# SO KHỚP NGỮ NGHĨA ĐỐI TƯỢNG CHO BÀI TOÁN CHÚ THÍCH HÌNH ẢNH BẰNG TIẾNG VIỆT

**Hứa Văn Sơn**

Trường Đại học Công Nghệ Thông Tin - Đại học Quốc Gia

Thành phố Hồ Chí Minh

## Lý do chọn đề tài

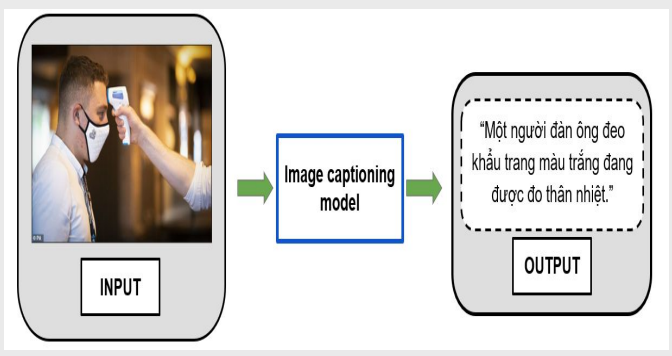
Đại dịch COVID-19 đã làm trầm trọng thêm tình trạng thiếu hụt nhân viên y tế đang diễn ra trên toàn cầu, đặt ra nhu cầu cấp thiết về những trợ lý thông minh có thể hợp tác hiệu quả với con người để lấp đầy khoảng trống. Để đạt được mục tiêu cuối cùng, dự án này nhằm mục đích nghiên cứu cách tiếp cận hiện đại đã đạt được kết quả cao trong tiếng Anh và áp dụng nó vào tiếng Việt để mô tả nội dung trực quan trong môi trường chăm sóc sức khỏe

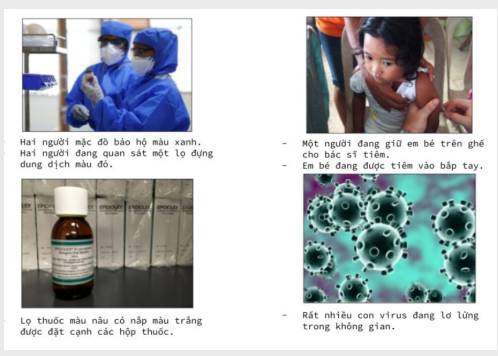
## Giới thiệu

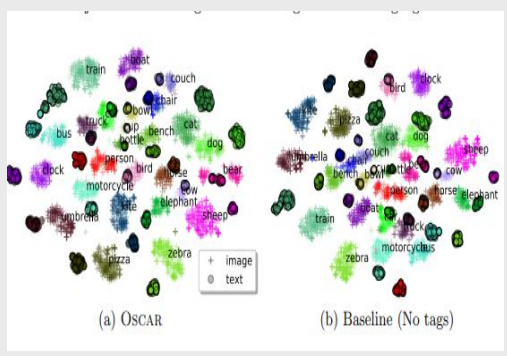
Nghiên cứu sẽ tập trung vào việc đánh giá sự hiệu quả của hướng tiếp cận mới dựa theo mô hình OSCAR sử dụng nhãn đối tượng phát hiện trong ảnh làm đầu vào để huấn luyện mô hình áp dụng vào bài toán tự động mô tả hình ảnh bằng tiếng việt. Kết quả nghiên cứu sẽ cho thấy sự ảnh hưởng của việc sử dụng các nhãn đối tượng để huấn luyện sẽ ảnh hưởng như thế nào đến hiệu suất của mô hình.

## Overview

Cài đặt mô hình → Huấn luyện trên VieCap4H → Đánh giá và phân tích




## Description

### 1. Nội Dung

- Cài đặt lại các cách tiếp cận trước đây trong bài báo "Show, Attend, and Tell: Neural Image Caption Generation with Visual Attention" và "Neural Baby Talk" và hướng tiếp cận mới được giới thiệu trong bài báo “Oscar: Object-Semantics Aligned Pre-training for Vision-Language Tasks”

- Thực hiện việc huấn luyện các mô hình và so sánh đánh giá trên bộ dữ liệu tiếng việt VieCap4H - đây là bộ dữ liệu liên quan đến lĩnh vực y tế được ra đời trong cuộc thi VLSP năm 2021 tổ chức bởi Đại học Công Nghệ Thông Tin, Đại học Quốc Gia TPHCM.

