**TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT**

**KHOA CÔNG NGHỆ THÔNG TIN**





ĐỒ ÁN KHOA HỌC MÁY TÍNH

**Đề tài:**

**Ứng dụng xử lý ảnh và học sâu trong**

**nhận diện và trích xuất biển số xe**

**Giáo viên hướng dẫn:** PGS.TSLê Hồng Anh

**Sinh viên:** Nguyễn Minh Tiến

**Mã sinh viên:** 2121050346

**Lớp:** Khoa học máy tính K66 A

HÀ NỘI, NĂM 2025

# MỤC LỤC

[MỤC LỤC I](#_Toc191678566)

[DANH MỤC CÁC HÌNH VẼ III](#_Toc191678567)

[DANH MỤC CÁC BẢNG BIỂU IV](#_Toc191678568)

[DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ V](#_Toc191678569)

[LỜI CAM ĐOAN VI](#_Toc191678570)

[LỜI CẢM ƠN VII](#_Toc191678571)

[MỞ ĐẦU 1](#_Toc191678572)

[CHƯƠNG 1 TỔNG QUAN VỀ ĐỀ TÀI 2](#_Toc191678573)

[1.1 Lý do chọn đề tài 2](#_Toc191678574)

[1.2 Mục tiêu đề tài 3](#_Toc191678575)

[1.3 Phạm vi nghiên cứu đề tài 4](#_Toc191678576)

[CHƯƠNG 2 CƠ SỞ LÝ THUYẾT 5](#_Toc191678577)

[2.1 Tổng quan về trí tuệ nhân tạo, học máy và học sâu 5](#_Toc191678578)

[2.1.1 Trí tuệ nhân tạo (AI) 5](#_Toc191678579)

[2.1.2 Học máy 5](#_Toc191678580)

[2.1.3 Học sâu 7](#_Toc191678581)

[2.2 Tổng quan về thị giác máy tính và xử lý ảnh 8](#_Toc191678582)

[2.2.1 Thị giác máy tính (Computer Vision) 8](#_Toc191678583)

[2.2.2 Xử lý ảnh 9](#_Toc191678584)

[2.3 Tổng quan CNN và mô hình YOLO 11](#_Toc191678585)

[2.3.1 Mạng nơ-ron tích chập (Convolution Neural Network - CNN) 11](#_Toc191678586)

[2.3.2 Mô hình YOLO (You Only Look Once) 16](#_Toc191678587)

[CHƯƠNG 3 PHÂN TÍCH BÀI TOÁN 17](#_Toc191678588)

[3.1 Tiêu đề mục 17](#_Toc191678589)

[3.1.1 Tiêu đề tiểu mục 17](#_Toc191678590)

[CHƯƠNG 4 KẾT QUẢ CÀI ĐẶT, THỬ NGHIỆM HOẶC KẾT QUẢ NGHIÊN CỨU 18](#_Toc191678591)

[4.1 Tiêu đề mục 18](#_Toc191678592)

[4.1.1 Tiêu đề tiểu mục 18](#_Toc191678593)

[KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 19](#_Toc191678594)

[4.1.2 Kết luận chung 19](#_Toc191678595)

[PHỤ LỤC 21](#_Toc191678596)

# DANH MỤC CÁC HÌNH VẼ

[Hình 1 – Mối quan hệ giữa Trí tuệ nhân tạo (AI), Học máy (Machine Learning) và Học sâu (Deep Learning) 7](#_Toc191244211)

# DANH MỤC CÁC BẢNG BIỂU

**No table of figures entries found.**

# DANH MỤC CÁC TỪ VIẾT TẮT VÀ GIẢI THÍCH CÁC THUẬT NGỮ

|  |  |  |
| --- | --- | --- |
| **Kí hiệu** | **Diễn giải** | **Ý nghĩa** |
| AI | Artificial intelligence | Trí tuệ nhân tạo |
| LLMs | Large language model | Mô hình ngôn ngữ lớn |
| Pixel | Picture element | Đơn vị nhỏ nhất trong mô hình ảnh số |
| CNN | Convolution Neural Network | Mạng nơ ron tích chập |
|  |  |  |

# LỜI CAM ĐOAN

# LỜI CẢM ƠN

# MỞ ĐẦU

Hiện nay, việc quản lý bãi đỗ xe tại Việt Nam vẫn còn nhiều hạn chế do quy trình thủ công, dễ xảy ra sai sót trong việc ghi nhận thời gian và tính phí, dẫn tới việc thất thoát doanh thu cũng như gây bất tiện cho người sử dụng. Đặc biệt, vào giờ cao điểm, việc kiểm soát phương tiện ra vào gặp nhiều khó khăn, có thể gây ùn tắc và mất thời gian. Thiếu tính minh bạch trong quá trình tính phí cũng dễ tạo ra tranh cãi giữa khách hàng và đơn vị quản lý. Những vấn đề này đặt ra nhu cầu cấp thiết về một giải pháp tự động hóa nhằm nâng cao hiệu quả quản lý và trải nghiệm người dùng.

