thấy vị trí bị lệch.
3. Vẽ lại bằng nhiều điểm bám theo vạch.
4. Zoom vào để cho lớp thấy sự khác biệt.

Lưu ý: Slide có 3 bước (chấm hai đầu, thêm điểm, zoom); lời thoại có thêm bước "chọn Polyline". Hai bên khớp ý. Chuyển sang CVAT thật khi nói đến đây, rồi quay lại slide 6.



Slide 6/10 · TV4 · 1:30 · bắt đầu 4:45

Từ 4 ảnh thành góc nhìn 360°

Vậy tại sao trên màn hình ô tô, mình lại nhìn được chiếc xe từ trên xuống giống như có một chiếc flycam bay phía trên?

Thực ra không có flycam nào cả.

Hệ thống lấy hình từ bốn camera, xử lý độ méo rồi ghép lại thành một hình nhìn từ trên xuống, gọi là Top View.

Với mặt đường và vạch sơn, cách này khá hợp lý. Nhưng với người, xe hoặc cột điện thì không hoàn toàn đúng.

Mọi người tưởng tượng lấy một tấm ảnh có cây cột rồi ép xuống mặt bàn. Cây cột có thể bị kéo dài hoặc biến dạng.

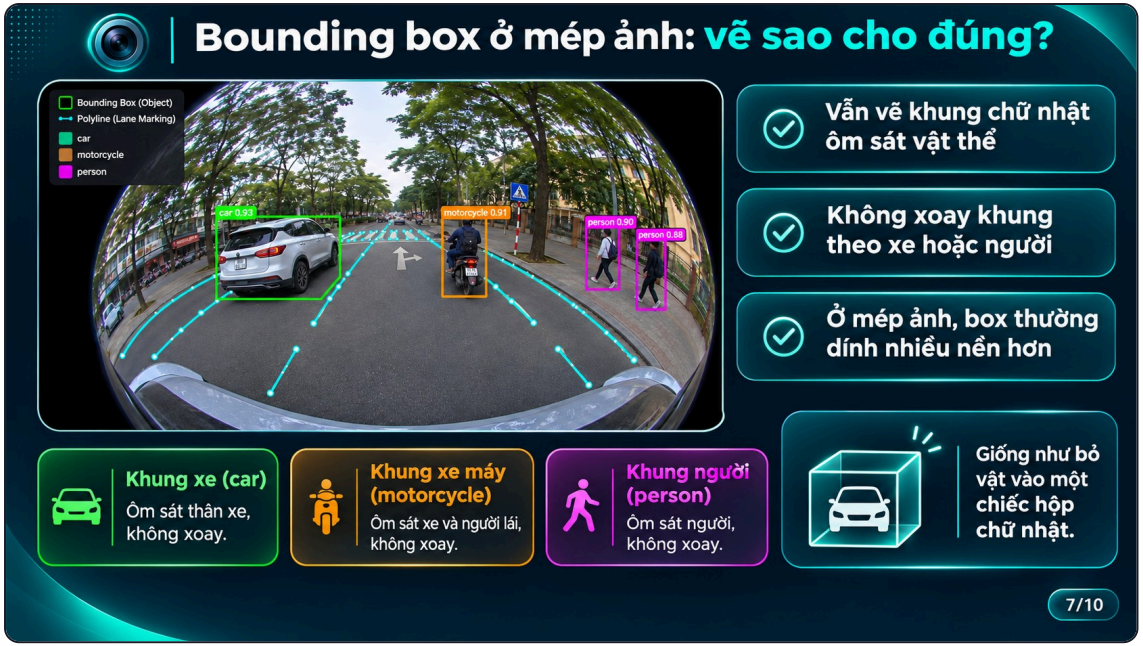
Vì vậy, trong bài toán gán nhãn xe và người của chúng ta, mình ưu tiên ảnh fisheye gốc. Còn 3D Bowl giống như thay mặt bàn bằng một lòng bát để góc nhìn trông tự nhiên hơn.

Điểm cần phân biệt

- Top View và 3D Bowl dùng để hiển thị; dữ liệu phải gán trên hệ tọa độ mà bài toán huấn luyện yêu cầu.
- Bài này quy định gán vật thể trên ảnh fisheye gốc.

Câu nhớ nhanh: Top View: Bốn camera ghép thành góc nhìn từ trên.

Lưu ý: Ảnh slide chỉ có Top View, chưa có hình 3D Bowl; câu về 3D Bowl nói bằng lời (hoặc mở viewer.html trong repo nếu cần minh họa).



