QUANTIFYING STATISTICAL SIGNIFICANCE FOR DEEP SEMI-SUPERVISED ANOMALY DETECTION VIA SELECTIVE INFERENCE

Đặng Quang Vinh - 23521786
Cao Lê Công Thành - 23521437
GVHD: PGS. TS Lê Đình Duy

Tóm tắt

- Lớp: CS519.Q11.KHTN
- Link Github của nhóm: https://github.com/vinh0406/CS519.Q11
- Link YouTube video: https://youtu.be/hmZpbOSwJJ4

Cao Lê Công Thành - 23521437

Đặng Quang Vinh - 23521786

Giới thiệu

Bối cảnh hiện nay:

Phát hiện bất thường (Anomaly Detection - AD): Là bài toán quan trọng trong nhiều lĩnh vực:

- Y tế: Phát hiện khối u.
- Tài chính: Phát hiện gian lận thẻ tín dụng.
- An ninh mạng: Phát hiện tấn công mạng.

Deep SAD (Deep Semi-supervised Anomaly Detection): Là phương pháp phát hiện bất thường SOTA (State-of-the-art) hiện nay, tận dụng lượng nhỏ dữ liệu có nhãn để học biểu diễn đặc trưng tốt hơn so với các phương pháp không giám sát truyền thống.

Vấn đề (Gap): Khi Deep SAD đưa ra dự đoán bất thường cho một mẫu dữ liệu,chúng ta không biết được nó có thực sự là bất thường hay chỉ do nhiễu gây ra.

Giới thiệu

❖ **Vấn đề (Gap Research):** Các mô hình SOTA như Deep SAD hoạt động như một "hộp đen", đưa ra quyết định nhị phân (Bình thường/Bất thường), thiếu thước đo tin cậy

→ Tỷ lệ Báo động giả (dương tính giả) không được kiểm soát.

❖ **Hệ quả nghiêm trọng của Báo động giả (dương tính giả) trong các lĩnh vực thiết yếu:**

➢ Y tế (Chẩn đoán ảnh): Một cảnh báo ung thư sai (Dương tính giả) dẫn đến các xét nghiệm xâm lấn đau đớn và hoảng loạn tâm lý không cần thiết cho bệnh nhân.

➢ Sản xuất công nghiệp: Báo động sai khiến dây chuyền phải dừng hoạt động để kiểm tra thủ công, gây lãng phí chi phí vận hành và thời gian.

❖ **Đóng góp của đề tài:**

➢ **Chuyển từ "Hộp đen" sang "Minh bạch hoá":**

■ Thay vì chỉ đưa ra kết quả nhị phân (Bình thường/Bất thường), hệ thống cung cấp thêm một thước đo tin cậy thống kê cho từng quyết định.

➢ **Cam kết về mặt Thống kê toán học:**

■ Phương pháp đề xuất sẽ đảm bảo tỷ lệ dương tính giả sẽ luôn được duy trì xấp xỉ một mức ý nghĩa α (ví dụ $\alpha = 0.05$), bất kể độ phức tạp của dữ liệu.

→ Cung cấp cơ sở định lượng để giúp người dùng (bác sĩ, kỹ sư vận hành,...) ra quyết định can thiệp chính xác, tránh báo động giả tràn lan.

Giới thiệu

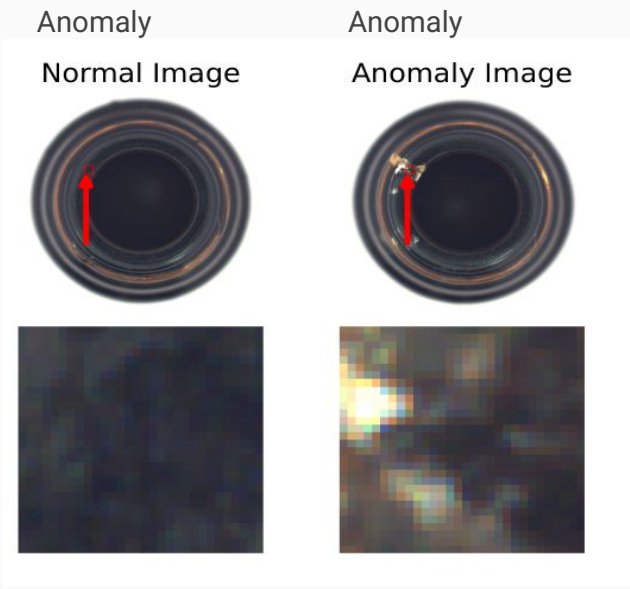
Input:

- Một mẫu dữ liệu dạng ảnh đã được mô hình DeepSAD 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**.