Đề tài "**Ứng dụng xử lý ảnh và học sâu trong nhận diện và trích xuất biển số xe**" tập trung nghiên cứu và phát triển một hệ thống nhận diện biển số xe dựa trên công nghệ xử lý ảnh và học sâu. Hệ thống này sẽ áp dụng các thuật toán và bộ lọc xử lý ảnh như Gaussian để giảm nhiễu, Canny Edge Detection để xác định biên cạnh, và Hough Line Transform để phát hiện đường thẳng, kết hợp với ngôn ngữ lập trình Python và thư viện OpenCV để xử lý hình ảnh chuyên sâu.

Ngoài ra, mô hình YOLO sẽ được huấn luyện để nhận diện biển số xe trong hình ảnh, trong khi mô hình CNN sẽ được sử dụng để nhận diện các ký tự trên biển số. Cuối cùng, hệ thống sẽ được tích hợp vào một ứng dụng web bằng Flask nhằm trình diễn khả năng nhận diện và trích xuất biển số xe từ hình ảnh một cách trực quan.

Với phương pháp tiếp cận này, hệ thống không chỉ hỗ trợ quản lý bãi đỗ xe thông minh mà còn có thể mở rộng ứng dụng vào các lĩnh vực như giám sát giao thông, kiểm soát an ninh và các hệ thống tự động hóa khác.

# TỔNG QUAN VỀ ĐỀ TÀI

## Lý do chọn đề tài

Trong thời đại công nghệ số phát triển mạnh mẽ, trí tuệ nhân tạo (AI) và thị giác máy tính (Computer Vision) đang dần trở thành những công cụ quan trọng, mang lại nhiều lợi ích trong cuộc sống. Một trong những ứng dụng thực tế quan trọng của các công nghệ này là nhận diện biển số xe, giúp tự động hóa quy trình giám sát và quản lý phương tiện, đặc biệt trong các bãi đỗ xe, khu đô thị và hệ thống giao thông thông minh.

Việc nhận diện và trích xuất thông tin từ biển số xe không chỉ giúp giảm thiểu sai sót, tiết kiệm thời gian và chi phí vận hành, mà còn tăng cường tính minh bạch, hạn chế gian lận trong việc kiểm soát phương tiện. Ngoài ra, hệ thống này còn hỗ trợ lực lượng chức năng trong công tác giám sát giao thông, phát hiện các phương tiện vi phạm hoặc bị đánh cắp, góp phần nâng cao an ninh và trật tự xã hội.

Bên cạnh đó, với xu hướng phát triển của thành phố thông minh, việc ứng dụng công nghệ vào giao thông là một bước tiến quan trọng, giúp xây dựng các hệ thống hiện đại, đồng bộ và tự động hóa. Thông qua đề tài này, em không chỉ mong muốn góp phần vào sự phát triển của các hệ thống nhận diện biển số xe mà còn có cơ hội tiếp cận, tìm hiểu và áp dụng các phương pháp tiên tiến trong lĩnh vực xử lý ảnh và học sâu vào thực tiễn.

Với những lý do trên, việc nghiên cứu và triển khai đề tài **"Ứng dụng xử lý ảnh và học sâu trong nhận diện và trích xuất biển số xe"** không chỉ mang lại giá trị học thuật mà còn có ý nghĩa thực tiễn, mở ra nhiều cơ hội ứng dụng trong tương lai.

## Mục tiêu đề tài

Đề tài **"Ứng dụng xử lý ảnh và học sâu trong nhận diện và trích xuất biển số xe"** được thực hiện với mục tiêu nghiên cứu và xây dựng một ứng dụng web cơ bản có khả năng nhận diện biển số xe từ hình ảnh. Hệ thống sẽ áp dụng các phương pháp xử lý ảnh và học sâu để xác định biển số xe, trích xuất và nhận diện các ký tự trên biển số, đảm bảo khả năng hoạt động trong điều kiện thực tế.

Cụ thể, đề tài đặt ra các mục tiêu sau:

* Nghiên cứu và áp dụng các thuật toán xử lý ảnh: Tìm hiểu và sử dụng các phương pháp như Gaussian (giảm nhiễu), Canny Edge Detection (xác định biên), Hough Line Transform (xác định đường thẳng) kết hợp với thư viện OpenCV để tiền xử lý ảnh.
* Huấn luyện mô hình nhận diện biển số xe: Tìm hiểu và sử dụng YOLO để phát hiện vị trí biển số xe trong ảnh.
* Huấn luyện mô hình nhận diện ký tự: Áp dụng mô hình CNN để nhận diện từng ký tự trên biển số xe.
* Xây dựng ứng dụng web cơ bản: Tích hợp hệ thống nhận diện vào một ứng dụng web sử dụng Flask, cho phép người dùng tải lên hình ảnh và nhận kết quả nhận diện biển số.
* Đánh giá kết quả và cải thiện mô hình: Kiểm tra độ chính xác của hệ thống, đánh giá hiệu suất nhận diện và đề xuất hướng cải tiến.

