Thông tin chung của nhóm

- Lớp: CS2205.CH190
- Link Github của nhóm: https://github.com/thonglb19/CS2205.CH190
- Link YouTube video: https://youtu.be/4bxITdHNF_s
- Họ và Tên: Lê Bá Thông
- MSSV: 240101078

Giới thiệu

Few-shot Object Detection (FSOD) là bài toán phát hiện đối tượng khi chỉ có rất ít mẫu huấn luyện. Các phương pháp hiện tại còn hạn chế trong việc sinh ảnh có bố cục và ngữ nghĩa hợp lý.

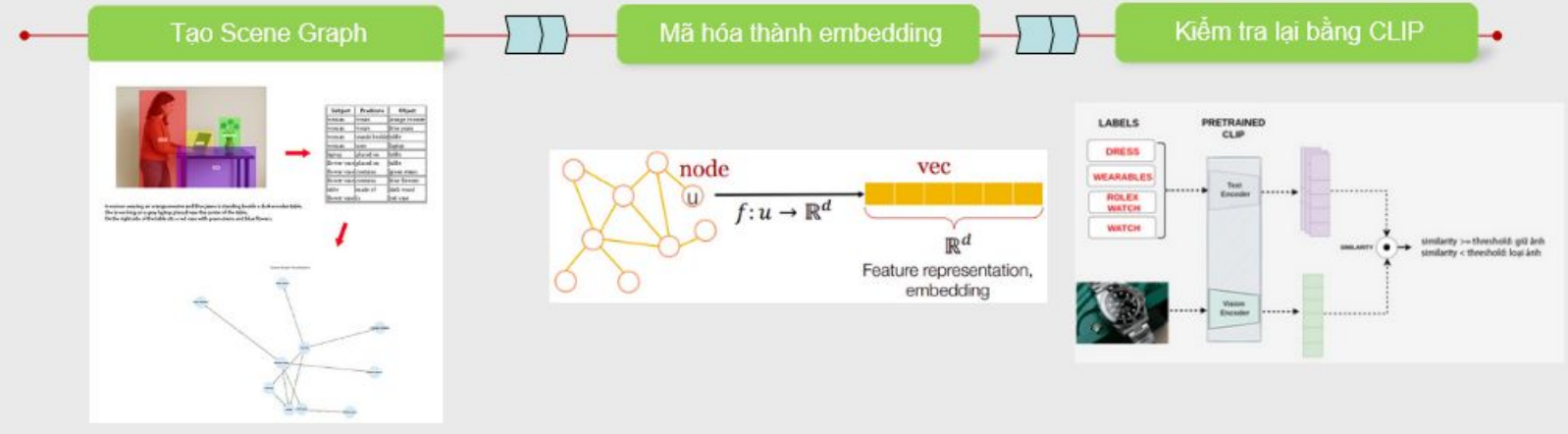
Đề tài đề xuất hệ thống sinh dữ liệu có kiểm soát, kết hợp:

- **GNN** để mã hóa scene graph từ bounding boxes, segmentation và mô tả từ LLM;
- **Diffusion model** để sinh ảnh đúng ngữ cảnh;
- **CLIP** để đánh giá và lọc ảnh tạo sinh.

Hệ thống giúp tạo tập dữ liệu tổng hợp chất lượng cao, hỗ trợ huấn luyện hiệu quả FSOD trên COCO-FS, PASCAL-FS.

Mục tiêu

- Đề xuất mô hình sinh dữ liệu có kiểm soát kết hợp Diffusion Models và GNN cho bài toán FSOD.
- Xây dựng scene graph từ bounding boxes, segmentation maps và mô tả từ LLM, làm điều kiện cho quá trình sinh ảnh.
- Tích hợp CLIP để đánh giá ảnh tạo sinh, chọn lọc mẫu có chất lượng và khớp ngữ



Nội dung và Phương pháp

1. Nội dung

- Khảo sát FSOD, mô hình Diffusion, GNN và CLIP để xây dựng cơ sở lý luận.
- Chuẩn bị dữ liệu đầu vào (bounding boxes, segmentation, mô tả từ LLM) và tạo scene graph.
- Áp dụng GNN để tạo embedding cảnh, làm điều kiện cho Diffusion Model sinh ảnh.
- Dùng CLIP để đánh giá, lọc ảnh tạo sinh khớp với mô tả.
- Tạo tập dữ liệu tổng hợp và huấn luyện mô hình FSOD, đánh giá so sánh với baseline.

