**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

TRƯỜNG ĐẠI HỌC BÁCH KHOA

KHOA ĐIỆN – ĐIỆN TỬ

**BỘ MÔN ĐIỆN TỬ –** **VIỄN THÔNG**

---------------o0o---------------

****

**TIỂU LUẬN**

**DỰ ĐOÁN TIÊU ĐỀ CHO ẢNH**

**GVHD: TS. VÕ TUẤN KIỆT**

**SVTH: NGUYỄN HÀ NHẤT PHƯƠNG**

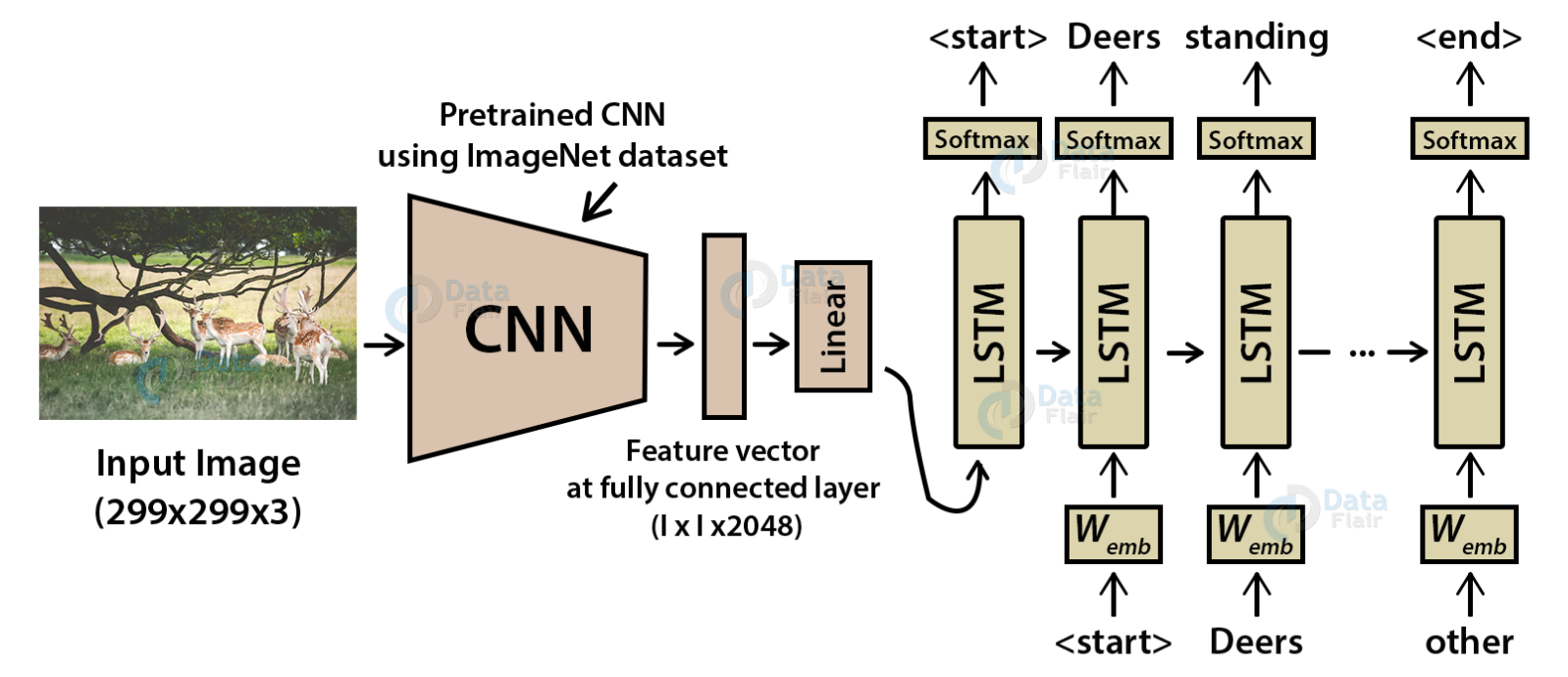
**MSSV: 1813622**

**TP. HỒ CHÍ MINH, THÁNG 12 NĂM 2021**

*YÊU CẦU*

Ở bài tiểu luận này chúng ta sẽ giải quyết bài toán Image captioning - kết hợp computer vision và NLP.

Một model kiểu mẫu cho bài toán này có thể được biểu diễn như sau:



Một lời giải cho bài toán này sẽ dựa vào một tutorial từ tensorflow: https://www.tensorflow.org/tutorials/text/image\_captioning. Đây cũng là nhiệm vụ cơ bản của bài tiểu luận này.