Đề tài này cũng là tiền đề giúp em phát triển đồ án tốt nghiệp trong tương lai. Thông qua quá trình thực hiện, em có cơ hội tìm hiểu sâu hơn về các thuật toán xử lý ảnh, mô hình học sâu cũng như cách tích hợp chúng vào một hệ thống thực tế. Giúp em rèn luyện kỹ năng lập trình, làm việc với các thư viện AI phổ biến và phát triển ứng dụng web, tạo nền tảng vững chắc để mở rộng nghiên cứu trong các lĩnh vực như thị giác máy tính, trí tuệ nhân tạo ứng dụng hoặc các hệ thống nhận diện thông minh. Đây sẽ là cơ sở quan trọng để em tiếp tục cải tiến, mở rộng chức năng của hệ thống và triển khai vào những bài toán thực tế lớn hơn trong đồ án tốt nghiệp sau này.

## Phạm vi nghiên cứu đề tài

Đề tài tập trung vào nghiên cứu và ứng dụng các kỹ thuật xử lý ảnh và học sâu để nhận diện và trích xuất biển số xe. Phạm vi nghiên cứu bao gồm các nội dung chính sau:

* Nghiên cứu về Thị giác máy tính (Computer Vision): Tìm hiểu các phương pháp phân tích và xử lý hình ảnh nhằm nhận diện biển số xe một cách hiệu quả.
* Áp dụng các kỹ thuật tiền xử lý ảnh:
* Gaussian Blur để giảm nhiễu và cải thiện chất lượng ảnh đầu vào.
* Canny Edge Detection để phát hiện và xác định biên ảnh.
* Hough Line Transform để nhận diện đường thẳng, mục đích hỗ trợ cho việc xoay biển số thẳng
* Phát hiện và nhận diện biển số xe:
  + Huấn luyện mô hình YOLO để xác định vùng chứa biển số xe trên ảnh.
  + Nhận diện ký tự trên biển số xe: Huấn luyện mô hình mạng nơ-ron tích chập (CNN) để nhận diện ký tự trong biển số xe sau khi đã tách từng ký tự riêng lẻ.
* Phát triển ứng dụng web cơ bản để thử nghiệm bằng Flask, cho phép người dùng tải ảnh lên và hiển thị kết quả.

# CƠ SỞ LÝ THUYẾT

## Tổng quan về trí tuệ nhân tạo, học máy và học sâu

### Trí tuệ nhân tạo (AI)

Trí tuệ nhân tạo (AI) là một công nghệ có khả năng giải quyết vấn đề như con người. Cách thức hoạt động của Trí tuệ nhân tạo dường như mô phỏng trí tuệ của con người – nó có thể nhận dạng hình ảnh, làm thơ và đưa ra dự đoán dựa trên dữ liệu. [1]

Lịch sử của trí tuệ nhân tạo (AI) bắt đầu từ những ý tưởng sơ khai vào thập kỷ 1940. Tuy nhiên, chỉ trong những năm gần đây, AI mới phát triển mạnh mẽ nhờ vào những tiến bộ đáng kể trong lĩnh vực máy học và sự sẵn có của nguồn dữ liệu lớn. Một số thành tựu nổi bật của AI bao gồm nhận diện giọng nói, nhận diện hình ảnh, xe tự hành, và khả năng chơi cờ vua ở cấp độ chuyên nghiệp.

Các ứng dụng của AI ngày càng mở rộng, xuất hiện trong hầu hết các lĩnh vực như y tế, tài chính, giáo dục, và dự báo thời tiết. AI đang thay đổi cách con người tương tác với công nghệ, mang đến những giải pháp thông minh cho các vấn đề phức tạp. Bên cạnh đó, AI cũng góp phần tạo ra những tiện ích hàng ngày như trợ lý ảo, hệ thống tự động hóa và nhiều ứng dụng thực tiễn khác.

### Học máy

Máy học hay học máy là môn khoa học nhằm phát triển những thuật toán và mô hình thống kê mà các hệ thống máy tính sử dụng để thực hiện các tác vụ dựa vào khuôn mẫu và suy luận mà không cần hướng dẫn cụ thể. Các hệ thống máy tính sử dụng thuật toán máy học để xử lý khối lượng lớn dữ liệu trong quá khứ và xác định các khuôn mẫu dữ liệu. Việc này cho phép chúng dự đoán kết quả chính xác hơn từ cùng một tập dữ liệu đầu vào cho trước. Ví dụ: các nhà khoa học dữ liệu có thể đào tạo một ứng dụng y tế chẩn đoán ung thư từ ảnh chụp X-quang bằng cách lưu trữ hàng triệu ảnh quét và chẩn đoán tương ứng. [2]

Về cơ bản, học máy (Machine Learning) là quá trình tạo ra các mô hình máy tính có khả năng tự điều chỉnh dựa trên kinh nghiệm và dữ liệu đã được cung cấp. Điểm quan trọng ở đây là học máy không chỉ đơn thuần áp dụng các quy tắc lập trình cứng nhắc, mà thay vào đó, nó tìm kiếm và xác định các mối liên kết, quy luật, cũng như biểu diễn từ dữ liệu, nhằm đưa ra dự đoán và quyết định cho dữ liệu mới.

