TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TRẦN VĂN THẢO - 52000805**

**LÊ ĐÀO TẤN THÀNH - 52000804**

**CHÚ THÍCH HÌNH ẢNH CHÓ VÀ MÈO**

**BÁO CÁO CUỐI KÌ MÔN NHẬP MÔN THỊ GIÁC MÁY TÍNH**

**TÍNH TOÁN THÔNG MINH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TRẦN VĂN THẢO - 52000805**

**LÊ ĐÀO TẤN THÀNH - 52000804**

**CHÚ THÍCH HÌNH ẢNH CHÓ VÀ MÈO**

**BÁO CÁO CUỐI KÌ MÔN NHẬP MÔN THỊ GIÁC MÁY TÍNH**

**TÍNH TOÁN THÔNG MINH**

Người hướng dẫn

**TS. Phạm Văn Huy**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**LỜI CẢM ƠN**

Chúng em xin chân thành cảm ơn thầy Phạm Văn Huy đã dạy em những kiến thức về thị giác máy tính qua đó giúp nhóm em hoàn thành bài báo cáo này.

*TP. Hồ Chí Minh, ngày 06 tháng 03 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Thảo*

*Trần Văn Thảo*

*Thành*

*Lê Đào Tấn Thành*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của TS. Phạm Văn Huy. 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 Dự 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 Dự án của mình**. Trường Đại học Tôn Đức Thắng 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 06 tháng 03 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Thảo*

*Trần Văn Thảo*

*Thành*

*Lê Đào Tấn Thành*

**CHÚ THÍCH HÌNH ẢNH CHÓ VÀ MÈO**

**TÓM TẮT**

Báo cáo tập trung vào mô tả hình ảnh cho chó và mèo, sử dụng mô hình CNN và LSTM. Nói về lý thuyết, xây dựng mô hình, và ứng dụng thực tế trong xây dựng trang web thú cưng. Báo cáo đánh giá hiệu suất mô hình và đề xuất hướng phát triển tương lai. Kết hợp giữa lý thuyết và ứng dụng, mang lại giá trị trong lĩnh vực trí tuệ nhân tạo và xử lý hình ảnh.

**CHÚ THÍCH HÌNH ẢNH CHÓ VÀ MÈO**

**ABSTRACT**

Bài báo cáo này tập trung vào việc giải quyết vấn đề mô tả hình ảnh cho chó và mèo, với sự áp dụng của lý thuyết và phương pháp xây dựng mô hình hiệu quả. Bước đầu tiên của bài báo cáo là giới thiệu về lý do chọn đề tài, với nhấn mạnh vào sự quan trọng của mô tả hình ảnh trong bối cảnh dữ liệu ngày càng lớn và ngày càng quan trọng của trí tuệ nhân tạo. Phần lớn của báo cáo tập trung vào lý thuyết và kỹ thuật xây dựng mô hình. Giải thích chi tiết về sự lựa chọn kiến trúc mạng nơ-ron, đặc biệt là việc sử dụng mô hình CNN và cách áp dụng LSTM để giải quyết các thách thức trong xử lý chuỗi dữ liệu hình ảnh. Bài báo cáo cũng bao gồm các phương pháp tiền xử lý dữ liệu và tối ưu hóa mô hình để đảm bảo hiệu suất tốt nhất. Ứng dụng thực tế của mô hình được mô tả chi tiết trong báo cáo, với những ví dụ cụ thể về cách nó đã giúp tự động mô tả hình ảnh chó và mèo một cách chính xác và tự nhiên. Bài báo cáo cũng bao gồm những ứng dụng tiềm năng trong các lĩnh vực khác như giáo dục, quảng cáo thú cưng và trải nghiệm tương tác động vật ảo. Cuối cùng, bài báo cáo đánh giá sự hiệu quả của mô hình và đề xuất những hướng phát triển tiếp theo. Tổng cộng, nó là một báo cáo toàn diện với sự kết hợp giữa lý thuyết, kỹ thuật và ứng dụng thực tế, mang lại giá trị lớn trong lĩnh vực trí tuệ nhân tạo và xử lý hình ảnh.

**MỤC LỤC**

[DANH MỤC HÌNH VẼ viii](#_Toc926112545)

[DANH MỤC BẢNG BIỂU ix](#_Toc1340329224)

[DANH MỤC CÁC CHỮ VIẾT TẮT x](#_Toc494695688)

[CHƯƠNG 1. MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI 1](#_Toc1538518104)

[1.1 Lý do chọn đề tài 1](#_Toc597052330)

[1.2 Mục tiêu thực hiện đề tài 1](#_Toc2135478106)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 2](#_Toc1853102094)

[2.1 Mạng nơ-ron tích chập (CNN) 2](#_Toc491690482)

[2.1.1 Tổng quan về mạng nơ-ron tích chập 2](#_Toc554864313)

[2.1.2 Xception 2](#_Toc353975324)

[2.2 Mạng nơ-ron hồi quy 2](#_Toc534443594)

[2.2.1 Tổng quan về mạng nơ-ron hồi quy 2](#_Toc1932730042)

[2.2.2 Long Short-term Memory 2](#_Toc859923220)

[CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT 3](#_Toc1409724819)

[3.1 Xây dụng mô hình kết hợp CNN - LSTM 3](#_Toc1688862910)

[3.1.1 Tổng quan 3](#_Toc770758157)

[3.1.2 Kiến trúc 3](#_Toc474934758)

[CHƯƠNG 4. THỰC NGHIỆM 3](#_Toc1783826849)

[4.1 Dữ liệu thực nghiệm 3](#_Toc462762226)

[4.2 Cài đặt thực nghiệm 3](#_Toc1078278155)

[4.2.1 Cài đặt mô hình cho đề tài 4](#_Toc1506467745)

