**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**

**======\*\*\*======**

**ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC**

**NGÀNH CÔNG NGHỆ THÔNG TIN**

**NGHIÊN CỨU VÀ XÂY DỰNG HỆ THỐNG MÔ TẢ ĐƯỜNG ĐI HỖ TRỢ NGƯỜI KHIẾM THỊ VÀ NGƯỜI CÓ THỊ LỰC KÉM DI CHUYỂN**

|  |  |
| --- | --- |
| **Cán bộ hướng dẫn** | **: ThS. Nguyễn Lan Anh** |
| **Sinh viên thực hiện** | **: Đặng Phan Anh** |
| **Mã số sinh viên** | **: 2021604178** |

**Hà Nội – Năm 2025**

**LỜI CẢM ƠN**

Trong bối cảnh cuộc cách mạng công nghệ thông tin đang diễn ra mạnh mẽ, việc ứng dụng công nghệ vào đời sống, đặc biệt là hỗ trợ người khuyết tật, trở nên cần thiết hơn bao giờ hết. Đề tài tốt nghiệp “Nghiên cứu và xây dựng hệ thống mô tả đường đi cho người bị khiếm thị và người có thị lực kém” là một bước tiến quan trọng, hướng đến việc tối ưu hóa khả năng di chuyển và tạo ra môi trường sống an toàn, thuận tiện cho người khiếm thị.

Chúng ta đều nhận thấy rằng việc hỗ trợ người khiếm thị không chỉ đòi hỏi kiến thức chuyên môn về công nghệ mà còn phải kết hợp với việc hiểu rõ nhu cầu và thách thức mà họ đối mặt hàng ngày. Hệ thống mô tả đường đi sẽ giúp chúng ta nắm bắt được những thông tin quan trọng, từ đó đưa ra các biện pháp phù hợp để tạo ra môi trường sống tích cực và an toàn cho người khiếm thị.

Em xin chân thành cảm ơn cô Nguyễn Lan Anh đã hướng dẫn em thực hiện đề tài này. Báo cáo đồ án tốt nghiệp đại học này đã giúp em củng cố và rèn luyện được rất nhiều kỹ năng trong việc xây dựng ứng dụng xử lý hình ảnh, ứng dụng trí tuệ nhân tạo và học sâu trong thực tiễn. Đây chắc chắn sẽ là hành trang rất hữu ích trong sự nghiệp của em trong tương lai.

Tuy nhiên, do thời gian và tài nguyên có hạn, hệ thống hiện tại vẫn còn nhiều thiếu sót như chỉ hỗ trợ một số tình huống cụ thể, chưa thể hoạt động hiệu quả trong mọi môi trường thực tế, và chưa có giao diện thân thiện để người dùng dễ dàng sử dụng. Vì vậy, em rất mong được các thầy cô đóng góp thêm ý kiến để em có thể cải thiện hệ thống này, tiến tới ứng dụng thực tiễn trong tương lai.

***Em xin chân thành cảm ơn!***

# DANH MỤC CÁC CHỮ VIẾT TẮT

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Viết tắt** | **Thuật ngữ** | **Ý nghĩa** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# DANH MỤC HÌNH ẢNH, HÌNH VẼ

# DANH MỤC BẢNG BIỂU

# MỞ ĐẦU

## 1. Lí do chọn đề tài

Công nghệ thông tin ngày nay phát triển vượt bậc và đóng vai trò thiết yếu trong đời sống hiện đại. Con người đã tạo ra những hệ thống thông minh có khả năng tự thu thập, xử lý và phân tích dữ liệu từ nhiều nguồn thiết bị ghi hình—từ camera điện thoại, máy tính bảng, máy ảnh chuyên dụng đến hệ thống giám sát CCTV—với độ chính xác ngày càng cao. Đặc biệt, những tiến bộ trong lĩnh vực trí tuệ nhân tạo, học máy và học sâu đã mở ra khả năng nhận dạng vật thể, nhận dạng chữ viết, định vị không gian và nhận dạng giọng nói, chuyển đổi văn bản thành giọng nói… tạo nền tảng cho các ứng dụng hỗ trợ người dùng yếu thị lực.

Với mục tiêu nâng cao chất lượng cuộc sống cho người khiếm thị và người có thị lực kém, đề tài “Nghiên cứu và xây dựng hệ thống mô tả đường đi cho người bị khiếm thị và người có thị lực kém” được lựa chọn. Hệ thống này sẽ tận dụng công nghệ xử lý ảnh và AI để xác định vật cản và đặc điểm lộ trình, sau đó chuyển đổi thành thông tin mô tả bằng âm thanh hoặc ký hiệu dễ hiểu. Kết quả kỳ vọng mang lại giải pháp di chuyển an toàn, độc lập và dễ tiếp cận, đóng góp tích cực cho cộng đồng người khuyết tật thị giác.

## 2. Mục tiêu của đề tài

Nghiên cứu, phân tích, thiết kế và cài đặt nhằm mục tiêu xây dựng thử nghiệm hệ thống mô tả đường đi cho người bị khiếm thị và người có thị lực kém dựa trên các đặc trưng hình ảnh. Hệ thống sẽ thu thập thông tin về khung cảnh, vật cản, địa hình, sau đó chuyển đổi thành hướng dẫn giọng nói. Từ đó, hỗ trợ người dùng bằng cách cung cấp mô tả về lộ trình và gợi ý người dùng cách di chuyển, giúp họ di chuyển an toàn, tự tin và độc lập, góp phần nâng cao chất lượng đời sống và khả năng hòa nhập cộng đồng.

