


# THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://www.youtube.com/watch?v=UGpcok5-vmo>
- Link slides (dạng .pdf đặt trên Github):  
<https://github.com/KanzNguyen/CS519.O21.KHTN/blob/main/Khang%20Nguy%E1%BB%85n%20Nguy%C3%AAn%20-%20CS519.O21.KHTN.DeCuong.FinalReport.Template.Slide.pdf>
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

|   |   |
|---|---|
| <ul style="list-style-type: none"><li>● Họ và Tên: Nguyễn Nguyên Khang</li><li>● MSSV: 22520623</li></ul>  | <ul style="list-style-type: none"><li>● Lớp: CS519.O21.KHTN</li><li>● Tự đánh giá (điểm tổng kết môn): 8.5/10</li><li>● Số buổi vắng: 1</li><li>● Số câu hỏi QT cá nhân: 6</li><li>● Số câu hỏi QT nhóm: 1</li><li>● Link Github: <a href="https://github.com/KanzNguyen">https://github.com/KanzNguyen</a></li></ul> |
|---|---|

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

BAYESIAN SELF-CRITICAL SEQUENCE TRAINING CHO BÀI TOÁN SINH  
BÁO CÁO Y KHOA

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

BAYESIAN SELF-CRITICAL SEQUENCE TRAINING FOR MEDICAL REPORT  
GENERATION

## TÓM TẮT (Tối đa 400 từ)

Các công trình nghiên cứu bài toán Sinh Báo cáo Y khoa (Medical Report Generation - MRG) gần đây đã đạt được những bước tiến đáng kể thông qua các cách tiếp cận bao gồm các mô hình RNN phân cấp, cơ chế Attention hoặc các thuật toán học tăng cường tuy nhiên vẫn còn tồn tại **một số điểm không nhất quán giữa quy trình huấn luyện và kiểm thử** ảnh hưởng đến hiệu quả mô hình: 1) Mô hình sử dụng hàm cross-entropy và cực đại hóa log-likelihood của dữ liệu huấn luyện trong khi sử dụng các hàm bất khả vi để đánh giá hiệu năng mô hình trong quá trình kiểm thử, 2) Trong lúc huấn luyện mô hình dự đoán từ tiếp theo dựa trên các từ thật sự (ground-truth) trong khi chưa từng tiếp xúc với các dự đoán của mình trước khi kiểm thử. Trong đề tài này, chúng tôi tập trung giải quyết ba vấn đề của bài toán MRG: Đầu tiên, tiến hành nâng cao hiệu năng mô hình VLCI [1] sinh báo cáo y khoa bằng cách sử dụng thuật toán B-SCST nhằm giải quyết hai vấn đề vừa nêu trên, tạo thành **kiến trúc VLCI with B-SCST** và so sánh hiệu năng thu được của mô hình mới này với phiên bản gốc. Thứ hai, chúng tôi khám phá và thử nghiệm một số phương pháp **tăng cường dữ liệu dựa trên các mô hình GAN** và so sánh với cách tiếp cận truyền thống, đây là một cách tiếp cận mới mẻ với hy vọng tạo ra nguồn dữ liệu chân thực và đầy đủ cho quá trình huấn luyện. Kết quả sẽ được đánh giá trên hai bộ dữ liệu sử dụng các chiến lược tăng cường dữ liệu vừa nêu và so sánh với mô hình gốc để đánh giá hiệu năng của các cải tiến. Cuối cùng chúng tôi **xây dựng một chương trình ứng dụng** để minh họa cho hiệu quả của những cải tiến trên mô hình trong môi trường thực tế.

