SEMINAR CS406.P11

Image Captioning

21521465 - Trần Ngọc Thiện

Giới thiệu

Mô tả bài toán

Bài toán image captioning (tạo chú thích cho hình ảnh) là một lĩnh vực rất hấp dẫn trong thị giác máy tính và xử lý ngôn ngữ tự nhiên. Đây là một bài toán phức tạp, đòi hỏi máy tính không chỉ hiểu được nội dung của hình ảnh mà còn có khả năng diễn đạt chúng bằng ngôn ngữ tự nhiên một cách chính xác và mạch lạc.

Mô tả bài toán

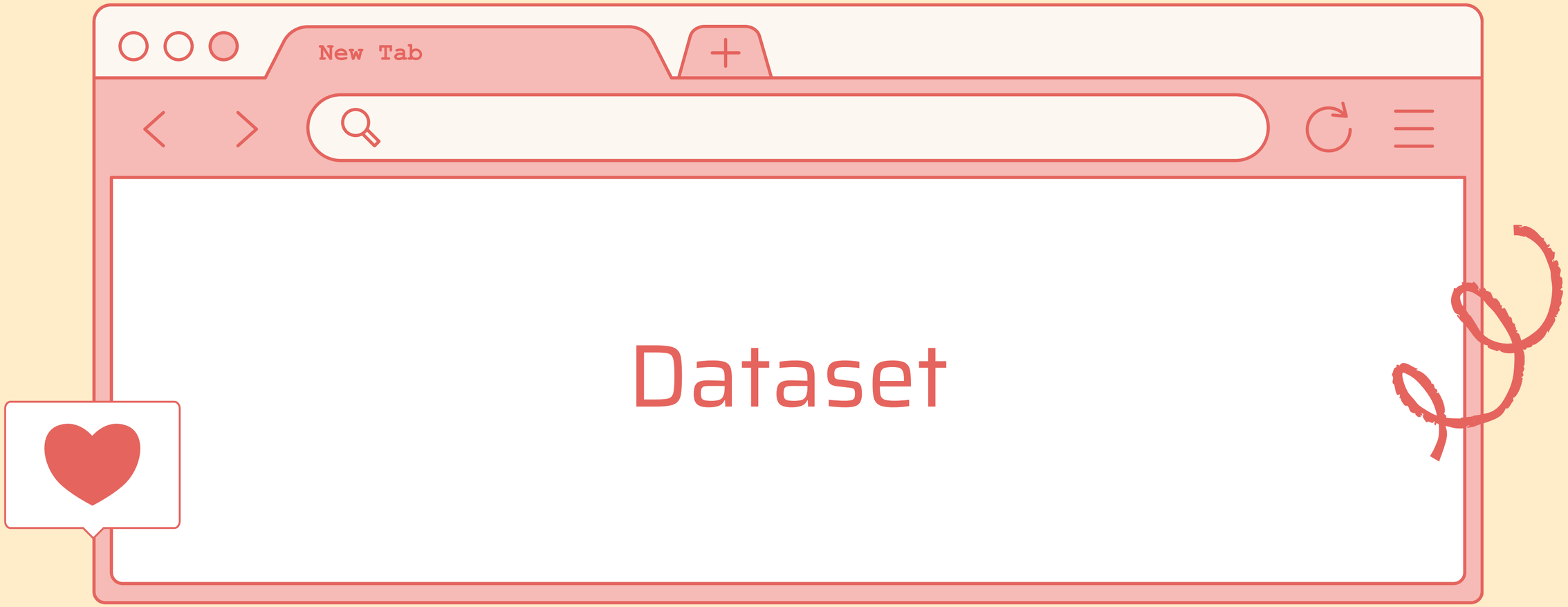
- Input: Hình ảnh
- Output: Câu mô tả nội dung hình ảnh

Ứng dụng

- Tìm kiếm hình ảnh dựa trên văn bản
- Hỗ trợ người khuyết tật
- Truyền thông xã hội
- Phân tích hình ảnh y tế

Thách thức

- Nhiều cách diễn đạt
- Chi tiết phức tạp
- Ngữ cảnh
- Hình ảnh có nhiều đối tượng, các mối quan hệ phức tạp giữa các đối tượng
- Hình ảnh có đối tượng nhỏ hoặc bị che khuất