## 3. Nội dung nghiên cứu

Khi nghiên cứu và xây dựng hệ thống mô tả đường đi hỗ trợ người khiếm thị và người có thị lực kém di chuyển, có thể áp dụng nhiều phương pháp và kỹ thuật cho bài toán phát hiện và phân tích đối tượng. Dưới đây là một số phương pháp phổ biến:

* Hệ thống định vị có giám sát: Trong phương pháp này, hệ thống được huấn luyện với dữ liệu đã được gán nhãn, chẳng hạn như các tuyến đường an toàn, các chướng ngại vật cần tránh. Mô hình học từ các mẫu dữ liệu này để cung cấp hướng dẫn chính xác cho người dùng. Việc này đòi hỏi một tập dữ liệu lớn và đa dạng để đảm bảo tính chính xác và hiệu quả của hệ thống.
* Hệ thống định vị bán giám sát: Ở đây, chỉ một phần dữ liệu được gán nhãn, phần còn lại không có nhãn. Hệ thống học từ dữ liệu có nhãn và cố gắng tự động học từ dữ liệu không có nhãn để cải thiện khả năng định vị và hướng dẫn. Phương pháp này hữu ích khi việc gán nhãn dữ liệu là tốn kém hoặc khó khăn.
* Hệ thống định vị không giám sát: Phương pháp này không sử dụng dữ liệu gán nhãn. Hệ thống tự học từ dữ liệu thu thập được, phát hiện các mẫu hoặc nhóm thông tin hữu ích cho việc định vị và hướng dẫn người dùng. Điều này đặc biệt hữu ích khi khám phá và hiểu rõ hơn về các môi trường mới hoặc chưa được biết đến.

Trên thực tế, việc hỗ trợ người khiếm thị di chuyển đã thu hút sự quan tâm của nhiều nhà nghiên cứu trên thế giới. Một số hệ thống đã được phát triển, chẳng hạn như hệ thống định vị thông minh kết hợp nhận dạng vật thể và phản hồi âm thanh để hướng dẫn người dùng. Ngoài ra, các thiết bị như gậy thông minh sử dụng cảm biến siêu âm để phát hiện chướng ngại vật cũng đã được triển khai.

Việc kết hợp các phương pháp và kỹ thuật như trí tuệ nhân tạo và học máy trong việc xây dựng hệ thống mô tả đường đi không chỉ giúp người khiếm thị di chuyển an toàn và độc lập hơn mà còn góp phần nâng cao chất lượng cuộc sống và hòa nhập cộng đồng cho họ.

## 4. Phạm vi đề tài

Phương pháp nghiên cứu tài liệu:

* Thu thập dữ liệu và đánh giá các công trình khoa học trong và ngoài nước: Tìm kiếm, tổng hợp và đánh giá các nghiên cứu đã công bố liên quan đến hệ thống hỗ trợ di chuyển cho người khiếm thị, bao gồm các công nghệ như trí tuệ nhân tạo, học máy, cảm biến và thiết bị hỗ trợ.​
* Phân tích và tổng hợp các nghiên cứu có trước: So sánh, đối chiếu các phương pháp, kỹ thuật và kết quả từ các nghiên cứu trước để xác định những điểm mạnh, điểm yếu và khoảng trống cần nghiên cứu thêm.​
* Phân loại và hệ thống các công bố theo từng nội dung nghiên cứu của đề tài: Tổ chức các tài liệu thu thập được thành các nhóm chủ đề như: công nghệ định vị, mô tả đường đi, giao diện người dùng, và các yếu tố ảnh hưởng đến hiệu quả của hệ thống hỗ trợ.

Phương pháp thực nghiệm:

* Tiến hành thực nghiệm để đánh giá hiệu năng của giải thuật đề xuất: Xây dựng các kịch bản thử nghiệm mô phỏng các tình huống di chuyển thực tế của người khiếm thị, đo lường các chỉ số hiệu suất như độ chính xác của mô tả đường đi, thời gian phản hồi.​

## 5. Kết quả dự kiến đạt được

Đề tài hướng tới việc phát triển một hệ thống hoàn chỉnh có khả năng hỗ trợ người khiếm thị và người có thị lực kém di chuyển an toàn và hiệu quả trong môi trường thực tế. Cụ thể, các kết quả dự kiến bao gồm:

* Xây dựng hệ thống mô tả đường đi thông minh: Phát triển một hệ thống sử dụng các thuật toán học sâu tiên tiến để nhận diện và mô tả các yếu tố trong môi trường xung quanh, như chướng ngại vật, lối đi, biển báo, và các điểm định hướng quan trọng.​
* Giao diện người dùng thân thiện: Thiết kế giao diện người dùng dễ sử dụng, cho phép người khiếm thị và người có thị lực kém dễ dàng tương tác với hệ thống thông qua các phương thức phù hợp như âm thanh, rung, hoặc giao diện cảm ứng đơn giản.​
* Khả năng hoạt động ổn định trong môi trường thực tế: Đảm bảo hệ thống hoạt động hiệu quả và ổn định trong các điều kiện môi trường khác nhau, từ trong nhà đến ngoài trời, và có thể thích ứng với các tình huống di chuyển đa dạng.​
* Áp dụng công nghệ: Ứng dụng công nghệ chuyển đổi văn bản thành giọng nói Text-to-speech để đảm bảo hệ thống có thể cung cấp các thông tin hữu ích và gợi ý chỉ dẫn hợp lý cho người khiếm thị và người có thị lực kém.​
* Tiềm năng mở rộng và tích hợp: Hệ thống có khả năng mở rộng để tích hợp thêm các chức năng mới, như nhận diện các hành vi nguy hiểm hoặc tích hợp với các hệ thống hỗ trợ khác, nhằm nâng cao chất lượng cuộc sống cho người khiếm thị và người có thị lực kém.​

Việc phát triển hệ thống này không chỉ giúp người khiếm thị và người có thị lực kém di chuyển an toàn và độc lập hơn mà còn góp phần thúc đẩy sự hòa nhập xã hội và nâng cao chất lượng cuộc sống cho họ.