Học máy có thể được chia thành bốn loại chính:

* Học có giám sát (Supervised Learning):
  + Trong phương pháp này, mô hình được huấn luyện trên một tập dữ liệu có nhãn, nghĩa là mỗi mẫu dữ liệu đã được gán nhãn với đầu ra mong muốn. Mục tiêu là học cách ánh xạ dữ liệu đầu vào đến các nhãn đầu ra để có thể dự đoán chính xác kết quả cho các mẫu dữ liệu mới.
* Học không giám sát (Unsupervised Learning):
  + Ở đây, mô hình làm việc với một tập dữ liệu không có nhãn, và nhiệm vụ của nó là tự tìm ra các cấu trúc, mẫu hình hoặc nhóm dữ liệu mà không cần biết trước đầu ra mong muốn. Phương pháp này thường được sử dụng trong phân cụm dữ liệu và giảm chiều dữ liệu.
* Học bán giám sát (Semi-supervised Learning):
  + Đây là phương pháp kết hợp giữa học có giám sát và học không giám sát, trong đó mô hình được huấn luyện với một lượng nhỏ dữ liệu có nhãn và một lượng lớn dữ liệu không có nhãn. Cách tiếp cận này giúp tận dụng tốt dữ liệu chưa dán nhãn để cải thiện hiệu suất mô hình.
* Học tăng cường (Reinforcement Learning):
  + Phương pháp này hoạt động dựa trên cơ chế thử - sai và phần thưởng. Mô hình tương tác với môi trường, học hỏi từ những hành động của mình và điều chỉnh chiến lược để tối ưu hóa phần thưởng nhận được. Không giống như các phương pháp học máy truyền thống, học tăng cường đặc biệt hữu ích trong các môi trường động và không chắc chắn, chẳng hạn như trò chơi điện tử, robot tự hành, hoặc giao dịch tài chính.

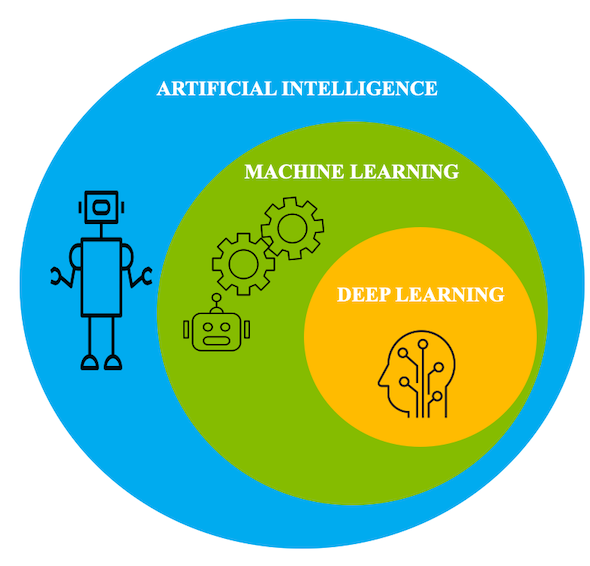
Học máy đã đạt được nhiều thành công trong thực tế, với các ứng dụng phổ biến như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, dự đoán chuỗi thời gian, và nhiều lĩnh vực khác. Nó đã làm thay đổi cách con người xử lý thông tin, ra quyết định, và mang lại sự tự động hóa cũng như hiệu suất vượt trội.

Tuy nhiên, một thách thức lớn của học máy là nó hoạt động kém hiệu quả đối với dữ liệu phức tạp có tính phi tuyến cao. Các mô hình tuyến tính truyền thống gặp khó khăn trong việc nắm bắt các mối quan hệ phức tạp giữa đầu vào và đầu ra, đòi hỏi các phương pháp tiên tiến hơn như mạng nơ-ron nhân tạo (Neural Networks) để cải thiện hiệu suất.

### Học sâu

Học sâu (Deep Learning) là một dạng của học máy (Machine Learning) có khả năng nhận diện các mẫu phức tạp và tạo ra mối liên kết theo cách tương tự con người. Khả năng của nó có thể bao gồm từ việc xác định vật thể trong ảnh, nhận diện giọng nói, đến điều khiển xe tự hành hoặc tạo ra hình ảnh minh họa. Về cơ bản, một mô hình học sâu là một chương trình máy tính có thể thể hiện trí thông minh nhờ vào cách tiếp cận phức tạp và tinh vi trong xử lý dữ liệu.