[4.2.2 Triển khai ứng dựng thực tế 4](#_Toc1606535447)

[CHƯƠNG 5. KẾT LUẬN 4](#_Toc1331722999)

[5.1 Kết luận 4](#_Toc657776341)

[5.2 Hướng phát triển 4](#_Toc2016197081)

[TÀI LIỆU THAM KHẢO 6](#_Toc70613995)

# DANH MỤC HÌNH VẼ

# DANH MỤC BẢNG BIỂU

# DANH MỤC CÁC CHỮ VIẾT TẮT

|  |  |
| --- | --- |
| BERT | Bidirectional Encoder Representations from Transformers |
| GEC | Grammatical Error Correction |
| MLM | Masked Language Model |
| NLP | Natural Language Processing |
| NSP | Next Sentence Prediction |

# MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI

## Lý do chọn đề tài

Đề tài "Mô tả hình ảnh cho chó và mèo" được lựa chọn với những lý do đáng chú ý. Đầu tiên, với sự bùng nổ của dữ liệu hình ảnh và video trên Internet, khả năng tự động mô tả hình ảnh đặc biệt quan trọng để hiểu và tận dụng nguồn thông tin lớn này. Mô hình mô tả hình ảnh cho chó và mèo không chỉ hỗ trợ trong việc tự động hóa quy trình này mà còn đưa ra những thách thức mới đối với lĩnh vực trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên.

## Mục tiêu thực hiện đề tài

Mục tiêu của nghiên cứu là xây dựng một mô hình có khả năng mô tả chính xác và tự nhiên cho hình ảnh chó và mèo. Ngoài ra, việc tìm hiểu về cách mô hình có thể tự động học được đặc điểm quan trọng của chó và mèo từ hình ảnh giúp mở rộng kiến thức về học máy và trí tuệ nhân tạo. Điều này không chỉ mang lại giá trị nghiên cứu mà còn có ứng dụng rộng rãi trong nhiều lĩnh vực như giáo dục, quảng cáo thú cưng, và tạo ra các ứng dụng di động hoặc trang web hữu ích cho người dùng. Đề tài còn mở ra cơ hội mở rộng sang việc mô tả hình ảnh cho nhiều loài động vật khác, đặt ra những thách thức mới và đưa ra những đóng góp quan trọng cho cộng đồng nghiên cứu và ứng dụng công nghệ.

# CƠ SỞ LÝ THUYẾT

## Mạng nơ-ron tích chập (CNN)

### Tổng quan về mạng nơ-ron tích chập

#### S

### Xception

#### S

## Mạng nơ-ron hồi quy

### Tổng quan về mạng nơ-ron hồi quy

#### S

### Long Short-term Memory

#### Sc

# MÔ HÌNH ĐỀ XUẤT

## Xây dụng mô hình kết hợp CNN - LSTM

### Tổng quan

### Kiến trúc

# THỰC NGHIỆM

## Dữ liệu thực nghiệm

## Cài đặt thực nghiệm

### Cài đặt mô hình cho đề tài

### Triển khai ứng dựng thực tế

# KẾT LUẬN

## Kết luận

Trong đề tài "Mô tả hình ảnh cho chó và mèo", chúng em đã đặt ra mục tiêu xây dựng một mô hình CNN - LSTM có khả năng mô tả chính xác và tự nhiên cho hình ảnh chó và mèo. Nghiên cứu này có ý nghĩa lớn trong bối cảnh tăng cường khả năng hiểu và tận dụng thông tin từ dữ liệu hình ảnh ngày càng lớn trên Internet. Qua quá trình nghiên cứu, chúng em đã áp dụng các kỹ thuật xử lý ngôn ngữ tự nhiên và học máy để đạt được mục tiêu đề ra. Kết quả của đề tài không chỉ giúp tự động hóa quy trình mô tả hình ảnh mà còn mang lại những đóng góp quan trọng cho lĩnh vực trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên.

## Hướng phát triển

Để mở rộng và cải thiện đề tài, có một số hướng phát triển có thể được xem xét:

* Mở rộng Loại Động vật: Nghiên cứu có thể được mở rộng để mô tả hình ảnh cho nhiều loài động vật khác nhau, tạo ra một hệ thống đa dạng và đa loài.
* Tích hợp Nguyên tắc Học Sâu: Áp dụng các nguyên tắc học sâu để cải thiện khả năng tự động học và hiểu sâu sắc hơn về đặc trưng của chó và mèo từ dữ liệu hình ảnh.
* Phân loại Tình cảm Động vật: Nghiên cứu mở rộng sang việc nhận diện tình cảm của động vật từ hình ảnh, tạo ra khả năng mô tả thêm về trạng thái tinh thần của chúng.

Những hướng phát triển này có thể giúp đưa đề tài vào một chiều sâu và ứng dụng rộng rãi hơn trong cả lĩnh vực nghiên cứu và thực tế.

# TÀI LIỆU THAM KHẢO

Tiếng Anh

https://www.analyticsvidhya.com/blog/2021/12/step-by-step-guide-to-build-image-caption-generator-using-deep-learning/

https://www.researchgate.net/profile/Jameer-Kotwal/publication/377852268\_IMAGE\_CAPTION\_GENERATOR\_USING\_CONVOLUTIONAL\_NEURAL\_NETWORKS\_AND\_LONG\_SHORT-TERM\_MEMORY/links/65bb6cdb1e1ec12eff684fd8/IMAGE-CAPTION-GENERATOR-USING-CONVOLUTIONAL-NEURAL-NETWORKS-AND-LONG-SHORT-TERM-MEMORY.pdf

https://www.ijitee.org/wp-content/uploads/papers/v10i3/C83830110321.pdf