RESEARCH AND APPLICATION OF PARAMETER-EFFICIENT ADAPTATION (LORA) ON VISION-LANGUAGE MODELS FOR INDUSTRIAL ANOMALY DETECTION

Hồ Đặng Thanh Hồ - 250101021

Tóm tắt

- Lớp: CS2205.CH201
- Link Github của nhóm: github-url
- Link YouTube video: youtube-url
- Họ và Tên: Hồ Đặng Thanh Hồ
- MSHV: 250101021

Giới thiệu

Bối cảnh

- Kiểm soát chất lượng là yếu tố sống còn trong sản xuất hiện đại, đảm bảo an toàn và uy tín thương hiệu.
- Các phương pháp thủ công gây ra các vấn đề về tốc độ kiểm tra và chi phí kiểm tra như tốn kém và chiếm nhiều thời gian
- Các giải pháp Tự động hoá & Thị giác Máy tính truyền thống thiếu đi sự linh hoạt khi có sản phẩm mới và đòi hỏi chi phí huấn luyện mô hình cao

Thách thức với AI hiện tại

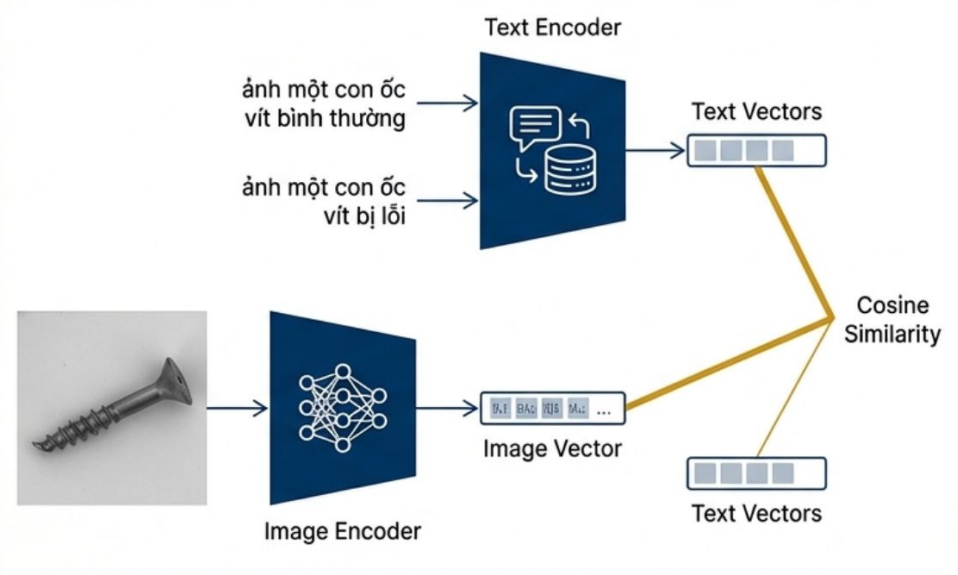
- **Yêu cầu dữ liệu lớn:** Các mô hình Deep Learning cần hàng ngàn ảnh được gán nhãn cho mỗi loại lỗi, một quá trình tốn kém và không thực tế.
- **Chi phí huấn luyện cao:** Huấn luyện lại toàn bộ mô hình cho mỗi dòng sản phẩm mới đòi hỏi tài nguyên tính toán khổng lồ.
- **Thiếu khả năng tổng quát hóa:** Mô hình được huấn luyện cho một sản phẩm thường không hoạt động tốt trên sản phẩm khác mà không có sự tinh chỉnh đáng kể.

Bước đột phá: Mô hình Ngôn ngữ–Thị giác (VLM) và năng lực Zero-Shot

Các mô hình VLM như **CLIP (Contrastive Language-Image Pre-training)** được huấn luyện trên hàng trăm triệu cặp (ảnh, văn bản) từ internet.

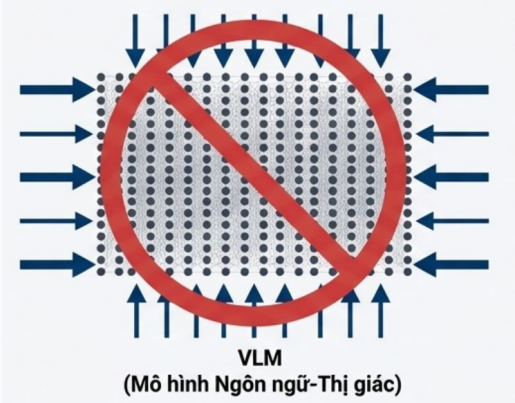
Chúng không chỉ "nhìn" thấy các pixel mà còn hiểu được các khái niệm trong ảnh thông qua ngôn ngữ tự nhiên.

Năng lực chính: Khả năng nhận dạng **"Zero-Shot"** – xác định các đối tượng hoặc khái niệm mà không cần bất kỳ ví dụ huấn luyện nào, chỉ bằng cách sử dụng một mô tả văn bản (prompt).