Dataset

Dataset

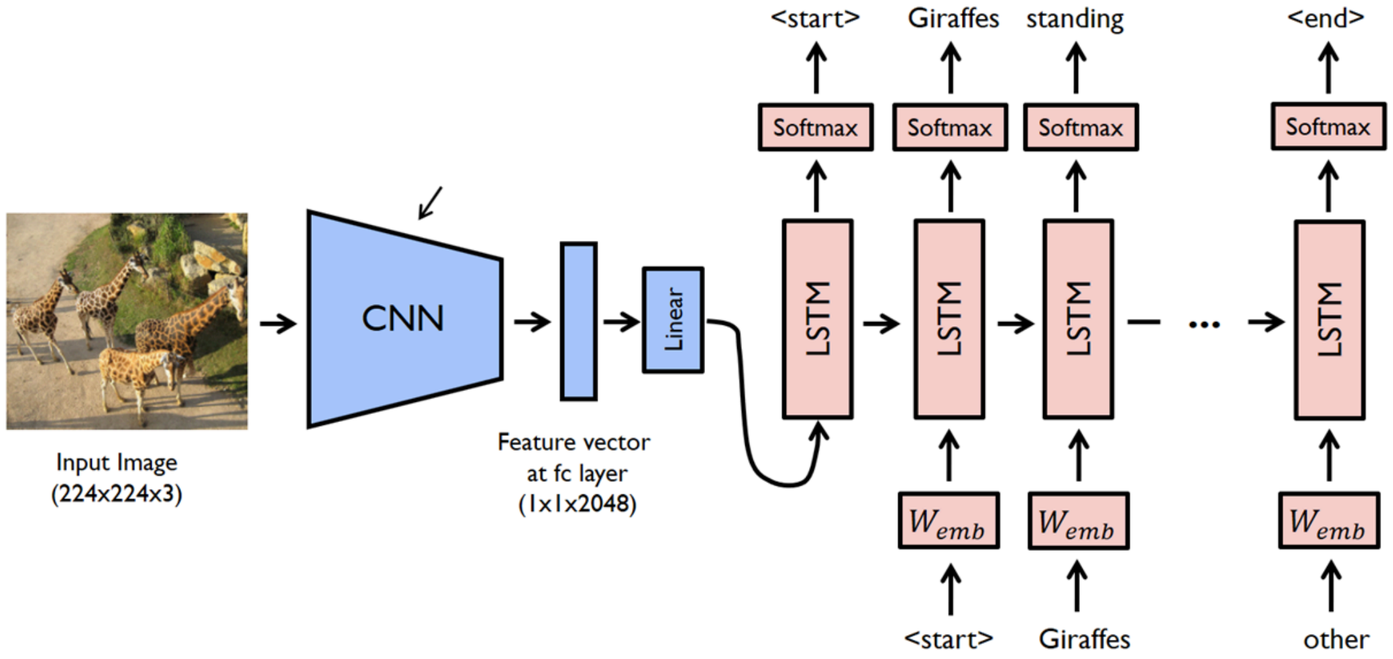
- Sử dụng UIT-ViIC (version 1.0)
- Bộ dữ liệu có 3850 hình ảnh chủ đề thể thao lấy từ MS-COCO với 19250 câu caption bằng tiếng Việt (5 câu cho mỗi hình ảnh)

- Một cậu bé đang chống đẩy trên quả banh.
- Một cậu bé đang chống đầu lên quả bóng đá hít đất.
- Một cậu bé đang hít đất đầu trên quả banh trên cỏ.
- Một cậu bé đang hít đất cùng quả bóng và một cậu bé đang quan sát.
- Một cậu bé đang chống đẩy trên một quả bóng đá.

