ĐỊNH LƯỢNG Ý NGHĨA THỐNG KÊ CHO PHƯƠNG PHÁP PHÁT HIỆN BẤT THƯỜNG DỰA TRÊN MÔ HÌNH HỌC SÂU BÁN GIÁM SÁT THÔNG QUA SUY DIỄN CHỌN LỌC

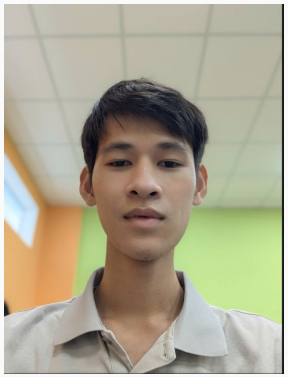
Đặng Quang Vinh - 23521786

Cao Lê Công Thành - 23521437

Tóm tắt

- Lớp: CS519.Q11.KHTN
- Link Github của nhóm: https://github.com/vinh0406/CS519.Q11
- Link YouTube video:

https://drive.google.com/drive/folders/1Zi6E0J2dwVSYIk9PKoesJWJJ_0V6lnMy?usp=sharing



Cao Lê Công Thành - 23521437

Đặng Quang Vinh - 23521786

Giới thiệu

Trong những năm gần đây, các phương pháp phát hiện bất thường dựa trên học sâu đã đạt được hiệu suất vượt trội, đặc biệt là các mô hình bán giám sát như Deep SAD (Deep Semi-Supervised Anomaly Detection) [1].

Tuy nhiên, hạn chế lớn của Deep SAD chỉ dừng lại ở kết quả dự đoán từ mô hình. Trong khi đó, các ứng dụng đòi hỏi độ an toàn cao như chẩn đoán y tế hay kiểm định công nghiệp, chỉ dựa vào kết quả dự đoán từ mô hình là chưa đủ vì thiếu cơ sở để đánh giá mức độ tin cậy. Vì vậy chúng ta cần một thước đo độ tin cậy của kết quả dự đoán từ mô hình,

Các phương pháp kiểm định thống kê truyền thống thường dẫn đến sai lệch do vấn đề thiên kiến chọn lựa. Điều này khiến cho tỷ lệ dương tính giả không được kiểm soát đúng mức ý nghĩa mong muốn.Chính vì thế vấn đề được đặt ra là làm thế nào để định lượng mức độ ý nghĩa thống kê chính xác thông qua việc tính toán p-value cho một mẫu dữ liệu khi biết rằng mẫu đó được mô hình Deep SAD dự đoán là bất thường.

Giới thiệu

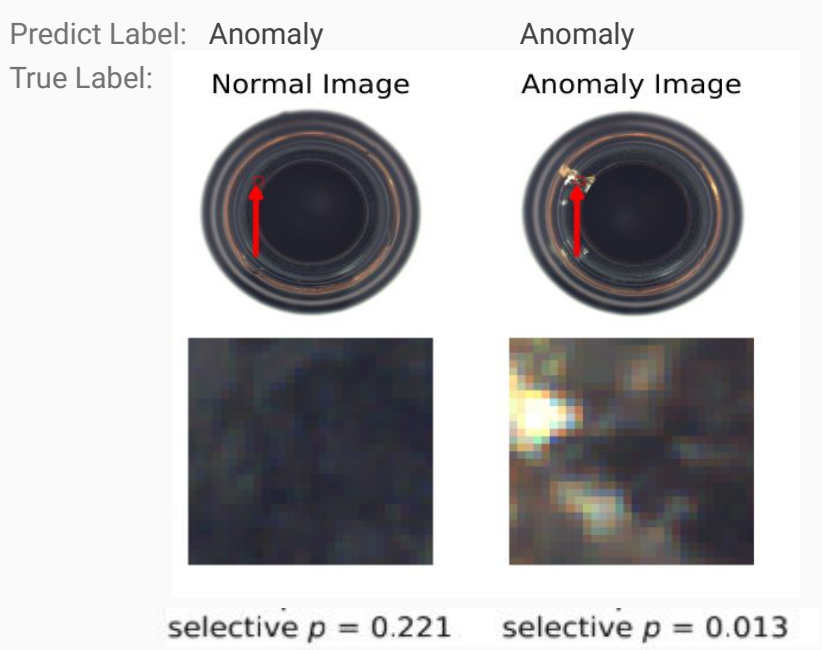
Input:

- Một mẫu dữ liệu đã được mô hình dự đoán là bất thường cần kiểm tra độ tin cậy.

Output :

- Giá trị p-value: biểu thị **xác suất quan sát được mẫu dữ liệu đang xét dưới giả thuyết rằng mẫu là bình thường, sau khi đã điều chỉnh các yếu tố do quá trình lựa chọn của mô hình gây ra**.

- Nhãn dự đoán: *Bình thường* hoặc *Bất thường*.