Ngoài ra, nhiệm vụ mở rộng của bài toán sẽ bao gồm việc khảo sát và thực thi những kiến trúc khác khả thi ở hai giai đoạn CNN và RNN/LSTM: Attention, Transformer ...

MỤC LỤC

[1 SỬ DỤNG OPTIMIZER ADAM 1](#_Toc92546193)

[2 SỬ DỤNG OPTIMIZER SGD 2](#_Toc92546194)

# SỬ DỤNG OPTIMIZER ADAM

Ở phần này (tệp *image\_captioning.ipynb*), cũng là phần dự đoán tốt nhất, em chọn các thông số như sau:

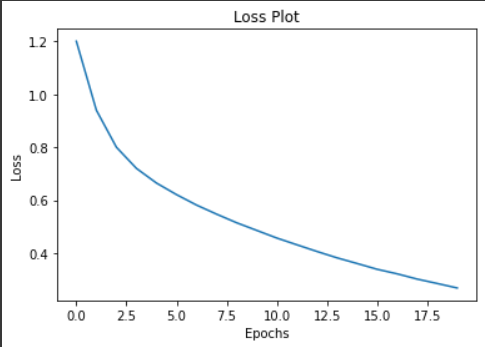
RNN\_decoder.attention: BahdanauAttention

Optimizer: adam

Loss function: SparseCategoricalCrossentropy

Epoch: 20

Đồ thị hàm loss:



Ảnh test và tiêu đề được đặt: a man surfing on a surfboard



# SỬ DỤNG OPTIMIZER SGD

Ở phần này, em thử 3 bộ thông số khác nhau.

* Trường hợp tốt nhất (tệp *image\_captioning3.ipynb*) là:

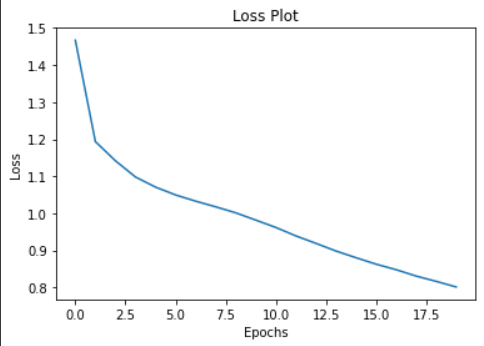
RNN\_decoder.attention: BahdanauAttention

Optimizer: SGD, learning rate = 0.02, momentum = 0.42

Loss function: SparseCategoricalCrossentropy

Epoch: 20

Đồ thị hàm loss:



Ảnh test và tiêu đề được đặt: a surfer that riding a dark board on a surfboard



* Trường hợp tệ thứ hai (tệp *image\_captioning4\_1.ipynb*):

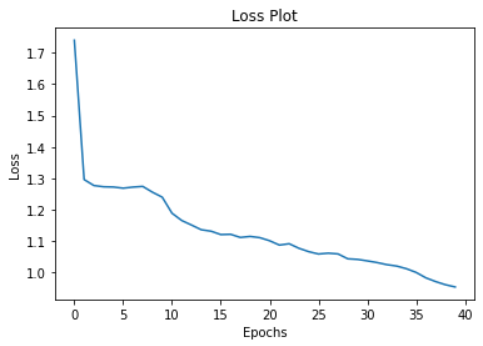
RNN\_decoder.attention: BahdanauAttention

Optimizer: SGD, learning rate = 0.015, momentum = 0.85

Loss function: SparseCategoricalCrossentropy

Epoch: 40

Đồ thị hàm loss:



Ảnh test và tiêu đề được đặt: dark with a racket



* Trường hợp tệ nhất (tệp *image\_captioning2.ipynb*):

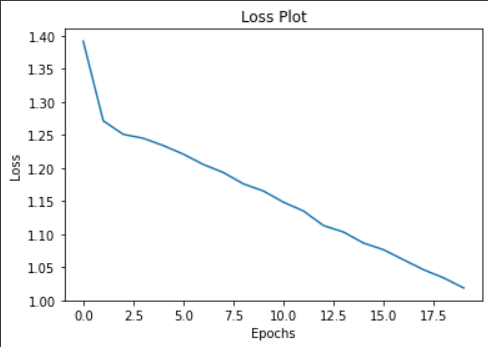
RNN\_decoder.attention: BahdanauAttention

Optimizer: SGD, learning rate = 0.002, momentum = 0.64

Loss function: SparseCategoricalCrossentropy

Epoch: 20

Đồ thị hàm loss:



Ảnh test và tiêu đề được đặt: a woman [UNK]