Phương pháp

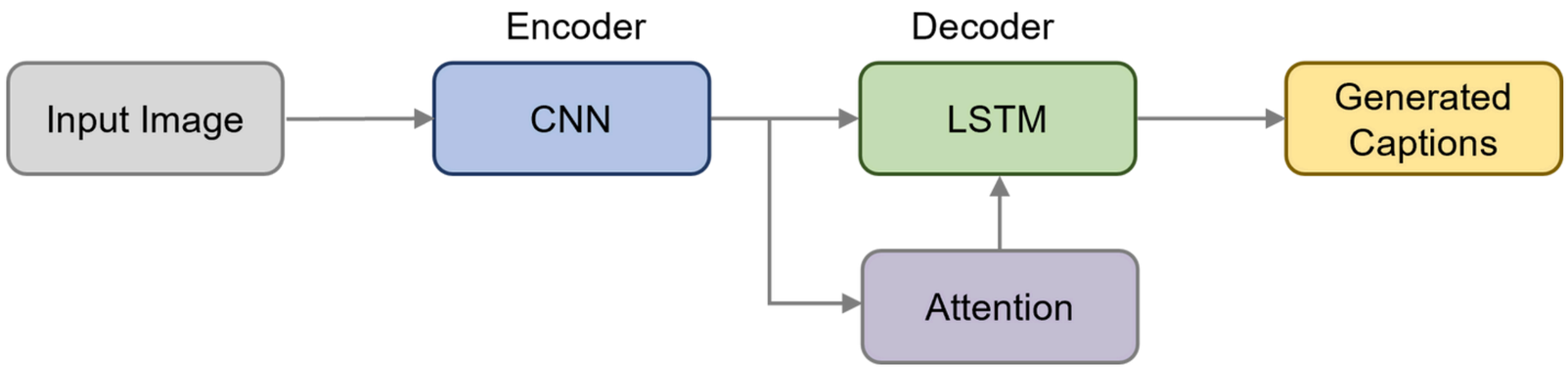
Phương pháp truyền thống

Sử dụng encoder (CNN) để xử lý ảnh và decoder (RNN, LSTM) để tạo caption



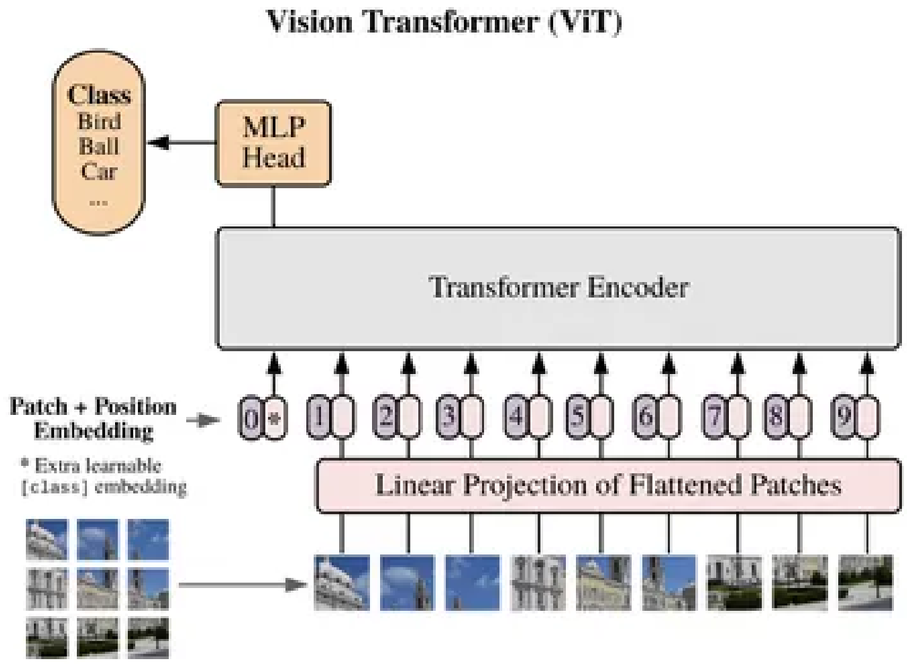
Phương pháp truyền thống

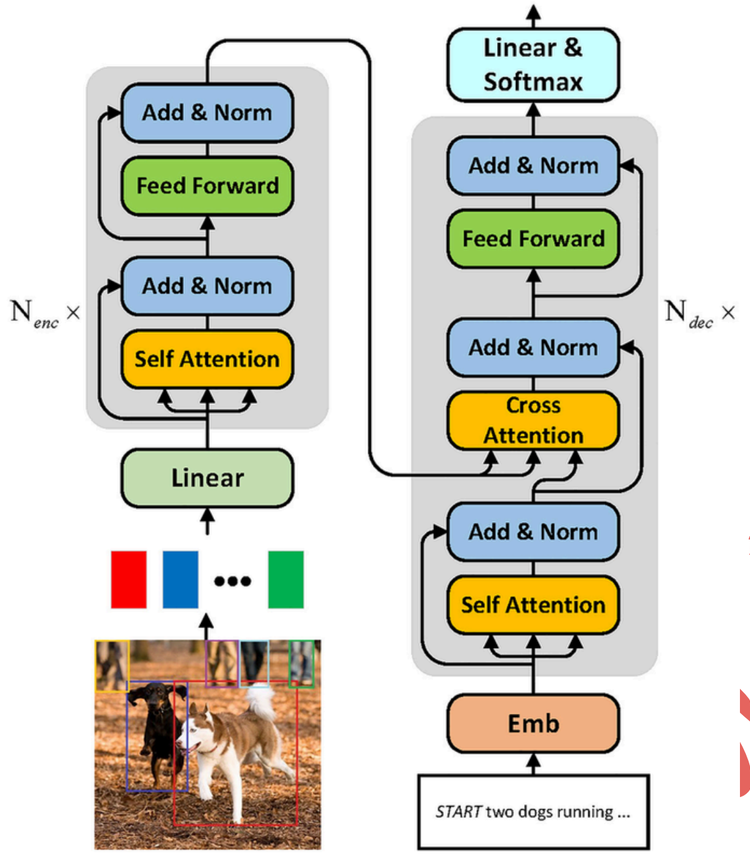
- Ưu điểm:
  - Kiến trúc dễ hiểu và triển khai.
  - Đã được chứng minh hiệu quả trên nhiều bài toán khác nhau.
- Nhược điểm:
  - Khó xử lý các chuỗi dài do vấn đề gradient vanishing/exploding.
  - Không tận dụng tốt thông tin toàn cục: RNN chỉ xử lý thông tin theo trình tự tuần tự, khó nắm bắt các mối quan hệ giữa các từ cách xa nhau trong câu.



Phương pháp transformer-based

Sử dụng mô hình vision transformer





Phương pháp transformer-based

- Ưu điểm:
  - Xử lý song song
  - Cơ chế attention giúp Transformer nắm bắt các mối quan hệ giữa các từ cách xa nhau trong câu.
  - Hiệu suất cao
- Nhược điểm:
  - Phức tạp
  - Tốn tài nguyên tính toán

Đánh giá

Độ đo BLEU Score