Nội dung và Phương pháp

2. Phương pháp

- Giai đoạn 1: Khảo sát các nghiên cứu liên quan về FSOD, diffusion, GNN và CLIP. Xác định bộ dữ liệu thực nghiệm (COCO-FS, PASCAL-FS).
- Giai đoạn 2: Chuẩn hóa đầu vào (bounding boxes, segmentation maps, mô tả từ LLMs). Xây dựng scene graph và huấn luyện GNN để tạo embedding cấu trúc cảnh.
- Giai đoạn 3: Tích hợp embedding từ GNN vào mô hình diffusion để sinh ảnh có kiểm soát theo bố cục và ngữ nghĩa.
- Giai đoạn 4: Dùng CLIP đánh giá ảnh tạo sinh, lọc các ảnh có độ tương đồng cao với mô tả văn bản.
- Giai đoạn 5: Sử dụng dữ liệu tạo sinh để huấn luyện mô hình FSOD, đánh giá hiệu quả so với baseline và hoàn thiện báo cáo nghiên cứu.

Kết quả dự kiến

- Cải thiện hiệu suất FSOD: Gia tăng đáng kể độ chính xác (mAP) trên COCO-FS, PASCAL-FS so với baseline nhờ sử dụng dữ liệu tạo sinh.
- Xây dựng quy trình tạo ảnh có kiểm soát: Sinh ảnh tổng hợp đa dạng, đúng bố cục và ngữ nghĩa theo điều kiện đầu vào.
- Tạo tập dữ liệu tăng cường chất lượng cao: Kết hợp dữ liệu thực và ảnh tạo sinh, được CLIP đánh giá và chọn lọc.
- Khẳng định vai trò của GNN và CLIP: GNN hỗ trợ ảnh tạo sinh có điều kiện, CLIP đảm bảo tính liên quan ngữ nghĩa của dữ liệu đầu ra.

Tài liệu tham khảo

[1] H. Fang, B. Han, S. Zhang, S. Zhou, C. Hu, and W.-M. Ye, “Data Augmentation for Object Detection via Controllable Diffusion Models,” 2024.

[2] A. Abdullah, N. Ebert, and O. Wasenmüller, “Boosting Few-Shot Detection with Large Language Models and Layout-to-Image Synthesis,” 2024. [Online]. Available: https://link.springer.com/conference/accv

[3] D. Sridhar, A. Peri, R. Rachala, and N. Vasconcelos, “Adapting Diffusion Models for Improved Prompt Compliance and Controllable Image Synthesis,” 2024.

Tóm tắt

Vấn đề: Few-shot object detection (FSOD) gặp khó khăn do thiếu dữ liệu huấn luyện, đặc biệt ở các lớp thiểu số.
Giải pháp đề xuất:
→ Kết hợp Diffusion Models và Graph Neural Networks (GNN) để sinh ảnh có kiểm soát.
→ Xây dựng scene graph từ: (1) bounding boxes,(2) segmentation maps,(3) mô tả từ LLMs.
→ GNN học embedding cảnh → dùng Diffusion Models có điều kiện để sinh ảnh.
→ Dùng CLIP đánh giá ảnh tạo sinh bằng cosine similarity để lọc giữ ảnh chất lượng cao.
Kết quả kỳ vọng:
→ Nâng cao chất lượng và đa dạng dữ liệu huấn luyện.
→ Cải thiện hiệu suất FSOD trên COCO-FS và PASCAL-FS.