**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA CÔNG NGHỆ THÔNG TIN**

----------



**BÁO CÁO ĐỒ ÁN TỐT NGHIỆP**

***Chủ đề: Nghiên cứu công nghệ Intelligent Character Recognition áp dụng vào ứng dụng chuyển tài liệu***

***viết tay thành văn bản điện tử***

|  |  |  |
| --- | --- | --- |
|  | **Giảng viên hướng dẫn:** | **Nguyễn Mạnh Hùng** |
|  | **Sinh viên thực hiện:** | **Phạm Trọng Biên** |
|  | **Mã sinh viên:** | **B20DCCN091** |

**HÀ NỘI, 2024**

**LỜI MỞ ĐẦU**

Nhận dạng ký tự viết tay luôn là một lĩnh vực nghiên cứu hấp dẫn trong lĩnh vực nhận dạng mẫu trong xử lý ảnh. Nó được áp dụng rộng rãi trong nhiểu lĩnh vực đặc biệt là trong việc lưu trữ và quản lý thông tin, khi mà trong thời đại hiện nay vấn đề này vô cùng quan trọng và cần thiết. Trong nhiều lĩnh vực từ giáo dục, y tế đến các doanh nghiệp, các tài liệu viết tay vẫn còn tồn tại phổ biến Những tài liệu này thường chứa đựng thông tin quan trọng, nhưng việc chuyển đổi chúng sang dạng văn bản điện tử để dễ dàng lưu trữ và tra cứu vẫn là một thách thức lớn. Việc này đòi hỏi nhiều công sức và thời gian nếu thực hiện thủ công, dễ dẫn đến sai sót và thiếu hiệu quả.

Phương pháp chuyển đổi tài liệu viết tay thành văn bản điện tử đã được áp dụng rộng rãi để giải quyết vấn đề trên. Trong báo cáo này, sẽ tập trung vào việc ứng dụng công nghệ Intelligent Character Recognition(ICR)

**MỤC LỤC**

**[Chương I: Bài toán nhận dạng chữ viết tay 4](#_Toc7278)**

[1. Giới thiệu 4](#_Toc14010)

[2. Tầm quan trọng 4](#_Toc20191)

[3. Lịch sử phát triển 4](#_Toc6980)

[4. Ứng dụng 6](#_Toc22988)

[5. Phương pháp và kỹ thuật 7](#_Toc19579)

[5.1 CapsNets 7](#_Toc23875)

[5.2 Multidimensional Recurrent Neural Networks (MDRNNs) 8](#_Toc21854)

[5.3 Connectionist Temporal Classification (CTC) 9](#_Toc32487)

[5.4 Transformer models 10](#_Toc22835)

[5.5 Encoder - Decoder and Attention Networks 11](#_Toc24649)

[5.6 Handwriting Text Generation 14](#_Toc32215)

[6. Thách thức và giải pháp 15](#_Toc32416)

**[Chương II: Công nghệ Optical Character Recognition(OCR) và Intelligent Character Recognition(ICR) 17](#_Toc13689)**

[1. Giới thiệu 17](#_Toc25725)

[2. Nguồn gốc 17](#_Toc7990)

[2.1 Những nỗ lực đầu tiên 17](#_Toc4200)

[2.2 Sự khởi đầu của OCR 17](#_Toc11948)

[2.3 OCR thế hệ đầu tiên 18](#_Toc15493)

[2.4 OCR thế hệ thứ hai 18](#_Toc24194)

[2.5 OCR thế hệ thứ ba 19](#_Toc17826)

[2.6 OCR ngày nay 19](#_Toc23019)

[2.7 Sự ra đời của ICR 20](#_Toc27973)

[3. Sự khác biệt của ICR so với OCR 20](#_Toc31500)

[4. Cơ chế hoạt động 20](#_Toc4325)

[4.1 Quét quang học(Optical scanning) 21](#_Toc3519)

[4.2 Xác định vị trí và phân đoạn 23](#_Toc10288)

[4.3 Tiền xử lý 24](#_Toc335)

[4.4 Trích xuất đặc trưng 25](#_Toc10834)

[4.5 Phân loại 29](#_Toc27440)

[4.6 Hậu xử lý 31](#_Toc22126)

[5. Đánh giá công nghệ ICR 32](#_Toc10610)

[5.1 Ưu điểm 32](#_Toc23940)

[5.2 Nhược điểm 33](#_Toc11987)

**[Chương III: Áp dụng công nghệ ICR để xây dựng ứng dụng chuyển tài liệu viết tay thành văn bản điện tử 34](#_Toc1286)**

[1. Giới thiệu chung về ứng dụng 34](#_Toc11256)

[2. Chức năng các thành phần 34](#_Toc25262)

[3. Kết quả cài đặt và thử nghiệm 34](#_Toc859)

[4. Đánh giá ứng dụng 34](#_Toc15510)

**[TÀI LIỆU THAM KHẢO 35](#_Toc16905)**

**Chương I: Bài toán nhận dạng chữ viết tay**

1. **Giới thiệu**

Nhận dạng chữ viết tay(Handwriting Recognition - HWR) hay còn được gọi là nhận dạng văn bản viết tay(Handwritten Text Recognition - HTR) là khả năng máy tính nhận và giải thích dữ liệu chữ viết tay từ các nguồn tài liệu như giấy, ảnh, màn hình cảm ứng và các thiết bị khác. Nói 1 cách đơn giản, đó là một quy trình tự động sử dụng kỹ thuật nhận dạng mẫu và học máy để nhận dạng các ký tự hoặc chữ viết.[1]

Ở mức khải niệm, kỹ thuật nhận dạng chữ viết tay được chia thành hai phương pháp chính. Nhận dạng chữ viết trực tiếp dựa trên thông tin động(Online handwriting recognition): nhận dạng ký tự hoặc chữ viết dựa trên thông tin thu được trong thời gian thực ngay lúc người dùng thực hiện hành động viết, những thông tin đó là tốc độ viết, áp lực khi viết và hướng viết. Phương pháp thứ hai là nhận dạng chữ viết gián tiếp dựa trên thông tin tĩnh(offline handwriting recognition): chương trình sẽ thông dịch các kí tự, các chữ hay các đoạn văn được viết trên các mẫu giấy hoặc các bề mặt khác mà chúng ta có thể thu thập thông tin về chúng thông qua hình ảnh thu được từ các bề mặt bằng cách chụp lại hình ảnh.[2] Trong báo cáo này khi đề cập đến nhận dạng chữ viết tay sẽ tập trung chủ yếu vào phương pháp thứ hai.

**2. Tầm quan trọng**

Chữ viết tay vẫn tiếp tục tồn tại như một phương tiện giao tiếp và ghi lại thông tin trong cuộc sống hàng ngay ngay cả khi có sự ra đời của các công nghệ mới. Do tính phổ biến của nó trong các giao dịch của con người, việc nhận dạng chữ viết tay bằng máy tính có ý nghĩa rất thực tế như trong việc đọc các ghi chú viết tay dưới dạng biểu mẫu, địa chỉ ghi trên phong bì thư, số tiền ghi trong séc ngân hàng,.. Nó không chỉ giúp tiết kiệm thời gian và công sức mà còn mang lại sự tiện lợi và linh hoạt trong quản lý thông tin cá nhân và công việc văn phòng.

**3. Lịch sử phát triển**

Trong quá trình phát triển của công nghệ, nhận dạng chữ viết tay đã trải qua một hành trình dài từ những nỗ lực ban đầu đến các tiến bộ hiện đại. Ban đầu, các nhà nghiên cứu tập trung vào việc xây dựng các hệ thống cơ bản dựa trên quy tắc và các phương pháp thống kê đơn giản, cùng với việc sử dụng các kỹ thuật tiền định dạng để chuẩn bị hình ảnh và giảm nhiễu trước khi quá trình nhận dạng chính thức bắt đầu. Sự ra đời của hệ thống nhận diện ký tự quang học (Optical Character Recognition - OCR) đã đóng vai trò quan trọng trong việc nhận dạng chữ viết tay, giúp tự động hóa và tăng tốc quá trình nhận dạng. Trích từ [3]:

Hệ thống nhận dạng ký tự quang học đã được nghiên cứu gần đây nhiều thập niên. Năm 1914, Emanuel Goldberg đã phát triển một hệ thống đọc ký tự và chữ số viết tay rồi chuyển đổi thành mã điện báo. Đồng thời Edmund Fournier d'Albe đã phát triển Optophon, một máy quét cầm tay quét trang in và tạo ra đầu ra. Goldberg tiếp tục phát triển hệ thống nhận dạng chữ viết tay để nhập dữ liệu. Sau một thời gian, anh ấy đề xuất ghép các hình ảnh với các mẫu có chứa thông tin nhận dạng mong muốn. Kỹ thuật này được gọi là phương pháp khớp mẫu.

Sau đó, Paul W.Handel cũng đề xuất bằng sáng chế của Hoa Kỳ về công nghệ chữ viết phù hợp với mẫu ở Hoa Kỳ vào năm 1933. Năm 1994, các kỹ sư của RCA đã đề xuất nhận dạng ký tự quang học kiểu máy tính nguyên thủy đầu tiên để giúp đỡ người mù. Nó được thiết kế để chuyển báo cáo viết tay thành thẻ đục lỗ để nhập vào máy tính nhằm hỗ trợ xử lý lô hàng 20-25 triệu cuốn sách trong một năm. Năm 1965, Reader’s Digest và RCA hợp tác xây dựng hệ thống nhận dạng ký tự quang học. Năm 1985, các phương pháp tiếp cận mang tính cấu trúc đã được đề xuất cùng với các phương pháp thống kê. Trong hệ thống này, các ký tự được chia thành tập hợp các mẫu như đường ngang, đường dọc và các đường cong khác nhau. Trong hệ thống phương pháp này tập trung vào hình dạng của các ký tự.