- BLEU (Bilingual Evaluation Understudy) là một trong những metric phổ biến nhất để đánh giá chất lượng của các mô hình tạo văn bản.

- Tại sao sử dụng BLEU?
  - Đơn giản
  - Hiệu quả
  - Phổ biến
- Hạn chế của BLEU
  - Chỉ đo lường sự trùng lặp, không đánh giá được tính lưu loát, ý nghĩa của câu.
  - Không đánh giá sự đa dạng
  - Nhạy cảm với các từ hiếm

Độ đo BLEU Score

- Precision 1-gram

Target Sentence: The guard arrived late because ~~it was raining~~

Predicted Sentence: The guard arrived late because of the rain

- Precision 2-gram

Target Sentence: The guard arrived late because it was raining

Predicted Sentence: The guard arrived late because of the rain

Độ đo BLEU Score

$$
\begin{aligned}
Geometric\ Average\ Precision\ (N) &= exp(\sum_{n=1}^{N} w_n\ log\ p_n) \\
&= \prod_{n=1}^{N} {p_n}^{w_n} \\
&= (p_1)^{\frac{1}{4}} \cdot (p_2)^{\frac{1}{4}} \cdot (p_3)^{\frac{1}{4}} \cdot (p_4)^{\frac{1}{4}}
\end{aligned}
$$

Độ đo BLEU Score

$$Brevity\ Penalty = \begin{cases} 1, & \text{if } c > r \\ e^{(1-r/c)}, & \text{if } c <= r \end{cases}$$

