

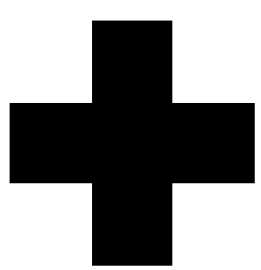
## Hướng tiếp cận mới trên bộ dữ liệu UIT-ViIC: Hỏi đáp trực quan

Trương Phước Bảo Khanh, Trần Thị Thu Hà  
{20520579, 20521273}@gm.uit.edu.vn  
Trường Đại học Công Nghệ Thông Tin – ĐHQG TP.HCM

### GIỚI THIỆU

Bài toán trả lời câu hỏi trực quan (Visual Question Answering) kết hợp hai lĩnh vực quan trọng của học máy (Machine Learning) là thị giác máy tính (ComputerVision) và xử lý ngôn ngữ tự nhiên (Natural Language Processing).

Trong bài báo này, chúng tôi đã thử nghiệm phương pháp sinh các cặp câu hỏi - câu trả lời từ những bộ dữ liệu chú thích ảnh tiếng Việt, hay ở đây là bộ dữ liệu UIT-ViIC.



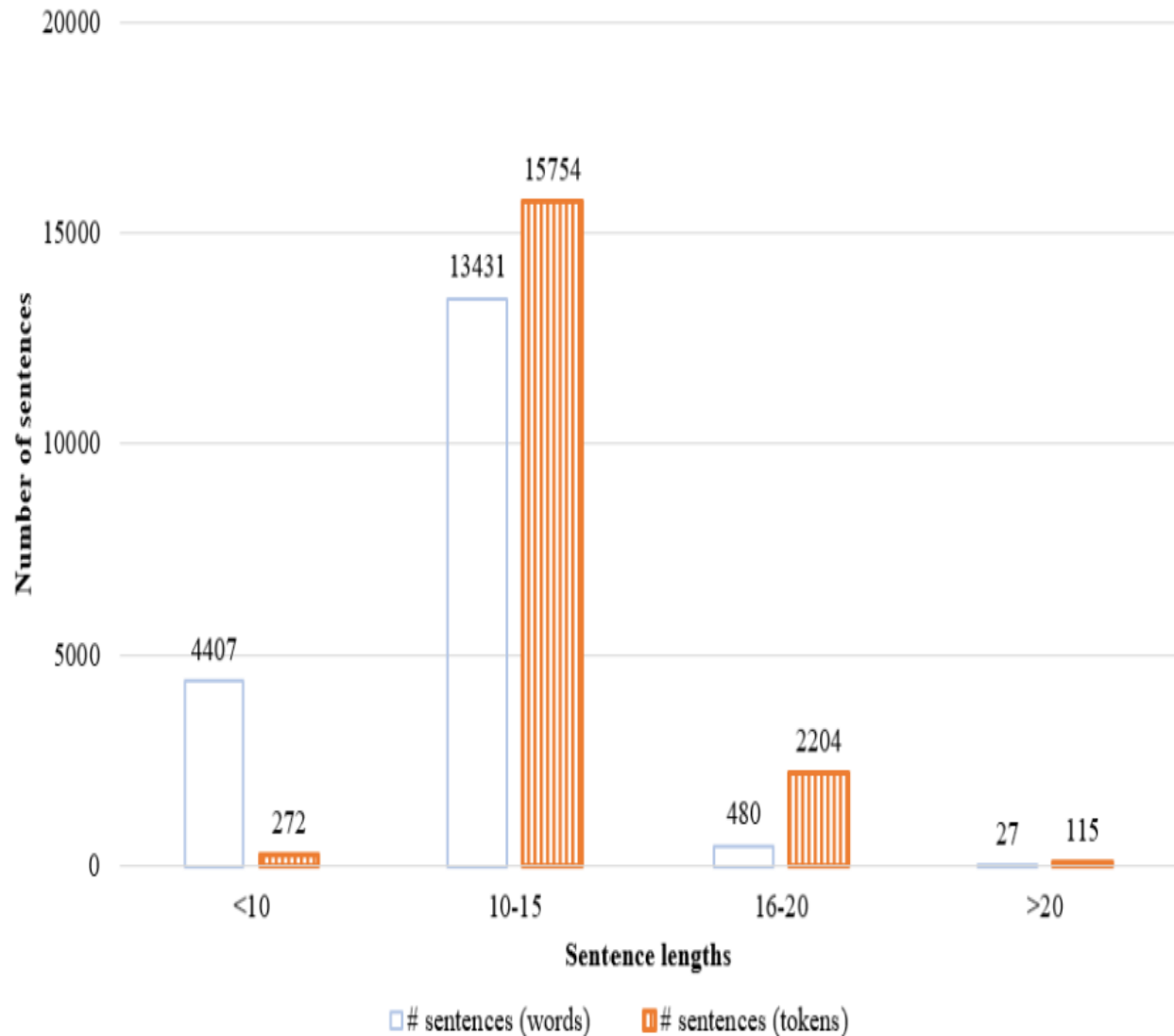
**Chú thích:** Vận động viên tennis nam đang cầm vợt đứng ngoài biên đánh bóng.