Sau năm 1990, tiến bộ thực sự đã đạt được bằng cách sử dụng các kỹ thuật và phương pháp mới trong xử lý hình ảnh và nhận dạng mẫu. Trong thế giới ngày nay, nhiều máy tính mạnh hơn và các thiết bị chính xác hơn như bút điện tử, máy quét và máy tính bảng được sử dụng. Nhiều phương pháp như HMM, mạng nơ-ron, thuật toán lan truyền ngược, mạng nơ-ron mờ đang được sử dụng để nhận dạng văn bản viết tay.

1. **Ứng dụng**

Công nghệ nhận dạng chữ viết tay đã đem lại nhiều ứng dụng quan trọng và tiện ích trong nhiều lĩnh việc khác nhau. Một trong những ứng dụng quan trọng nhất của công nghệ này là trong việc chuyển đổi và quản lý thông tin từ dạng viết tay sang dạng điện tử. Theo [4], một số ứng dụng điển hình bao gồm:

+ *Số hóa tài liệu:* Các doanh nghiệp và tổ chức thường có kho lưu trữ rộng rãi các tài liệu viết tay, thứ cần được chuyển đổi sang văn bản điện tử để có thể dễ dàng lưu trữ, truy xuất và phân tích. Hệ thống nhận dạng chữ viết tay có thể chuyển đổi chính xác những tài liệu viết tay này thành văn bản có thể đọc được bằng máy, tiết kiệm thời gian và công sức.

+ *Bảo quản tài liệu lịch sử:* Trong lĩnh vực di sản văn hóa, nhận dạng chữ viết tay đóng vai trò quan trọng trong việc bảo tồn các tài liệu lịch sử. Nhiều bản thảo và văn bản cổ được viết bằng tay nên rất khó để tiếp cận, đọc và nghiên cứu. Công nghệ nhận diện chữ viết tay cho phép các nhà sử học và nhà nghiên cứu ghi lại và phân tích các tài liệu này hiệu quả hơn, góp phần bảo tồn di sản văn hóa.

+ *Nhập dữ liệu và xử lý biểu mẫu:* Các doanh nghiệp xử lý các biểu mẫu viết tay và khảo sát được hưởng lợi từ các giải pháp nhập dữ liệu dựa trên công nghệ nhận dạng chữ viết tay. Hệ thống nhận dạng chữ viết tay có thể tự động trích xuất dữ liueej từ các mẫu viết tay, giảm lỗi nhập dữ liệu thủ công và cải thiện hiệu quả quy trình làm việc. Ứng dụng này đặc biệt có giá trị trong ngành chăm sóc sức khỏe, tài chính và dịch vụ khách hàng.

+ *Tổ chức ghi chú cá nhân:* Trong thời đại kỹ thuật số, các cá nhân thường xuyên sử dụng sổ ghi chép kỹ thuật số, nhưng một số người vẫn thích ghi chú viết tay hơn. Các ứng dụng hỗ trợ có thể chuyển đổi các ghi chú viết tay thành văn bản điện tử để dễ dàng tìm kiếm và chỉnh sửa, mang đến cho người dùng những điều tốt nhất của cả 2 cách ghi chú. Ứng dụng này phổ biến trong giới sinh viên, chuyên gia và nhà sáng tạo.

+ *Sắp xếp thư viết tay:* Các công ty dịch vụ bưu chính và cung ứng sử dụng công nghệ nhận diện chữ viết tay để tự động hóa việc phân loại địa chỉ viết tay trên phong bì và gói hàng. Điều này không chỉ đẩy nhanh quá trình gửi thư mà còn giảm sai sót trong việc định tuyến thư.

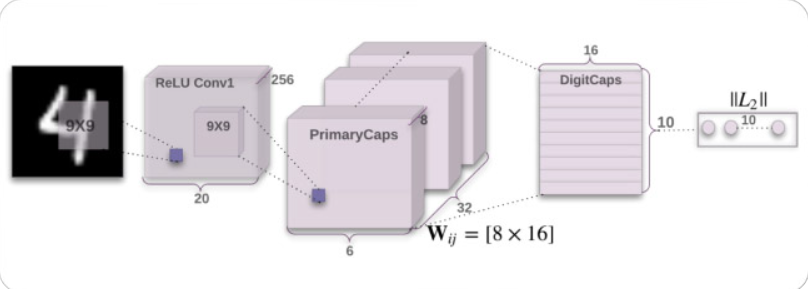
1. **Phương pháp và kỹ thuật**

Tham khảo từ [5]:

* 1. **CapsNets**

Capsule Networks(CapsNets) là một kiến trúc mạng nơ-ron mới mẻ và tiên tiến, được thiết kể để cải thiện và vượt qua những hạn chế của mạng nơ-ron tích chập(CNNs), đặc biệt là trong việc nhận dạng hình ảnh và văn bản. Một trong những điểm yếu chính của CNNs là lớp pooling, lớp này được sử dụng để giảm kích thước dữ liệu nhưng nó dẫn đến mất mát thông tin không gian và vị trí của đối tượng trong hình ảnh. Điều này gây ra sự không chính xác trong việc xác định vị trí chính xác của đối tượng, đặc biệt khi đối tượng bị thay đổi góc quay, vị trí hoặc tỷ lệ.

CapsNets giải quyết vấn đề này bằng cách sử dụng các “capsule”, đơn vị xử lý đa chiều để biểu diễn đặc trưng của hình ảnh. Mỗi capsule lưu trữ thông tin về vị trí, góc quay, tỉ lệ và các thuộc tính khác của đối tượng trong không gian vector đa chiều. Quan trong hơn định tuyến động(dynamic routing) giữa các capsule cho phép mạng tự động học cách kết hợp thông tin từ các capsule để tạo ra biểu diễn không gian phức tạp và đồng nhất với các biến đổi vị trị của đối tượng

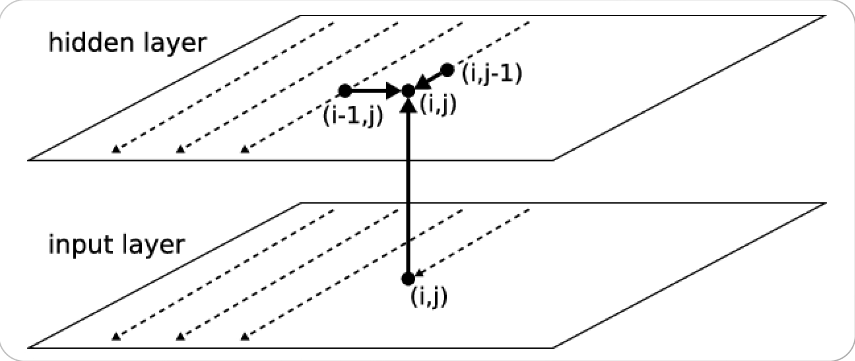


Hình 1: Định tuyến động giữa các capsules[5]

Mặc dù CapsNets cung cấp sự cải thiện đáng kể so với CNNs trong việc nhận dạng và xác định vị trí của đối tượng trong hình ảnh, nhưng chúng vẫn đối mặt với một số nhược điểm, bao gồm tính toán phức tạp và yêu cầu dữ liệu huấn luyện lớn hơn. Tuy nhiên, với khả năng giảm lượng dữ liệu huấn luyện cần thiết so với CNNs, CapsNets có thể là lựa chọn hứa hẹn trong các ứng dụng OCR và nhận dạng hình ảnh khác.

* 1. **Multidimensional Recurrent Neural Networks (MDRNNs)**

Multidimensional Recurrent Neural Networks (MDRNNs) là một phương pháp mới trong học sâu, được thiết kế để xử lý dữ liệu có chiều đa dạng, như hình ảnh, thay vì chỉ có thể áp dụng cho dữ liệu tuần tự như văn bản như RNN/LSTM truyền thống. Đặc điểm chính của MDRNNs là khả năng xử lý dữ liệu đa chiều bằng cách thay thế các kết nối tuần tự đơn trong RNN tiêu chuẩn bằng nhiều đơn vị tuần tự tương ứng với số chiều trong dữ liệu



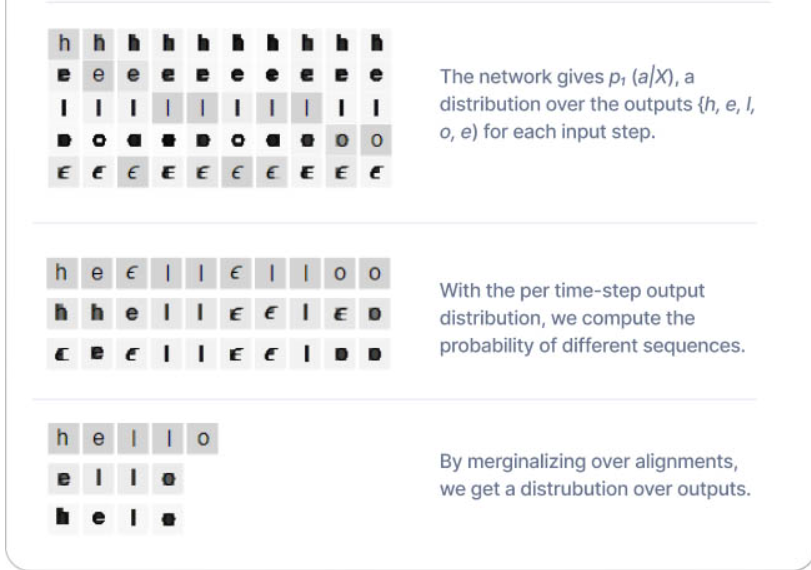
Hình 2: MDRNN 2 chiều[5]

Đặc điểm chính của MDRNNs là khả năng xử lý dữ liệu đa chiều bằng cách thay thế các kết nối tuần tự đơn trong RNN tiêu chuẩn bằng nhiều đơn vị tuần tự tương ứng với số chiều trong dữ liệu. Trong quá trình lan truyền, mỗi điểm trong chuỗi dữ liệu, lớp ẩn của mạng nhận cả một đầu vào bên ngoài và các kích hoạt của chính nó từ một bước trở lại theo tất cả các chiều của dữ liệu. MDRNNs giải quyết vấn đề chính trong hệ thống nhận dạng bằng cách chuyển đổi hình ảnh hai chiều thành các chuỗi nhãn một chiều. Điều này được thực hiện bằng cách đưa dữ liệu đầu vào qua một loạt các lớp MDRNN, với các khối hàm kích hoạt ở giữa sau mỗi lớp RNN. Các lớp này được thiết kế để dần dần thu gọn hình ảnh hai chiều thành các chuỗi một chiều, sau đó lớp đầu ra có thể gán nhãn.