## GIỚI THIỆU (Tối đa 1 trang A4)

Các dạng hình ảnh y khoa (X-Ray, MRI...) cung cấp nhiều thông tin phục vụ cho việc chẩn đoán bệnh nhưng việc nhận biết các chi tiết bất thường và viết báo cáo một cách chính xác và hiệu quả lại tiêu tốn rất nhiều thời gian, kể cả với những y bác sĩ giàu kinh nghiệm. Hơn nữa, đối với các y bác sĩ chưa có nhiều kinh nghiệm thì rất thường xuyên xảy ra trường hợp bỏ sót một số chi tiết nhỏ trong quy trình viết báo cáo do những chi tiết này có yêu cầu rất cao về kiến thức chuyên môn. Để giải quyết vấn đề này, bài toán Sinh Báo cáo Y khoa (MRG) đã ra đời và thu hút nhiều sự chú ý từ cộng đồng. Bài toán trích xuất những thông tin có ích từ hình ảnh y khoa và sinh ra báo cáo tương ứng khá tương đồng với bài toán image captioning. Một cách tương tự, bài toán MRG cũng gặp phải các vấn đề: Đầu tiên, trong quá trình huấn luyện mô hình, mô hình tìm cách tối ưu hàm mất mát cross-entropy trong khi quá trình kiểm thử lại sử dụng các hàm bất khả vi để đánh

giá như: ROUGE, CIDER, BLEU,... Sự không đồng nhất này có thể dẫn đến kết quả không tốt khi xét trên các độ đo này. Thứ hai, trong quá trình huấn luyện, mô hình tìm cách cực đại hóa log-likelihood của từ dự đoán với các từ thật sự (ground-truth) từ dữ liệu huấn luyện trong khi tại thời điểm tiến hành kiểm thử, mô hình không được tiếp xúc với các từ thật sự này mà phải dựa vào các từ trước đó mà mô hình đã dự đoán cho nên nếu mô hình mắc một sai sót nhỏ trong quá trình sinh báo cáo thì lỗi sai này sẽ tích tụ và lan truyền, ảnh hưởng đến các từ tiếp theo [4].

Trong đề tài này, chúng tôi chủ trương giải quyết 2 vấn đề nêu trên bằng một thuật toán dựa trên phương pháp học tăng cường – thuật toán Bayesian Self-critical Sequence Training (B-SCST) [2] áp dụng lên mô hình đã được huấn luyện bằng cách sử dụng cross-entropy để tiến hành huấn luyện tăng cường sử dụng hàm điểm thưởng (reward function) dựa trên độ đo ROUGE-L, giúp đồng nhất độ đo đánh giá trong quá trình huấn luyện và kiểm thử. Phương pháp sử dụng kỹ thuật Monte-Carlo dropout để tạo ra các phiên bản của mô hình sử dụng các cách thức drop-out khác nhau và sinh ra các bản báo cáo khác nhau để tính điểm thưởng và tính giá trị kỳ vọng baseline, mô hình sẽ tính toán mức chênh lệch về phần thưởng giữa các báo cáo đầu ra so với baseline để cập nhật tham số mô hình dựa trên các so sánh này, từ đó khuyến khích mô hình tạo ra các kết quả chất lượng hơn trong tương lai.

Một vấn đề khác tồn tại trong lớp bài toán MRG là chất lượng của các bộ dữ liệu sẵn có không phải lúc nào cũng có lượng dữ liệu đủ lớn phục vụ cho việc huấn luyện mô hình (đặc biệt là các hình ảnh xét nghiệm liên quan đến các trường hợp bệnh hiếm gặp có số lượng dữ liệu khá ít). Bên cạnh đó các tiến bộ gần đây của các kiến trúc GAN có khả năng sinh hình ảnh đã đạt được nhiều kết quả ấn tượng, là một hướng tiếp cận tiềm năng ít được khám phá trong bài toán tăng cường dữ liệu y khoa [6] nên trong đề tài này chúng tôi sẽ khám phá hiệu quả của các mô hình GAN để tăng cường dữ liệu, áp dụng trên 2 bộ dữ liệu ảnh chụp X-quang ngực là IU-Xray và MIMIC-CXR, kết quả đầu ra sẽ được so sánh với cách tiếp cận truyền thống (lật ảnh, phóng to, thu nhỏ, dịch chuyển,...).

