SINH MÔ TẢ CHO ẢNH VỚI MẠNG RESNET VÀ CƠ CHẾ ATTENTION

Lê Thành Đạt - 19521332
Vũ Tuấn Anh - 19521228
Cao Tuấn Anh - 19520008

Tóm tắt

- Lớp: CS519.M11.KHCL
- Link Github của nhóm: https://github.com/ledat1205/CS519.M11.KHCL
- Link YouTube video:
- Ảnh + Họ và Tên của các thành viên:

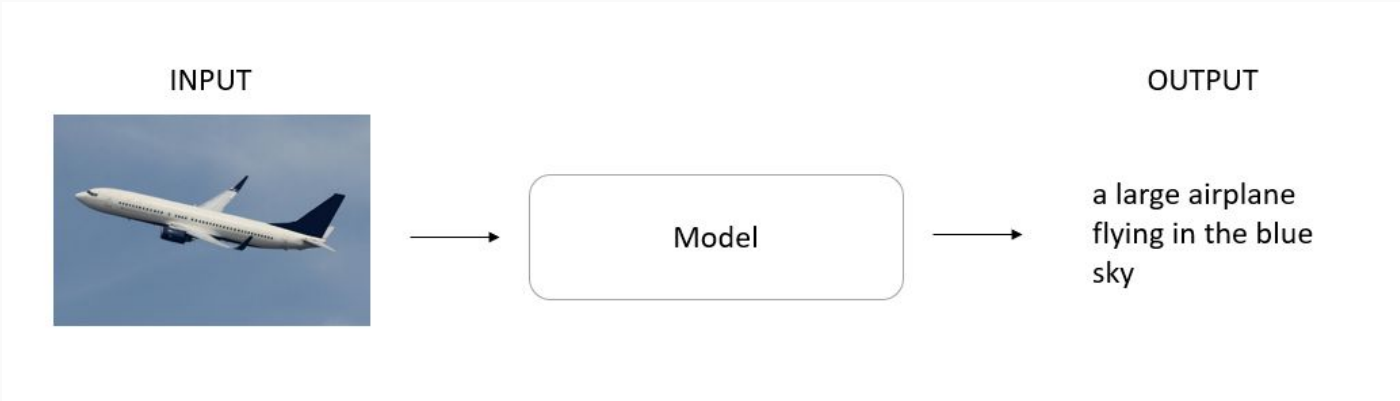
Lê Thành Đạt

Vũ Tuấn Anh

Cao Tuấn Anh

Giới thiệu

- Bài toán image captioning là một bài toán kết hợp giữa 2 cả lĩnh vực trong AI là thị giác máy tính và xử lý ngôn ngữ tự nhiên.
- Có nhiều ứng dụng trong thực tế như: hệ thống đề xuất trong các công cụ chỉnh sửa, xây dựng trợ lý ảo, cho các mạng xã hội



Giới thiệu

- Đề tài của chúng tôi tập trung cải tiến mô hình show and tell [1] trong bài toán image captioning.
- Thừa hưởng từ ý tưởng là mạng ResNet [3] và Attention [2]. Chúng tôi kết hợp nó nhằm cải thiện hiệu suất của mô hình show and tell ban đầu.
- Kết quả của mô hình sẽ được đánh giá trên độ đo BELU với 3 tập dữ liệu là: Flickr8k, Flickr30k và MSCOCO.

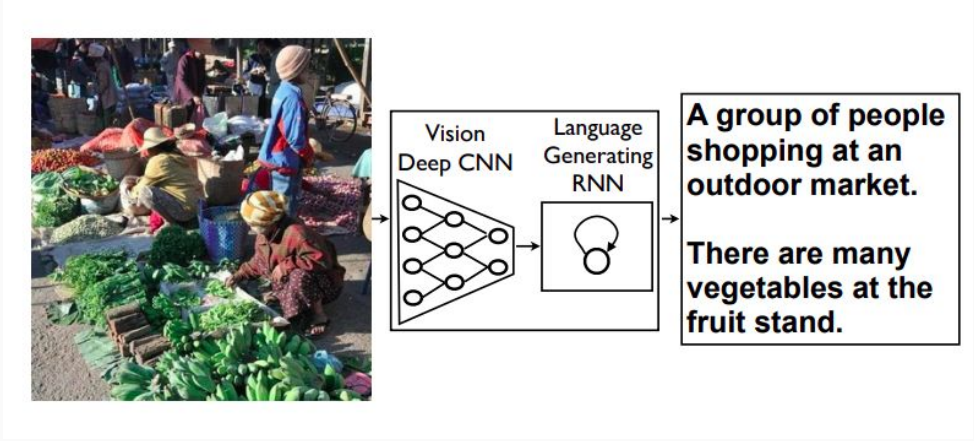
Mục tiêu

- Đề xuất một giải pháp end to end cho bài toán image captioning được cải tiến từ mô hình show and tell. Cụ thể là cải tiến mô show and tell với mạng ResNet và cơ chế Attention.
- Xây dựng mô hình với cả 2 hướng Attention là “Hard” Attention và “Soft” Attention.
- Tìm ra mô hình cải tiến với hiệu suất cao hơn dựa trên độ đo BELU trên 3 bộ dataset là: Flickr8k, Flickr30k và MSCOCO

Nội dung và Phương pháp

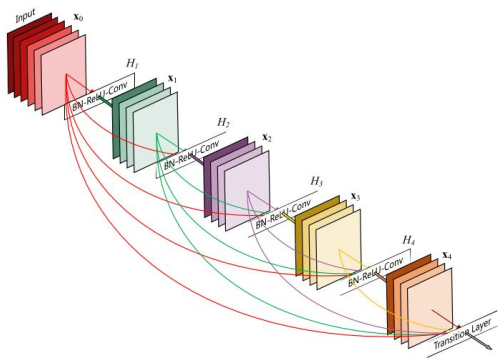
- Mô hình show and tell:

Mô hình sử dụng kiến trúc encoder-decoder với một mạng CNN đóng vai trò là encoder và mạng RNN với vai trò decoder để sinh dòng mô tả cho ảnh (cụ thể là mô hình LSTM).