Ưu điểm của MDRNNs là làm cho mô hình ngôn ngữ chịu được các biến dạng cục bộ trên mọi kết hợp của các chiều đầu vào (như xoay và nghiêng hình ảnh, sự mơ hồ của nét vẽ và các phong cách viết khác nhau) và cho phép mô hình mô phỏng một ngữ cảnh đa chiều linh hoạt. Tuy nhiên, nhược điểm của MDRNNs có thể bao gồm độ phức tạp tính toán cao và yêu cầu lượng dữ liệu huấn luyện lớn để đạt được hiệu suất tốt.

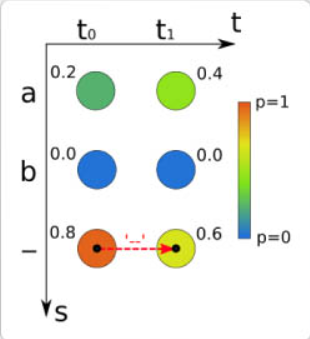
* 1. **Connectionist Temporal Classification (CTC)**

Connectionist Temporal Classification (CTC) là một thuật toán xử lý các tác vụ như nhận dạng giọng nói, nhận dạng chữ viết,… trong đó toàn bộ dữ liệu đầu vào được ánh xạ tới lớp đầu ra. Nhận dạng văn bản viết tay liên quan đến việc ánh xạ hình ảnh tới văn bản tương ứng. CTC giải quyết vấn đề của việc không biết cách các phần của hình ảnh sẽ được căn chỉnh với các ký tự cụ thể nào trong quá trình nhận dạng.



Hình 3: Mô hình hóa trình tự với CTC[5]

Không cần sự căn chỉnh trực tiếp, CTC tìm kiếm sự ánh xạ bằng cách tổng hợp xác suất của tất cả các căn chỉnh có thể giữa dữ liệu đầu vào và chuỗi ký tự đầu ra. Các mô hình được đào tạo bằng CTC thường sử dụng mạng RNN để ước tính xác suất theo từng bước thời gian khi RNN tính đến ngữ cảnh trong đầu vào. Nó xuất ra điểm số ký tự cho từng thành phần chuỗi, được biểu thị bằng ma trận.



Hình 4: Chiến lược giải mã đường dẫn tốt nhất [5]

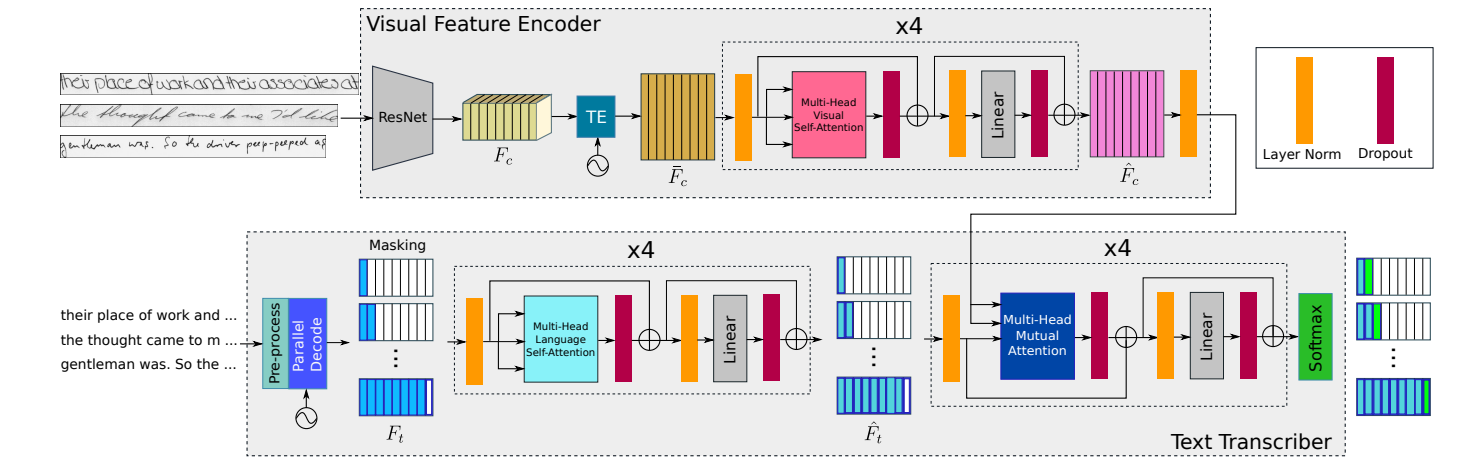
Để giải mã, có thể sử dụng:

+ Giải mã theo đường đi tốt nhất: Dự đoán từ bằng cách ghép ký tự có xác suất cao nhất tại mỗi thời điểm để tạo thành từ hoàn chỉnh. Trong vòng lặp tiếp theo của quá trình huấn luyện, các ký tự trùng lặp và khoảng trắng được loại bỏ để giải mã văn bản tốt hơn.

+ Giải mã bằng cách tìm kiếm theo đám mây: Đề xuất nhiều đường đi đầu ra với xác suất cao nhất. Các đường đi có xác suất thấp hơn sẽ bị loại bỏ để duy trì kích thước đám mây cố định. Kết quả thu được thông qua phương pháp này thường chính xác hơn và thường được kết hợp với các mô hình ngôn ngữ để đưa ra kết quả ý nghĩa.

* 1. **Transformer models**

Transformer models là một phương pháp khác biệt so với mạng nơ-ron tuần tự (RNNs) trong xử lý dữ liệu tuần tự như văn bản. Trong khi RNNs xuất sắc trong việc bắt các phụ thuộc thời gian, nhưng chúng mắc phải vấn đề huấn luyện chậm do xử lý tuần tự và hạn chế bộ nhớ. Ngược lại, các mô hình transformer sử dụng cơ chế tự chú ý để xử lý toàn bộ chuỗi cùng một lúc. Đối với việc nhận dạng chữ viết tay, mô hình transformer cung cấp một phương pháp không tuần tự. Bằng cách tích hợp các lớp tự chú ý đa đầu trong cả hai miền hình ảnh và văn bản, chúng có thể học các phụ thuộc trong các chuỗi ký tự để giải mã.



Hình 5: Tổng quan về kiến trúc của mô hình[5]

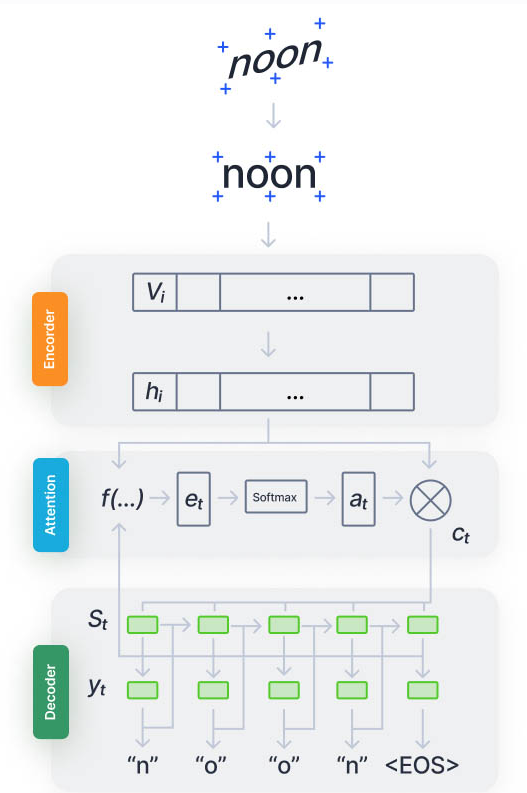
Kiến trúc chia ra làm 2 phần:

+ Text transcriber: dành cho đầu ra của các ký tự được giải mã bằng cách cùng tham gia vào các tính năng liên quan đến hình ảnh và ngôn ngữ.

+ Visual feature encoder: nhằm mục đích trích xuất thông tin liên quan từ hình ảnh văn bản viết tay bằng cách tập trung vào vị trí ký tự khác nhau và thông tin ngữ cảnh chung của chúng.

Ưu điểm chính của mô hình transformer nằm ở khả năng nhúng kiến thức ngôn ngữ trực tiếp vào mô hình, loại bỏ nhu cầu phải sử dụng mô hình ngôn ngữ phụ thuộc sau khi xử lý. Chúng cũng có khả năng dự đoán các đầu ra nằm ngoài từ vựng. Tuy nhiên, mô hình transformer cũng có nhược điểm, bao gồm sự phức tạp trong việc hiểu và triển khai, đặc biệt là khi xử lý các tập dữ liệu lớn và trong việc điều chỉnh siêu tham số để đạt được hiệu suất tốt nhất.

* 1. **Encoder - Decoder and Attention Networks**



Hình 6: Kiến trúc của mô hình mã hóa - giải mã kết hợp với mạng chú ý[5]

Các hệ thống nhận dạng chữ viết trong huần luyện luôn gặp phải tình trạng khan hiếm dữ liệu huấn luyện vì không thể tạo ra một bộ có tất cả các tổ hợp ngôn ngữ, kiểu nét,… Để giải quyết vấn đề, phương pháp này tận dụng các vectơ đặc trưng được đào tạo trước của văn bản làm điểm bắt đầu. Các mô hình hiện đại gợi ý sử dụng cơ chế chú ý kết hợp với RNN để tập trung vào các đặc điểm hữu ích tại mỗi dấu thời gian. Kiến trúc mô hình hoản chỉnh có thể được chia thành 4 giai đoạn:

+ ***Chuyển đổi****:* Sử dụng mạng CNN để định vị, học tọa độ của các điểm dấu hiệu để ghi lại hình dạng của văn bản, và chuẩn hóa hình ảnh từ các từ viết tay bằng cách áp dụng các biến đổi.