Về tổng quát, trong đề tài này những đóng góp của chúng tôi sẽ bao gồm:

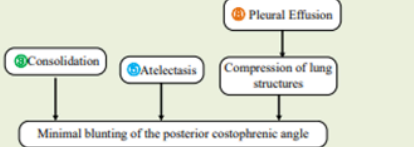
- Chúng tôi khám phá hiệu năng của mô hình VLCI – một mô hình dựa trên kiến trúc transformer phổ biến có khả năng giảm thiểu ảnh hưởng của các mối quan hệ nhân-quả không chính xác [1] kết hợp với thuật toán B-SCST [2] cho bài toán MRG và đánh giá hiệu quả của sự kết hợp này so với mô hình gốc.
- Chúng tôi khám phá hiệu quả của các mô hình GAN phổ biến cho bài toán tăng cường cho dữ liệu hình ảnh y khoa, đánh giá hiệu quả của cách tiếp cận này với cách tiếp cận truyền thống.
- Chúng tôi xây dựng một chương trình ứng dụng minh họa cho những hiệu năng đạt được từ những cải tiến và cách tiếp cận đã đề xuất cho bài toán MRG áp dụng trên ảnh chụp X-quang vùng ngực.

### Input:

Bộ dữ liệu huấn luyện đã gán nhãn bao gồm các hình ảnh y khoa và báo cáo y khoa tương ứng với ảnh đó. Một ảnh xét nghiệm y khoa chưa được viết báo cáo chẩn đoán.

**Output:** Báo cáo chẩn đoán dựa trên ảnh xét nghiệm y khoa được cung cấp

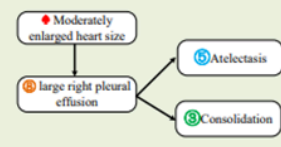
(a) ⑤③⑧



**Ground Truth:** Chest PA and lateral radiograph demonstrates a markedly elevated right hemidiaphragm with adjacent compressive ⑤atelectasis or ⑧consolidation. Minimal blunting of the posterior costophrenic angle ⑤may indicate a small right pleural effusion. Left lung is clear. ♦Cardiomediastinal borders are unremarkable.

**VLCI:** frontal and lateral views of the chest were obtained. ⑤there is a small right pleural effusion ⑤with overlying atelectasis ⑤underlying consolidation is not excluded. the left lung is clear. ♦the cardiac silhouette is not enlarged. the aorta is calcified and tortuous. ⑦no pneumothorax is seen.

(b) ⑤③⑧♦



**Ground Truth:** ♦The heart size is moderately enlarged. ♦The mediastinal silhouette and hilar contours are unchanged. A moderate to ⑤large right-sided pleural effusion is slightly increased in volume compared to prior examination with collapse of much of the right lower lobe and right middle lobe. ⑤There is also some consolidation at the base of the right upper lobe which could be due to ⑤compressive atelectasis. ⑥There is no left effusion. ⑦There is no pneumothorax.

**VLCI:** there are low lung volumes which the bronchovascular markings. ⑤there is a large right pleural effusion with ⑤overlying atelectasis. ⑤underlying consolidation is not excluded. ⑦there is no pneumothorax. ♦the enlarged cardiac silhouette is due to the large right pleural effusion. mediastinal contours are grossly stable. no overt pulmonary edema is seen.

## MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

Nghiên cứu hiệu suất của mô hình VLICI [1] và phiên bản cải tiến của VLICI sử dụng thuật toán B-SCST trên các độ đo CIDER, BLEU, ROUGE-L, METEOR và CE

Nghiên cứu hiệu suất của 2 mô hình trên 2 bộ dữ liệu IU-Xray và MIMIC-CXR sử dụng cách tiếp cận tăng cường dữ liệu truyền thống và sử dụng mô hình GAN.

Phát triển một chương trình ứng dụng để minh họa hiệu năng của mô hình cải tiến.