Câu hỏi	Câu trả lời
Ai đang chơi bóng ở ngoài biên đánh bóng?	Vận động viên tennis nam
Vận động viên tennis nam đang làm gì?	Cầm vợt đánh bóng
Ai đang làm gì với vợt đỡ bóng?	Vận động viên

### BỘ DỮ LIỆU

Chúng tôi sử dụng bộ dữ liệu bộ dữ liệu UIT-ViIC bộ dữ liệu này bao gồm 19,250 chú thích tiếng Việt cho 3,850 hình ảnh từ bộ dữ liệu Microsoft COCO liên quan đến các môn thể thao chơi với bóng.



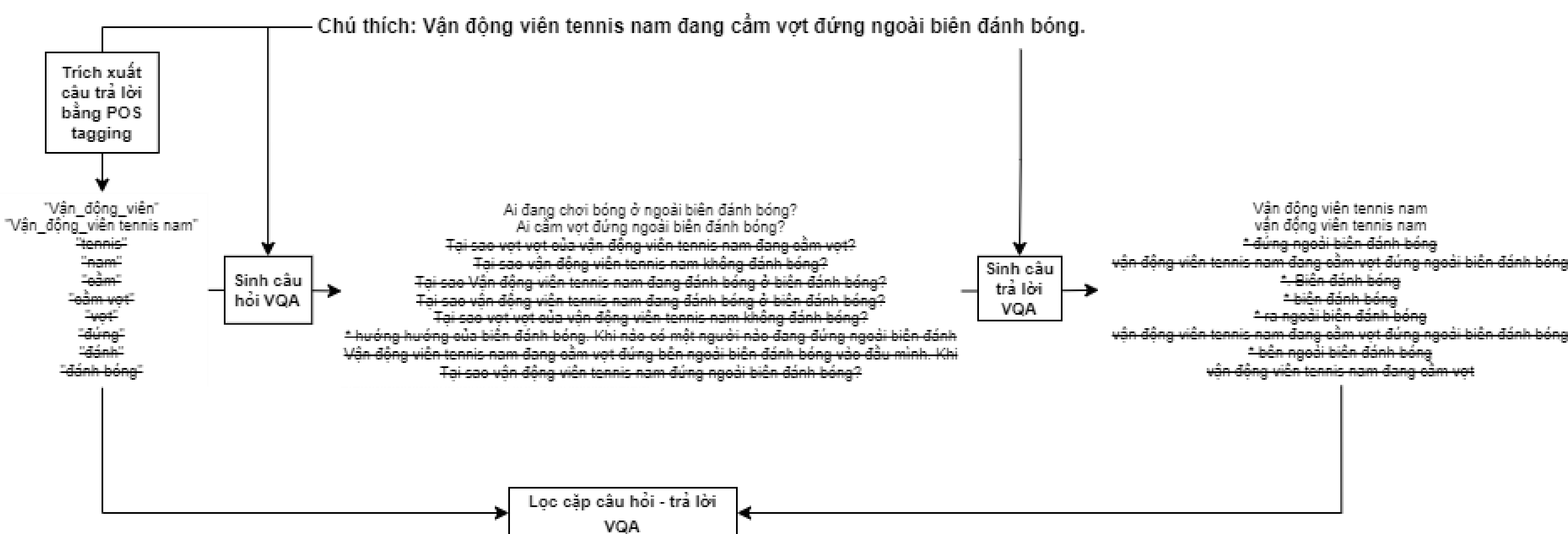
- Vận động viên tennis nam đang cầm vợt đứng ngoài biên đánh bóng.
- Vận động viên tennis nam đang vung vợt đánh bóng.
- Vận động viên tennis nam đang cầm vợt đỡ bóng.
- Vận động viên tennis đang cầm vợt thi đấu trên sân.
- Người đàn ông áo vàng đang chơi tennis trên sân.