Thách thức thích nghi và giải pháp hiệu quả: Kỹ thuật LoRA

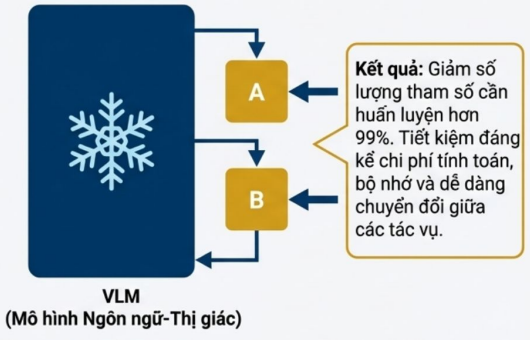
Vấn đề: Rào cản của việc Tinh chỉnh Toàn bộ (Full Fine-Tuning)



Các mô hình VLM có hàng tỷ tham số.

- Việc cập nhật tất cả các tham số này đòi hỏi tài nguyên tính toán cực lớn, nguy cơ “quên kiến thức cũ”, và phải lưu trữ một bản sao đầy đủ của mô hình cho mỗi tác vụ.

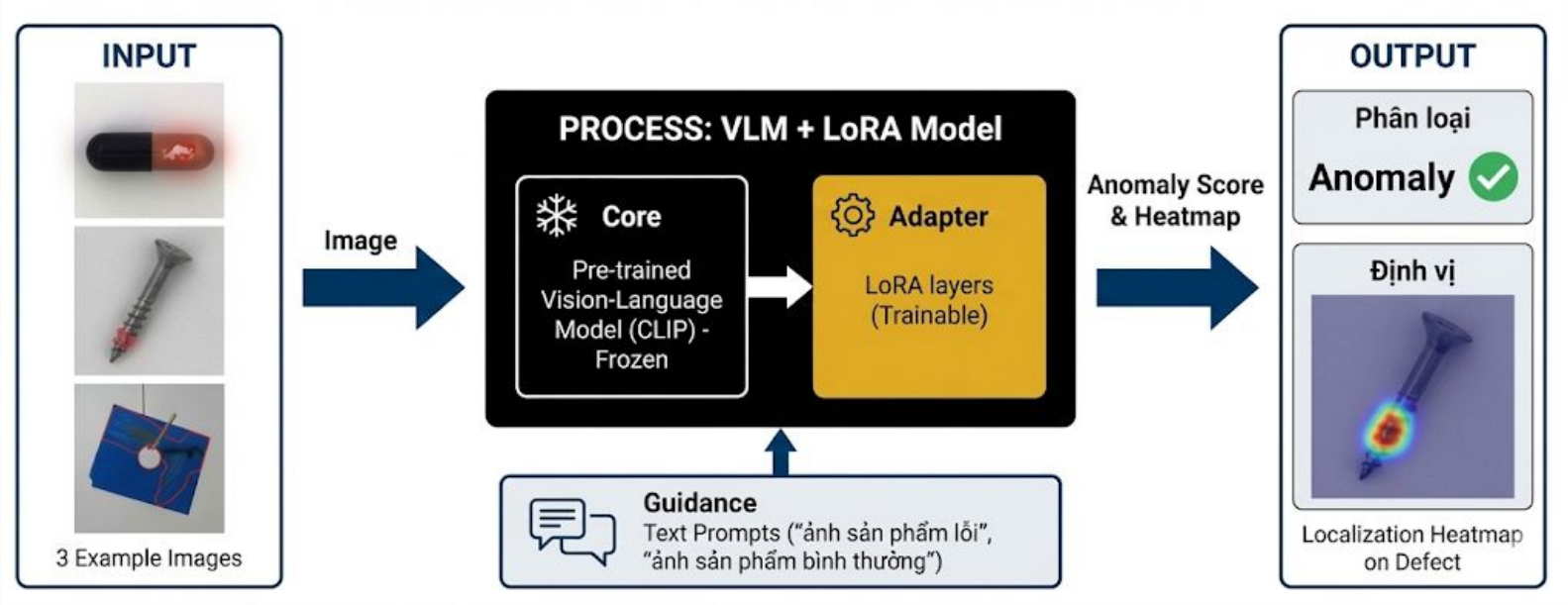
Giải pháp: Thích nghi Tham số Hiệu quả với LoRA



Chỉ huấn luyện ~0.1% tổng số tham số.

Đề xuất của đề tài: Hệ thống phát hiện bất thường linh hoạt và hiệu quả

Đề tài này nghiên cứu và ứng dụng kỹ thuật LoRA để thích nghi hiệu quả một mô hình Ngôn ngữ–Thị giác (VLM) cho bài toán phát hiện và định vị bất thường trên các sản phẩm công nghiệp, hướng tới một giải pháp có khả năng mở rộng, tiết kiệm chi phí và dễ dàng tùy chỉnh cho các dây chuyền sản xuất khác nhau.