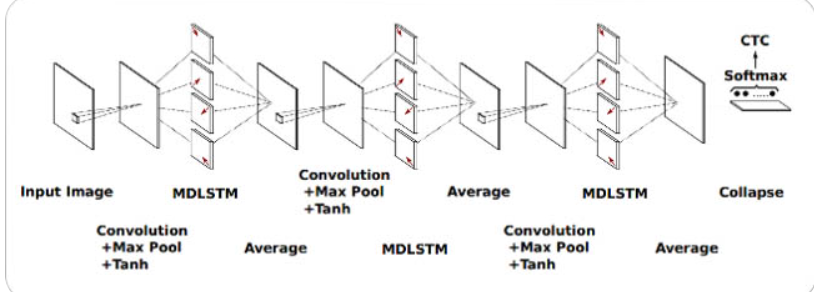
+ ***Trích xuất đặc trưng***: Sử dụng một kiến trúc tương tự ResNet để mã hóa hình ảnh đầu vào thành một bản đồ đặc trưng hình ảnh 2D.

+ ***Mô hình hóa chuỗi***: Sử dụng các đặc trưng trích xuất từ bước trước để giải mã bằng một LSTM hai chiều để giữ lại thông tin ngữ cảnh trong một chuỗi từ cả hai phía và nhận diện từng ký tự một cách độc lập.

+***Dự đoán***: Chuyển các vector đầu ra từ bộ giải mã cuối cùng thành từ, và sau đó áp dụng hàm softmax để nhận được điểm xác suất cho mỗi từ trong từ vựng.

**5.5.1 Scan, Attend and Read**

Scan, Attend and Read là một phương pháp đề xuất cho việc nhận dạng chữ viết tay từ đầu đến cuối trang bằng cách sử dụng cơ chế chú ý. Phương pháp này không phụ thuộc vào việc phân đoạn trước một từ hoặc một dòng. Nó sử dụng một kiến trúc LSTM đa chiều (MDLSTM) như một bộ trích xuất đặc trưng tương tự như đã mô tả ở trên, với một sự khác biệt duy nhất ở lớp cuối cùng, nơi các bản đồ đặc trưng được rút gọn theo chiều dọc và áp dụng hàm kích hoạt softmax để nhận dạng văn bản tương ứng.



Hình 7: Thuật toán Scan As You Read để nhận dạng chữ viết tay, xen kẽ các lớp LSTM theo kiểu đa chiều[5]

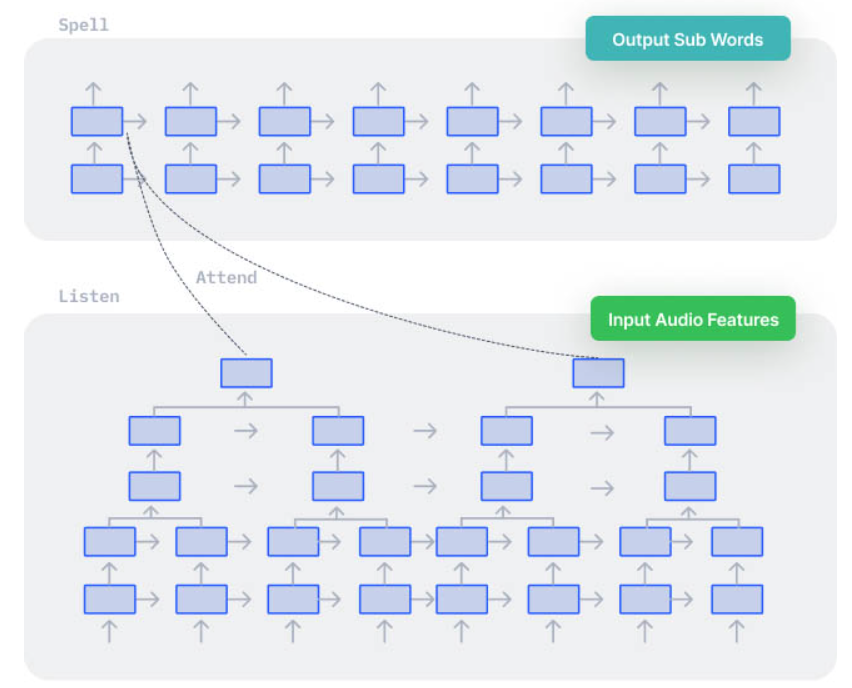
**5.5.2 Convolve, Attend and Spell**

Nhận dạng chữ viết tay được kết nối với nhận dạng mẫu(pattern recognition) theo nhiều cách. Mạng lưới thần kinh tuần tự được hỗ trợ bằng cơ chế chú ý có thể trở thành một kỹ thuật tiên tiến để nhận dạng chữ viết tay. Convolve, Attend and Spell là mô hình theo trình tự để nhận dạng từ viết tay dựa trên cơ chế chú ý. Kiến trúc có ba phần chính:

+ Một bộ mã hóa, bao gồm CNN và GRU hai chiều

+ Một cơ chế chú ý tập trung vào các đặc điểm quan trọng

+ Một bộ giải mã được hình thành bởi GRU một chiều, có thể đánh vần từ tương ứng từng ký tự một.



Hình 8: Hoạt động của mô hình Listen, Attend and Spell để nhận dạng chữ viết[5]

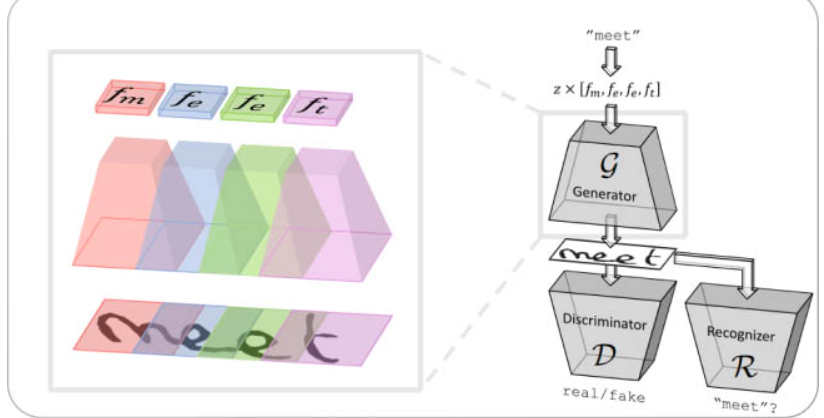
Bộ mã hóa bao gồm một mạng nơ-ron tích chập (CNN) để trích xuất các đặc trưng hình ảnh từ văn bản viết, được mã hóa tuần tự bởi một mạng nơ-ron tái lặp (RNN). Bộ giải mã là một RNN khác nhau giải mã từng ký tự một, từ đó xây dựng từ hoàn chỉnh và viết ra. Một cơ chế chú ý kết nối bộ mã hóa và bộ giải mã để cung cấp một vector ngữ cảnh cao tương quan tập trung vào các đặc điểm của mỗi ký tự tại mỗi bước giải mã.

* 1. **Handwriting Text Generation**

Tạo chữ viết tay tổng hợp là nhiệm vụ tạo ra văn bản viết tay trông như thật. Nó có thể được sử dụng để tăng cường các bộ dữ liệu hiện có. Các mô hình học sâu đòi hỏi rất nhiều dữ liệu để huấn luyện và thu thập được một lượng lớn hình ảnh chữ viết tay có chú thích cho các ngôn ngữ khác nhau là một nhiệm vụ khó khăn. Chúng ta có thể sử dụng Generative Adversarial Networks để tạo dữ liệu đào tạo nhằm giải quyết vấn để này.

**ScrabbleGAN**

Nhận dạng văn bản viết tay có phạm vi hạn chế trong dữ liệu đào tạo vì mỗi người có một phong cách viết riêng. Việc thu thập một bộ dữ liệu đa dạng là rất tốn kém và việc chú thích văn bản thậm chí còn khó khăn hơn. Để giảm thiểu nhu cầu thu thập dữ liệu và chú thích dữ liệu viết tay, học bán giám sát là một lựa chọn phù hợp. Nó sử dụng kết hợp các mẫu dữ liệu được gán nhãn và không gán nhãn để cải thiện hiệu suất của các mô hình. So với các mô hình được giám sát hoàn toàn, nó học cách xác định các đặc điểm tốt hơn và thích ứng với hình ảnh không nhìn thấy tốt hơn.



Hình 9: Generative Adversial networks cho việc tạo dữ liệu văn bản[5]

ScrabbleGAN là một phương pháp bán giám sát để tổng hợp hình ảnh văn bản viết tay. Nó dựa trên một mô hình tổng quát có thể tạo ra hình ảnh của các từ có độ dài tùy ý bằng cách sử dụng mạng tích chập hoàn toàn. Hơn nữa, trình tạo đủ thông minh để thao tác kiểu và nét văn bản thu được. Ngoài bộ phân biệt D, hình ảnh thu được cũng được đánh giá bởi mạng nhận dạng văn bản R. Trong khi D khuyến khích các kiểu viết tay trông thực tế, R khuyến khích kết quả có thể đọc được và đúng với văn bản đầu vào.

1. **Thách thức và giải pháp**

Bất chấp những tiến bộ đáng kể trong việc cải thiện công nghệ nhận dạng chữ viết tay những năm gần đây, vẫn còn một số thách thức cần phải vượt qua để cải thiện độ chính xác và hiệu quả của nó. Dưới đây là một số thách thức và vai trò của học máy trong việc giải quyết chúng[6]:

+ Một trong những thách thức chính của việc nhận dạng chữ viết tay là sự đa dạng của kiểu chữ viết tay. Không giống như văn bản in tuân theo phông chữ chuẩn, chữ viết tay có thể thay đổi đáng kể tùy thuộc vào độ tuổi, trình độ học vấn và thậm chí cả tâm trạng của người viết. Sự thay đổi này có thể gây khó khăn cho việc nhận dạng và diễn giải chính xác văn bản viết tay, đặc biệt khi xử lý chữ viết ẩu hoặc chữ viết cách điệu. Tuy nhiên, các thuật toán học máy có thể được đào tạo trên các tập dữ liệu lớn về mẫu chữ viết tay để nhận dạng và thích ứng tốt hơn với các kiểu chữ viết tay khác nhau.