## 6. Bố cục đề tài

Nội dung chính của đề tài gồm có 3 chương:

**Chương 1:** Tổng quan về Deep Learning và các mô hình Convolutional Neural Network, Recurrent neural networks. Nội dung giới thiệu những ứng dụng đột phá của công nghệ này trong cuộc sống hàng ngày, từ nhận dạng khuôn mặt trên điện thoại, chuyển đổi giọng nói thành văn bản, sinh mô tả cho ảnh cho đến chuyển đổi ngôn ngữ ký hiệu thành văn bản.

**Chương 2:** Phương pháp mô tả và gợi ý đường đi. Trình bày các phương pháp nhận diện và sinh mô tả cho đường đi, phương pháp sinh gợi ý cho đường đi.

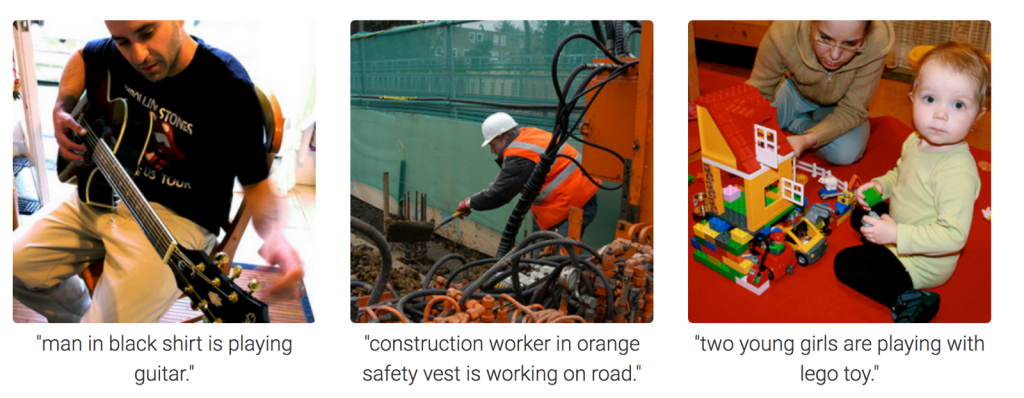
**Chương 3:** Xây dựng chương trình thực nghiệm và đánh giá hệ thống mô tả đường đi và gợi ý chỉ dẫn cho người khiếm thị và người có thị lực kém. Đánh giá hiệu quả của hệ thống trong việc sinh mô tả cho đường đi và gợi ý chỉ dẫn.

# CHƯƠNG 1: TỔNG QUAN VỀ DEEP LEARNING VÀ MÔ HÌNH CNN

## Deep Learning là gì?

​ Deep learning (học sâu) là một nhánh của học máy (machine learning), sử dụng các mạng nơ-ron nhân tạo nhiều lớp để mô phỏng cách bộ não con người xử lý dữ liệu. Nhờ khả năng tự học từ dữ liệu lớn, deep learning được ứng dụng rộng rãi trong các lĩnh vực như nhận diện hình ảnh, xử lý ngôn ngữ tự nhiên và điều khiển tự động.​

Một trong những thành tựu lớn nhất và tiêu biểu của việc áp dụng các mô hình học sâu như CNN và RNN với các mô hình học sâu khác là trong lĩnh vực nhận dạng hình ảnh và mô tả nội dung trực quan. Sự kết hợp giữa mạng nơ-ron tích chập (CNN) và mạng nơ-ron hồi tiếp (RNN) đã tạo nên các hệ thống có khả năng tự động phân tích và mô tả nội dung hình ảnh hoặc video bằng ngôn ngữ tự nhiên. Điều này mở ra nhiều ứng dụng thực tiễn như hỗ trợ người khiếm thị tiếp cận thông tin hình ảnh, cải thiện tìm kiếm hình ảnh theo ngữ nghĩa, và nâng cao hiệu quả trong các hệ thống giám sát an ninh.

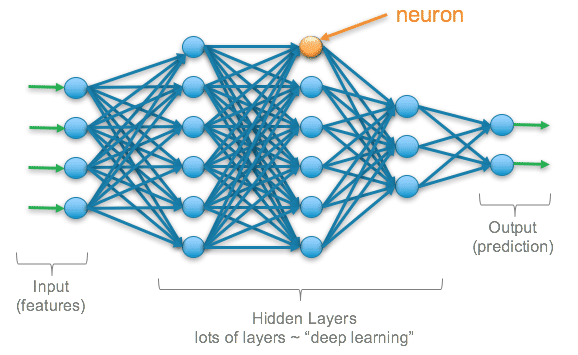


## Deep Learning hoạt động như thế nào?

Deep Learning sử dụng mạng nơ-ron nhân tạo nhiều lớp (deep neural networks) để mô phỏng cách bộ não con người xử lý và học hỏi từ dữ liệu. Phương pháp này cho phép máy tính tự động học từ dữ liệu mà không cần lập trình rõ ràng từng bước, đặc biệt hiệu quả với dữ liệu phi cấu trúc như hình ảnh, âm thanh và văn bản

Học sâu hoạt động bằng cách khám phá các cấu trúc phức tạp trong dữ liệu mà chúng thu thập. Các giai đoạn quá trình làm việc của Deep Learning như sau:

* Dữ liệu đầu vào (Input): Dữ liệu mới, như một hình ảnh hoặc văn bản, được cung cấp cho mô hình.
* Tiến bộ về phía trước (Forward Pass): Dữ liệu đi qua các lớp mạng, sử dụng các trọng số đã được huấn luyện trước đó, để tạo ra đầu ra. Quá trình này không bao gồm tính toán loss hoặc cập nhật trọng số, chỉ đơn thuần là xử lý dữ liệu.
* Đầu ra (Output): Mô hình sản xuất kết quả dự đoán chưa được khám phá bằng cách sử dụng những kiến thức và kinh nghiệm trước đó của nó, có thể là một nhãn phân loại (ví dụ: "mèo" hoặc "chó"), một giá trị hồi quy (ví dụ: giá nhà), hoặc một đầu ra khác tùy thuộc vào nhiệm vụ.