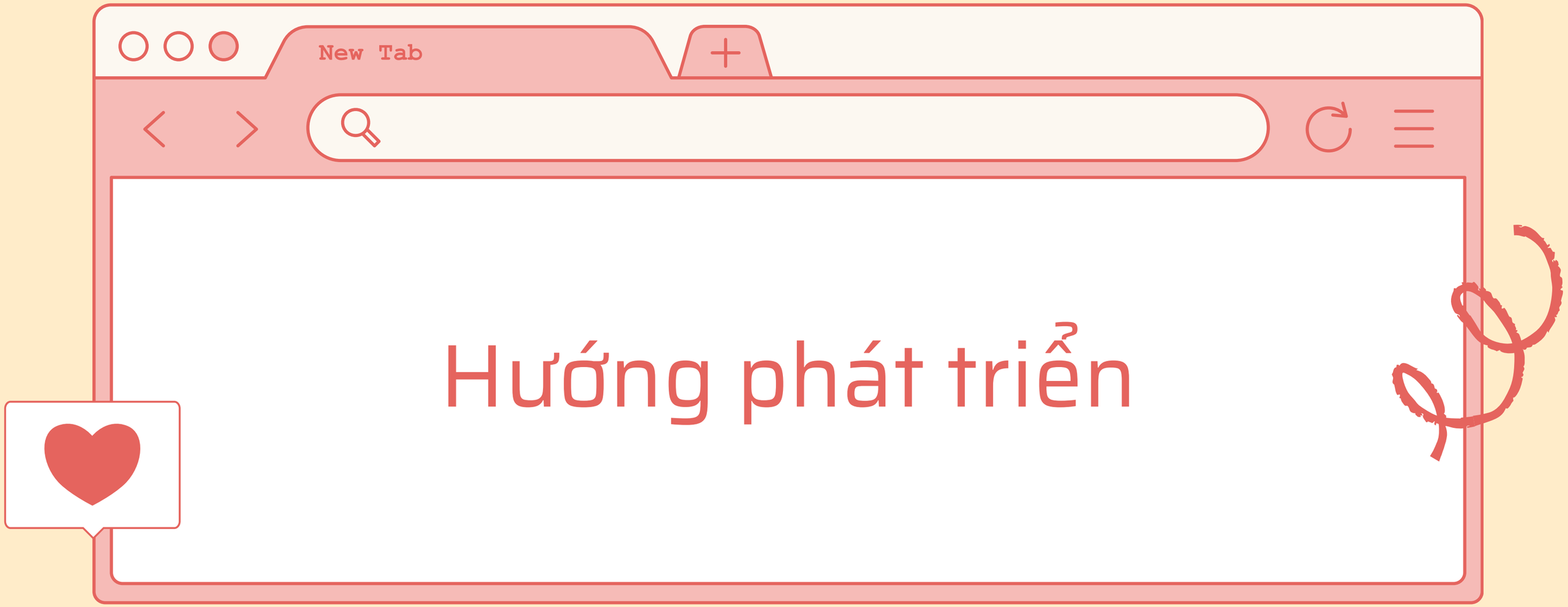
- c là predicted length = số lượng từ có trong predicted sentence
- r là target length = số lượng từ có trong target sentence

Độ đo BLEU Score

$$Bleu\ (N) = Brevity\ Penalty \cdot Geometric\ Average\ Precision\ Scores\ (N)$$

Bleu Score có thể tính toán cho nhiều giá trị N khác nhau. Cụ thể trong trường hợp N = 4.

- BLEU-1 sử dụng unigram Precision Score.
- BLEU-2 sử dụng geometric average of unigram and bigram Precision Score.
- BLEU-3 sử dụng geometric average of unigram, bigram, and trigram Precision Score.
- Và cứ tiếp tục như thế cho đến BLEU-N.



Hướng phát triển

Hướng phát triển

- Cải thiện chất lượng chú thích (đa dạng, chi tiết, theo ngữ cảnh)
- Xử lý trên dữ liệu đa dạng và phức tạp hơn
- Video captioning
- Multimodal learning: Kết hợp thông tin từ nhiều modal khác nhau (ví dụ: văn bản, âm thanh) để tạo ra các mô tả phong phú hơn.

Cảm ơn thầy và các bạn đã lắng nghe