Học sâu là một nhánh của trí tuệ nhân tạo (AI) và đóng vai trò cốt lõi trong cách nhiều dịch vụ và mô hình AI hoạt động. Các mô hình ngôn ngữ lớn (LLMs) như ChatGPT, Bard và Bing Chat, cùng với các công cụ tạo hình ảnh như Midjourney và DALL-E, dựa vào học sâu để học ngôn ngữ, ngữ cảnh và tạo ra các phản hồi thực tế. Các mô hình AI dự đoán cũng sử dụng học sâu để rút ra kết luận từ kho dữ liệu lịch sử khổng lồ. [3]



Hình – Mối quan hệ giữa Trí tuệ nhân tạo (AI), Học máy (Machine Learning) và Học sâu (Deep Learning)

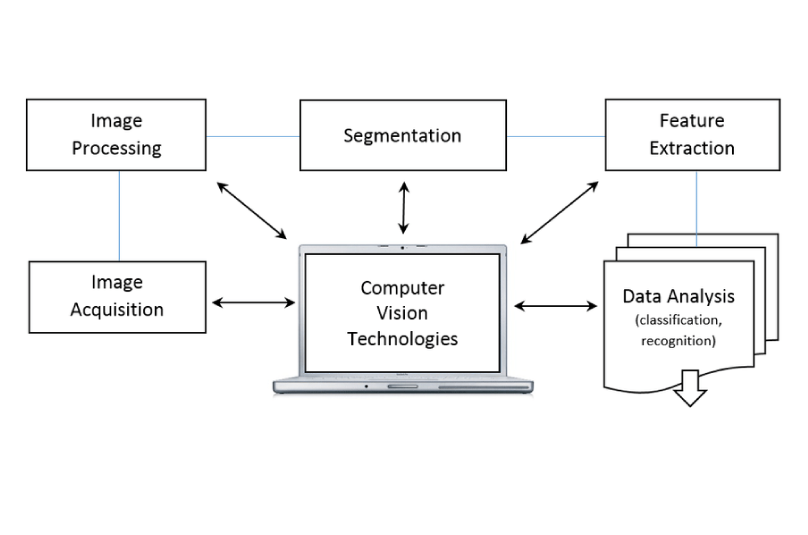
Các mô hình này, còn được gọi là mạng nơ-ron nhân tạo, được thiết kế với nhiều lớp liên kết mô phỏng cách hoạt động của bộ não con người. Trong quá trình huấn luyện, mạng nơ-ron học hỏi từ dữ liệu thông qua hàm đánh giá, dần dần tích lũy kinh nghiệm. Những kinh nghiệm này giúp mô hình có khả năng ghi nhớ và sử dụng để dự đoán kết quả cho các dữ liệu chưa từng gặp trước đó.

## Tổng quan về thị giác máy tính và xử lý ảnh

### Thị giác máy tính (Computer Vision)

Thị giác máy tính (Computer Vision) là một lĩnh vực nghiên cứu tập trung vào việc thu nhận, xử lý và phân tích hình ảnh kỹ thuật số nhằm trích xuất thông tin có ý nghĩa. Nó bao gồm các phương pháp tự động để nhận diện, phân loại và diễn giải dữ liệu hình ảnh từ thế giới thực, giúp máy tính có khả năng hiểu và phản hồi với môi trường xung quanh thông qua dữ liệu trực quan.

Thị giác máy tính là một nhánh của trí tuệ nhân tạo và khoa học máy tính, tập trung vào việc giúp máy tính thu nhận, phân tích và diễn giải thông tin từ hình ảnh và video, nhằm mô phỏng khả năng nhìn và hiểu thế giới như con người.



Hình – Khái quát về cơ chế hoạt động của Thị giác máy tính

**Ứng dụng:** Thị giác máy tính được áp dụng trong nhiều lĩnh vực như tái tạo cảnh 3D, nhận diện hình ảnh, phát hiện sự kiện, theo dõi đối tượng trong video, phân tích chuyển động, khôi phục và cải thiện chất lượng ảnh, cũng như xử lý ảnh nâng cao.

### Xử lý ảnh

Xử lý ảnh là một nhánh của khoa học máy tính và kỹ thuật, chuyên nghiên cứu các phương pháp phân tích, biến đổi và thao tác hình ảnh để trích xuất thông tin có ý nghĩa. Quá trình này có thể bao gồm việc cải thiện chất lượng ảnh, nhận diện mẫu, hoặc chuẩn bị dữ liệu đầu vào cho các hệ thống thị giác máy tính. Trong bài toán nhận diện biển số xe, xử lý ảnh đóng vai trò quan trọng trong việc làm sạch dữ liệu đầu vào, phát hiện vùng chứa biển số và tách ký tự phục vụ cho quá trình nhận dạng.

A diagram of a flowchart

AI-generated content may be incorrect.