Ví dụ: Một mô hình Deep Learning mạng nơ ron hồi quy có thể được huấn luyện của các loại văn bản(lên đến hàng triệu văn bản) và mạng này học hỏi từ các sắc thái, mẫu câu có được trong đoạn văn bản. Mỗi câu trong văn bản được biểu diễn dưới dạng chuỗi các từ, và RNN xử lý từng từ theo thứ tự để nắm bắt ngữ cảnh. Sau khi xử lý toàn bộ câu, mạng sẽ dự đoán nhãn phù hợp, ví dụ như "tích cực" hoặc "tiêu cực" .​

## 1.3. Ưu và nhược điểm của Deep Learning

Một số ưu và nhược điểm của Deep Learning có thể kể đến như sau:

|  |  |
| --- | --- |
| **Ưu điểm** | **Nhược điểm** |
| * Tự động học tính năng: Giảm nhu cầu kỹ sư tính năng thủ công, học từ dữ liệu thô. * Xử lý dữ liệu lớn và phức tạp: Xử lý hiệu quả dữ liệu không cấu trúc như hình ảnh, văn bản, âm thanh. * Độ chính xác cao: Đạt hiệu suất cao trong nhận diện hình ảnh, giọng nói, NLP. * Khả năng mở rộng : Mở rộng cho dữ liệu lớn, triển khai trên đám mây hoặc thiết bị biên. * Mô hình dự đoán: Dự đoán xu hướng, sự kiện dựa trên dữ liệu lịch sử. * Xử lý dữ liệu thiếu: Bổ sung giá trị thiếu, vẫn đưa ra dự đoán chính xác. * Đa dạng trong học tập: Hỗ trợ nhiều phương pháp học như giám sát, không giám sát, chuyển giao. * Khả năng tạo sinh: Tạo nội dung mới như văn bản, hình ảnh, nhạc. | * Chi phí tính toán cao: Yêu cầu GPU mạnh mẽ, tốn kém và tiêu tốn năng lượng. * Nguy cơ quá khớp (overfitting): Dễ bị quá khớp nếu không chuẩn hóa đúng cách hoặc dữ liệu huấn luyện nhỏ. * Thiếu khả năng giải thích: Khó hiểu cách ra quyết định, được coi là "hộp đen". * Phụ thuộc vào dữ liệu: Hiệu suất phụ thuộc vào chất lượng và số lượng dữ liệu huấn luyện. * Mối quan ngại về quyền riêng tư: Xử lý dữ liệu lớn gây ra vấn đề bảo mật và quyền riêng tư. * Yêu cầu chuyên môn: Cần kiến thức chuyên sâu để triển khai hiệu quả. * Hậu quả không lường trước: Có thể dẫn đến thiên kiến và vấn đề đạo đức. * Khả năng tổng quát hạn chế: Không hoạt động tốt trên dữ liệu khác biệt với dữ liệu huấn luyện. * Quá trình huấn luyện tốn thời gian: Huấn luyện mất nhiều thời gian, đặc biệt với dữ liệu lớn và mô hình phức tạp. |

## 1.4. Ứng dụng của Deep Learning vào đời sống

​Deep learning (học sâu) đang ngày càng hiện diện trong đời sống hàng ngày, mang lại nhiều tiện ích và cải tiến đáng kể. Một trong những ứng dụng nổi bật là trong các trợ lý ảo như Siri, Google Assistant hay Alexa, giúp người dùng thực hiện các tác vụ như đặt lịch, tra cứu thông tin và điều khiển thiết bị thông minh bằng giọng nói . Trong lĩnh vực y tế, deep learning hỗ trợ chẩn đoán hình ảnh như phát hiện tế bào ung thư, phân tích ảnh chụp cắt lớp, giúp bác sĩ đưa ra quyết định chính xác hơn . Ngoài ra, công nghệ này còn được ứng dụng trong xe tự lái, nhận diện khuôn mặt để mở khóa điện thoại, cá nhân hóa nội dung trên các nền tảng như Netflix hay Spotify, và phát hiện tin giả trên mạng xã hội. Nhờ khả năng học từ dữ liệu lớn và xử lý thông tin phức tạp, deep learning đang góp phần thay đổi cách chúng ta sống, làm việc và tương tác với công nghệ..

### 1.4.1 Ứng dụng trong Digital Marketing

Deep learning đang cách mạng hóa lĩnh vực marketing kỹ thuật số thông qua khả năng phân tích dữ liệu lớn và hành vi người dùng. Các thương hiệu như Delta Air Lines và Mars sử dụng AI để tối ưu hóa hiệu suất quảng cáo và kết nối dữ liệu quảng cáo với doanh số bán hàng. Ví dụ, Delta đã sử dụng mạng nơ-ron để xác định rằng tài trợ Olympic của họ đóng góp 30 triệu USD vào doanh số bán hàng. Ngoài ra, các công ty như Unilever sử dụng nền tảng Omniverse của Nvidia để tạo hình ảnh sản phẩm nhanh hơn và tiết kiệm chi phí hơn. AI cũng tăng tốc sản xuất quảng cáo, như chiến dịch do AI tạo ra của Forever 21 đã đạt được lợi tức đầu tư cao hơn 66%.