+ Một thách thức khác là sự hiện diện của nhiễu và biến dạng trong văn bản viết tay. Các vết ố, nếp nhăn và vết rách đều có thể ảnh hưởng đến mức độ dễ đọc của văn bản viết tay, gây khó khăn trong việc phân biệt giữa những biến dạng này và chữ viết tay thực tế để nhận dạng chính xác văn bản. Các thuật toán học máy có thể được đào tạo để xác định, lọc nhiễu và biến dạng, từ đó có thể cải thiện thêm độ chính xác.

+ Thách thức thứ ba là sự hạn chế của dữ liệu đào tạo. Không giống như văn bản in, có thể dễ dàng tạo và dãn nhán, các mẫu chữ viết tay khó lấy và dãn nhãn hơn nhiều. Điều này có thể gây khó khăn cho việc đào tạo các mô hình học máy để nhận dạng chữ viết tay, đặc biệt là khi xử lý các phong cách viết hiếm hoặc chuyên biệt. Tuy nhiên, những tiến bộ gần đây trong kỹ thuật tăng cường dữ liệu đã giúp tạo ra các mẫu chữ viết tay tổng hợp để bổ sung cho dữ liệu đào tạo và cải thiện độ chính xác.

Nhận dạng chữ viết tay là một nhiệm vụ quan trọng, và đã phát triển được một chặng đường dài trong những năm gần đây. Tuy nhiên những thách thức như sự thay đổi của kiểu chữ viết tay, sự hiện diện của nhiễu và biến dạng cũng như tính sẵn có hạn chế của dữ liệu đào tạo phải được giải quyết để cải thiện độ chính xác và hiệu quả. Với sự trợ giúp của các thuật toán học máy chẳng hạn như CNN được sử dụng để phân loại các ký tự riêng lẻ trong văn bản viết tay, RNN được sử dụng để nhận dạng toàn bộ từ và câu, … chúng ta có thể mong đợi những tiến bộ hơn nữa trong lĩnh vực nhận dạng chữ viết tay và các ứng dụng của nó.

**Chương II: Công nghệ Optical Character Recognition(OCR) và Intelligent Character Recognition(ICR)**

1. **Giới thiệu**
2. **Nguồn gốc**

Về mặt phương pháp, nhận dạng ký tự là một tập con của lĩnh vực nhận dạng mẫu. Tuy nhiên, chính việc nhận dạng ký tự đã tạo động lực lớn cho việc nhận dạng mẫu và phân tích hình ảnh trở thành những lĩnh vực khoa học lớn.Theo [7], công nghệ OCR có nguồn gốc như sau:

* 1. **Những nỗ lực đầu tiên**

Để tái tạo các chức năng của con người bằng máy móc, giúp máy có thể thực hiện các nhiệm vụ như đọc sách là 1 giấc mơ từ xa xưa. Nguồn gốc của nhận dạng ký tự thực sự có thể được tìm thấy vào năm 1870. Đây là năm mà C.R.Carey ở Boston Massachusetts đã phát minh ra máy quét võng mạc là một hệ thống truyền hình ảnh sử dụng khảm các tế bào quang điện. Hai tập kỷ sau, P.Nipkow người Ba Lan đã phát minh ra máy quét tuần tự, đây là một bước đột phá lớn cho cả đọc máy và tivi hiện đại. Trong những thập kỷ đầu tiên của thế kỷ 19, một số nỗ lực đã được thực hiện nhằm phát triển các thiết bị hỗ trợ người mù thông qua các thí nghiệm với OCR. Tuy nhiên, phiên bản hiện đại của OCR không xuất hiện cho đến giữa những năm 1940 cùng với sự phát triển của máy tính kỹ thuật số. Động lực phát triển từ đó trở đi là những ứng dụng khả thi trong thế giới kinh doanh.

* 1. **Sự khởi đầu của OCR**

Đến những năm 1950, cuộc cách mạng công nghệ đang tiến triển với tốc độ cao và xử lý dữ liệu điện tử đang trở thành một lĩnh vực quan trọng. Việc nhập dữ liệu được thực hiện thông qua thẻ đục lỗ và cần có cách xử lý hiệu quả về mặt chi phí khi lượng dữ liệu ngày càng tăng. Đồng thời, công nghệ đọc máy đã đủ trưởng thành để ứng dụng, và đến giữa những năm 1950 máy OCR đã được thương mại hóa. Máy đọc OCR thực sự đầu tiên được lắp đặt tại Reader’s Digest vào năm 1954. Thiết bị này được sử dụng để chuyển các báo cáo bán hàng được đánh từ máy đánh chữ thành thẻ đục lỗ để nhập vào máy tính.

* 1. **OCR thế hệ đầu tiên**

Các hệ thống OCR thương mại xuất hiện trong giai đoạn từ 1960 đến 1965 có thể được gọi là thế hệ đầu tiên của OCR. Thế hệ máy OCR này chủ yếu có đặc điểm là bị hạn chế hình dạng chữ cái có thể đọc. Các ký hiệu được thiết kế đặc biệt để máy đọc và những ký hiệu đầu tiên thậm chí trông không tự nhiên cho lắm. Theo thời gian, các máy đa phông chữ bắt đầu xuất hiện, có thể đọc tới mười phông chữ khác nhau. Số lượng phông chữ bị giới hạn bởi phương pháp nhận dạng mẫu được áp dụng, khớp mẫu, so sánh hình ảnh ký tự với thư viện hình ảnh nguyên mẫu cho từng ký tự của tứng phông.

* 1. **OCR thế hệ thứ hai**

Máy đọc thế hệ thứ hai xuất hiện vào giữa những năm 1960 và đầu những năm 1970. Các hệ thống này có thể nhận dạng các ký tự được in bằng máy và cũng có khả năng nhận dạng các ký tự được viết bằng tay. Khi xem xét các ký tự viết in(hand printed), bộ ký tự bị giới hạn ở các chữ số và một số chữ cái và ký hiệu. Hệ thống đầu tiên và nổi tiếng thuộc loại này là IBM 1287. được trưng bày tại triển lãm World Fair ở New York vào năm 1965. Ngoài ra, trong giai đoạn này Toshiba đã phát triển máy phân loại chữ cái tự động đầu tiên cho mã số bưu điện và Hitachi đã chế tạo máy OCR đầu tiên cho hiệu suất cao và chi phí thấp. Trong giai đoạn này, công việc quan trọng đã được thực hiện trong lĩnh vực tiêu chuẩn hóa. Năm 1966, một nghiên cứu kỹ lưỡng về các yêu cầu OCR đã được hoàn thành và bộ ký tự OCR tiêu chuẩn của Mỹ đã được xác định: OCR-A. Phông chữ này được thiết kế và cách điệu hóa cao để hỗ trợ nhận dạng quang học, mặc dù thế con người vẫn có thể đọc được. Một tiêu chuẩn chữ châu Âu cũng được thiết kế: OCR-B, có nhiều phông chữ tự nhiên hơn tiêu chuẩn Mỹ. Một số nỗ lực đã được thực hiện để hợp nhất hai tiêu chuẩn này thành một tiêu chuẩn, nhưng thay vào đó máy có thể đọc cả 2 tiêu chuẩn đã xuất hiện.



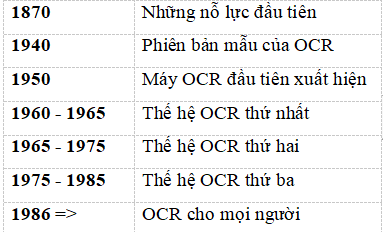
Hình 10: Tiêu chuẩn OCR-A(ở trên) và OCR-B(ở dưới)[7]

* 1. **OCR thế hệ thứ ba**

Đối với hệ thống OCR thế hệ thứ 3, xuất hiện vào giữa những năm 1970, thách thức ở thời điểm này là các tài liệu có chất lượng kém và bộ ký tự in và viết tay lớn. Chi phí thấp và hiệu suất cao cũng là những mục tiêu quan trọng, điều này được cải thiện nhờ những tiến bộ vượt bậc trong công nghệ phần cứng. Mặc dù các máy OCR phức tạp hơn đã bắt đầu xuất hiện trên thị trường nhưng các thiết bị OCR đơn giản vẫn rất hữu ích. Trong thời kỳ trước khi máy tính cá nhân và máy in laser bắt đầu thống trị lĩnh vực sản xuất văn bản, việc đánh máy là một lĩnh vực đặc biệt dành cho OCR. Khoảng cách in đồng đều và đa số lượng phông chữ nhỏ khiến các thiết bị OCR được thiết kế đơn giản trở nên rất hữu ích. Bản nháp thô có thể được tạo trên máy đánh chữ thông thường và đưa vào máy tính thông qua thiết bị OCR để chỉnh sửa lần cuối. Bằng cách này, bộ xử lý văn bản vốn là một nguồn tài nguyên đắt tiền vào thời điểm đó, có thể hỗ trợ nhiều người và chi phí cho thiết bị có thể được cắt giảm.

* 1. **OCR ngày nay**

Mặc dù máy OCR đã được thương mại hóa từ những năm 1950 nhưng chỉ có vài nghìm hệ thống được bán trên toàn thế giới cho đến năm 1986. Lý do chính cho điều này là chi phí của hệ thống. Tuy nhiên, khi phần cứng ngày càng rẻ hơn và hệ thống OCR bắt đầu có sẵn dưới dạng gói phần mềm, doanh số bàn hàng đã tăng lên đáng kể. Ngày nay, vài nghìn là số lượng hệ thống được bán ra mỗi tuần và chi phí cho một OCR đa phông chữ đã giảm với hệ số 10 mỗi năm trong 6 năm gần đây.