- Kết quả dự đoán: *Bình thường* hoặc *Bất thường*.



Mục tiêu

1. Phát triển một phương pháp suy diễn thống kê cho mô hình Deep SAD có khả năng định lượng được ý nghĩa thống kê chính xác cho các dự đoán bất thường từ mô hình .

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 1: Tìm hiểu tổng quan đề tài và cơ sở lý thuyết**
  - Phương pháp:
    - Tìm hiểu về mô hình Deep SAD: Đọc và phân tích các bài báo khoa học liên quan đến Deep Semi-Supervised Anomaly Detection (Deep SAD) để hiểu cơ chế hoạt động, ưu điểm và hạn chế của nó trong việc phát hiện bất thường.
    - Nghiên cứu về Suy diễn chọn lọc (Selective Inference): Tìm hiểu nguyên lý toán học của phương pháp này và cách nó được ứng dụng để kiểm định độ tin cậy trong các mô hình học máy khác.
    - Khảo sát các phương pháp hiện có: Tìm hiểu các kỹ thuật kiểm định thống kê truyền thống (như Naive p-value, Bonferroni correction) để làm cơ sở so sánh với phương pháp mới của chúng tôi.
- **Nội dung 2: Nghiên cứu và đề xuất phương pháp mới**
  - Phương pháp:
    - Phân tích cấu trúc của mạng của mô hình Deep SAD để xác định các yếu tố ảnh hưởng đến kết quả dự đoán.
    - Thiết kế quy trình tính toán độ tin cậy dựa trên lý thuyết Suy diễn chọn lọc, đảm bảo kết quả phản ánh đúng độ tin cậy của mô hình.

Nội dung và Phương pháp

- **Nội dung 3: Hiện thực hoá phương pháp đề xuất**
  - Phương pháp:
    - Sử dụng ngôn ngữ Python và các thư viện học sâu (Deep Learning frameworks) để xây dựng module.
    - Lập trình chức năng: Nhận vào hình ảnh → Mô hình phát hiện bất thường → Tính toán độ tin cậy → Trả kết quả cuối cùng.
- **Nội dung 4: Thực nghiệm và đánh giá kết quả**
  - Phương pháp:
    - Thu thập và xử lý dữ liệu: Sử dụng bộ dữ liệu chuẩn trong công nghiệp như MVTec AD (gồm ảnh các sản phẩm lỗi và không lỗi) để làm dữ liệu kiểm thử.
    - Thiết lập kịch bản kiểm thử: Tiến hành chạy thực nghiệm các phương pháp kiểm định hiện có và phương pháp do chúng tôi đề xuất.
    - Thống kê kết quả, so sánh hiệu năng và khả năng kiểm soát tỷ lệ báo động giả trong quá trình thực nghiệm giữa các phương pháp

Kết quả dự kiến

1. Báo cáo khoa học hoàn chỉnh về lý thuyết của phương pháp đề xuất 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 của phương pháp đề xuất và so sánh hiệu năng và khả năng kiểm soát tỷ lệ báo động giả của phương pháp đề xuất với các phương pháp kiểm định thống kê hiện có.

2. Module Python:

- Module hiện thực hoá phương pháp Suy diễn chọn lọc cho mô hình Deep SAD với chức năng cung cấp nhãn dự đoán và kèm theo một **thước đo độ tin cậy** 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: 31627–31639

[5]. Daiki Miwa, Vo Nguyen Le Duy, Ichiro Takeuchi: Valid P-Value for Deep Learning-Driven Salient Region. ICLR 2023

[6]. Vo Nguyen Le Duy, Hiroki Toda, Ryota Sugiyama, Ichiro Takeuchi: Computing Valid p-value for Optimal Changepoint by Selective Inference using Dynamic Programming. NeurIPS 2020: 11356-11367

[7]. 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