Hình – Mô tả về cơ chế hoạt động của xử lý ảnh

**Các phương pháp và phép toán phổ biến trong xử lý ảnh:**

* **Chuyển đổi ảnh**
  + Thay đổi độ phân giải hoặc kích thước ảnh để phù hợp với yêu cầu xử lý.
  + Chuyển đổi không gian màu (ví dụ: từ RGB sang grayscale để giảm độ phức tạp hoặc sang HSV để dễ dàng xử lý theo màu sắc).
  + Thực hiện các phép biến đổi để thay đổi đặc điểm của ảnh, hỗ trợ trong quá trình nhận diện đối tượng.
* **Làm mịn và tăng độ nét ảnh**
  + Sử dụng các bộ lọc giúp giảm nhiễu hoặc làm nổi bật đặc trưng trong ảnh:
    - Bộ lọc Gaussian: Làm mịn ảnh và giảm nhiễu mà không làm mất quá nhiều chi tiết quan trọng.
    - Bộ lọc Laplacian: Nhấn mạnh đường biên và các vùng có sự thay đổi cường độ sáng lớn, giúp làm rõ các đặc trưng của ảnh.
* **Trích xuất đặc trưng**
  + Phát hiện biên ảnh:
    - Thuật toán Canny Edge Detection giúp xác định các cạnh rõ ràng trong ảnh, hỗ trợ trong việc tìm vùng chứa biển số.
  + Phát hiện điểm đặc trưng:
    - Harris Corner Detection giúp nhận diện các góc hoặc điểm có tính chất đặc trưng, hỗ trợ trong việc xác định hình dạng của đối tượng.
* **Biến đổi hình học**
  + Thực hiện các phép biến đổi như xoay, co giãn để điều chỉnh góc nhìn hoặc hình dạng của ảnh.
  + Áp dụng chuyển đổi affine và non-affine để thay đổi cấu trúc hình học của ảnh mà vẫn giữ nguyên các đặc trưng quan trọng.
* **Phân đoạn ảnh**
  + Dựa trên màu sắc: Chia ảnh thành các vùng khác nhau dựa vào thông tin màu sắc, hữu ích trong việc tách đối tượng khỏi nền.
  + Dựa trên đường biên: Sử dụng đặc trưng đường biên để xác định và phân vùng ảnh, giúp nhận diện các thành phần cụ thể như biển số xe.
* **Lọc và xử lý nhiễu**
  + Bộ lọc trung vị (Median Filter): Loại bỏ nhiễu xung mà vẫn giữ được đường biên rõ ràng, phù hợp với ảnh biển số có nhiều chi tiết nhỏ.
  + Bộ lọc Bilateral: Làm mịn ảnh nhưng vẫn giữ nguyên các đường biên, giúp tăng độ rõ nét của biển số xe mà không làm mờ ký tự.
* **Biến đổi Fourier**
  + Chuyển ảnh từ miền không gian sang miền tần số và ngược lại, hỗ trợ trong các tác vụ như nén ảnh, lọc tần số, hoặc cải thiện độ sắc nét của ảnh.
* **Histogram Equalization:** Cải thiện độ tương phản của ảnh bằng cách phân phối lại các giá trị pixel, làm cho màu sắc phong phú và đa dạng hơn.

**Các bước triển khai**

* Thu nhận ảnh: Tiến hành chụp và số hóa hình ảnh từ camera hoặc các nguồn dữ liệu khác.
* Tiền xử lý: Giảm nhiễu, cân chỉnh độ tương phản, điều chỉnh độ sáng và thay đổi kích thước ảnh để tối ưu hóa chất lượng đầu vào.
* Phân tích và trích xuất đặc trưng: Xác định các yếu tố quan trọng của ảnh như cạnh, đường biên, hình dạng và kết cấu.
* Phân đoạn ảnh: Chia ảnh thành các khu vực hoặc đối tượng riêng biệt nhằm hỗ trợ quá trình xử lý tiếp theo.
* Nhận diện: Phát hiện và phân loại đối tượng trong ảnh bằng cách sử dụng các thuật toán trí tuệ nhân tạo hoặc học máy.

**Ứng dụng**

* Trong y học: Xử lý ảnh X-quang, MRI để chẩn đoán
* Thị giác máy tính: Dùng trong xe tự lái, nhận diện khuôn mặt, nhận dạng ký tự quang học (OCR).
* Đồ họa và nhiếp ảnh số: Cải thiện chất lượng ảnh, tạo hiệu ứng

## Tổng quan CNN và mô hình YOLO

### Mạng nơ-ron tích chập (Convolution Neural Network - CNN)

CNN là một loại mạng nơ-ron chuyên xử lý dữ liệu dạng hình ảnh. Thay vì làm việc trực tiếp với từng pixel, CNN sử dụng các bộ lọc (filter/kernel) để trích xuất đặc trưng quan trọng từ ảnh như cạnh, góc, đường nét, giúp nhận diện ký tự trên biển số hiệu quả hơn.

Kiến trúc CNN sử dụng các tầng tích chập (Convolutional Layers) để trích xuất các đặc trưng cục bộ từ dữ liệu đầu vào. Nhờ khả năng tự động học đặc trưng, CNN đã đạt được những thành tựu đáng kể trong lĩnh vực thị giác máy tính, đặc biệt trong các bài toán như phân loại ảnh, nhận diện đối tượng và phân đoạn ảnh.

A diagram of a diagram of a building

AI-generated content may be incorrect.

Hình – Ví dụ về mô hình CNN

**Các lớp cơ bản của mạng CNN:**

#### Lớp tích chập (Convolutional Layers)

Đây là lớp quan trọng nhất trong CNN, có nhiệm vụ thực hiện phép tích chập (convolution) để trích xuất các đặc trưng cục bộ từ hình ảnh. Lớp này sử dụng các bộ lọc (filter), còn gọi là kernel, để quét qua toàn bộ ảnh và phát hiện các đặc trưng quan trọng.