### 1.4.2. Ứng dụng trong chế tạo Robot

Trong lĩnh vực chế tạo robot, deep learning giúp robot học hỏi và thích nghi với môi trường phức tạp. Các công ty như Agility Robotics và Tesla đang phát triển robot hình người có khả năng mang vật thể và di chuyển trên hai chân, nhờ vào tiến bộ trong AI và xử lý ngôn ngữ tự nhiên. Điều này giúp robot dễ dàng được đào tạo và thực hiện các nhiệm vụ trong môi trường như nhà kho và nhà máy, nhằm giải quyết tình trạng thiếu lao động và giảm rủi ro cho con người.

### 1.4.3. Ứng dụng trong hệ thống lái xe tự động

Một trong các thành tựu nổi bật nhất của Deep learning là nền tảng cho công cuộc xây dựng hệ thống lái xe tự động

Deep learning là nền tảng cho sự phát triển của xe tự lái. Các thuật toán học sâu giúp xe nhận diện và phân tích hình ảnh từ camera, radar và lidar để hiểu môi trường xung quanh, phát hiện vật cản, biển báo giao thông và người đi bộ. Điều này cho phép xe đưa ra quyết định lái xe an toàn và hiệu quả trong thời gian thực, giảm thiểu tai nạn và cải thiện lưu thông giao thông.

### 1.4.4. Ứng dụng trong trợ lý ảo

Trợ lý ảo đã xuất hiện và tồn tại trong một khoảng thời gian rất lâu. Chúng đã trở nên thân thuộc và không còn xa lạ gì với công chúng.

Trợ lý ảo như Siri, Google Assistant và Alexa sử dụng deep learning để hiểu và phản hồi các lệnh thoại của người dùng. Các mô hình ngôn ngữ lớn (LLMs) giúp trợ lý ảo hiểu ngữ cảnh, nhận diện ý định và cung cấp phản hồi tự nhiên hơn. Điều này cải thiện trải nghiệm người dùng và mở rộng khả năng của trợ lý ảo trong việc hỗ trợ các tác vụ hàng ngày.

Trợ lý ảo sử dụng Deep Learning để biết thêm về người dùng, từ các thói quen sử dụng, sở thích, các lịch hẹn. Các công nghệ này sẽ học tập hành vi người dùng, qua đó gợi ý cho người dùng các chức năng hữu ích được tích hợp trong hệ thống. Ngoài ra chúng, chúng có thể nhận lệnh từ người dùng thông qua công nghệ xử lý ngôn ngữ tự nhiên để thực hiện các tác vụ một cách thuận tiện.

Trợ lý ảo sử dụng Deep Learning thể hiện xuất sắc ở các tác vụ như: đặt lịch hẹn tự động, thông báo lịch hẹn, dịch văn bản tự động, tự động tìm kiếm tra khảo tài liệu, cung cấp thông tin dự báo thời tiết, cung cấp lộ trình di chuyển

### 1.4.5. Ứng dụng trong nghiên cứu y học

Deep learning đang được ứng dụng tích cực vào các hoạt động nghiên cứu y học và phẫu thuật chẩn đoán hiện nay.

Deep learning đang thúc đẩy tiến bộ trong nghiên cứu y học. Công ty Insitro, do Daphne Koller sáng lập, sử dụng AI để phân tích mẫu gen và phát hiện các cơ chế gây bệnh tiềm ẩn, mở đường cho việc phát triển thuốc mới cho các bệnh như ALS, ung thư và xơ cứng ống thận . Ngoài ra, một nghiên cứu cho thấy AI có thể dự đoán nguy cơ mắc bệnh tim trong 10 năm chỉ bằng hình ảnh X-quang ngực, với độ chính xác tương đương các phương pháp truyền thống. Các hệ thống AI cũng được phát triển để mô phỏng và dự đoán hình dạng, vị trí và cấu trúc của hàm răng và bộ mặt sau khi tiến hành phẫu thuật.

### 1.4.6. Ứng dụng trong nhận diện giọng nói

Deep Learning được sử dụng trong nhận diện giọng nói không chỉ nhằm phục vụ cho mục đích bảo mật mà còn được ứng dụng việc phân đoạn âm thành.

Deep learning cải thiện đáng kể khả năng nhận dạng giọng nói, cho phép các hệ thống chuyển đổi lời nói thành văn bản với độ chính xác cao. Điều này được ứng dụng trong các dịch vụ như tự động ghi chú cuộc họp, hỗ trợ khách hàng qua điện thoại và điều khiển thiết bị bằng giọng nói. Các mô hình học sâu có thể học từ nhiều giọng nói và ngữ điệu khác nhau, giúp hệ thống hiểu và phản hồi chính xác hơn

Tuy nhiên, thách thức đặt ra đối với Deep Learning khi nhận diện khuôn mặt là làm thế nào để xác định được chính xác giọng nói của một người ngay cả khi giọng nói của người đó đang bị biến dạng, hoặc trong lúc đang có quá nhiều tiếng ồn, hoặc là khi chất lượng âm thanh nhận vào kém.