- So sánh hiệu suất của các mô hình và Đánh giá sự hiệu quả khi dùng nhãn đối tượng được phát hiện trong ảnh để làm đầu vào khi huấn luyện

### 3. Kết quả mong đợi

- Xây dựng một mô hình tự động mô tả hình ảnh tiếng việt

- Cho thấy việc sử dụng nhãn đối tượng hiệu quả, từ đó mở rộng nghiên cứu bằng cách tập trung tối ưu xây dựng các mô hình phát hiện đối tượng tốt có độ chính xác cao làm để sử dụng các nhãn đối tượng cho tác vụ tự động mô tả ảnh

### 2. Phương pháp

- Khác với các mô hình tự động tạo văn bản trước đây khi chỉ nhận đặc trưng của hình ảnh đầu vào để sinh ra văn bản mô tả thì OSCAR nhận thêm một đầu khác là nhãn đối tượng được phát hiện trong ảnh, điều này được thúc đẩy bởi sự quan sát rằng các đối tượng có trong ảnh cũng sẽ thường được miêu tả trong văn bản

- Để chứng minh được sự nổi trội của hướng tiếp cận này chúng tôi sẽ huấn luyện cả 3 mô hình trên bộ dữ liệu VieCap4H và đánh giá dựa trên BLEU

- Phân tích về khả năng sinh ảnh của các phương pháp (giữa việc có sử dụng nhãn đối tượng và không sử dụng nhãn đối tượng)


*Figure 4.* Một số ví dụ về các mô hình OSCAR mô tả ảnh

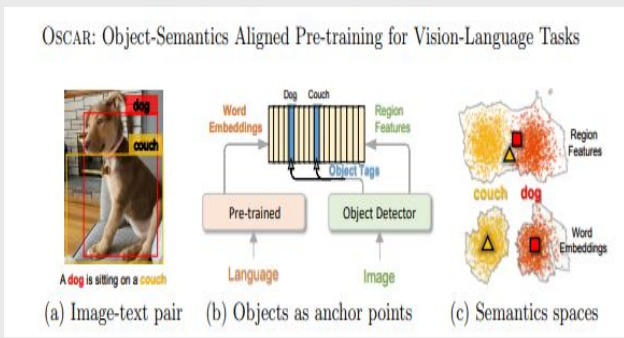



*Figure 1.* Minh họa về quá trình Oscar biểu diễn một cặp hình ảnh-văn bản

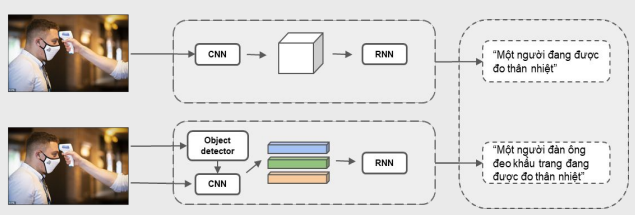



*Figure 2.* (a) Kiến trúc mô hình Show and tell. IEEE2015
(b) Kiến trúc mô hình Neural baby talk. IEEE2018

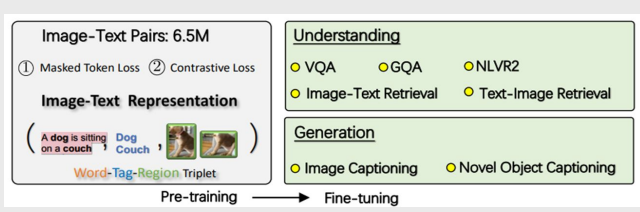



*Figure 3.* Kiến trúc mô hình pretrained OSCAR và được tinh chỉnh cho các tác vụ trong đó có tự động mô tả hình ảnh

NII

**Hứa Văn Sơn - Đại học Công Nghệ Thông Tin**

**TEL :0855166647 Email : sonhv.17@grad.uit.edu.vn**