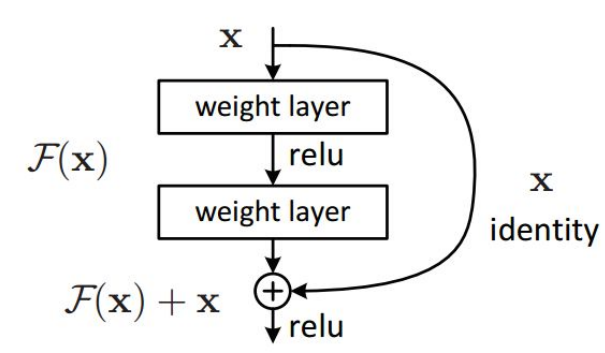


Nội dung và Phương pháp

- Mạng ResNet: Nghiên cứu về kiến trúc và khả năng trích xuất đặc trưng ảnh, kỹ thuật skip connection trong ResNet.
- Sử dựng mạng ResNet152 đóng vai trò là encoder trích xuất đặc trưng ảnh đầu vào.



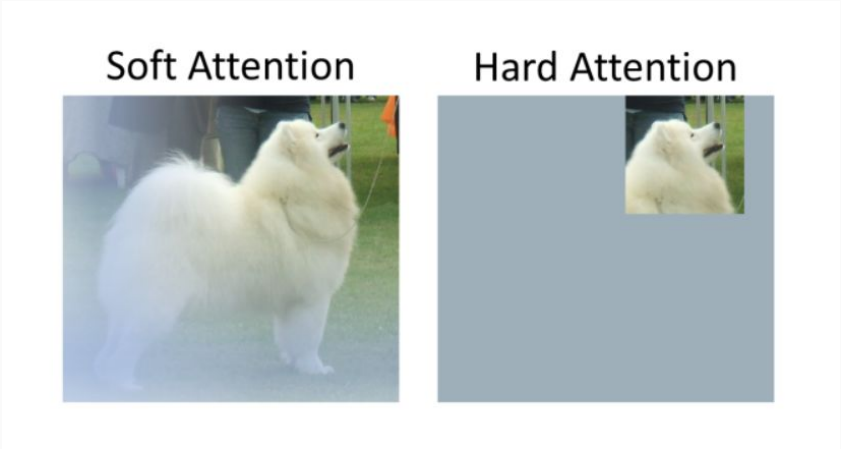
Kiến trúc ResNet



Skip connection trong ResNet

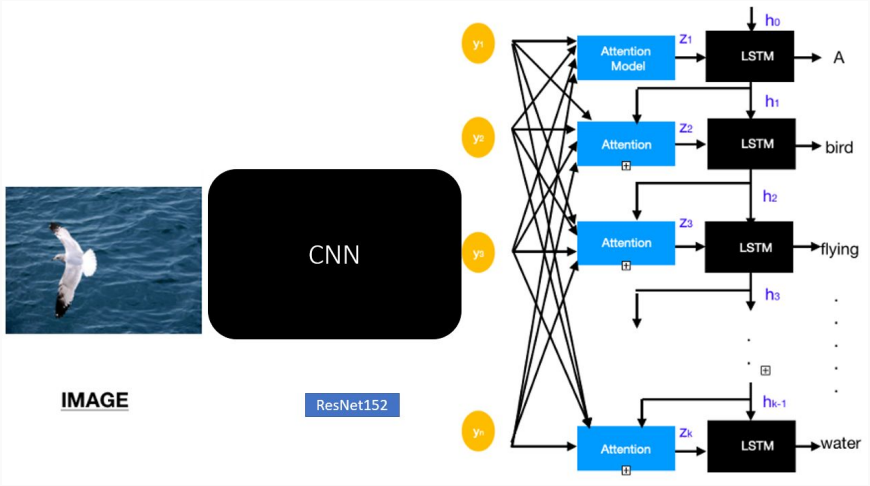
Nội dung và Phương pháp

- Cơ chế Attention: Tìm hiểu cấu trúc và cơ chế của Attention trên feature map của ảnh
- “Soft” Attention và “Hard” Attention: Tìm hiểu và áp dụng 2 concept attention trên ảnh



Nội dung và Phương pháp

- Độ đo BELU: Tìm hiểu cách đánh giá cho bài toán image captioning trên độ BELU, ý nghĩa và cách tính toán.



Cấu trúc của mô hình cải tiến.

Kết quả dự kiến

- Tài liệu báo cáo các phương pháp và kỹ thuật chi tiết để cải tiến mô hình show and tell ban đầu.
- Kết quả thực nghiệm, so sánh giữa các mô hình cải tiến với nhau và với mô hình ban đầu trên độ đo BELU với 3 tập dữ liệu: Flickr8k, Flickr30k và MSCOCO
- Mô hình cải tiến có kết quả thực nhiệm tốt hơn mô hình show and tell ban đầu

Tài liệu tham khảo

- [1]. Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan:
- Show and Tell: A Neural Image Caption Generator. CVPR 2015.
- [2]. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention Is All You Need. ArXiV 2017.
- [3]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition. CVPR 2016.
- [4]. Andrea Galassi, Marco Lippi, Paolo Torroni: Attention in Natural Language Processing. IEEE 2021.
- [5]. Ralf C. Staudemeyer, Eric Rothstein Morris: Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks. ArXiv, 2019