### 1.4.7. Ứng dụng trong An ninh mạng

Deep learning (học sâu) đang trở thành công cụ cốt lõi trong lĩnh vực an ninh mạng, giúp phát hiện và phản ứng với các mối đe dọa ngày càng tinh vi. Dưới đây là những ứng dụng cụ thể và chi tiết của deep learning trong lĩnh vực này:​

* Hệ thống phát hiện xâm nhập (IDS) nâng cao: Các hệ thống IDS truyền thống dựa vào chữ ký (signature) để phát hiện các mối đe dọa đã biết. Tuy nhiên, deep learning cho phép phát hiện các mối đe dọa chưa từng thấy bằng cách học và nhận diện các mẫu hành vi bất thường trong lưu lượng mạng. Ví dụ, mạng nơ-ron tích chập (CNN) có thể phân tích lưu lượng mạng như hình ảnh, phát hiện các mẫu dữ liệu lạ mà hệ thống truyền thống có thể bỏ sót. ​
* Phân tích hành vi người dùng và thực thể (UEBA): Deep learning, đặc biệt là mạng nơ-ron hồi tiếp (RNN), giúp phân tích chuỗi hành vi của người dùng theo thời gian để phát hiện các hành vi bất thường. Điều này rất hữu ích trong việc phát hiện các mối đe dọa nội bộ hoặc tài khoản bị xâm nhập mà các hệ thống dựa trên quy tắc khó nhận ra. ​
* Phát hiện phần mềm độc hại và mã độc: Deep learning có khả năng phân tích mã phần mềm để phát hiện các phần mềm độc hại, kể cả những biến thể mới chưa có trong cơ sở dữ liệu chữ ký. Các mô hình như CNN có thể học từ các mẫu mã độc và phân biệt chúng với phần mềm hợp pháp, giúp phát hiện nhanh chóng các mối đe dọa mới. ​
* Phát hiện tấn công từ chối dịch vụ (DDoS): Deep learning có thể phân tích lưu lượng mạng để phát hiện các mẫu tấn công DDoS bằng cách nhận diện các đột biến bất thường trong lưu lượng. Điều này cho phép hệ thống phản ứng kịp thời và giảm thiểu tác động của các cuộc tấn công. ​
* Phát hiện lừa đảo (phishing) và thư rác: Các mô hình deep learning có thể phân tích nội dung và siêu dữ liệu của email để phát hiện các cuộc tấn công lừa đảo. Bằng cách học từ các mẫu email lừa đảo trước đó, hệ thống có thể nhận diện và ngăn chặn các email tương tự trong tương lai. ​
* Phát hiện và ngăn chặn tấn công mạng tiên tiến:Deep learning giúp phát hiện các cuộc tấn công mạng tiên tiến bằng cách phân tích hành vi và phát hiện các mẫu bất thường trong hệ thống. Điều này cho phép phát hiện sớm các cuộc tấn công và giảm thiểu thiệt hại. ​

## 1.5. Mạng học sâu CNN

### 1.5.1. Khái niệm chung

Mạng nơ-ron tích chập (Convolutional Neural Network – CNN) là một kiến trúc mạng nơ-ron sâu (Deep Neural Network) được thiết kế đặc biệt để xử lý dữ liệu có cấu trúc lưới, như hình ảnh, video hoặc chuỗi thời gian. CNN mô phỏng cơ chế xử lý thị giác của con người, cho phép máy tính tự động học và trích xuất các đặc trưng quan trọng từ dữ liệu đầu vào mà không cần can thiệp thủ công.

CNN hoạt động bằng cách áp dụng các bộ lọc lên ảnh đầu vào để tạo ra các bản đồ đặc trưng, sau đó sử dụng các tầng gộp để giảm kích thước và trích xuất các đặc trưng quan trọng. Cuối cùng, các tầng kết nối đầy đủ sử dụng các đặc trưng này để thực hiện các nhiệm vụ như phân loại hoặc nhận diện.

### 1.5.2. Các kiểu tầng trong mạng CNN

**Convolution Layer**

Tầng Convolutional là nền tảng của CNN, thực hiện phép tích chập (convolution) giữa đầu vào (thường là hình ảnh) và các bộ lọc (filters hoặc kernels). Mỗi bộ lọc có kích thước nhỏ (ví dụ 3x3 hoặc 5x5) và quét qua toàn bộ hình ảnh, tạo ra bản đồ đặc trưng (feature maps).

Chức năng: Trích xuất các đặc trưng cục bộ như cạnh, góc, hoặc hoa văn, giữ nguyên mối quan hệ không gian.

Tham số chính:

* Số lượng bộ lọc (K): Quyết định số lượng bản đồ đặc trưng đầu ra.
* Kích thước bộ lọc (F): Thường là 3x3 hoặc 5x5.
* Stride (S): Khoảng cách bộ lọc di chuyển, ví dụ stride = 1 di chuyển từng pixel.
* Padding (P): Thêm các pixel xung quanh đầu vào để giữ nguyên kích thước đầu ra, ví dụ padding = 1.

Công thức tính kích thước đầu ra:

* Chiều rộng đầu ra:
* Chiều cao đầu ra:
* Chiều sâu:
* Các hàm phi tuyến :

Các hàm phi tuyến (activation functions) là các hàm được sử dụng trong các lớp của mạng nơ-ron để biến đổi đầu ra của một neuron trước khi nó được truyền sang lớp tiếp theo. Chúng giới thiệu tính phi tuyến vào mô hình, cho phép mạng học các mối quan hệ phức tạp trong dữ liệu, điều mà các mô hình tuyến tính không thể làm được. Trong bài viết này, chúng ta sẽ khám phá chi tiết các hàm phi tuyến phổ biến nhất trong mạng nơ-ron, bao gồm cả các hàm đã được đề cập (ReLU, tanh, sigmoid) và một số hàm khác thường được sử dụng.