* **Nguyên lý hoạt động:** Bộ lọc là một ma trận nhỏ (thường có kích thước 3×3 hoặc 5×5), di chuyển lần lượt trên từng vùng của ảnh và thực hiện phép tích chập với giá trị pixel tại vị trí đó.
* **Filter Map:** Là tập hợp các ma trận ba chiều chứa tham số, được điều chỉnh trong quá trình huấn luyện. Mỗi bộ lọc sẽ học một đặc trưng khác nhau của ảnh, chẳng hạn như cạnh, góc hoặc các hình dạng phức tạp hơn.
* **Stride:** Xác định khoảng dịch chuyển của bộ lọc trên ảnh. Nếu stride = 1, bộ lọc di chuyển từng pixel một, còn nếu stride = 2, bộ lọc sẽ nhảy hai pixel mỗi lần quét, giúp giảm kích thước của đầu ra (feature map).

A white rectangular box with a arrow pointing to the left

AI-generated content may be incorrect.

Hình – Lớp tích chập hoạt động với stride là 2

* **Padding:** Để tránh mất thông tin ở các vùng biên của ảnh khi thực hiện phép tích chập, người ta thường thêm các giá trị 0 xung quanh ảnh trước khi áp dụng bộ lọc. Việc này giúp giữ nguyên kích thước của ảnh đầu vào và đảm bảo rằng các đặc trưng quan trọng ở rìa ảnh vẫn được phát hiện hiệu quả.

A screenshot of a grid

AI-generated content may be incorrect.

Hình – Minh họa về Zero Padding (Padding 0)

* **Feature Map:** Sau khi bộ lọc quét qua toàn bộ ảnh, kết quả của mỗi lần tích chập được lưu trong một ma trận gọi là feature map. Ma trận này chứa các đặc trưng quan trọng đã được trích xuất, giúp mạng CNN nhận diện được các chi tiết như cạnh, góc và hình dạng trong ảnh.

#### Lớp kích hoạt (Activation Layer)

Lớp kích hoạt đóng vai trò quan trọng trong việc đưa tính phi tuyến vào CNN, giúp mô hình học các đặc trưng phức tạp hiệu quả hơn. Nếu không có lớp này, CNN chỉ có thể biểu diễn các mối quan hệ tuyến tính, làm hạn chế khả năng nhận dạng hình ảnh.

**ReLU** (Rectified Linear Unit) Là một hàm kích hoạt phi tuyến phổ biến trong CNN, giúp mạng học được các đặc trưng phức tạp hơn bằng cách đưa tính phi tuyến vào mô hình, do phép tích chập vốn dĩ là một phép toán tuyến tính.

* **Công thức:**

A diagram of a function

AI-generated content may be incorrect.

Hình – Cơ chế hoạt động của hàm phi tuyến tính – ReLU

Hàm này giữ nguyên giá trị đầu vào nếu nó dương và đặt về 0 nếu âm, giúp mô hình tránh vấn đề vanishing gradient và tăng hiệu suất huấn luyện. Ngoài ReLU, lớp kích hoạt còn có các hàm khác như Sigmoid, Tanh,… tùy vào yêu cầu của mô hình mà ta sử dụng.

#### Lớp lấy mẫu (Pooling Layer)

Lớp này giúp giảm kích thước feature map để giảm số lượng tham số, tránh overfitting.

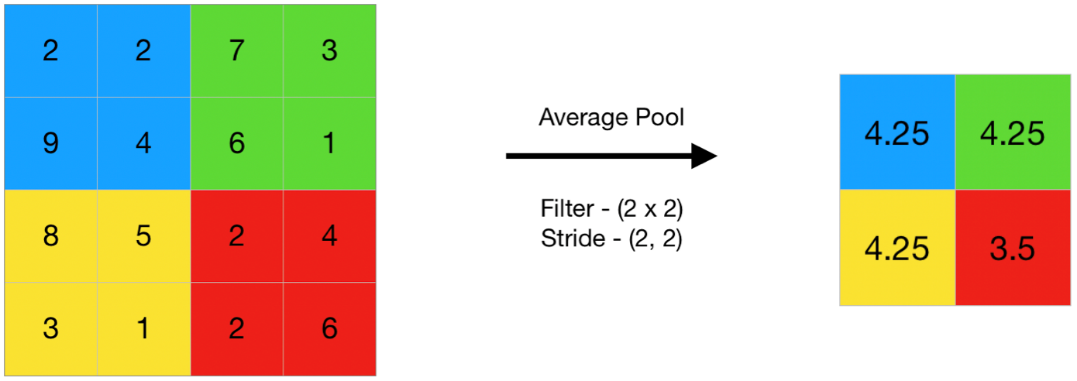
Các loại Pooling Layer:

* **Max Pooling:** Giữ lại giá trị lớn nhất trong từng vùng nhỏ.
* **Average Pooling:** Tính trung bình các giá trị trong vùng nhỏ.

A diagram of a graph

AI-generated content may be incorrect.