Mục tiêu

1. Phát triển một phương pháp suy diễn thống kê cho mô hình Deep SAD: Xây dựng nền tảng toán học chặt chẽ cho bộ khung thuật toán Selective Inference for Deep SAD có khả năng định lượng được ý nghĩa thống kê chính xác cho các dự đoán bất thường từ mô hình học sâu bán giám sát.

2. Đánh giá thực nghiệm chứng minh phương pháp đề xuất mang lại hiệu quả: Kiểm thử phương pháp trên các tập dữ liệu chuẩn như MVTec AD. Chứng minh tỷ lệ dương tính giả (**False Positive Rate**) của phương pháp đề xuất duy trì ổn định xấp xỉ mức ý nghĩa α (khắc phục tình trạng sai lệch của phương pháp Naive p-value). So sánh sức mạnh kiểm định (**Power - True Positive Rate**) với các phương pháp hiệu chỉnh truyền thống (như Bonferroni correction) để khẳng định phương pháp đề xuất thể hiện sức mạnh kiểm định lớn hơn.

Nội dung và Phương pháp

Nội dung:

- Nghiên cứu lý thuyết Suy diễn chọn lọc (Selective Inference) để giải quyết vấn đề thiên kiến chọn lọc (Selection Bias) trên các kết quả đầu ra của mô hình Deep SAD.

- Phân tích và mô hình hóa các ràng buộc hình học của cấu trúc mạng nơ-ron (Leaky ReLU, MaxPool) và biên quyết định hình siêu cầu của Deep SAD để xác định chính xác không gian sự kiện lựa chọn.

- Xây dựng phương pháp ước lượng phân phối xác suất có điều kiện và tính toán giá trị p-value nhằm định lượng độ tin cậy cho từng mẫu bất thường được phát hiện.

Phương pháp đề xuất: Phương pháp tích hợp mô hình Deep SAD và Suy diễn Chọn lọc (Selective Inference), được chia làm 2 giai đoạn chính (Two-stage approach).

Nội dung và Phương pháp

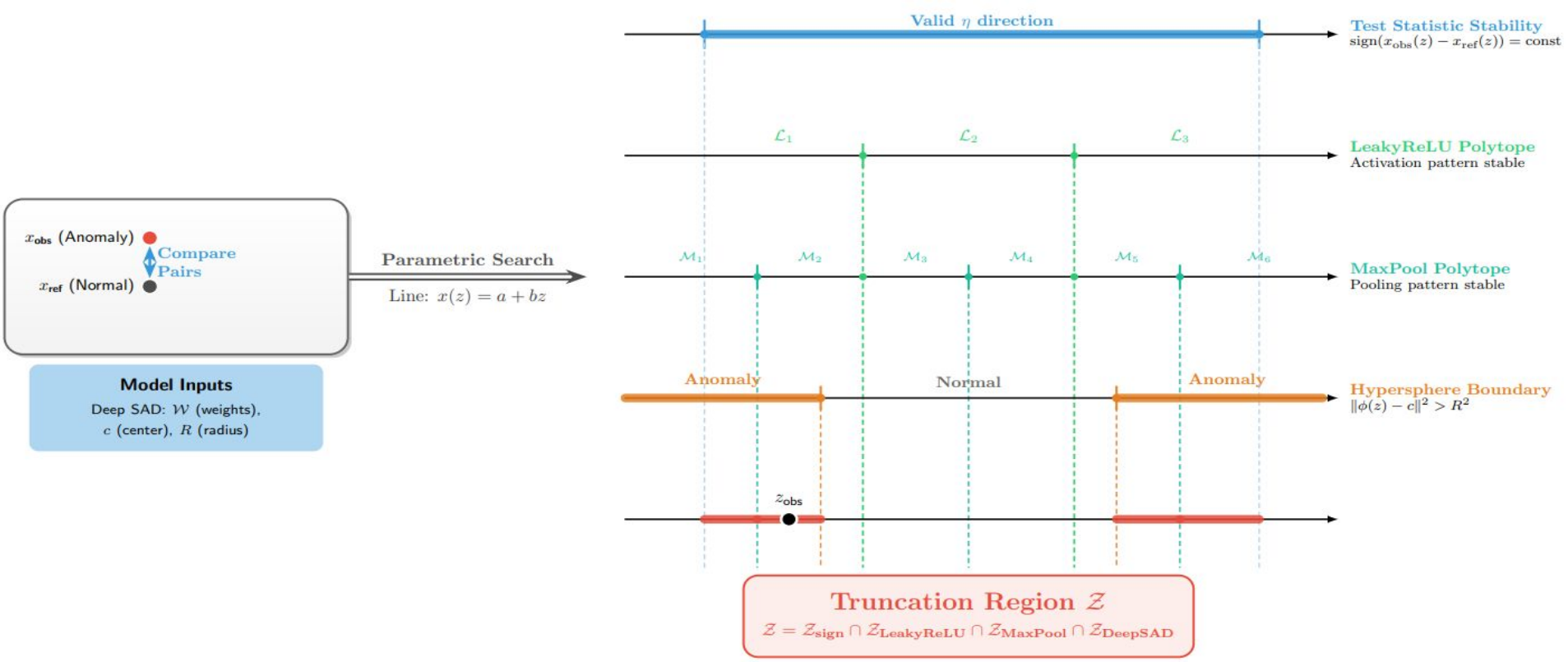
Giai đoạn 1: Huấn luyện Mô hình & Sàng lọc (Selection Procedure)