## **NỘI DUNG VÀ PHƯƠNG PHÁP**

*(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)*

### **a. Nội dung:**

Tìm hiểu câu trả lời cho câu hỏi: “**Liệu việc ứng dụng thuật toán B-SCST có khả năng huấn luyện nâng cao từ mô hình nền tảng được áp dụng thành công cho các bài toán image captioning, speech recognition và question answering [3][4][5] vào một bài toán hoàn toàn mới phức tạp hơn: sinh báo cáo y khoa từ ảnh chụp xét nghiệm có mang lại hiệu quả hơn so với kiến trúc ban đầu không?**” và câu hỏi “**Liệu cách tiếp cận mới trong bài toán tăng cường dữ liệu sử dụng các mô hình GAN trên tập dữ liệu huấn luyện có mang lại hiệu năng vượt trội hơn so với cách tiếp cận truyền thống hay không?**”.

Nghiên cứu hiệu quả của mô hình VLICI with B-SCST cho bài toán sinh báo cáo y khoa (MRG) so với hiệu quả của mô hình ban đầu.

Nghiên cứu cách ứng dụng các mô hình GAN phổ biến cho bài toán tăng cường dữ liệu hình ảnh y khoa bao gồm GAN truyền thống, DCGAN và CycleGAN [7][8][9].

Huấn luyện mô hình VLICI-base và VLICI with B-SCST trên 2 bộ dữ liệu IU-Xray và MIMIC-CXR, so sánh hiệu quả của các chiến lược tăng cường dữ liệu theo cách truyền thống và cách sử dụng mô hình GAN.

Xây dựng chương trình ứng dụng.

### **b. Phương pháp:**

Nghiên cứu kiến trúc và quy trình vận hành của mô hình VLICI có khả năng giảm thiểu ảnh hưởng hoặc thậm chí loại bỏ những thiên kiến thị giác và ngôn ngữ gây ra những mối quan hệ tương quan nhân quả không chính xác (do các confounders gây ra) trong quá trình huấn luyện.

Nghiên cứu cách áp dụng thuật toán B-SCST, kỹ thuật này dựa trên thuật toán REINFORCE và SCST để tiến hành huấn luyện nâng cao cho một mô hình đã được huấn luyện trước đó để đồng nhất cách đánh giá hiệu năng mô hình trong quá trình huấn luyện và kiểm thử sử dụng các độ đo tương ứng là các hàm bất khả vi. Mô hình VLICI sau khi đã được huấn luyện sẽ được nối với một sampling module hoạt động theo nguyên tắc của B-SCST để tạo ra các chuỗi con và tính toán baseline, các chuỗi này sẽ được so sánh với baseline và dùng kết quả so sánh này để cập nhật tham số của mô hình dựa trên hàm mất mát được quy định bởi B-SCST.

Phân chia 2 tập dữ liệu thành các tập huấn luyện và kiểm thử, tạo ra các phiên bản khác nhau trên tập huấn luyện sử dụng phương pháp tăng cường dữ liệu truyền thống, huấn luyện các mô hình GAN để tạo ra phiên bản dữ liệu huấn luyện có tăng cường dữ liệu sử dụng các lớp mô hình GAN.

