BÁO CÁO TUẦN 6 (05/05 - 12/05)

Mục tiêu tuần:

- Hoàn thiện hệ thống nhận diện bối cảnh bằng cách tích hợp mô hình phát hiện đối tượng YOLOv8 đã huấn luyện với mô hình mô tả ảnh BLIP.
- Tạo mô tả ngữ cảnh ảnh tự động từ các đối tượng đã phát hiện.

✓ Các công việc đã hoàn thành:

1. Huấn luyện mô hình YOLOv8:

- Sử dụng tập dữ liệu COCO để fine-tune YOLOv8.
- Mô hình đã phát hiện chính xác các đối tượng phổ biến như người, động vật,
 xe, đồ vật, cây cối...

2. Tích hợp BLIP (Bootstrapped Language Image Pretraining):

- o BLIP được sử dụng để tạo chú thích ảnh (caption) từ ảnh gốc.
- Caption tao ra phản ánh tổng thể nôi dung ảnh ở mức ngữ nghĩa.

3. Kết hợp YOLOv8 và BLIP để tạo ngữ cảnh:

- Sau khi YOLO phát hiện các đối tượng, kết quả được kết hợp với caption từ
 BLIP để mô tả ngữ cảnh cụ thể hơn.
- Ví dụ: "Một người đang cưỡi ngựa trong cánh đồng" trong đó YOLO xác định "người", "ngựa", "cánh đồng" và BLIP bổ sung phần mô tả hành động và mối quan hệ.

4. Tạo giao diện người dùng:

 Giao diện bằng Tkinter đơn giản cho phép chọn ảnh, hiển thị kết quả YOLO và BLIP.

5. Tự động dịch mô tả sang tiếng Việt:

 Sử dụng Google Translate API để dịch mô tả từ tiếng Anh sang tiếng Việt theo phương pháp dùng thư viện googletrans.

🧠 Kết quả đạt được:

- Hoàn thành hệ thống AI nhận diện bối cảnh trong ảnh hoạt động ổn định.
- Hệ thống đầu vào là ảnh, đầu ra là ngữ cảnh bằng tiếng Việt.
- Giao diện dễ sử dụng, phù hợp người dùng không chuyên.

🚧 Khó khăn gặp phải:

- Lỗi khi tải mô hình BLIP do không có mạng hoặc lỗi tên miền (huggingface.co).
- Thiếu dung lượng khi sử dụng Git để đẩy code lên GitHub, dẫn đến phải loại bỏ thư mục .venv.