- **Mô hình:** Deep SAD (Ruff *et al.*) với hàm mục tiêu cực tiểu hóa khoảng cách mẫu bình thường đến tâm c và cực đại hóa khoảng cách mẫu bất thường.
- **Điều kiện phát hiện:** Một mẫu x được coi là bất thường nếu điểm số $Score(x) = ||f(x) - c||^2 > R^2$.
  - $f(x)$: Biểu diễn trên không gian ẩn (latent space representation).
  - R: Bán kính quyết định (tính từ quantile của tập normal và tập unlabeled trong tập train).

Giai đoạn 2: Suy diễn chọn lọc (Selective Inference)

- **Kỹ thuật:** Parametric Selective Inference [3].
- **Quy trình tính p-value:**
  1. **Tham số hóa:** Xây dựng đường thẳng $X(z) = a + bz$ đi qua mẫu nghi ngờ (x_{test})
  2. **Tìm kiếm ràng buộc (Constraints Tracking):** Tìm tập hợp các khoảng giá trị của z sao cho:
     - Bảo toàn dấu thống kê: Đảm bảo dấu của Test statistic giữa x_{test} và x_{ref} không thay đổi để vector hướng kiểm định η cố định - Giải hệ bất phương trình tuyến tính.
     - Cấu trúc mạng không đổi (các neuron Leaky ReLU và MaxPool giữ nguyên trạng thái kích hoạt - giải hệ bất phương trình tuyến tính).
     - Mô hình vẫn dự đoán là bất thường ($Score(X(z)) > R^2$ - giải bất phương trình bậc 2).
  3. **Tính toán thống kê:** Tính xác suất có điều kiện (p-value) dựa trên phân phối Gaussian bị cắt cụt (Truncated Gaussian) trong các khoảng tìm được ở bước 2.

Nội dung và Phương pháp



Selective Inference for Deep SAD

Kết quả dự kiến

1. Báo cáo khoa học hoàn chỉnh về phương pháp luận và kiểm chứng thực nghiệm:

- **Về mặt lý thuyết:** Tài liệu trình bày chi tiết cơ sở lý thuyết và chứng minh toán học của phương pháp đề xuất.
- **Về mặt thực nghiệm:** Cung cấp các kết quả định lượng để minh chứng cho tính đúng đắn và hiệu quả của phương pháp:
  - *Tính hợp lệ thống kê:* Biểu đồ phân phối p-value trên tập kiểm tra tuân theo phân phối đều *Uniform[0,1]* dưới giả thuyết H_0 (dữ liệu bình thường).
  - *Kiểm soát rủi ro:* Kết quả đo lường Tỷ lệ dương tính giả (False Positive Rate) trên các tập dữ liệu chuẩn xấp xỉ mức ý nghĩa α.
  - *Hiệu năng:* Kết quả so sánh cho thấy Sức mạnh kiểm định (Power) của phương pháp đề xuất vượt trội so với các phương pháp truyền thống (như Naive, Bonferroni).

2. Module Python:

- Module hiện thực hoá phương pháp Suy diễn chọn lọc cho Deep SAD với chức năng cung cấp nhãn dự đoán và kèm theo một **thước đo thống kê (p-value)** giúp người dùng đánh giá độ tin cậy của từng quyết định phát hiện bất thường từ mô hình Deep SAD.

Tài liệu tham khảo

[1]. Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, Marius Kloft: Deep Semi-Supervised Anomaly Detection. ICLR 2020

[2]. Mizuki Niihori, Shuichi Nishino, Teruyuki Katsuoka, Tomohiro Shiraishi, Kouichi Taji, Ichiro Takeuchi: Quantifying Statistical Significance of Deep Nearest Neighbor Anomaly Detection via Selective Inference. NeurIPS 2025

[3]. Vo Nguyen Le Duy, Ichiro Takeuchi: More Powerful Conditional Selective Inference for Generalized Lasso by Parametric Programming. JMLR. 23: 300:1-300:37 (2022)

[4]. Vo Nguyen Le Duy, Shogo Iwazaki, Ichiro Takeuchi: Quantifying Statistical Significance of Neural Network-based Image Segmentation by Selective Inference. NeurIPS 2022

[5]. Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, Marius Kloft: Deep One-Class Classification. ICML 2018: 4390-4399