Mỗi bộ lọc (kernel) học một đặc trưng cụ thể, và nhiều kernel giúp mạng học nhiều thuộc tính khác nhau. Kết quả là một tensor 3D, sau đó đi qua hàm kích hoạt trước khi truyền sang lớp tiếp theo. Mỗi kernel, như một ma trận nhỏ, trượt qua hình ảnh để phát hiện đặc trưng, chẳng hạn như cạnh hoặc kết cấu. Với k kernel, mạng tạo ra k bản đồ đặc trưng (feature maps), mỗi bản đồ là một ma trận 2D. Những bản đồ này được xếp chồng thành tensor 3D, với chiều sâu bằng k, lưu trữ thông tin từ tất cả các kernel. Sau khi tạo tensor, nó đi qua hàm kích hoạt, như ReLU, để giới thiệu tính phi tuyến, giúp mạng học các mẫu phức tạp. Tensor này sau đó trở thành đầu vào cho lớp tiếp theo, hỗ trợ mạng xây dựng các đặc trưng ngày càng phức tạp.

**Pooling layer**

Pooling layer là một thành phần quan trọng trong mạng nơ-ron tích chập (CNN), giúp giảm kích thước không gian của các bản đồ đặc trưng (feature maps) từ lớp tích chập. Nó được thiết kế để downsample (giảm kích thước không gian) và tổng hợp thông tin từ nhiều vector thành ít vector hơn. Nó hoạt động bằng cách áp dụng các phép toán tổng hợp, như lấy giá trị lớn nhất (max pooling) hoặc trung bình (average pooling), trên các vùng nhỏ của bản đồ đặc trưng.

Vai trò:

* Tăng hiệu quả tính toán, làm mạng nơ-ron hoạt động nhanh hơn, đặc biệt trong các mạng sâu.
* Cải thiện khả năng không biến đổi (translation invariance), giúp mô hình nhận diện đối tượng ngay cả khi chúng dịch chuyển nhẹ trong ảnh.
* Ngăn ngừa overfitting bằng cách giảm số lượng tham số, giúp mô hình tổng quát hóa tốt hơn trên dữ liệu mới.
* Xây dựng biểu diễn đặc trưng phân cấp, nơi các lớp sâu hơn học các đặc trưng phức tạp hơn dựa trên các đặc trưng cơ bản từ lớp trước.

Các hàm gộm phổ biến thường được sử dụng:

* Max Pooling (Lấy giá trị lớn nhất)
* Cách hoạt động: Chọn giá trị lớn nhất trong mỗi vùng nhỏ (ví dụ: 2×2) của bản đồ đặc trưng.
* Ưu điểm: Giữ lại các đặc trưng nổi bật nhất, giúp mô hình tập trung vào các đặc điểm quan trọng.
* Ứng dụng: Thường được sử dụng sau các lớp convolution để giảm kích thước và tăng tính không đổi đối với dịch chuyển nhỏ trong ảnh.​
* Average Pooling (Lấy giá trị trung bình)
* Cách hoạt động: Tính trung bình các giá trị trong mỗi vùng nhỏ của bản đồ đặc trưng.
* Ưu điểm: Giữ lại thông tin tổng quát, làm mượt bản đồ đặc trưng.
* Ứng dụng: Phù hợp khi cần giảm nhiễu và giữ lại thông tin tổng thể của ảnh.​
* Global Pooling (Pooling toàn cục)
* Global Max Pooling (GMP): Lấy giá trị lớn nhất trên toàn bộ bản đồ đặc trưng cho mỗi kênh.
* Global Average Pooling (GAP): Tính trung bình toàn bộ bản đồ đặc trưng cho mỗi kênh.
* Ưu điểm: Giảm số lượng tham số, tránh overfitting, và thường được sử dụng trước lớp fully connected hoặc softmax trong các mô hình phân loại.​
* Các kỹ thuật pooling khác
* Mixed Pooling: Kết hợp giữa max pooling và average pooling bằng cách sử dụng trọng số.
* Lp Pooling: Tổng quát hóa max và average pooling bằng cách sử dụng chuẩn Lp.
* Stochastic Pooling: Chọn ngẫu nhiên một giá trị trong vùng pooling dựa trên xác suất tỷ lệ với giá trị đó.
* Softmax Pooling: Sử dụng hàm softmax để tính trọng số cho các giá trị trong vùng pooling.
* Spatial Pyramid Pooling: Áp dụng pooling ở nhiều mức độ phân giải khác nhau và kết hợp kết quả.

**Fully connected layers**

Lớp Fully Connected (FC), hay còn gọi là lớp kết nối đầy đủ hoặc lớp Dense, là một thành phần quan trọng trong kiến trúc của mạng nơ-ron tích chập (Convolutional Neural Network - CNN). Lớp này thường được đặt ở cuối mạng, sau các lớp tích chập (convolution) và lớp gộp (pooling), và đóng vai trò chuyển đổi các đặc trưng đã trích xuất thành kết quả đầu ra như phân loại hoặc dự đoán.​

* Cấu trúc và nguyên lý hoạt động

Trong lớp Fully Connected, mỗi neuron được kết nối với tất cả các neuron ở lớp trước đó. Điều này có nghĩa là mọi đặc trưng được trích xuất đều có thể ảnh hưởng đến quyết định cuối cùng của mạng.​

Quy trình hoạt động:

* Flattening: Sau khi dữ liệu đi qua các lớp tích chập và gộp, đầu ra thường là một tensor nhiều chiều. Để đưa vào lớp FC, tensor này được "làm phẳng" thành một vector một chiều.​
* Tính toán đầu ra: Mỗi neuron trong lớp FC thực hiện phép biến đổi tuyến tính trên vector đầu vào, sau đó áp dụng hàm kích hoạt phi tuyến (như ReLU, sigmoid hoặc softmax) để tạo ra đầu ra.​
* Vai trò:
* Tổng hợp đặc trưng: Kết hợp các đặc trưng cục bộ từ các lớp trước để tạo ra thông tin toàn cục, giúp mạng hiểu được mối quan hệ phức tạp giữa các đặc trưng.​
* Phân loại hoặc dự đoán: Chuyển đổi các đặc trưng đã tổng hợp thành kết quả đầu ra cụ thể, chẳng hạn như xác định hình ảnh thuộc về lớp nào trong bài toán phân loại.