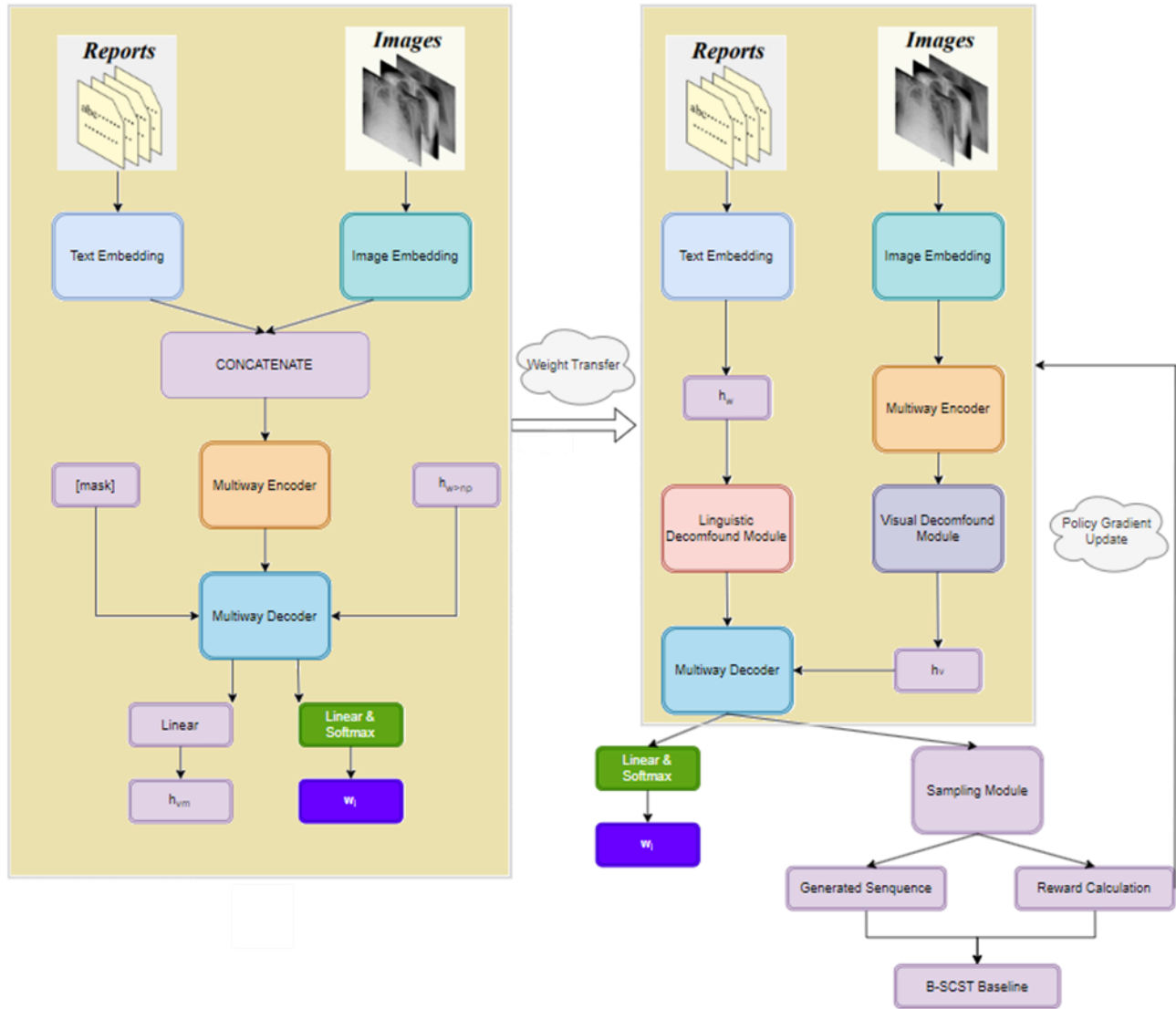
Nghiên cứu cách đánh giá mô hình tạo báo cáo y khoa trên các độ đo ROUGE-L, CIDER, BLEU, METEOR và CE.

Huấn luyện mô hình VLICI-base trên 2 bộ dữ liệu IU-Xray và MIMIC-CXR trong cả trường hợp sử dụng phương pháp tăng cường dữ liệu truyền thống và phương pháp sử dụng các mô hình GAN.