### KẾT QUẢ

Chúng tôi đã tạo ra được bộ câu hỏi và câu trả lời tương ứng cho bộ dữ liệu UIT-ViIC. Chúng tôi tiến hành lọc và đánh giá bằng phương pháp F1 score token-level trong khoảng [0; 1] và trên 0.65 là đạt.

Câu trả lời	Câu hỏi đã tạo	Câu trả lời đã được	Kết quả đánh giá
Vận động viên	Ai đang chơi bóng ở ngoài biên đánh bóng?	Vận động viên tennis nam	0.7 (đạt)
Vận động viên tennis nam	Ai đang làm gì với vợt đỡ bóng?	Vận động viên tennis nam	1 (đạt)
Đánh bóng	Vận động viên tennis nam đang làm gì?	Vận động viên tennis nam đang cầm vợt	0 (không đạt)
Cầm vợt	Tại sao vận động viên tennis nam đang cầm vợt đỡ bóng?	vận động viên tennis nam đang cầm vợt đỡ bóng	0 (không đạt)
Cầm vợt đánh bóng	Vận động viên tennis đang làm gì?	Biên đánh bóng	0.5 (không đạt)

### PHƯƠNG PHÁP TIẾP CẬN



#### 1. Trích xuất câu trả lời VQA từ câu chú thích của ảnh

##### a. Gán nhãn từ loại (POS tagging)

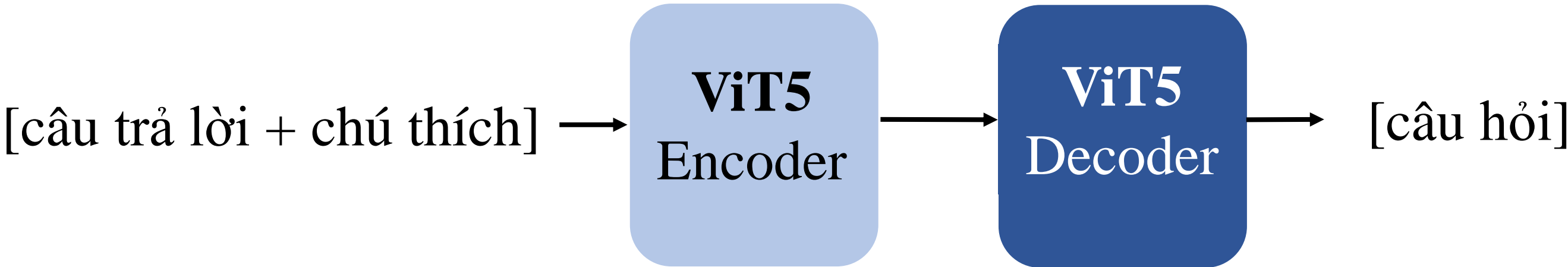
Chúng tôi đã sử dụng thư viện VNCORE NLP

##### b. Cấu trúc cây (Phrase structure tree)

Chúng tôi đã sử dụng cây cấu trúc để lấy các cụm danh từ, cụm động từ làm câu trả lời và đảm bảo các cụm từ đó đã bao gồm các cây con của các cụm từ.

#### 2. Sinh câu hỏi VQA

Chúng tôi đã tiến hành fine-tuning mô hình pre-trained ViT5 base dựa trên kiến trúc Transformer.



#### 3. Sinh câu trả lời VQA

Chúng tôi sử dụng cùng một phương pháp với cách tạo sinh câu hỏi VQA, bằng cách thay đổi đầu vào là cặp câu hỏi - ngữ cảnh.

### KẾT LUẬN

Dự định trong tương lai sẽ áp dụng các mô hình VQA baselines và SOTA multi modal fusion để có thể đánh giá phương pháp, bộ dữ liệu chúng tôi được chính xác và khách quan hơn.

Trong ngữ cảnh nghiên cứu bằng tiếng Việt, một trở ngại chung là thiếu các bộ dữ liệu huấn luyện và kiểm thử bằng tiếng Việt. Vì khác biệt về ngữ pháp và văn phạm, không khả thi để huấn luyện trên các bộ dữ liệu tiếng Anh hoặc các ngôn ngữ khác rồi áp dụng vào tiếng Việt. Tổng kết lại, bài báo đã giới thiệu về hướng tiếp cận mới trong việc xây dựng bộ dữ liệu VQA dựa trên bộ dữ liệu UIT-ViIC, hay mở rộng ra hơn là các bộ dữ liệu sinh chú thích ảnh trong tiếng Việt.