

ỨNG DỤNG MÔ HÌNH CNN & LSTM GÁN NHÃN CHO HÌNH ẢNH

Trương Thanh Luân - 220201016

Tóm tắt



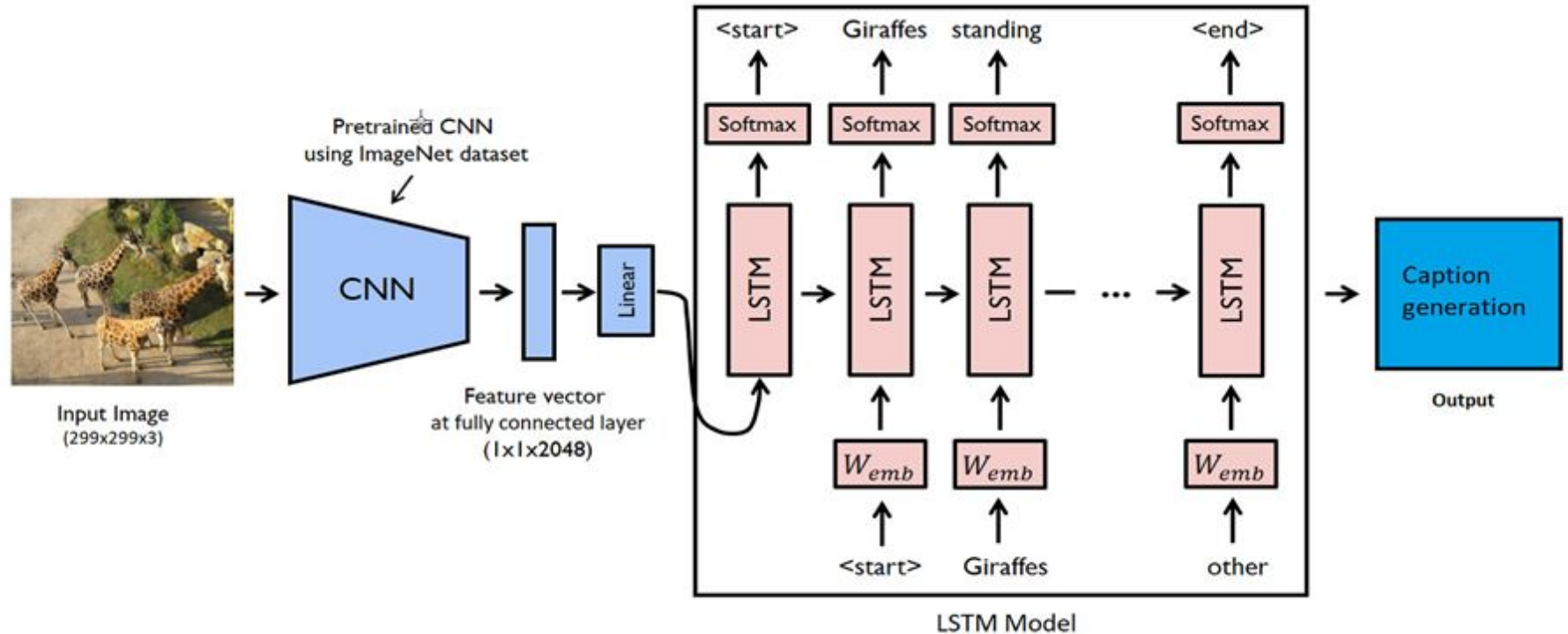
TRƯƠNG THANH LUÂN
MSHV: 220201016

- Link Github:
<https://github.com/thanhluanpy/CS2205.APR2023>
- Link YouTube video:
<https://youtu.be/9FkeVSkDSho>

Giới thiệu

- Bài toán:
 - Giúp những người khiếm thính hoặc các em nhỏ có thể biết được cảnh vật xung quanh (Image -> text -> voice).
 - Quản lý và tìm kiếm hình ảnh.
- Giải pháp: xây dựng mô hình CNN và LSTM để gán nhãn cho hình ảnh.
 - CNN được sử dụng để trích xuất các đặc trưng hình ảnh.
 - LSTM sử dụng thông tin từ CNN để tạo mô tả cho hình ảnh.

Giới thiệu



Mục tiêu

- Đạt độ chính xác cao: có khả năng nhận diện và phân loại đối tượng, vật thể và hành động trong hình ảnh một cách chính xác và đáng tin cậy.
- Tự động hóa quy trình gán nhãn: giúp tiết kiệm thời gian và công sức so với việc gán nhãn thủ công.
- Mở rộng ứng dụng: ứng dụng rộng lớn trong các lĩnh vực như quảng cáo, thương mại điện tử, y tế và nhiều lĩnh vực khác.

Nội dung và Phương pháp

- **Nội dung**

- Trích xuất đặc trưng hình ảnh bằng mô hình CNN.
- Chuyển đặc trưng hình ảnh thành ngôn ngữ tự nhiên bằng LSTM.
- Gán nhãn cho hình ảnh.

- **Phương pháp**

- Chuẩn bị dữ liệu huấn luyện Flickr8K.
- Xây dựng mô hình CNN & LSTM.
- Huấn luyện và đánh giá mô hình.
- Gán nhãn cho hình ảnh:
 - Input: tập hình ảnh bất kỳ.
 - Output: tập hình ảnh đã được gán nhãn.

Kết quả dự kiến

- Độ chính xác cao.
- Tự động hóa quá trình gán nhãn.
- Xử lý hình ảnh phức tạp và chuỗi hình ảnh.
- Tính linh hoạt và ứng dụng rộng rãi.



"man in black shirt is playing guitar."



"construction worker in orange safety vest is working on road."



"two young girls are playing with lego toy."

Tài liệu tham khảo

- [1] Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan: Show and Tell: A Neural Image Caption Generator. In arXiv:1411.4555, 2014. 1, 2
- [2] Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, Trevor Darrell: Long-term Recurrent Convolutional Networks for Visual Recognition and Description. In arXiv:1411.4389, 2014. 1, 2, 3, 4
- [3] Ayush Kumar Poddar, Dr. Rajneesh Rani: Hybrid Architecture using CNN and LSTM for Image Captioning in Hindi Language. In ICCECE, 2020.
- [4] Nguyễn Thanh Tuấn: Sách Deep Learning Cơ Bản. Tái bản lần 2 - Tháng 08, Năm 2020.
- [5] Dataset: Flickr8K Dataset & Flickr8K Text. Url:
<https://academictorrents.com/details/9dea07ba660a722ae1008c4c8afdd303b6f6e53b>
- [6] Data Flair Blog. Url:
<https://data-flair.training/blogs/python-based-project-image-caption-generator-cnn>
- [7] Basic CNN in Deep Learning. Url:
<https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning>