**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**ĐỒ ÁN CUỐI KÌ**

**CÔNG NGHỆ MỚI TRONG PHÁT TRIỂN ỨNG DỤNG**

*Người thực hiện*: **TĂNG QUANG NHẬT NAM – 21082941**

**HOÀNG THANH TÚ – 21105251**

Lớp**: DHKHDL17A**

Khoá**: 17**

*Người hướng dẫn*: **TS. BÙI THANH HÙNG**

**Ths. TRƯƠNG VĨNH LINH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**ĐỒ ÁN CUỐI KÌ**

**CÔNG NGHỆ MỚI TRONG PHÁT TRIỂN ỨNG DỤNG**

*Người thực hiện*: **TĂNG QUANG NHẬT NAM – 21082941**

**HOÀNG THANH TÚ – 21105251**

Lớp**: DHKHDL17A**

Khoá**: 17**

*Người hướng dẫn*: **TS. BÙI THANH HÙNG**

**Ths. TRƯƠNG VĨNH LINH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025**

**LỜI CẢM ƠN**

Đồ án được sự thực hiện bởi 2 thành viên (Hoàng Thanh Tú và Tăng Quang Nhật Nam) lớp DHKHDL17A dưới sự hướng dẫn của Thầy Bùi Thanh Hùng và Thầy Trương Vĩnh Linh giảng viên môn Công nghệ mới trong phát triển ứng dụng. Thầy đã tạo điều kiện và giúp đỡ về kiến thức và hỗ trợ các thắc mắc về các vấn đề trong quá trình thực hiện để chúng em hoàn thành đồ án. Chúng em xin chân thành cảm ơn Thầy.

Với toàn bộ tâm huyết và cố gắng trong suốt thời gian qua, cùng với những kiến thức đã học được và một số kinh nghiệm tích lũy trong quá trình học tập đã giúp chúng em hoàn thành bài đồ án cuối kì này. Tuy nhiên, bài đồ án này vẫn còn nhiều thiếu sót trong việc thực hiện nghiên cứu, trình bày và đánh giá. Chúng em rất mong nhận được sự thông cảm và đóng góp ý kiến của quý thầy cô và các bạn.

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng chúng tôi và được sự hướng dẫn của TS. Đặng Thị Phúc. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo. Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Công nghiệp TP Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Tăng Quang Nhật Nam *

*Hoàng Thanh Tú*

**

# PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

# TÓM TẮT

Trong lĩnh vực xử lý ngôn ngữ tự nhiên và thị giác máy tính, bài toán tạo chú thích tự động cho hình ảnh (Image Captioning) là một thách thức lớn, kết hợp giữa việc phân tích hình ảnh và ngôn ngữ. Bài toán này yêu cầu mô hình phải hiểu nội dung của hình ảnh và diễn đạt chúng thành các câu văn tự nhiên.

Vấn đề nghiên cứu: Nghiên cứu tập trung vào việc phát triển các mô hình hiệu quả để tạo chú thích tự động cho hình ảnh, nhằm cải thiện độ chính xác và tự nhiên của các câu chú thích. Mục tiêu chính là tạo ra các mô hình có thể hiểu ngữ cảnh của hình ảnh và sinh ra các câu văn có nghĩa, mạch lạc. [1]

Các hướng tiếp cận: Mô hình CNN+ LSTM và Mô hình Transformer

Cách giải quyết vấn đề: Nghiên cứu triển khai và so sánh hai mô hình: CNN+LSTM và Transformer. Mô hình CNN+LSTM kết hợp mạng CNN để trích xuất đặc trưng hình ảnh và mạng LSTM để sinh ra câu chú thích. Mô hình Transformer sử dụng cơ chế tự chú ý để học các mối quan hệ dài hạn và tạo ra câu chú thích mạch lạc hơn. [2]

Kết quả đạt được và những phát hiện cơ bản: Mô hình Transformer vượt trội về chỉ số BLEU so với mô hình CNN+LSTM, đặc biệt là BLEU-1, cho thấy khả năng sinh ra câu chú thích chính xác và tự nhiên hơn. Transformer có khả năng nắm bắt ngữ cảnh dài hạn và sinh ra các câu chú thích mạch lạc hơn.

Kết luận: Nghiên cứu khẳng định việc sử dụng mô hình Transformer trong bài toán tạo chú thích tự động cho hình ảnh mang lại hiệu quả cao hơn so với CNN+LSTM. Đây là bước tiến quan trọng trong phát triển hệ thống AI có khả năng hiểu và mô tả nội dung hình ảnh tự nhiên và chính xác, mở ra hướng đi mới cho ứng dụng Transformer trong các tác vụ kết hợp giữa thị giác máy tính và xử lý ngôn ngữ tự nhiên.

MỤC LỤC

[LỜI CẢM ƠN 2](#_Toc183974315)

[TÓM TẮT 5](#_Toc183974317)

[MỤC LỤC 6](#_Toc183974318)

[DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT 7](#_Toc183974319)

[DANH MỤC CÁC HÌNH VẼ 8](#_Toc183974320)

[SINH CHÚ THÍCH MỘT ẢNH BẰNG CÁC PHƯƠNG PHÁP HỌC SÂU 9](#_Toc183974321)

[1.1 Giới thiệu về bài toán 9](#_Toc183974322)

[1.2 Phân tích yêu cầu của bài toán 10](#_Toc183974323)

[1.2.1 Yêu cầu của bài toán 10](#_Toc183974324)

[1.2.2 Các phương pháp giải quyết bài toán 10](#_Toc183974325)

[1.2.3 Phương pháp đề xuất giải quyết bài toán 12](#_Toc183974326)

[1.3 Phương pháp giải quyết bài toán 13](#_Toc183974327)

[1.3.1 Mô hình tổng quát 13](#_Toc183974328)

[1.3.2 Đặc trưng của mô hình đề xuất 15](#_Toc183974329)

[1.4 Thực nghiệm 17](#_Toc183974330)

[1.4.1 Dữ liệu 17](#_Toc183974331)

[1.4.2 Xử lý dữ liệu 18](#_Toc183974332)

[1.4.3 Công nghệ sử dụng 18](#_Toc183974333)

[1.4.4 Cách đánh giá 19](#_Toc183974334)

[1.5 Kết quả đạt được 20](#_Toc183974335)

[1.6 Kết luận 25](#_Toc183974336)

[LÀM VIỆC NHÓM 27](#_Toc183974337)

[TÀI LIỆU THAM KHẢO 29](#_Toc183974338)

[TỰ ĐÁNH GIÁ (Bài nhóm) 31](#_Toc183974340)

# DANH MỤC KÍ HIỆU VÀ CHỮ VIẾT TẮT

**CÁC CHỮ VIẾT TẮT**

CNN Mạng nơ-ron tích chập (Convolutional Neural Network)

LSTM Mạng nơ-ron hồi tiếp dài ngắn hạn (Long Short-Term Memory)

BLUE Chỉ số đánh giá chất lượng bản dịch (Bilingual Evaluation Understudy)

# 

# DANH MỤC CÁC HÌNH VẼ

[Hình 1.3.1.1: Mô hình tổng quát giải quyết bài toán sinh chú thích cho ảnh 13](#_Toc167186234)

[Hình 1.5.1: Sự mất mát trong quá trình huấn luyện và xác thực 23](#_Toc167186236)

[Hình 1.5.2: Sự phân phối độ dài của các chú thích 22](#_Toc167186235)

[Hình 1.5.3: Sự mất mát trong quá trình huấn luyện và xác thực 23](#_Toc167186236)

## CHƯƠNG 1

## PHÂN TÍCH, THIẾT KẾ

* 1. **Mô tả bài toán**

Trong các tòa nhà văn phòng, trung tâm đào tạo hoặc khu cho thuê, việc quản lý thiết bị và bảo trì hiện nay chủ yếu còn thủ công, gây mất thời gian, khó kiểm soát và thiếu minh bạch.

**BMSync** (Building Management – Synchronization) một nền tảng quản lý thiết bị thông minh, tích hợp chatbot hỗ trợ truy vấn thông tin nhanh, phân quyền rõ ràng, và tự động hóa toàn bộ quy trình bảo trì – từ lúc thiết bị gặp sự cố đến khi xử lý xong và xác nhận hoàn tất.

Hệ thống hướng tới việc:

* Giúp người thuê dễ dàng báo lỗi thiết bị và theo dõi tiến độ xử lý.
* Giúp quản lý và admin theo dõi thiết bị toàn tòa nhà, phân công hợp lý và giám sát hiệu quả bảo trì.
* Giúp nhân viên kỹ thuật nhận thông báo công việc, cập nhật tiến độ, và xác nhận hoàn thành nhanh chóng.
* Tích hợp chatbot thông minh để truy vấn: "Hôm nay có ai đang bảo trì không?", "Có bao nhiêu thiết bị ở tầng 5?", "Thiết bị này đã bảo trì bao giờ chưa?" – trả về bảng dữ liệu trực quan và link tới chức năng chi tiết.
* Hỗ trợ thông báo tự động khi có thiết bị cần xử lý, tài khoản mới cần xác nhận, hay công việc vừa hoàn tất.

**1.2. Sơ đồ chức năng tổng quát**

**1.3. Biểu đồ trường hợp sử dụng Usercase**

**1.4. Biểu đồ hoạt động**

**1.5. Biểu đồ trình tự**

**1.6. Biểu đồ Lớp (Class diagram)**

**1.7. Biểu đồ luồng dữ liệu Database diagram**

**1.8. Biểu đồ mối quan hệ giữa các dữ liệu**

**1.9. Thiết kế giao diện (các giao diện chính)**

**1.10. Thiết kế giải thuật (nếu có, nếu làm về xây dựng ứng dụng sử dụng học máy/học sâu thì trình bày chi tiết về tổng quan mô hình đề xuất, rút trích đặc trưng và giải thuật học ở mục này)**

**1.11. Thiết kế cách tiến hành Test**

Bài toán sinh chú thích cho ảnh là một trong những bài toán quan trọng trong lĩnh vực trí tuệ nhân tạo, nằm ở giao điểm của thị giác máy tính và xử lý ngôn ngữ tự nhiên. Mục tiêu của bài toán này là phát triển các mô hình máy học có khả năng tự động tạo ra mô tả mô tả hình ảnh một cách tự nhiên từ một ảnh đầu vào. Điều này không chỉ đòi hỏi máy tính hiểu được nội dung của hình ảnh mà còn phải có khả năng diễn đạt chúng bằng ngôn ngữ tự nhiên. Điều này tạo ra một thách thức đa chiều đòi hỏi tính đa ngữ cả về hình ảnh và ngôn ngữ. [3]

Ý nghĩa của bài toán này là rất đa dạng và sâu sắc. Một trong những mặt quan trọng nhất là trong lĩnh vực của trí tuệ nhân tạo, bài toán này đóng vai trò như một bước tiến quan trọng trong việc đưa máy tính gần hơn với khả năng hiểu và tương tác với thế giới xung quanh một cách tự nhiên và thông minh.

Ứng dụng của bài toán này cũng rất rộng lớn. Ví dụ, trong lĩnh vực y tế, một hệ thống có khả năng mô tả hình ảnh có thể cung cấp thông tin chi tiết về hình ảnh y tế mà các chuyên gia y tế có thể sử dụng để đưa ra chẩn đoán chính xác và kế hoạch điều trị. Trong lĩnh vực giáo dục, một ứng dụng có khả năng mô tả hình ảnh có thể hỗ trợ việc học tập và giảng dạy bằng cách cung cấp mô tả chi tiết và đa dạng về các hình ảnh đồ họa.

Ngoài ra, việc nghiên cứu và phát triển các phương pháp hiệu quả để sinh chú thích cho ảnh cũng đóng góp vào sự tiến bộ của lĩnh vực trí tuệ nhân tạo và máy học. Các phương pháp và công nghệ phát triển từ bài toán này có thể được áp dụng vào nhiều lĩnh vực khác như robotica, xe tự lái và công nghệ thông tin.

## 1.2 Phân tích yêu cầu của bài toán

### 1.2.1 Yêu cầu của bài toán

Xác định ảnh đầu vào: Yêu cầu cơ bản của bài toán là có một ảnh đầu vào mà chúng ta muốn sinh chú thích cho nó. Ảnh này có thể được biểu diễn dưới dạng một ma trận pixel, với mỗi pixel có các giá trị màu tương ứng.

Xác định dữ liệu huấn luyện: Để huấn luyện một mô hình sinh chú thích, chúng ta cần có một tập dữ liệu huấn luyện chứa các cặp ảnh và chú thích tương ứng. Dữ liệu này thường được gán nhãn bởi con người, nghĩa là mỗi ảnh được kèm theo một hoặc nhiều câu chú thích.

Định dạng chú thích: Yêu cầu tiếp theo là xác định định dạng của các chú thích. Chúng có thể là các câu ngắn mô tả nội dung của ảnh, ví dụ: "Một người đàn ông đang lái xe đạp trên con đường". Đối với mỗi ảnh, có thể có nhiều chú thích khác nhau phản ánh các góc nhìn và mô tả khác nhau.

Yêu cầu về độ chính xác: Một yếu tố quan trọng khác là đánh giá độ chính xác của các chú thích được sinh ra bởi mô hình. Điều này có thể được đo lường bằng các phương pháp đánh giá như BLEU score, METEOR score, CIDEr score, hoặc thông qua đánh giá của con người.

Khả năng tùy chỉnh: Trong một số trường hợp, bài toán cũng có thể yêu cầu khả năng tùy chỉnh mô hình để phù hợp với nhu cầu cụ thể của ứng dụng, chẳng hạn như tùy chỉnh mô hình để tạo ra các chú thích chuyên biệt cho một lĩnh vực cụ thể như y học, du lịch, hoặc thể thao.

### 1.2.2 Các phương pháp giải quyết bài toán

**Phương pháp Encoder-Decoder**

- Bài báo tham khảo: "Show and Tell: A Neural Image Caption Generator" (Oriol Vinyals et al., 2015). [4]

- Phương pháp: Sử dụng một mạng CNN để mã hóa thông tin từ ảnh và một mạng RNN để giải mã và sinh chú thích tương ứng.

- Dữ liệu thực nghiệm: Tác giả sử dụng tập dữ liệu MS COCO, một tập dữ liệu lớn chứa hình ảnh và các chú thích tương ứng.

- Kết quả: Mô hình đạt được kết quả tốt trên các tiêu chí đánh giá như BLEU score, METEOR score, và CIDEr score. Tuy nhiên, mô hình có thể gặp khó khăn khi mô tả các ảnh phức tạp và có nhiều đối tượng.

**Phương pháp Attention-Based**

- Bài báo tham khảo: "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention" (Kelvin Xu et al., 2015). [5]

- Phương pháp: Mở rộng phương pháp Encoder-Decoder bằng cách thêm cơ chế attention, giúp mô hình tập trung vào các phần quan trọng của ảnh khi sinh chú thích.

- Dữ liệu thực nghiệm: Sử dụng tập dữ liệu MS COCO hoặc các tập dữ liệu tương tự.

- Kết quả: Mô hình attention-based cải thiện khả năng sinh chú thích, đặc biệt là trong việc mô tả các ảnh phức tạp và đa đối tượng. Tuy nhiên, cơ chế attention có thể làm tăng độ phức tạp tính toán của mô hình.

**Phương pháp Transformer**

- Bài báo tham khảo: "Image Transformer" (Niki Parmar et al., 2018). [6]

- Phương pháp: Sử dụng kiến trúc transformer mạnh mẽ từ lĩnh vực dịch máy và áp dụng nó trực tiếp vào bài toán sinh chú thích cho ảnh.

- Dữ liệu thực nghiệm: Thường sử dụng các tập dữ liệu lớn như MS COCO hoặc Flickr30k.

- Kết quả: Mô hình transformer đạt được kết quả tốt và có khả năng học được các mối quan hệ phức tạp giữa các đối tượng trong ảnh. Tuy nhiên, yêu cầu tài nguyên tính toán lớn hơn so với các phương pháp truyền thống.

### Phương pháp đề xuất giải quyết bài toán

**Sử dụng CNN và LSTM trong mô hình sinh chú thích cho ảnh**

- Mạng CNN để mã hóa thông tin từ ảnh:

* CNN đã được chứng minh là một công cụ hiệu quả trong việc trích xuất đặc trưng từ ảnh nhờ khả năng phát hiện các mẫu không gian như cạnh, góc, và đối tượng.
* Ảnh được chuyển đổi thành vector đặc trưng, đại diện cho nội dung quan trọng của ảnh.

- Mạng LSTM để sinh chú thích dựa trên thông tin từ CNN:

* LSTM, một dạng mạng RNN, nổi bật với khả năng ghi nhớ và xử lý các chuỗi dài.
* LSTM nhận vector đặc trưng từ CNN và sinh chuỗi chú thích dựa trên ngữ cảnh đã tạo.

- Lý giải lựa chọn hướng giải quyết này:

* Sự kết hợp CNN và LSTM tạo ra một mô hình tận dụng thế mạnh của cả hai: CNN xử lý ảnh tốt, còn LSTM xử lý chuỗi văn bản hiệu quả.
* Hướng này đã chứng minh sự thành công trong các ứng dụng khác như xử lý ngôn ngữ tự nhiên, dịch máy, và thị giác máy tính.

**Sử dụng Transformer trong mô hình sinh chú thích cho ảnh**

- Mạng CNN để mã hóa thông tin từ ảnh: CNN tiếp tục được sử dụng để trích xuất vector đặc trưng từ ảnh, tương tự như phương pháp CNN + LSTM.

- Mô hình Transformer để sinh chú thích dựa trên thông tin từ CNN:

* Transformer, đặc biệt với cơ chế Attention, cho phép mô hình tập trung vào các phần quan trọng của ảnh, đồng thời ghi nhớ ngữ cảnh dài hạn.
* Thay vì sử dụng LSTM để xử lý chuỗi tuần tự, Transformer có khả năng xử lý song song, giúp tăng hiệu suất và khả năng mở rộng.

- Lý giải lựa chọn hướng giải quyết này:

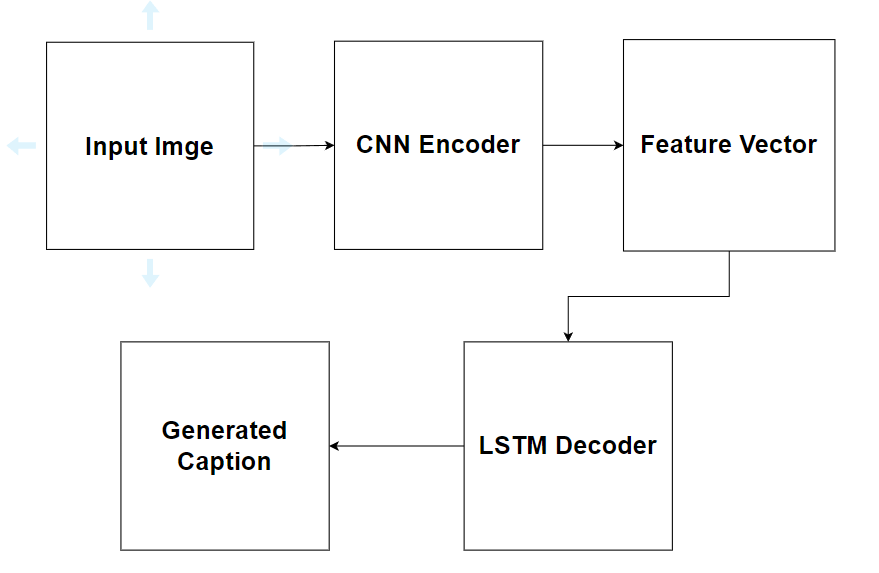
* Transformer vượt trội so với LSTM nhờ cơ chế Attention, giúp mô hình không chỉ ghi nhớ thông tin mà còn tập trung vào các đặc trưng quan trọng trong ảnh.
* Tính hiệu quả về thời gian huấn luyện của Transformer rất phù hợp với các bài toán lớn và phức tạp như sinh chú thích cho ảnh.
* Hướng tiếp cận này đã được chứng minh hiệu quả qua các mô hình nổi tiếng như Vision Transformer (ViT) và các hệ thống sinh văn bản GPT.

## 1.3 Phương pháp giải quyết bài toán

### 1.3.1 Mô hình tổng quát

**Mô hình CNN và LSTM sinh chú thích cho ảnh**

Mô hình tổng quát để giải quyết bài toán sinh chú thích cho ảnh bao gồm hai phần chính: bộ mã hóa CNN (CNN Encoder) để trích xuất đặc trưng từ ảnh và bộ giải mã LSTM (LSTM Decoder) để sinh chú thích từ thông tin đã mã hóa.

****

Hình 1.3.1.1: Mô hình tổng quát giải quyết bài toán sinh chú thích cho ảnh

- Input Image (Ảnh đầu vào): Ảnh đầu vào được cung cấp cho mô hình. Đây là các hình ảnh kỹ thuật số mà mô hình sẽ sinh chú thích cho chúng.

- CNN Encoder (Bộ mã hóa CNN): Bộ mã hóa CNN chịu trách nhiệm trích xuất các đặc trưng từ ảnh đầu vào và chuyển đổi chúng thành một vector đặc trưng.

- Feature Vector (Vector đặc trưng): Vector đặc trưng này đại diện cho toàn bộ nội dung của ảnh dưới dạng một vector số.

- LSTM Decoder (Bộ giải mã LSTM): Bộ giải mã LSTM sử dụng vector đặc trưng từ CNN để sinh ra chuỗi chú thích. Mô hình LSTM sẽ dự đoán từng từ một trong chuỗi dựa trên thông tin từ vector đặc trưng và các từ đã dự đoán trước đó.

- Generated Caption (Chú thích sinh ra): Chú thích sinh ra là chuỗi các từ mô tả nội dung của ảnh đầu vào

**Mô hình Transformer trong sinh chú thích cho ảnh**

Mô hình sinh chú thích cho ảnh sử dụng Transformer bao gồm hai thành phần chính: CNN Encoder (Bộ mã hóa CNN) để trích xuất đặc trưng từ ảnh đầu vào, chuyển đổi thành vector số để làm đầu vào cho mô hình. Transformer Decoder (Bộ giải mã Transformer) đểin chuỗi chú thích dựa trên đặc trưng ảnh và các từ dự đoán trước đó.

- Ảnh đầu vào: được đưa qua các bước tiền xử lý (chuẩn hóa kích thước, chuyển đổi định dạng).

- Bộ mã hóa CNN: Trích xuất đặc trưng từ ảnh đầu vào qua mạng EfficientNet hoặc ResNet, tạo vector đặc trưng.

- Vector đặc trưng: Vector số hóa biểu diễn nội dung ảnh được đưa vào bộ giải mã.

- Bộ giải mã Transformer: Tạo ra chuỗi chú thích thông qua kiến trúc Attention, tích hợp thông tin ảnh và dự đoán từ vựng tiếp theo.

- Chú thích sinh ra: Chuỗi các từ mô tả nội dung ảnh, tạo thành kết quả đầu ra.

### Đặc trưng của mô hình đề xuất

**Mô hình CNN và LSTM sinh chú thích cho ảnh**

- Input Image (Ảnh đầu vào):

* Mô tả: Đây là ảnh đầu vào mà mô hình sẽ sinh chú thích. Nó được đưa vào mô hình để trích xuất các đặc trưng.
* Đặc điểm: Các ảnh đầu vào thường được chuyển đổi thành một định dạng cố định và tiền xử lý trước khi đưa vào mạng CNN Encoder.

- CNN Encoder (Bộ mã hóa CNN):

* Mô tả: Bộ mã hóa CNN được sử dụng để trích xuất các đặc trưng từ ảnh đầu vào và biến đổi chúng thành một vector đặc trưng.
* Chi tiết:
  + Lớp đầu vào (Input Layer): Nhận ảnh đầu vào dưới dạng ma trận các giá trị pixel.
  + Lớp tích chập (Convolutional Layers): Áp dụng các bộ lọc tích chập để trích xuất các đặc trưng từ ảnh.
  + Lớp pooling (Pooling Layers): Giảm kích thước của các đặc trưng trích xuất để giảm chi phí tính toán và giữ lại thông tin quan trọng.
  + Lớp hoàn tất (Fully Connected Layers): Kết hợp các đặc trưng trích xuất thành một vector đặc trưng cuối cùng.

- LSTM Decoder (Bộ giải mã LSTM):

* Mô tả: Bộ giải mã LSTM nhận vector đặc trưng từ CNN Encoder và sử dụng nó để sinh ra chuỗi chú thích.
* Chi tiết:
  + Lớp đầu vào (Input Layer): Nhận vector đặc trưng từ CNN Encoder và từ đầu tiên của chuỗi chú thích.
  + Lớp nhúng (Embedding Layer): Chuyển đổi các từ trong chuỗi chú thích thành các vector nhúng có kích thước cố định.
  + Lớp LSTM (LSTM Layer): Xử lý chuỗi đầu vào và sinh ra một vector đặc trưng cho mỗi bước thời gian.
  + Lớp kết hợp (Add Layer): Kết hợp vector đặc trưng từ CNN Encoder và LSTM Decoder.
  + Lớp kích hoạt (Activation Layer): Áp dụng hàm kích hoạt softmax để sinh ra xác suất cho từng từ trong từ vựng.
  + Lớp đầu ra (Output Layer): Xuất ra từ có xác suất cao nhất là từ tiếp theo trong chuỗi chú thích.

- Feature Vector (Vector đặc trưng): Vector đặc trưng này là kết quả cuối cùng của CNN Encoder, đại diện cho toàn bộ nội dung của ảnh dưới dạng một vector số.

- Generated Caption (Chú thích sinh ra): Chú thích sinh ra là chuỗi các từ mô tả nội dung của ảnh đầu vào. Nó được sinh ra bởi LSTM Decoder dựa trên vector đặc trưng từ CNN Encoder và từ đầu tiên của chuỗi chú thích.

**Mô hình Transformer trong sinh chú thích cho ảnh**

- CNN Encoder:

* Lớp tích chập (Convolution): Trích xuất đặc trưng cục bộ từ ảnh.
* Lớp Reshape: Đưa vector đặc trưng về dạng phù hợp làm đầu vào cho mô hình.

- Transformer Decoder:

* Kiến trúc chính:
  + Positional Embedding: Thêm thông tin vị trí vào từ vựng.
  + Multi-Head Attention: Kết hợp thông tin từ vector ảnh và ngữ cảnh trước đó.
  + Feed Forward Layers: Biến đổi thông tin ngữ nghĩa để dự đoán từ tiếp theo.
* Chi tiết các lớp:
  + Lớp Attention đầu tiên: Tìm hiểu mối quan hệ giữa các từ đã sinh và vector ảnh.
  + Lớp Dense: Dự đoán xác suất từ tiếp theo.

## 1.4 Thực nghiệm

### Dữ liệu

Dữ liệu được sử dụng trong mô hình là bộ dữ liệu từ Unsplashgồm 10,000 ảnh của 4 chủ đề sport, animal, human\_activitve, nature mỗi ảnh có 5 chú thích, với chú thích được thực hiện thủ công và sinh tự động để cải thiện tính đa dạng.

**Mô tả:**

- Ảnh: Tập dữ liệu bao gồm khoảng 10,000 ảnh phân chia thành các tập huấn luyện và kiểm tra.

- Chú thích: Mỗi ảnh đi kèm với năm chú thích mô tả nội dung của ảnh đó. Chú thích được thu thập và được đảm bảo có mối quan hệ với nội dung của ảnh.

- Đa dạng: Tập dữ liệu bao gồm nhiều loại hình ảnh khác nhau về cảnh đẹp, con người, động vật, vật phẩm, thuộc 4 chủ để sport, animal, human\_activitve, nature.

- Phân phối: Dữ liệu được chia thành tập huấn luyện và tập kiểm tra để đảm bảo tính công bằng trong quá trình đánh giá hiệu suất của mô hình.

- Chất lượng: Dữ liệu được đánh giá là có chất lượng tốt, với các chú thích phản ánh đầy đủ và chính xác nội dung của ảnh. Tuy nhiên, cũng có một số chú thích có thể không hoàn hảo do sự khác biệt trong cách diễn đạt của người dùng.