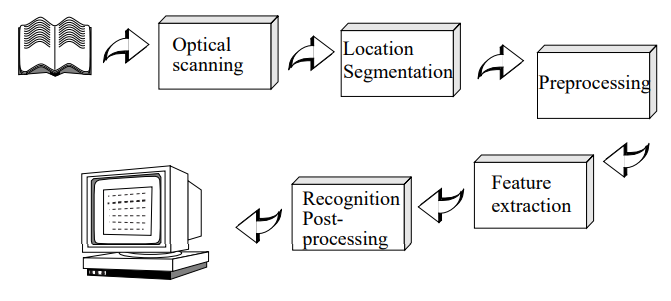


Hình 11: Tóm tắt niên đại của OCR[7]

**2.7 Sự ra đời của ICR**

1. **Sự khác biệt của ICR so với OCR**
2. **Cơ chế hoạt động**

Nguyên tắc chính trong nhận dạng mẫu tự động trước tiên là dạy cho biết loại mẫu nào có thể xuất hiện và trông chúng như thế nào. Trong OCR, các mẫu là câc chữ cái, số và một số ký hiệu đặc biệt như dấu phẩy, dấu chấm, … trong khi các lớp khác nhau tương ứng với các ký tự khác nhau. Việc dạy máy được thực hiện bằng cách hiển thị các mẫu máy về các ký tự của tất cả các lớp khác nhau. Dựa trên những mẫu này, máy sẽ xây dựng một nguyên mẫu hoặc mô tả từng loại ký tự. Sau đó trong quá trình nhận dạng, các ký tự chưa biết sẽ được so sánh với các mô tả thu được trước đó và được chỉ định lớp cho kết quả phù hợp nhất. Trong hầu hết các hệ thống nhận dạng ký tự thương mại, quá trình huần luyện đã được thực hiện trước. Tuy nhiên, một số hệ thống có bao gồm việc huấn luyện thêm các ký tự mới được phát hiện.[7]



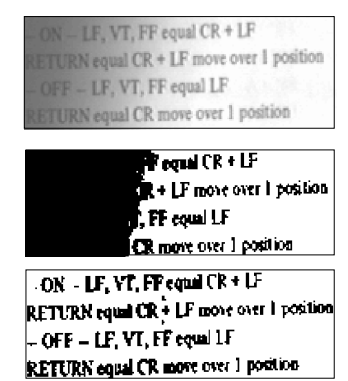
Hình 12: Các thành phần của hệ thống OCR[7]

Trong hình trên một thiết lập chung được minh họa. Bước đầu tiên trong quy trình là số hóa tài liệu tương tự bằng máy quét quang học. Khi xác định được các vùng chứa văn bản, mỗi ký hiệu sẽ được trích xuất thông qua một quá trình phần đoạn. Sau đó, các ký hiệu được trích xuất có thể được xử lý trước, loại bỏ nhiều, để tạo điều kiện thuận lợi cho việc trích xuất các đặc trưng ở bước tiếp theo.Việc nhận dạng từng ký hiệu được tìm thấy bằng cách so sánh các đặc điểm được trích xuất với mô tả về các lớp ký hiệu thu được qua giai đoạn học trước đó. Cuối cùng, thông tin theo ngữ cảnh được sử dụng để tái tạo lại các từ và số của văn bản gốc.[7]

Trích từ [7], các bước ở trên và một số phương pháp liên quan sẽ được mô tả chi tiết hơn như sau:

* 1. **Quét quang học(Optical scanning)**

Thông qua quá trình quét, hình ảnh kỹ thuật số của tài liệu gốc sẽ được ghi lại. Máy quét quang học OCR được sử dụng, thường bao gồm cơ chế vận chuyển cộng với các thiết bị cảm biến chuyển đổi cường độ ánh sáng thành mức xám. Tài liệu in thường bao gồm bản in màu đen trên nền trắng. Do đó, khi thực hiện OCR, thông thường là chuyển đổi hình ảnh đa mức thành hình ảnh hai mức trắng đen. Thông thường quá trình này, được gọi là phân ngưỡng, được thực hiện trên máy quét để tiết kiệm dung lượng toàn bộ nhớ và công sức tính toán. Quá trình phân ngưỡng rất quan trọng vì kết quả nhận dạng sau đây hoàn toàn phụ thuộc vào hình ảnh của hai mức. Tuy nhiên, việc xác định ngưỡng được thực hiện trên máy quét thường rất đơn giản. Một ngưỡng cố định được sử dụng, trong đó các mức xám dưỡi ngưỡng này được gọi là màu đen và các mức trên được gọi là màu trắng. Đối với tài liệu có độ tương phản cao trên với nền đồng nhất, ngưỡng cố định được chọn trước có thể là phù hợp. Tuy nhiên, rất nhiều tài liệu gặp trong thực tế có phạm vi tương phản khá lớn. Trong những trường hợp này, cần có những phương pháp xác định ngưỡng phức tạp hơn để đạt được kết quả tốt.



Hình 13: Vấn đề của phân ngưỡng[7]

Trên cùng: Hình ảnh gốc, giữa: hình ảnh được phân ngưỡng bằng phương pháp chung, dưới cùng: hình ảnh được phân ngưỡng bằng phương pháp thích ứng. Các phương pháp tốt nhất để xác định ngưỡng thường là những phương pháp có thể thay đổi ngưỡng trên tài liệu để thích ứng với các đặc tính cục bộ như độ tương phản và độ sáng. Tuy nhiên, các phương pháp OCR - Optical Character Recognition(nhận dạng ký tự quang học) thường phụ thuộc vào việc quét tài liệu nhiều cấp độ, đòi hỏi nhiều bộ nhớ và khả năng tính toán cao hơn. Do đó, những kỹ thuật như vậy hiếm khi được sử dụng cùng với hệ thống OCR, mặc dù chúng mang lại hình ảnh tốt hơn.

* 1. **Xác định vị trí và phân đoạn**

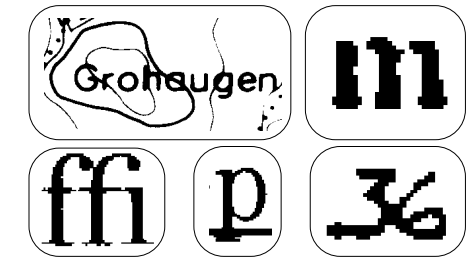
Phân đoạn là một quá trình xác định các thành phần của hình ảnh. Cần xác định vị trí các vùng của tài liệu nơi dữ liệu được in và phân việt chúng với các hình vẽ và đồ họa. Ví dụ khi thực hiện phân loại thư tự động, địa chỉ phải được định vị và tách biệt với các bản in khác trên phong bì như tem và logo công ty, trước khi được nhận dạng. Áp dụng cho văn bản, phân đoạn là sự cô lập các ký tự hoặc từ. Phần lớn các thuật toán nhận dạng ký tự quang học phân chia các từ thành các ký tự riêng biệt được nhận dạng riêng lẻ. Thông thường việc phân đoạn này được thực hiện bằng cách cô lập từng thành phần được kết nối, tức là từng vùng màu đen được kết nối. Kỹ thuật này dễ thực hiện nhưng sẽ xảy ra vấn đề nếu các ký tự chạm vào nhau hoặc nếu các ký tự bị phân mảnh và bao gồm nhiều phần. Các vấn để phân đoạn có thể được chia thành 4 nhóm chính:

+ Trính xuất các ký tự chạm và rời rạc: Những biến dạng như vậy có thể dẫn đến việc một số ký tự chung được hiểu là một số ký tự đơn lẻ hoặc một phần của ký tự được cho là toàn bộ ký hiệu. Sự ghép nối sẽ xảy ra nếu tài liệu là bản sao tối màu hoặc nếu nó được quét ở ngưỡng thấp. Ngoài ra, nếu các mối nối cũng phổ biến nếu phông chữ có dạng serif. Các ký tự có thể bị tách ra nếu tài liệu bắt nguồn từ bản sao sáng màu hoặc được quét ở ngưỡng cao.

+ Phân biệt nhiễu từ văn bản: Dấu chấm và dấu trọng âm có thể bị nhầm lẫn với nhiễu và ngược lại.

+ Nhầm đồ họa hoặc hình vẽ với văn bản: Điều này dẫn đến việc dữ liệu không phải văn bản sẽ được gửi đi để nhận dạng.

+ Mất văn bản khi kết nối cùng hỉnh vẽ hoặc đồ họa: Trong trường hợp này, văn bản sẽ không được chuyển sang giai đoạn nhận dạng. Điều này thường xảy ra nếu các ký tự được kết nối với đồ họa.



Hình 14: Những trường hợp gây nhầm lẫn[7]

* 1. **Tiền xử lý**

Hình ảnh thu được từ quá trình quét có thể chứa một lượng nhiễu nhất định. Tùy thuộc và độ phân giải của máy quét và sự thành công của kỹ thuật phân ngưỡng được áp dụng, các ký tự có thể bị lem hoặc bị hỏng. Một số lỗi này, sau này có thể gây ra tỷ lệ nhận dạng kém, có thể được loại bỏ bằng cách sử dụng bộ tiền xử lý để làm mịn các ký tự được số hóa.

Việc làm mịn bao gồm cả việc làm đầy và làm mỏng. Việc làm đầy giúp loại bỏ các khoảng ngắt, khoảng trống và lỗ hổng nhỏ trong các ký tự được số hóa, trong khi đó làm mỏng sẽ làm giảm độ rộng của ký tự, loại bỏ các mối nối thừa thãi. Các kỹ thuật làm mịn phổ biến nhất là di chuyển một cửa sổ qua hình ảnh nhị phân của ký tự, áp dụng các quy tắc nhất định cho nội dung của cửa sổ.