Hình – Cơ chế hoạt động của Max Pooling với filter 2x2 và stride = 2



Hình – Cơ chế hoạt động của Average Pooling với filter 2x2 và stride = 2

#### Lớp kết nối đầy đủ (Fully Connected Layer - FC)

**Fully Connected Layer** là lớp cuối cùng của mạng CNN. Lớp này giúp biến các đặc trưng (feature maps) đã được trích xuất và giảm kích thước từ các tầng trước thành một vector và thực hiện phân loại.

**Cách hoạt động:**

* Mỗi neuron trong tầng này được kết nối với toàn bộ neuron của tầng trước.
* Các đặc trưng cục bộ (đã trích xuất qua các tầng tích chập và pooling) được tổng hợp thành thông tin tổng quát hơn.
* Tầng cuối thường sử dụng Softmax (cho phân loại nhiều lớp) hoặc Sigmoid (cho bài toán nhị phân) để đưa ra dự đoán.

A diagram of a network

AI-generated content may be incorrect.

Hình - Quá trình làm phẳng và đưa feature maps vào lớp kết nối đầy đủ (FC)

### Mô hình YOLO (You Only Look Once)

YOLO (You Only Look Once) là một thuật toán phát hiện đối tượng dựa trên mạng nơ-ron tích chập (CNN).

A diagram of a graph

AI-generated content may be incorrect.

Hình – Minh họa về mô hình YOLO

Khác với các phương pháp trước đây như R-CNN, Faster R-CNN (chạy nhiều bước), YOLO có thể dự đoán vị trí và loại đối tượng trong một lần quét ảnh, giúp tăng tốc độ đáng kể.

A diagram of a dog and a bicycle

AI-generated content may be incorrect.

Hình – Minh họa về cách YOLO hoạt động

#### Cấu trúc của YOLO

* **Input:** một ảnh đầu vào có kích thước cố định.
* **Feature Extraction:** YOLO sử dụng CNN để trích xuất đặc trưng từ ảnh.
* **Phân chia ảnh thành lưới:** YOLO chia ảnh thành SxS ô lưới (ví dụ: 7x7).
* **Dự đoán bounding box:**
  + Mỗi ô lưới dự đoán B bounding box (khung giới hạn) kèm theo confidence score.
  + Mỗi bounding box gồm: tọa độ (x, y, w, h) + độ tin cậy + nhãn dự đoán.
* Sử dụng Non-Maximum Suppression (NMS) để loại bỏ các bounding box chồng lấn.

#### Công thức tính độ tin cậy

Công thức này dùng để đánh giá mức độ tin cậy của hộp giới hạn (bounding box) dự đoán.

Trong đó:

* **PObject** là xác xuất có đối tượng trong một ô lưới.
  + Nếu ô lưới không có đối tượng, xác xuất sẽ gần bằng 0
  + Nếu có đối tượng, xác xuất sẽ gần bằng 1
* **IOU (Intersection over Union)** là độ trùng khớp giữa hộp dự đoán và hộp thực tế (ground truth).
  + Công thức tính IOU
    - Nếu hộp dự đoán và hộp thực tế hoàn toàn khớp nhau → IOU = 1
    - Nếu hộp dự đoán lệch nhiều so với hộp thực tế → IOU ≈ 0.

#### Hàm mất mát (Loss function)

Hàm mất mát (Loss Function) trong YOLO giúp tối ưu hóa mô hình bằng cách điều chỉnh các dự đoán về vị trí, kích thước hộp, và lớp đối tượng sao cho sát với thực tế nhất. Nó gồm nhiều thành phần để xử lý các lỗi khác nhau, giúp cải thiện độ chính xác khi nhận diện đối tượng.

Cấu trúc gồm 3 phần chính:

* **Loss định vị (Localization Loss):** Đánh giá độ chính xác của hộp giới hạn (bounding box) so với thực tế, dựa trên tọa độ trung tâm, chiều rộng và chiều cao.
* **Loss tin cậy (Confidence Loss):** Xác định độ tin cậy của hộp dự đoán, giúp mô hình phân biệt giữa vùng có đối tượng và vùng nền.
* **Loss phân loại (Classification Loss):** Đánh giá xem mô hình có dự đoán đúng loại đối tượng hay không trong trường hợp có nhiều lớp (multi-class classification).

Công thức tổng quát:

# PHÂN TÍCH BÀI TOÁN

## Tiêu đề mục

### Tiêu đề tiểu mục

# KẾT QUẢ CÀI ĐẶT, THỬ NGHIỆM HOẶC KẾT QUẢ NGHIÊN CỨU

## Tiêu đề mục

### Tiêu đề tiểu mục

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

### Kết luận chung

TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | A. W. S. (AWS), "AI là gì? – Giải thích về trí tuệ nhân tạo," Amazon Web Services (AWS), [Online]. Available: https://aws.amazon.com/vi/what-is/artificial-intelligence/. |
| [2] | A. W. S. (AWS), "Công nghệ Máy học là gì?," Amazon Web Services, [Online]. Available: https://aws.amazon.com/vi/what-is/machine-learning/. |

# PHỤ LỤC

Các mã nguồn, các bảng biểu lớn, các phụ lục cho vào mục này.