### 1.5.3. Thuật toán lan truyền ngược (Backpropagation)

Thuật toán lan truyền ngược (Backpropagation) là một phương pháp quan trọng trong huấn luyện mạng nơ-ron nhân tạo, đặc biệt trong học sâu (deep learning). Nó cho phép tính toán gradient của hàm mất mát đối với các trọng số của mạng, từ đó cập nhật trọng số để giảm sai số dự đoán

**Các bước chính của thuật toán Backpropagation:**

* Lan truyền tiến (Forward Pass):
* Dữ liệu đầu vào được truyền qua các lớp của mạng nơ-ron để tính toán đầu ra dự đoán.
* Tại mỗi neuron, thực hiện phép biến đổi tuyến tính và áp dụng hàm kích hoạt.​
* Tính toán hàm mất mát (Loss Function): So sánh đầu ra dự đoán với giá trị thực tế để tính toán sai số.
* Lan truyền ngược (Backward Pass):
* Tính toán gradient của hàm mất mát đối với từng trọng số bằng cách áp dụng quy tắc chuỗi đạo hàm.
* Bắt đầu từ lớp đầu ra, tính toán gradient và lan truyền ngược qua các lớp để cập nhật trọng số.​
* Cập nhật trọng số (Weight Update):

Sử dụng thuật toán tối ưu (như Gradient Descent) để cập nhật trọng số:

Trong đó:

w: Trọng số (weight) hiện tại trong mạng nơ-ron.

:=: Gán giá trị mới cho w (tức là sau khi cập nhật).

η (eta): Tốc độ học (learning rate), là một hằng số nhỏ (ví dụ: 0.01), điều chỉnh mức độ thay đổi của trọng số trong mỗi lần cập nhật.

: Đạo hàm riêng của hàm mất mát L theo trọng số w, cho biết mức độ ảnh hưởng của w đến sai số dự đoán.

L: Hàm mất mát (loss function), đo sự khác biệt giữa đầu ra dự đoán và nhãn thực tế.

## 1.6. Mạng học sâu RNN và LSTM

### 1.6.1. Khái niệm chung

Mạng nơ-ron hồi quy (Recurrent Neural Network – RNN) là một kiến trúc mạng nơ-ron sâu được thiết kế đặc biệt để xử lý dữ liệu tuần tự, như văn bản, âm thanh hoặc chuỗi thời gian. Khác với các mạng nơ-ron truyền thống, RNN có khả năng ghi nhớ thông tin từ các bước trước đó thông qua cơ chế phản hồi, cho phép mô hình tận dụng ngữ cảnh của dữ liệu trước để dự đoán hoặc suy luận dữ liệu hiện tại.

RNN hoạt động bằng cách duy trì một trạng thái ẩn, hoạt động như một bộ nhớ, ghi lại thông tin về những gì đã được xử lý cho đến nay. Tại mỗi bước trong chuỗi, mạng sẽ lấy đầu vào hiện tại và trạng thái ẩn trước đó để tạo ra đầu ra và cập nhật trạng thái ẩn của nó.

RNN thường được sử dụng trong các ứng dụng như xử lý ngôn ngữ tự nhiên, dịch máy, nhận diện giọng nói và dự đoán chuỗi thời gian. Tuy nhiên, RNN cũng gặp phải một số hạn chế như vấn đề độ dốc biến mất khi xử lý các chuỗi dài, điều này có thể làm giảm hiệu quả học tập của mô hình.

### 1.6.2. Các kiểu tầng trong mạng RNN

### 1.6.3. Ví dụ minh họa mô hình RNN

# CHƯƠNG 2: PHƯƠNG PHÁP MÔ TẢ ĐƯỜNG ĐI VÀ GỢI Ý HƯỚNG DẪN NGƯỜI KHIẾM THỊ VÀ NGƯỜI CÓ THỊ LỰC KÉM

## 2.1. Bài toán sinh mô tả cho ảnh

## 2.2. Bài toán phân loại ảnh

## 2.3. Bài toán về mô tả đường đi và gợi ý hướng dẫn người khiếm thị và người có thị lực kém.

## 2.5. Xây dựng mô tả đường đi và gợi ý hướng dẫn người khiếm thị và người có thị lực kém.

# CHƯƠNG 3: THỰC NGHIỆM VÀ ĐÁNH GIÁ HỆ THỐNG MÔ TẢ ĐƯỜNG ĐI VÀ GỢI Ý HƯỚNG DẪN NGƯỜI KHIẾM THỊ VÀ NGƯỜI CÓ THỊ LỰC KÉM

## 3.1. Cài đặt chương trình huấn luyện mô hình

### 3.1.1. Tập dữ liệu

### 3.1.2. Cài đặt chương trình huấn luyện

**3.1.3. Kết quả huấn luyện**

## 3.2. Thử nghiệm và đánh giá trong môi trường thực tế

# KẾT LUẬN

**Kết quả đạt được**

**Hướng phát triển**

# TÀI LIỆU THAM KHẢO

[1] TS. Trần Hùng Cường – ThS. Nguyễn Phương Nga, Giáo trình Trí tuệ nhân tạo, Nhà xuất bản Trường Đại học Công nghiệp Hà Nội