Ngoài việc làm mịn, quá trình tiền xử lý thường bảo gồm việc chuẩn hóa. Việc chuẩn hóa được áp dụng để thu được các ký tự có kích thước, độ nghiêng và xoay đồng đều. Để có thể hiệu chỉnh góc quay thì phải tìm được góc quay. Đối với các trang và dòng văn bản được xoay, các biến thể của biến đổi Hough thường được sử dụng để phát hiện độ lệch. Tuy nhiên, không thể tìm được góc quay của một ký hiệu cho đến khi ký hiệu đó được nhận dạng.



Hình 15: Chuẩn hóa và làm mịn ký tự[7]

* 1. **Trích xuất đặc trưng**

Mục tiêu của việc trích xuất đặc trưng là nắm bắt các đặc điểm cơ bản của ký hiệu và người ta thường chấp nhận rằng đây là một trong những vấn đề khó khăn nhất trong nhận dạng mẫu. Cách đơn giản nhất để mô tả một ký tự là bằng hình ảnh raster thực tế. Một cách tiếp cận khác là trích xuất một số đặc điểm nhất định vẫn đặc trưng cho các ký hiệu nhưng loại bỏ các thuộc tính không quan trọng. Các kỹ thuật trích xuất các đặc điểm như vậy thường được chia thành ba nhóm chính, trong đó các đặc trưng được tìm thấy từ:

+ Việc phân bố điểm

+ Các phép biến đổi và mở rộng chuỗi

+ Phân tích cấu trúc

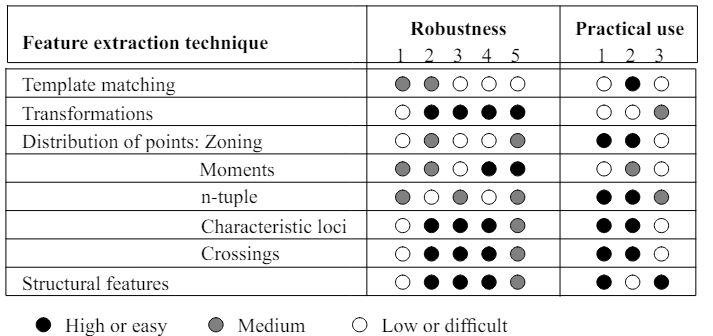
Các nhóm đặc trưng khác nhau có thể được đánh giá theo độ nhạy của chúng với nhiễu và biến dạng cũng như tính dễ thực hiện và sử dụng. Các tiêu chí được sử dụng trong đánh giá này như sau:

**Độ chắn chắn(Robustness):**

1. Nhiễu: Nhạy cảm với các đường bị ngắt kết nối, va chạm, khoảng trống, vòng lặp đầy,...
2. Sự biến dạng: Nhạy cảm với các biến đổi cục bộ như góc tròn, phần nhô ra không tự nhiên, độ giãn và độ co.
3. Biến thể kiểu dáng: Nhạy cảm với sự thay đổi trong kiểu dáng như việc sử dụng các hình dạng khác nhau để thể hiện cùng một ký tự hoặc việc sử dụng các nét chữ, nét nghiêng,…
4. Sự dịch chuyển: Nhạy cảm với sự di chuyển của toàn bộ ký tự hoặc các thành phần của nó.
5. Xoay: Nhạy cảm với sự thay đổi hướng của các ký tự.

**Thực tế sử dụng(Practical use):**

1. Tốc độ nhận dạng
2. Sự phức tạp trong cài đặt
3. Độc lập: Sự cần thiết của các kỹ thuật bổ sung



Hình 16: Đánh giá các kỹ thuật trích xuất đặc trưng[7]

* + 1. **Kỹ thuật so khớp mẫu và tương quan(Template-matching and correlation)**

Kỹ thuật này khác với những kỹ thuật khác ở chỗ không có đặc điểm nào thực sự được trích xuất. Thay vào đó, ma trận chứa hình ảnh của các ký tự đầu vào được khớp trực tiếp với một tập hợp các ký tự nguyên mẫu đại diện cho từng lớp cụ thể. Khoảng cách giữa mẫu và từng nguyên mẫu được tính toán và lớp của nguyên mẫu phù hợp nhất sẽ được gán cho mẫu. Kỹ thuật này đơn giản và dễ thực hiện trong phần cứng và đã được sử dụng trong nhiều máy OCR thương mại. Tuy nhiên, kỹ thuật này nhạy cảm với các biến thể vể nhiễu và kiểu dáng cũng như không có cách nào để xử lý các ký tự bị xoay.

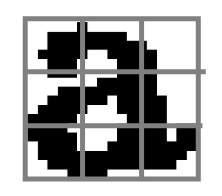
* + 1. **Các kỹ thuật dựa trên đặc trưng**

Trong các phương pháp này, các phép đo quan trọng được tính toán và trích xuất từ một ký tự và so sánh với các mô tả về các lớp ký tự thu được trong giai đoạn huấn luyện. Mô tả phù hợp nhất sẽ mang lại sự công nhận. Các đặc điểm được đưa ra dưới dạng số trong một vector đặc trưng và vector đặc trưng này được sử dụng để biểu diễn ký hiệu.

**2.4.2.a) Phân phối điểm(Distribution of points)**

Danh mục này bao gồm các kỹ thuật trích xuất đặc trưng dựa trên phân phối điểm thông kê. Những đặc điểm này thường có khả năng chịu được những biến dạng và biến đổi về kiểu dáng. Một số kỹ thuật điển hình trong lĩnh vực này được liệt kê dưới đây:

+ ***Zoning***: Hình chữ nhật bao quanh ký tự được chia thành nhiều vùng chồng chéo hoặc không chồng chéo và mật độ điểm đen trong các vùng này được tính toán và sử dụng làm đặc trưng.



Hình 17: Mô tả kỹ thuật zoning[7]

+ ***Moments***: Các chuyển động của các điểm đen về một tâm đã chọn, ví dụ như trọng tâm hoặc hệ tọa độ đã chọn, được sử dụng làm đặc trưng.

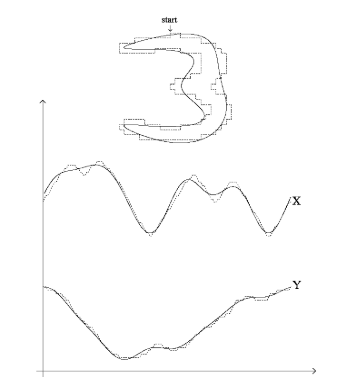
+ ***Crossings and distances:*** Trong kỹ thuật cắt ngang, các đặc điểm được tìm thấy dựa trên số lần hình dạng ký tự được vector cắt ngang theo các hướng nhất định. Kỹ thuật này thường được sử dụng bởi các hệ thống thương mại vì nó có thể được thực hiện ở tốc độ cao và yêu cầu độ phức tạp thấp. Khi sử dụng kỹ thuật khoảng cách, các độ dài nhất định dọc theo các vector đi qua hình dạng ký tự sẽ được đo. Ví dụ: độ dài của vector trong ranh giới của ký tự.

+ ***N-tuples:*** Sự xuất hiện chung tương đối của các điểm đen và trắng theo thứ tự nhất định được chỉ định được sử dụng làm đặc trưng

+ ***Characteristic loci:*** Đối với mỗi điểm trong nền của ký tự, các vector dọc và ngang được tạo ra. Số lần các đoạn dòng mô tả ký tự được giao nhau với các vector này được sử dụng làm đặc trưng.

**2.4.2.b) Các phép biến đổi và khai triển chuỗi**

Những kỹ thuật này giúp giảm chiều của vector đặc trưng và các đặc điểm được trích xuất có thể trở thành bất biến đối với các biến dạng tổng thể như dịch chuyển và xoay. Các phép biến đổi được sử dụng có thể là Fourier, Walsh, Haar, Hadamard, Karhunen-Loeve, Hough, biến đổi trục chính,…

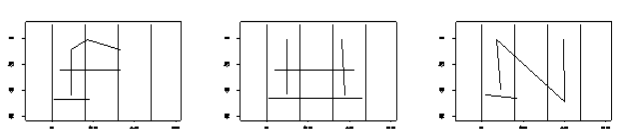


Hình 18: Mô tả Fourier hình elip[7]

Nhiều phép biến đổi trong số này dựa trên đường cong mô tả đường viền của các ký tự. Điều này có nghĩa là các tính năng này rất nhạy cảm với nhiễu ảnh hưởng đến đường viền của ký tự như những khoảng trống ngoài ý muốn trên đường viền. Tuy nhiên, chúng có khả năng chịu được nhiễu ảnh hưởng đến bên trong ký tự và những biến dạng.

**2.4.2.c) Phân tích cấu trúc**

Trong quá trình phân tích cấu trúc, các đặc điểm mô tả cấu trúc hình học và cấu trúc liên kết của ký hiệu sẽ được trích xuất. Bằng những đặc điểm này, người ta cố gắng mô tả cấu trúc vật lý của ký tự và một số đặc điểm thường được sử dụng là nét, đường, điểm cuối, giao điểm giữa các đường và vòng lặp. So với các kỹ thuật khác, phân tích cấu trúc mang lại đặc trưng có khả năng chịu nhiễu và biến đổi kiểu dáng cao. Tuy nhiên, các đặc trưng này chỉ có khả năng chịu xoay và dịch chuyển ở mức độ vừa phải. Thật không may, việc trích xuất những đặc điểm này không hề đơn giản và ở một mức độ nào đó vẫn là một lĩnh vực cần nghiên cứu thêm.



Hình 19: Các nét được trích xuất từ chữ in hoa F, H, N

* 1. **Phân loại**

Phân loại là quá trình xác định từng ký tự và gán cho nó lớp ký tự chính xác. Có 2 cách tiếp cận chính để phân loại trong nhận dạng ký tự. Đầu tiên là tiếp cận theo lý thuyết quyết định(decision-theoretic). Phương thức này được sử dụng khi ký tự được biển diễn dưới dạng một vector đặc trưng. Chúng ta cũng có thể có những đặc điểm khuôn mẫu bắt nguồn từ cấu trúc vật lý của ký tự mà không dễ dàng định lượng được. Trong những trường hợp này, mối quan hệ giữa các đặc điểm có thể có tầm quan trọng khi quyết định rằng một ký tự bao gồm một nét dọc và một nét ngang, thì nó có thể là chữ “L” hoặc chữ “T” và cần có mối quan hệ giữa hai nét để phân biệt ký tự. Khi đó cần có một cách tiếp cận mang tính cấu trúc.

