TẠO SINH DỮ LIỆU CÓ KIỂM SOÁT CHO PHÁT HIỆN ĐỐI TƯỢNG ÍT MẪU BẰNG MÔ HÌNH DIFFUSION

Lê Bá Thông[1]

[1] Trường ĐH Công Nghệ Thông Tin

What ?

Đề xuất một mô hình sinh dữ liệu có kiểm soát kết hợp **Diffusion Models** tạo ra dữ liệu huấn luyện chất lượng cao cho bài toán **phát hiện đối tượng ít mẫu (FSOD)**

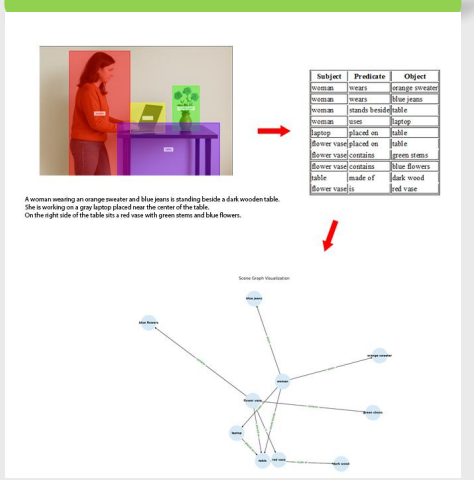
- Xây dựng **scene graph** từ bounding boxes, segmentation maps và mô tả từ LLM
- GNN mã hóa scene graph thành embedding → dùng làm điều kiện cho **Diffusion Models** tạo sinh ảnh.
- CLIP đánh giá ảnh tạo sinh → chọn lọc các ảnh khớp ngữ nghĩa để huấn luyện mô hình FSOD.

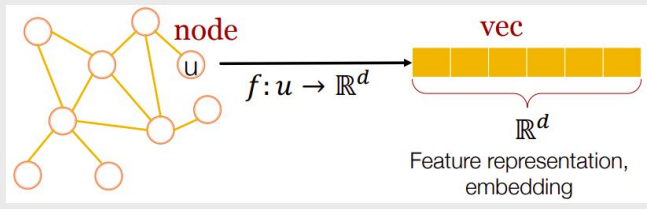
Why ?

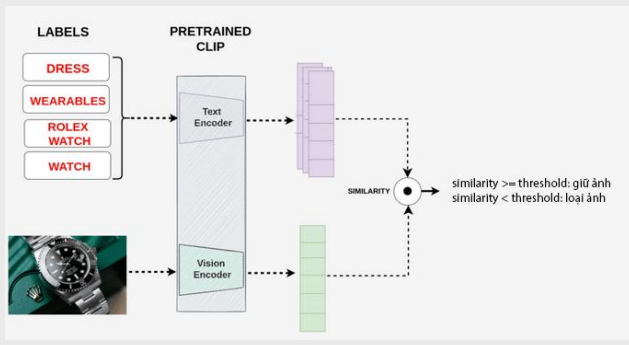
- Trong thực tế, nhiều lớp đối tượng chỉ có rất ít ảnh được gán nhãn, gây khó khăn cho việc huấn luyện mô hình học sâu. FSOD là một hướng tiếp cận đầy hứa hẹn nhưng phụ thuộc lớn vào hiệu quả của dữ liệu đầu vào.
- Các kỹ thuật augmentation truyền thống hoặc mô hình diffusion không kiểm soát tốt bố cục, mối quan hệ ngữ nghĩa và chất lượng ảnh tạo sinh, dẫn đến dữ liệu huấn luyện kém hiệu quả.
- Nhu cầu về một **quy trình sinh ảnh** có **kiểm soát** và **đánh giá được chất lượng**