### 1.4.2 Xử lý dữ liệu

Việc tiền xử lý dữ liệu là bước quan trọng để chuẩn bị dữ liệu đầu vào cho mô hình, việc này giúp cải thiện hiệu suất và chất lượng của mô hình bằng cách loại bỏ nhiễu, chuẩn hóa dữ liệu và chuẩn hóa văn bản, cũng như chuẩn hóa kích thước ảnh và xử lý ảnh nếu cần thiết. [8]

Tiền xử lý dữ liệu bao gồm các bước sau:

1. Loại bỏ nhiễu: Loại bỏ các ký tự đặc biệt không mong muốn từ chú thích và ảnh. Bỏ các ảnh có chú thích quá ngắn (<5 từ) hoặc quá dài (>25 từ)
2. Chuẩn hóa văn bản: Chuyển đổi tất cả các từ trong chú thích về chữ thường và loại bỏ các ký tự đặc biệt không cần thiết.
3. Thêm thẻ startseq và endseq, token hóa và chuyển đổi các chuỗi thành số (CNN+LSTM).
4. Tách từ (tokenization): Tách chuỗi văn bản thành các từ riêng lẻ hoặc token để mô hình có thể hiểu và xử lý dễ dàng hơn.
5. Chuẩn hóa kích thước ảnh: Đồng nhất kích thước và độ phân giải của ảnh để mô hình có thể học từ dữ liệu đồng nhất.

### 1.4.3 Công nghệ sử dụng

Ngôn ngữ lập trình: Python.

Hệ thống: GPU NVIDIA Tesla hỗ trợ TensorFlow 2.13.0 để tăng tốc huấn luyện

Thư viện sử dụng: Các thư viện phổ biến như TensorFlow, PyTorch, và Keras, Pandas, tqdm, numpy.

Công cụ: Các công cụ phát triển phần mềm như Jupyter Notebook, Google Colab, Kaggle và trình biên tập mã như Visual Studio Code**.**

### 1.4.4 Cách đánh giá

#### 1.4.4.1 Đại lượng BLEU

**- Tính tỷ lệ n-gram chung**

Đầu tiên, BLEU tính tỷ lệ của các từ hoặc n-gram (chuỗi gồm n từ liên tiếp) xuất hiện trong câu được sinh ra bởi mô hình và câu tham chiếu. Cụ thể, BLEU sử dụng các n-gram từ 1 đến m (với m là số nguyên dương, thường là 4) để tính toán. Điều này giúp đánh giá được sự chính xác của các từ hoặc cụm từ mà mô hình đã sinh ra. [9]

**- Tính toán độ chính xác và sự đa dạng**

BLEU tính toán tỷ lệ của các n-gram giống nhau trong câu sinh ra bởi mô hình và câu tham chiếu. Điều này cho biết mức độ chính xác của câu sinh ra. Tuy nhiên, BLEU cũng tính toán một yếu tố điều chỉnh để đảm bảo rằng mô hình không được lợi thế nếu chỉ sinh ra các câu ngắn hơn. Điều này thúc đẩy sự đa dạng trong các câu sinh ra.

**- Kết hợp các điểm n-gram và điều chỉnh**

Điểm BLEU cho mỗi câu được tính bằng cách kết hợp các điểm n-gram và một điều chỉnh để đảm bảo sự cân bằng giữa chính xác và đa dạng. Cụ thể, điểm BLEU được tính bằng cách sử dụng công thức:



**Trong đó:**

⦁ BP (Brevity Penalty) là yếu tố điều chỉnh cho độ ngắn của câu được sinh ra so với câu tham chiếu. Nó giảm tỷ lệ BLEU nếu câu sinh ra quá ngắn.

⦁ là trọng số của các n-gram, thường được thiết lập để tăng dần với n.

⦁ là tỷ lệ của các n-gram giống nhau giữa câu được sinh ra và câu tham chiếu.

⦁ Hàm mũ giúp tăng trọng số của các n-gram được đánh giá ở mức cao hơn.

## 1.5 Kết quả đạt được

**Phương pháp LSTM-CNN**

- Mô hình Encoder:

* Đầu vào hình ảnh: Shape (1920,)
* Tỉ lệ Dropout: 0.5
* Lớp Dense: Units = 256, Activation = 'relu'

- Các lớp đặc trưng của chuỗi:

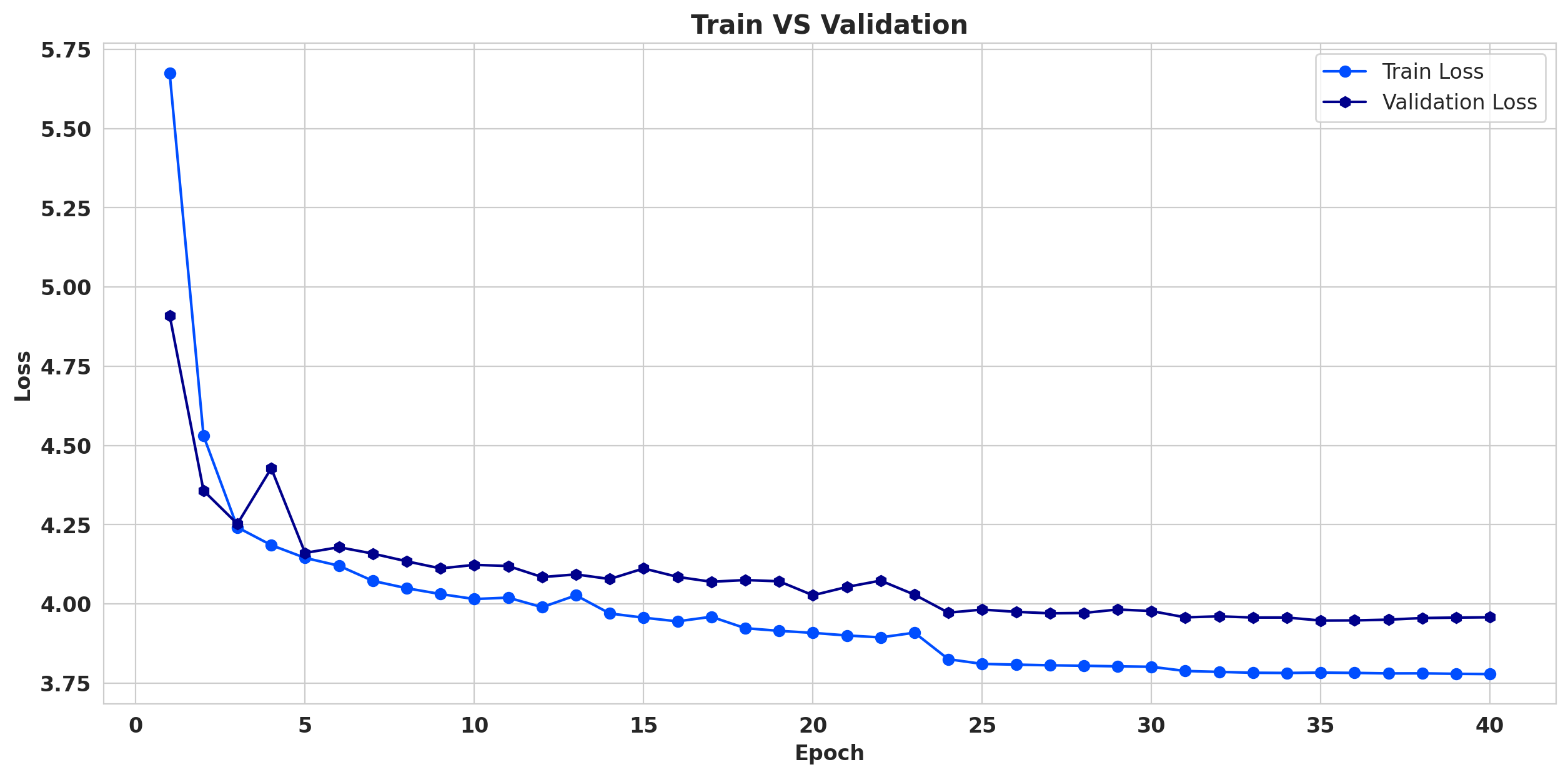
* Đầu vào chuỗi chú thích: Shape (max\_length,)
* Lớp Nhúng: Vocabulary size = 2664, Vector size = 256
* Tỉ lệ Dropout: 0.5
* Lớp LSTM: Units = 256

- Mô hình Decoder:

* Merge Layer: Kết hợp đầu ra của Encoder và LSTM
* Lớp Dense: Units = 128, Activation = 'relu'
* Lớp Dense cuối cùng: Units = Vocabulary size, Activation = 'softmax'

- Thông số Huấn luyện:

* Số epochs: 50
* Kích thước batch: 64
* Vocabulary size: 2664
* Max length: 45



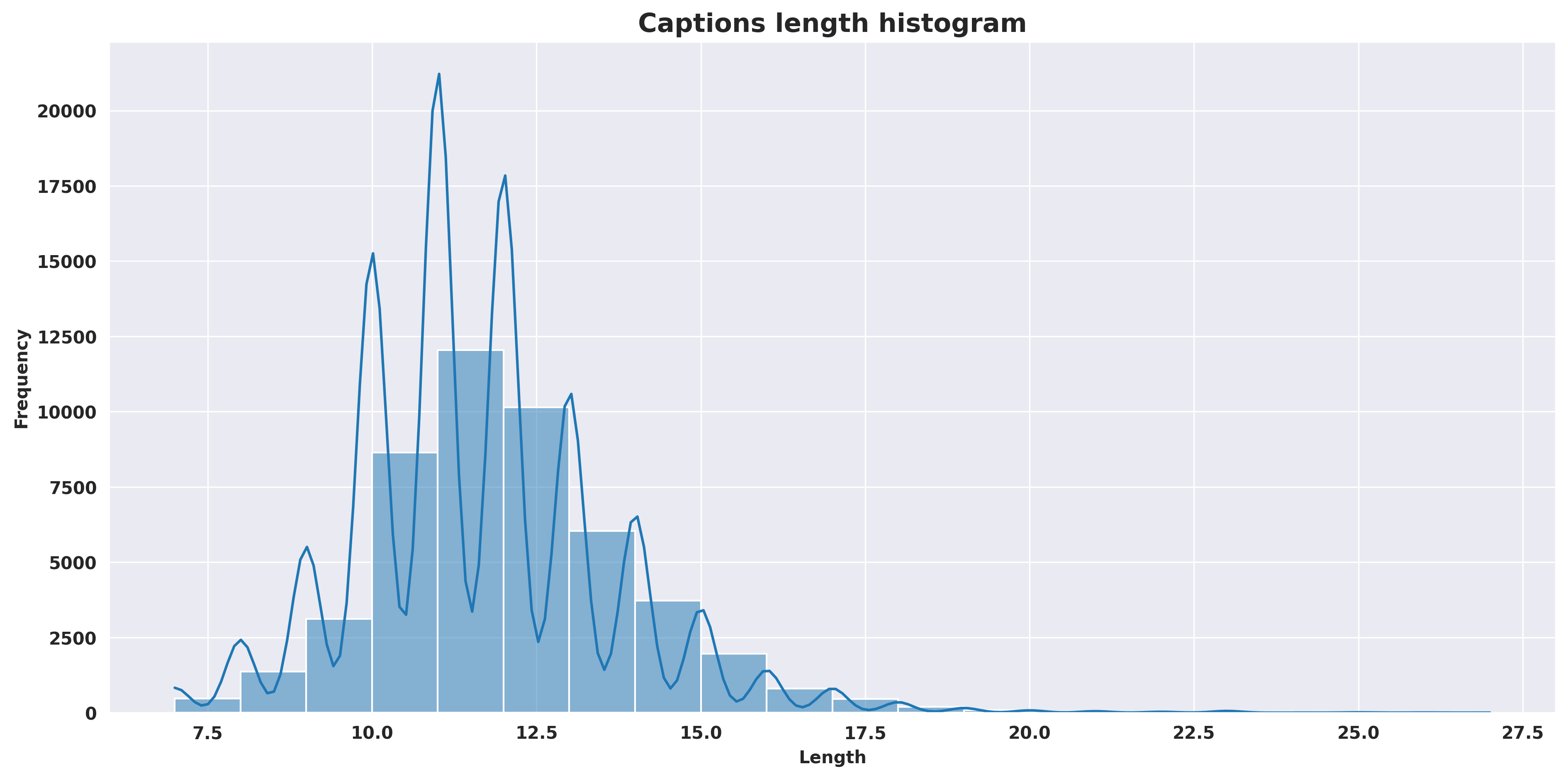
Hình 1.5.1: Sự mất mát trong quá trình huấn luyện và xác thực

- Kết quả đánh giá hiệu suất của mô hình trên tập dữ liệu kiểm tra được thể hiện qua các độ đo BLEU:

* BLEU-1: 0.44974
* BLEU-2: 0.39588
* BLEU-3: 0.49918
* BLEU-4: 0.57348

Kết quả cho thấy mô hình đạt được khả năng sinh chú thích cho ảnh với độ chính xác và tương đồng cao với các chú thích thực tế trên tập kiểm tra. Tuy nhiên, hiệu suất có thể được cải thiện thêm thông qua việc tăng cường dữ liệu hoặc sử dụng các kỹ thuật như Attention.

**Phương pháp Transformer**

****

Hình 1.5.2: Sự phân phối độ dài của các chú thích

- Encoder (CNN):

* Đầu vào CNN: Shape (None, 299, 299, 3).
* Đầu ra CNN: Shape (None, 100, 1280).

- Encoder Layer:

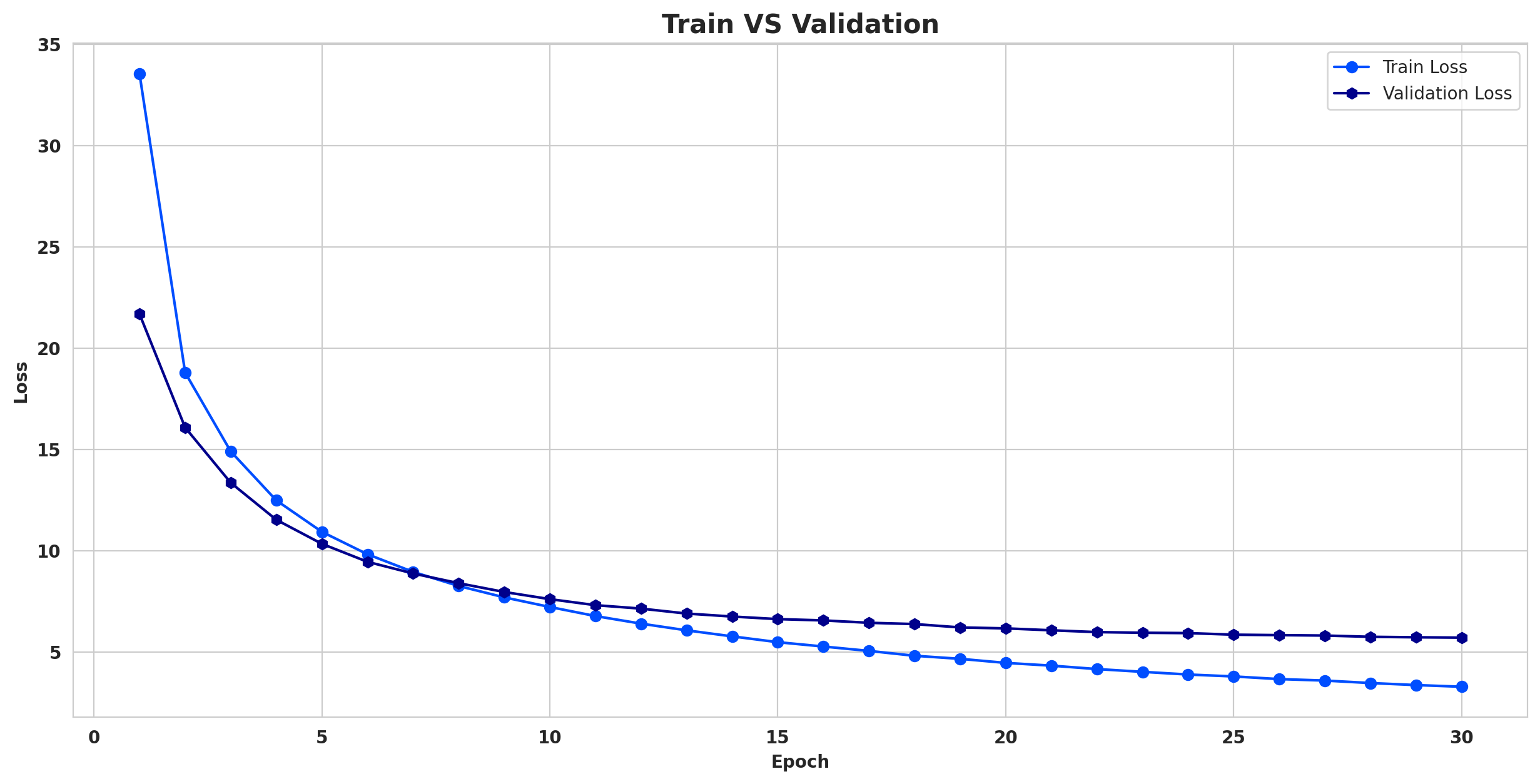
* Đầu vào: Shape (None, 100, 1280).
* Lớp Dense: Chuyển đổi đặc trưng ảnh từ CNN sang không gian nhúng 512 chiều: Shape (None, 100, 512).
* Đầu ra Encoder: Shape (None, 100, 512).

Decoder (Transformer):

* Đầu vào 1 (Caption): Chuỗi chú thích đầu vào có Shape (None, 24).
* Nhúng vị trí (Positional Embedding): Chuỗi đầu vào sau khi nhúng có Shape (None, 24, 512).
* Đầu vào 2 (Embedded Image Features): Đặc trưng ảnh nhúng từ Encoder với Shape (None, 100, 512).
* Lớp tương tác Multi-Head Cross-Attention: Kết hợp đặc trưng ảnh và ngữ cảnh chú thích, đầu ra Shape (None, 24, 512).
* Lớp Dense: Dự đoán các từ trong chú thích, đầu ra cuối cùng Shape (None, 24, 10000).

**Thông số Huấn luyện:**

* Kích thước batch: 512
* Số epochs: 30
* Optimizer: Adam với Learning Rate Scheduler để tự động điều chỉnh tốc độ học.



Hình 1.5.3: Sự mất mát trong quá trình huấn luyện và xác thực

Kết quả đánh giá của mô hình Transformer trên tập dữ liệu kiểm tra là như sau:

- Điểm BLEU: Các điểm BLEU đo lường sự chính xác của các n-gram giữa chú thích sinh ra và chú thích tham chiếu. Mô hình đạt được các chỉ số BLEU sau:

* BLEU-1: 0.34993
* BLEU-2: 0.58863
* BLEU-3: 0.70407
* BLEU-4: 0.76627

- Ý nghĩa: LEU-4 đạt điểm cao nhất, cho thấy mô hình sinh ra các chú thích dài có độ chính xác cao và ngữ nghĩa phù hợp với chú thích tham chiếu.

- Minh họa sự mất mát trong huấn luyện và xác thực: Đồ thị loss giữa tập huấn luyện và tập xác thực cho thấy quá trình huấn luyện hội tụ tốt, không xuất hiện overfitting.

- Nhận xét:

* Mô hình Transformer đã thể hiện khả năng sinh chú thích với độ chính xác và mạch lạc cao, với điểm BLEU-4 đạt 0.76627.
* Kết quả BLEU đủ để chứng minh rằng mô hình có thể tạo ra các chú thích hợp lý, tương đồng với các chú thích tham chiếu trên tập kiểm tra.

**Nguyên nhân khác biệt giữa các phương pháp:**

1. Cơ chế Attention với LSTM:

* Mô hình Transformer sử dụng cơ chế Attention, giúp mô hình tập trung vào các phần quan trọng của chuỗi đầu vào khi sinh ra từ, trong khi LSTM sử dụng trạng thái ẩn để nhớ thông tin. Điều này giúp Transformer có thể sinh ra các chuỗi từ chính xác hơn trong nhiều trường hợp.

1. Cấu trúc Mô hình:

* Transformer có nhiều lớp và mỗi lớp có các cơ chế Attention và feed-forward riêng biệt, giúp nó có khả năng xử lý thông tin một cách phức tạp hơn. Điều này có thể giải thích tại sao nó có điểm BLEU-4 cao hơn.

1. Thời gian và Dung lượng Huấn luyện:

* Số lượng epochs, kích thước batch, và các tham số huấn luyện khác cũng ảnh hưởng lớn đến hiệu suất của mô hình. Mô hình Transformer được huấn luyện với batch size lớn hơn và ít epochs hơn, điều này có thể ảnh hưởng đến khả năng học của nó so với LSTM-CNN.

## 1.6 Kết luận

**Kết quả đạt được của bài toán trên:**

* Xây dựng mô hình LSTM-CNN: Đã xây dựng thành công một mô hình kết hợp giữa mạng neural hồi quy dài (LSTM) và mạng neural tích chập (CNN) để sinh ra các chú thích cho ảnh. Mô hình này có khả năng trích xuất đặc trưng từ ảnh thông qua CNN và sử dụng thông tin này để sinh ra các chú thích phù hợp thông qua LSTM.
* Hiệu suất đánh giá: Mô hình LSTM-CNN đã đạt được một số kết quả khả quan trên tập dữ liệu kiểm tra. Điều này được đánh giá qua các độ đo như BLEU-4, với các điểm số đáng chú ý cho sự chính xác và đa dạng của các chú thích sinh ra.
* So sánh với mô hình Transformer: Cũng đã so sánh mô hình LSTM-CNN với một phương pháp tiên tiến khác là mô hình Transformer. Kết quả cho thấy mô hình Transformer vượt trội hơn về chất lượng chú thích, đặc biệt là trong việc sử dụng BLEU-4 để đo lường độ chính xác.

**Hạn chế của phương pháp giải quyết bài toán**

Mô hình LSTM-CNN:

* Phức tạp trong huấn luyện: Cần tinh chỉnh nhiều tham số và cấu trúc mạng, yêu cầu nhiều thời gian và tài nguyên.
* Giới hạn trong xử lý ngữ cảnh dài: LSTM có thể gặp khó khăn khi phải ghi nhớ thông tin trong khoảng thời gian dài, ảnh hưởng đến chất lượng chú thích đối với các câu dài và phức tạp.