Slide 7/10 · TV5 · 0:40 · bắt đầu 6:15

Bounding box ở mép ảnh: vẽ sao cho đúng?

Mọi người tưởng tượng bỏ một chiếc xe đang nằm nghiêng vào một chiếc thùng carton hình chữ nhật.

Dù mình đóng thùng sát nhất có thể, bốn góc vẫn thừa khoảng trống.

Bounding Box trên ảnh fisheye cũng vậy.

Ở gần tâm ảnh, chiếc xe nhìn tương đối bình thường. Nhưng ở sát mép, xe có thể nghiêng và méo, khiến khung hình chữ nhật chứa nhiều nền hơn.

Mình vẫn vẽ khung thẳng, ôm sát phần xe nhìn thấy. Không xoay và không cố nới khung.

Câu nhớ nhanh: Bounding Box: Như đóng thùng carton quanh vật.



Slide 8/10 · TV5 · 0:35 · bắt đầu 6:55

Khi nào bật edge_zone?

Nếu tâm khung nằm ngoài vùng 80% bán kính ảnh theo guideline, mình bật cờ edge_zone, để hệ thống biết đây là vật thể ở vùng dễ bị méo.

> **Bổ sung:** Chỉ bật cờ theo guideline, không đổi kiểu box: box vẫn thẳng và ôm sát vật.

Câu nhớ nhanh: edge_zone: Đánh dấu vật nằm ở vùng dễ méo.

Lưu ý: Ngưỡng 80% là quy định của bài học. Câu "bổ sung" lấy từ chữ trên slide để đủ thời lượng 35 giây.



Slide 9/10 · TV6 · 1:15 · bắt đầu 7:30

3 lỗi dễ gặp khi gán nhãn fisheye (Demo B)

Ở đây chúng ta có ba loại đối tượng: ô tô, xe máy và người. Mỗi đối tượng được vẽ bằng một khung hình chữ nhật.

Còn vạch đường được biểu diễn bằng Polyline, vì nó dài và có thể cong.

Sau khi vẽ xong, có ba việc cần kiểm tra: vẽ có đúng vị trí không, đã chọn đúng camera chưa, và vật ở mép có được đánh dấu đúng không.

Nhãn đẹp mà sai loại camera thì kết quả vẫn có thể sai.

Sau khi xuất dữ liệu, chúng ta cũng phải thử mở lại để chắc chắn không mất nhãn.

> **Bổ sung:** Ba lỗi hay gặp nhất: quên hoặc chọn sai tag camera; vẽ polyline thành đường thẳng; gán nhãn lên thân xe mình hoặc vùng đen ngoài ảnh. Chỉ cần tránh ba lỗi này là chất lượng tăng lên rất nhiều.

Thao tác CVAT

- Chỉ vào xe ô tô: Bounding Box màu xanh.
- Chỉ vào xe máy: Bounding Box màu cam.
- Chỉ vào người: Bounding Box màu hồng.
- Chỉ vào vạch đường: Polyline màu cyan.
- Mở metadata để kiểm tra tag camera và thuộc tính edge_zone.

Câu nhớ nhanh: QC: Vẽ đúng, chọn đúng camera, kiểm tra trước khi nộp.

Lưu ý: Ảnh slide nói về "3 lỗi", lời thoại gốc là Demo B. Gợi ý: mở CVAT làm phần Demo B trước, quay lại slide này để chốt 3 lỗi (đoạn bổ sung ở cuối).



Slide 10/10 · TV7 · 1:15 · bắt đầu 8:45

6 điều cần nhớ

Để kết thúc, mọi người chỉ cần nhớ sáu điều này.

Đầu tiên là chọn đúng camera. Tiếp theo, thấy đường cong thì vẽ theo đường cong, thấy xe hay người thì dùng khung ôm sát.

Vật ở mép ảnh cần kiểm tra thêm edge_zone. Những phần như thân xe mình và vùng đen ngoài ống kính thì không gán.

Cuối cùng, luôn kiểm tra lại trước khi nộp.

Nếu chỉ nhớ một câu sau bài này, thì đó là: **Camera thấy thế nào, mình gán đúng như thế ấy, theo đúng quy định của từng camera.**

> **Lưu ý:** Cần thống nhất: ảnh có điều "Vẽ theo ảnh gốc, không vẽ theo tưởng tượng" nhưng không có "Kiểm tra nhãn và file xuất trước khi nộp"; lời thoại gốc có điều đó. Câu chốt trên ảnh là "camera thấy sao, mình gán nhãn như vậy".