Overview

Tạo Scene Graph → Mã hóa thành embedding → Kiểm tra lại bằng CLIP



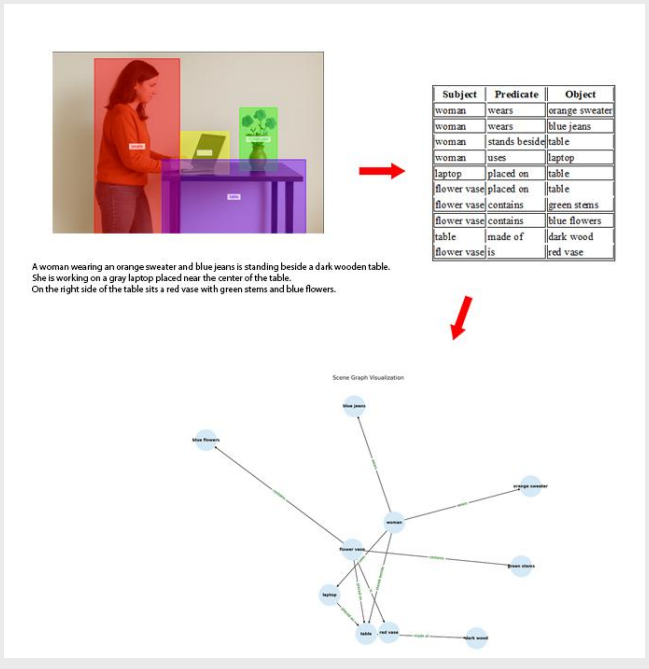




Description

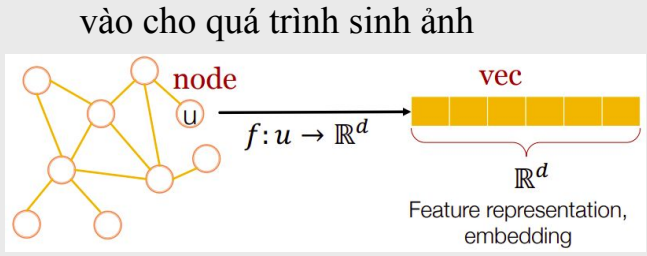
1. Tạo Scene Graph

- Từ bounding boxes, bản đồ phân đoạn ngữ nghĩa và mô tả cảnh từ LLMs tạo ra Scene Graph



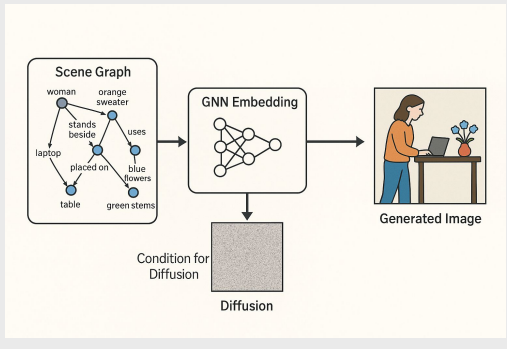
2. Mã hóa thành embedding

- Scene graph được GNN xử lý để tạo ra embedding làm điều kiện đầu vào cho quá trình sinh ảnh



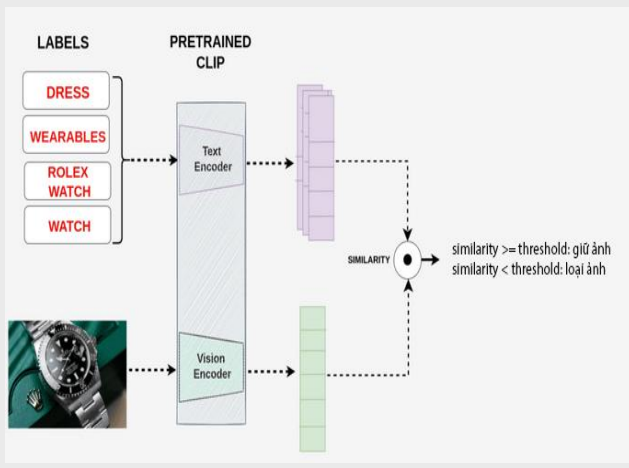
3. Sinh ảnh bằng Diffusion Models

- Embedding từ GNN sẽ được sử dụng để làm điều kiện cho mô hình Diffusion tạo sinh ảnh



4. Kiểm tra lại bằng CLIP

- Tích hợp cơ chế lọc ảnh tự động dựa trên mô hình CLIP, nhằm đánh giá mức độ khớp giữa ảnh sinh và mô tả văn bản thông qua phép đo cosine similarity



- Nếu similarity >= threshold: giữ ảnh
- Nếu similarity < threshold: loại ảnh

NII

Lê Bá Thông – Trường Đại học Công Nghệ Thông Tin
TEL : 0903153956 Email : thonglb.19@grad.uit.edu.vn