Mô hình Transformer:

* Yêu cầu tài nguyên cao: Đòi hỏi nhiều tài nguyên tính toán hơn so với LSTM-CNN do số lượng lớn các tham số và các phép tính phức tạp liên quan đến cơ chế Attention.
* Phụ thuộc vào dữ liệu huấn luyện: Hiệu suất của Transformer phụ thuộc mạnh mẽ vào lượng và chất lượng của dữ liệu huấn luyện. Nếu dữ liệu không đủ phong phú hoặc đa dạng, hiệu suất của mô hình có thể bị giảm.

**Hướng phát triển trong tương lai của bài toán sinh chú thích cho ảnh có thể tập trung vào hai khía cạnh chính: tăng cường dữ liệu huấn luyện và cải tiến mô hình.**

Tăng cường dữ liệu huấn luyện: Một trong những cách tiếp cận hiệu quả để cải thiện hiệu suất của mô hình là tăng cường dữ liệu huấn luyện. Điều này có thể bao gồm việc sử dụng các tập dữ liệu lớn hơn và đa dạng hơn như Flickr30k hoặc MS COCO.

Cải tiến mô hình: Một phần quan trọng của hướng phát triển này là cải tiến mô hình để tạo ra các chú thích chính xác và phong phú hơn. Cải tiến mô hình có thể bao gồm việc sử dụng các kiến trúc mạng nơ-ron sâu phức tạp hơn như Transformer, các phương pháp tự chú ý (self-attention), hoặc các biến thể của mạng LSTM và CNN. Ngoài ra, việc kết hợp các phương pháp học sâu với các phương pháp xử lý ngôn ngữ tự nhiên và thị giác máy tính cũng có thể mang lại kết quả tốt hơn.

# LÀM VIỆC NHÓM

**Cách thức làm việc nhóm:**

- Xác định mục tiêu: Nhóm đã đặt ra mục tiêu cụ thể là hoàn thành đồ án, đáp ứng yêu cầu và tiêu chuẩn của giảng viên.

- Phân chia nhiệm vụ: Nhóm đã phân chia các nhiệm vụ và công việc cho từng thành viên dựa trên sở trường và kỹ năng của nhóm. Nhóm cũng xác định rõ vai trò và trách nhiệm của mỗi thành viên.

- Thiết lập kênh giao tiếp: Nhóm đã sử dụng các kênh giao tiếp như email, github, tin nhắn và cuộc họp trực tuyến để giữ liên lạc và cập nhật tiến độ công việc.

- Thống nhất tiêu chuẩn và định dạng: Nhóm đã thống nhất về định dạng và nội dung bài tập, bao gồm định dạng tài liệu, thông tin cần bao gồm và thời hạn nộp.

- Phản hồi và đề xuất cải tiến: Nhóm đã đưa ra ý kiến, phản hồi và đề xuất để cải thiện kết quả và đạt được sự đồng thuận giữa các thành viên.

- Kiểm tra và đánh giá: Sau khi hoàn thành bài tập, nhóm đã kiểm tra và đánh giá kết quả, rút ra kinh nghiệm và học hỏi từ các sai sót để cải thiện trong lần tiếp theo.

**Phân chia công việc của các thành viên trong nhóm**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Họ và Tên** | **MSSV** | **Công việc** |
| 1 | Tăng Quang  Nhật Nam | 21082941 | - Phân tích nghiệp vụ  - Xây dựng Backend và thiết kế Database, xử lý logic. - Xây dựng Frontend và giao diện.  - Tạo chatbot và tạo tập câu hỏi mẫu  - Triển khai hệ thống  - Thực hiện kiểm thử và xử lý lỗi - Viết báo cáo |
| 2 | Hoàng Thanh Tú | 21105251 | - Phân tích nghiệp vụ  - Xây dựng Backend và thiết kế Database, xử lý logic. - Xây dựng Frontend và giao diện.  - Phối hợp xây dựng chatbot - Quản lý mã nguồn  - Thực hiện kiểm thử và xử lý lỗi - Viết báo cáo |

Tổng số lần gặp nhau (tính theo buổi): 12 buổi

Tổng thời gian gặp nhau (tính theo giờ): 50 giờ

# TÀI LIỆU THAM KHẢO

|  |  |  |
| --- | --- | --- |
| [1] |  | Yiyu Wang, Jungang Xu, Yingfei Sun, End-to-End Transformer Based Model for Image Captioning, 2022. |
| [2] |  | Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, Attention Is All You Need, 2023. |
| [3] |  | Yue Ming, Nannan Hu, Chunxiao Fan, Fan Feng, Jiangwan Zhou, Hui Yu, Visuals to Text: A Comprehensive Review on Automatic Image Captioning, IEEE, 2022. |
| [4] |  | Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan, Show and tell: A neural image caption generator, IEEE, 2015. |
| [5] |  | Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio, Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, 2015. |
| [6] |  | Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Łukasz Kaiser, Noam Shazeer, Alexander Ku, Dustin Tran, Image Transformer, 2018. |
| [7] |  | Phan Thanh Sơn, Dương Tử Cường, Trích chọn các tham số đặc trưng tiếng nói cho hệ thống tổng hợp tiếng Việt dựa vào mô hình Markov ẩn, Tạp chí Tin học và Điều khiển học, T.29, S.1, 55-65, 2013. |
| [8] |  | Trần Hùng Cường, Trần Thanh Hùng, Giáo trình khai phá dữ liệu, NXB Thống kê , 2017. |
| [9] |  | Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu, Bleu: a Method for Automatic Evaluation of Machine Translation, 2002. |
| [10] |  | C.-Y. Lin, ROUGE: A Package for Automatic Evaluation of Summaries, 2004, p. 74–81. |

# PHỤ LỤC

Phần này bao gồm những nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung đồ án như số liệu, biểu mẫu, tranh ảnh, … nếu sử dụng những câu trả lời cho một *bảng câu hỏi thì bảng câu hỏi mẫu này phải được đưa vào phần Phụ lục ở dạng nguyên bản* đã dùng để điều tra, thăm dò ý kiến; **không được tóm tắt hoặc sửa đổi**. Các tính toán mẫu trình bày tóm tắt trong các biểu mẫu cũng cần nêu trong Phụ lục của luận văn. Phụ lục không được dày hơn phần chính của đồ án

**TỰ ĐÁNH GIÁ (Bài nhóm)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Nội dung | Điểm chuẩn | Tự chấm | Ghi chú |
| 1 | **Phân tích, Thiết kế** | 4đ |  |  |
| 2 | **Hiện thực** | 4đ |  |  |
| 3 | **Kết luận** | 0.5đ |  |  |
| 4 | **Báo cáo (**chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng**)** | 1đ |  |  |
| 5 | **Điểm nhóm** (chú ý trả lời các câu hỏi trong mục làm việc nhóm) | 0.5đ |  |  |
| **Tổng điểm** | | |  |  |

**TỰ ĐÁNH GIÁ (Bài cá nhân)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Nội dung | Điểm chuẩn | Tự chấm | Ghi chú |
| 1 | **Phân tích, Thiết kế** | 4đ |  |  |
| 2 | **Hiện thực** | 4.5đ |  |  |
| 3 | **Kết luận** | 0.5đ |  |  |
| 4 | **Báo cáo (**chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng**)** | 1đ |  |  |
| **Tổng điểm** | | |  |  |