

BÁO CÁO TUẦN 4

Tên đề tài: Nhận diện bối cảnh trong khung ảnh sử dụng mô hình YOLOv8 và BLIP

Thời gian: Tuần 4 (từ ngày 27/4 đến ngày 4/5)

1. Công việc đã thực hiện

- Tối ưu hóa tham số YOLOv8:**
 - Tiến hành tối ưu hóa tham số mô hình YOLOv8 như learning rate, batch size và số epoch để nâng cao độ chính xác.
 - Sử dụng các kỹ thuật **data augmentation** và **learning rate scheduling** để giúp mô hình học tốt hơn.
- Tiền xử lý dữ liệu:**
 - Tiến hành chuẩn hóa dữ liệu đầu vào (cắt ảnh, thay đổi kích thước, normalization) để phù hợp với yêu cầu của mô hình YOLOv8.
 - Tiến hành **detection** đối tượng trong ảnh bằng YOLOv8, sau đó sử dụng **mô hình BLIP** để sinh mô tả cho các bối cảnh trong ảnh.
- Sử dụng mô hình BLIP:**
 - Sau khi YOLOv8 nhận diện đối tượng và bối cảnh, kết quả sẽ được đưa vào mô hình **BLIP** để sinh mô tả ngắn gọn về bối cảnh của ảnh.
 - Ví dụ, ảnh của một quán cà phê có thể được mô tả là “Một quán cà phê đông đúc với mọi người ngồi uống cà phê”.

2. Kết quả đạt được

- Mô hình YOLOv8** đã cải thiện độ chính xác lên khoảng **88-90%** sau khi tối ưu hóa tham số.
- Mô hình BLIP** sinh ra các mô tả ảnh chính xác hơn, cung cấp mô tả rõ ràng về bối cảnh và các đối tượng trong ảnh.
- Ví dụ, đối với một ảnh của một quán cà phê, mô hình đã sinh ra mô tả “Một quán cà phê với nhiều người ngồi xung quanh bàn”, giúp hệ thống hiểu rõ hơn về bối cảnh.

- Kết hợp giữa YOLOv8 và BLIP đã tạo ra một hệ thống mạnh mẽ giúp nhận diện và mô tả bối cảnh trong ảnh.
-

3. Khó khăn gặp phải

- Một số bối cảnh phức tạp vẫn khó mô tả chính xác, nhất là khi có quá nhiều đối tượng trong cùng một không gian hoặc ánh sáng không đồng đều.
 - Mô hình BLIP đôi khi không thể sinh ra mô tả chính xác khi không có đủ ngữ cảnh về các đối tượng trong ảnh.
 - Cần tài nguyên tính toán mạnh để chạy đồng thời cả YOLOv8 và BLIP, gây ảnh hưởng đến thời gian huấn luyện và kiểm tra.
-

4. Kế hoạch tuần 5

- Tiếp tục thu thập thêm dữ liệu cho các bối cảnh thiếu và cải thiện mô hình để mô tả chính xác hơn trong những tình huống khó khăn.
- Tinh chỉnh mô hình BLIP để sinh mô tả ảnh chính xác hơn, đặc biệt trong những bối cảnh phức tạp.
- Tạo giao diện giúp người dùng có thể upload ảnh và nhận diện bối cảnh kèm mô tả tự động.
- Tối ưu hóa quá trình huấn luyện để giảm thời gian và tài nguyên cần thiết cho việc sử dụng cả YOLOv8 và BLIP.