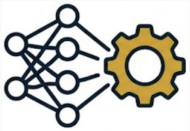
Nội dung và Phương pháp thực nghiệm

1. Chuẩn bị Dữ liệu & Môi trường

Dataset: Sử dụng bộ dữ liệu MVTec AD – một tiêu chuẩn công nghiệp với 15 loại đối tượng và texture khác nhau.

Môi trường: Thiết lập môi trường phát triển với PyTorch, Transformers, và thư viện PEFT.

2. Lựa chọn & Tích hợp Mô hình



Mô hình nền: CLIP ViT-L/14@336px (độ hiệu suất mạnh mẽ đã được chứng minh).

Tích hợp LoRA: Áp dụng các adapter LoRA vào các lớp attention của Vision Encoder và Text Encoder.

3. Thiết kế Prompt & Huấn luyện

Prompting: Áp dụng chiến lược object-agnostic với các cặp prompt.

Hàm mất mát (Loss Function): Sử dụng hàm loss kết hợp global loss để tối ưu hóa cả ở cấp độ ảnh và cấp độ pixel.

4. Đánh giá & Phân tích

Định lượng: Đo lường Image AUROC, Pixel AUROC, và PRO score.

Định tính: Trực quan hóa các heatmaps của vùng bất thường.

Phân tích hiệu quả: So sánh số lượng tham số và thời gian huấn luyện.

Mục tiêu nghiên cứu cụ thể

Mục tiêu 1: Nghiên cứu và Xây dựng Nền tảng

Tìm hiểu sâu về kiến trúc các mô hình VLM và các kỹ thuật Parameter-Efficient Fine-Tuning (PEFT), tập trung vào LoRA.

Phân tích và lựa chọn mô hình VLM nền tảng phù hợp nhất cho bài toán phát hiện bất thường công nghiệp.

Nghiên cứu các phương pháp thiết kế prompt hiệu quả, đặc biệt là các prompt object-agnostic để tăng tính tổng quát.

Mục tiêu 2: Hiện thực và Thử nghiệm

Cài đặt và tích hợp kỹ thuật LoRA vào mô hình VLM đã chọn.

Xây dựng quy trình huấn luyện (training pipeline) sử dụng bộ dữ liệu công nghiệp tiêu chuẩn (ví dụ: MVTec AD, VisA).

Huấn luyện và tinh chỉnh các tham số của LoRA để tối ưu hóa hiệu suất phát hiện và định vị lỗi.

Mục tiêu 3: Đánh giá và Tổng kết

Đánh giá hiệu suất của mô hình trên tập dữ liệu thử nghiệm bằng các độ đo tiêu chuẩn (AUROC, AP, PRO).

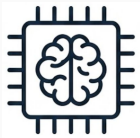
So sánh kết quả về độ chính xác và hiệu quả (số lượng tham số, thời gian huấn luyện) với phương pháp fine-tuning toàn bộ và các phương pháp SOTA khác.

Viết báo cáo tổng kết, công bố mã nguồn và mô hình đã huấn luyện.

Kết quả dự kiến và Đóng góp của đề tài

Sản phẩm hữu hình (Tangible Outcomes)

Mã nguồn mở (Source Code): Một repo trên GitHub bao gồm toàn bộ quy trình từ tiền xử lý dữ liệu, huấn luyện đến đánh giá mô hình.



Mô hình đã huấn luyện (Trained Models): Các trọng số LoRA đã được huấn luyện trên bộ dữ liệu MVTec AD, sẵn sàng để tái sử dụng và kiểm thử.

Báo cáo Khóa luận (Thesis Report): Một tài liệu khoa học chi tiết, mô tả toàn bộ quá trình nghiên cứu, phương pháp, kết quả và phân tích.

Đóng góp khoa học (Scientific Contributions)

1. **Chứng minh tính khả thi:** Khẳng định LoRA là một phương pháp hiệu quả và tiết kiệm để thích nghi các VLM cho bài toán phát hiện bất thường trong công nghiệp.
2. **Tăng cường khả năng tiếp cận:** Cung cấp một giải pháp nhẹ, giúp các doanh nghiệp vừa và nhỏ dễ dàng ứng dụng AI vào kiểm soát chất lượng mà không cần đầu tư hạ tầng tính toán lớn.
3. **Nền tảng cho nghiên cứu tương lai:** Mở ra hướng nghiên cứu mới về việc áp dụng các kỹ thuật PEFT khác cho các bài toán thị giác máy tính trong lĩnh vực công nghiệp.

Tài liệu tham khảo

[1]. Radford, A., et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. (CLIP Paper).

[2]. Hu, E. J., et al. (2022). LoRA: Low-Rank Adaptation of Large Language Models. (ICLR 2022).

[3]. Jeong, J., et al. (2023). WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation. (CVPR 2023).

[4]. Zhou, Q., et al. (2024). AnomalyCLIP: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection. (ICLR 2024).

[5]. Bergmann, P., et al. (2019). MVTec AD – A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. (CVPR 2019).