Huấn luyện mô hình VLICI with B-SCST từ mô hình VLICI-base đã được huấn luyện trước đó trên các phiên bản áp dụng tăng cường dữ liệu khác nhau của 2 bộ dữ liệu IU-Xray và MIMIC-CXR, hàm điểm thưởng sử dụng độ đo ROUGE-L đo sâu con chung dài chất giữa chuỗi dự đoán và chuỗi tham chiếu, nhằm khuyến khích mô hình tiến hành cập nhật tham số để đạt được kết quả cao hơn trên độ đo này tức tạo ra chuỗi dự đoán khớp với chuỗi tham chiếu, đồng thời đồng nhất quá trình huấn luyện và kiểm thử.

Đánh giá hiệu năng của mô hình VLICI-with B-SCST so với mô hình VLICI-base trên 2 bộ dữ liệu, trong cả trường hợp tăng cường dữ liệu theo hướng truyền thống và sử dụng GAN.

Xây dựng chương trình ứng dụng để khảo sát hiệu năng của mô hình trên các ảnh chụp X-quang ngực để tạo ra các báo cáo y khoa tương ứng.



## KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

Kết quả thực nghiệm của mô hình VLCI-with B-SCST mà chúng tôi phát triển ứng dụng trong bài toán tạo báo cáo y khoa, so sánh trên các độ đo ROUGE-L, CIDER, BLEU, CE với mô hình gốc, trên hai bộ dữ liệu với các phiên bản sử dụng tăng cường dữ liệu bằng các tiếp cận truyền thống và các mô hình GAN.

Một chương trình ứng dụng để minh họa hiệu năng của mô hình.

## TÀI LIỆU THAM KHẢO

- [1] Cross-Modal Causal Intervention for Medical Report Generation Weixing Chen, Yang Liu, Member, IEEE, Ce Wang, Jiarui Zhu, Shen Zhao, Guanbin Li, Member, IEEE, Cheng-Lin Liu, Fellow, IEEE, and Liang Lin, Fellow 28 Feb 2024.
- [2] B-SCST: Bayesian Self-Critical Sequence Training for Image Captioning Shashank Bujimalla\* Intel Corporation shashankbvs@gmail.com Mahesh Subedar\* Intel Labs mahesh.subedar@intel.com Omesh

Tickoo Intel Labs omesh.tickoo@intel.com 28 Jun 2020

[3] SELF-CRITICAL SEQUENCE TRAINING FOR AUTOMATIC SPEECH RECOGNITION Chen Chen, Yuchen Hu, Nana Hou, Xiaofeng Qi, Heqing Zou, Eng Siong Chng 13 Apr 2022

[4] Self-critical Sequence Training for Image Captioning Steven J. Rennie<sup>1</sup>, Etienne Marcheret<sup>1</sup>, Youssef Mroueh, Jerret Ross and Vaibhava Goel<sup>1</sup> Watson Multimodal Algorithms and Engines Group IBM T.J. Watson Research Center, NY, USA 16 Nov 2017

[5] Explainable Semantic Communication for Text Tasks Chuanhong Liu, Caili Guo, Senior Member, IEEE, Yang Yang, Wanli Ni, Yanquan Zhou, Lei Li, and Tony Q.S. Quek, Fellow, IEEE 17 May 2024

[6] A scoping review of transfer learning research on medical image analysis using ImageNet Mohammad Amin Morid, PhD<sup>1</sup>, Alireza Borjali, PhD<sup>2,3</sup>, Guilherme Del Fiol, MD, PhD<sup>4</sup> 13 Nov 2020

[7] UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS Alec Radford & Luke Metz indico Research Boston, MA {alec,luke}@indico.io Soumith Chintala Facebook AI Research New York, NY soumith@fb.com 7 Jan 2016

[8] Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks Jun-Yan Zhu\* Taesung Park\* Phillip Isola Alexei A. Efros Berkeley AI Research (BAIR) laboratory, UC Berkeley 24 Aug 2020

[9] Generative Adversarial Nets Ian J. Goodfellow, Jean Pouget-Abadie\*, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair†, Aaron Courville, Yoshua Bengio‡ Departement d'informatique et de recherche opérationnelle Université de Montreal Montreal, QC H3C 3J7 10 Jun 2014