* + 1. **Phương pháp lý thuyết quyết định**

Các cách tiếp cận chính để nhận dạng theo lý thuyết quyết định là phân loại khoảng cách tối thiểu, phân loại thống kê và mạng lưới thần kinh. Mỗi kỹ thuật phân loại sẽ được mô tả sau đây:

**2.5.1.a) Matching**

Matching(so khớp) bao gồm các nhóm kỹ thuật dựa trên các thước đo tương tự trong đó tính toán khoảng cách giữa các vector đặc trưng, mô tả ký tự được trích xuất và mô tả của từng lớp. Có thể sử dụng các thước đo khác nhau nhưng điểm chung là khoảng cách Euclide. Bộ phân loại khoảng cách tối thiểu này hoạt động tốt khi các lớp được phân tách tốt, tức là khi khoảng cách trung bình giữa các lớp lớn so với độ dàn trải của từng lớp.

Khi toàn bộ ký tự được sử dụng làm đầu vào cho phân loại và không có đặc điểm nào được trích xuất(khớp mẫu), phương pháp tương quan sẽ được sử dụng. Ở đây khoảng cách giữa hình ảnh ký tự và hình ảnh nguyên mẫu đại diện cho từng lớp được tính toán.

**2.5.1.b) Optimum statistical classifiers**

Trong phân loại thống kê, phương pháp xác suất để nhận biết được áp dụng. Ý tưởng là sử dụng một sơ đồ phân loại tối ưu bằng cách tính trung bình, việc sử dụng nó mang lại xác suất mắc lỗi phân loại thấp nhất.

Bộ phân loại giảm thiểu tổng mất mát trung bình được gọi là bộ phân loại Bayes. Cho một ký hiệu chưa biết được mô tả bằng vector đặc trưng của nó, xác suất ký hiệu đó thuộc lớp c được tính cho tất cả các lớp c=1…N. Ký hiệu sau đó được gán lớp cho xác suất tối đa.

Để cơ chế này được tối ưu, các hàm mật độ xác suất của các ký hiệu của mỗi lớp phải được biết cùng với xác suất xuất hiện của mỗi lớp. Trường hợp thứ 2 thường được giải quyết bằng cách giả định rằng tất cả các lớp đều có xác suất như nhau. Hàm mật độ thường được giả định là có phân phối chuẩn và giả định này càng gần với thực tế thì bộ phân loại của Bayes càng tiến gần đến hành vi tối ưu.

Bộ phân loại khoảng cách tối thiểu được mô tả ở trên được xác định hoàn toàn bởi vector trung bình của từng lớp và bộ phân loại Bayes cho các lớp Gaussian được xác định hoàn toàn bởi vector trung bình và ma trận hiệp phương sai của mỗi lớp. Các tham số chỉ định bộ phân loại này có được thông qua quá trình huấn luyện. Trong quá trình này, các mẫu huấn luyện của mỗi lớp được sử dụng để tính toán các tham số này và thu được mô tả của từng lớp.

**2.5.1.c) Neural networks**

Gần đây, việc sử dụng mạng lưới thần kinh để nhận dạng ký tự(và các loại mẫu khác) đã xuất hiện trở lại. Xem xét mạng lan truyền ngược, mạng này bao gồm nhiều lớp phần tử được kết nối với nhau. Một vector đặc trưng đi vào mạng ở lớp đầu vào. Mỗi phần tử của lớp tính toán tính tổng trọng số của đầu vào và biến nó thành đầu ra bằng hàm phi tuyến. Trong quá trình huấn luyện, trọng số tại mỗi kết nối được điều chỉnh cho đến khi đạt được đầu ra mong muốn. Một vấn đề của mạng lưới thần kinh trong OCR có thể là khả năng dự đoán và tính tổng quát hạn chế của chúng, trong khi ưu điểm là tính chất thích ứng của chúng.

* + 1. **Phương pháp cấu trúc**

Trong lĩnh vực nhận dạng cấu trúc, phương pháp cú pháp là một trong những phương pháp phổ biến nhất. Có những kỹ thuật khác nhưng chúng ít tổng quát hơn và sẽ không được đề cập ở báo cáo này.

**Syntactic methods(Phương pháp cú pháp):** Các thước đo về sự tương đồng dựa trên mối quan hệ giữa các thành phần cấu trúc có thể được hình thành bằng cách sử dụng các khái niệm ngữ pháp. Ý tưởng là mỗi lớp có ngữ pháp riêng xác định thành phần của ký tự. Một ngữ pháp có thể được biểu diễn dưới dạng chuỗi hoặc cây và các thành phần cấu trúc được trích xuất từ một ký tự chưa biết sẽ khớp với ngữ pháp của từng lớp. Giả sử chúng ta có hai lớp ký tự khác nhau có thể được tạo ra bởi hai ngữ pháp G1 và G2 tương ứng. Với mỗi ký tự không xác định, chúng ta nói rằng nó giống với lớp thứ nhất hơn nếu nó được tạo ra bởi ngữ pháp G1 chứ không phải bởi ngữ pháp G2.

* 1. **Hậu xử lý**
     1. **Nhóm**

Kết quả của việc nhận dạng ký tự đơn giản trên tài liệu là một tập hợp các ký hiệu riêng lẻ. Tuy nhiên, bản thân những ký tự này thường không chứa đủ thông tin. Thay vào đó, chúng tôi muốn liên kết các ký tự riêng lẻ thuộc cùng một chuỗi với nhau, tạo thành các từ và số. Quá trình thực hiện liên kết các ký tự này thành chuỗi thường được gọi là nhóm. Việc nhóm các ký tự thành chuỗi dựa trên vị trí của các ký tự trong tài liệu. Các ký tự được cho là đủ gần nhau sẽ được nhóm lại với nhau.

Đối với các phông chữ có độ cao cố định(fixed pitch), quá trình nhóm khá dễ dàng vì vị trị của từng ký tự đều được biết. Đối với các ký tự sắp chữ(typeset characters), khoảng cách giữa các từ thường lớn hơn đáng kể so với khoảng cách giữa các ký tự và do đó vẫn có thể nhóm được. Các vấn đề thực sự xảy ra đối với các ký tự viết tay hoặc khi văn bản bị lệch.

* + 1. **Phát hiện và sửa lỗi**

Cho đến khi việc nhóm từng ký tự được xử lý riêng biệt và bối cảnh mà mỗi ký tự xuất hiện thường chưa được khai thác. Tuy nhiên, trong các bài toán nhận dạng văn bản quang học nâng cao, hệ thống chỉ bao gồm nhận dạng một ký tự sẽ không đủ. Ngay cả những hệ thống nhận dạng chính xác 100% cho tất cả các ký tự, nhưng một số lỗi này có thể được phát hiện hoặc thậm chí sửa chữa bằng cách sử dụng ngữ cảnh.

Có hai cách tiếp cận chính, trong đó cách đầu tiên sử dụng khả năng các chuỗi ký tự xuất hiện cùng nhau. Điều này có thể được thực hiện bằng cách sử dụng các quy tắc xác định cú pháp của từ, chằng hạn bằng cách nói rằng sau dấu chấm thường phải có chữ in hoa. Ngoài ra, đối với các ngôn ngữ khác nhau, xác suất của hai hoặc nhiều ký tự xuất hiện cùng nhau trong một chuỗi có thể được tính toán và có thể được sử dụng để phát hiện lỗi. Ví dụ: trong tiếng Anh, xác suất chữ “k” xuất hiện sau chữ “h” trong một từ là bằng 0 và nếu phát hiện thấy sự kết hợp như vậy thì sẽ có lỗi.

Một cách tiếp cận khác là sử dụng từ điển, phương pháp này đã được chứng minh là phương pháp hiệu quả nhất để phát hiện và sửa lỗi. Với một từ có thể có lỗi, từ đó sẽ được tra cứu trong từ điển. Nếu từ đó không có trong từ điển thì đã phát hiện ra lỗi và có thể sửa bằng cách đổi từ đó thành từ giống nhất. Xác suất thu được từ việc phân loại có thể giúp xác định ký tự đã được phân loại sai. Nếu từ đó có trong từ điển thì rất tiếc điều này không chứng minh được rằng không có lỗi xảy ra. Một lỗi có thể đã biến đổi từ này thành 1 từ khác có trong từ điển và những lỗi đó không thể phát hiện được bằng quy trình này. Một nhược điểm khác của phương pháp từ điển là việc tìm kiếm và so sánh ngụ ý rất tốn thời gian.

1. **Đánh giá công nghệ ICR**

Công nghệ Optical Character Recognition(OCR) giúp phân biệt sự hiện diện của các ký tự theo hình dạng của chúng. Dù mang lại nhiều ưu điểm như tiết kiệm thời gian, tăng năng suất làm việc nhưng nó cũng không tránh khỏi những hạn chế như độ chính xác,… Theo [8], công nghệ OCR có ưu nhược điểm như sau:

* 1. **Ưu điểm**

+ Tăng hiệu quả và năng suất: Một trong những ưu điểm chính của công nghệ OCR là khả năng tăng hiệu quả và năng suất. Công nghệ OCR có thể được sử dụng để tự động hóa quá trình nhập và truy xuất dữ liệu, điều này có thể tiết kiệm một lượng thời gian và tiền bạc đáng kể. Điều này đặc biệt hữu ích cho các doanh nghiệp phải xử lý một khối lượng lớn dữ liệu một cách thường xuyên.

+ Độ chính xác của dữ liệu được cải thiện: Một ưu điểm khác của công nghệ OCR là khả năng cải thiện độ chính xác của dữ liệu. Công nghệ OCR có khả năng nhận dạng và trích xuất văn bản từ tài liệu ịn hoặc viết tay với độ chính xác cao. Điều này có thể đặc biệt hữu ích trong các lĩnh vực như chăm sóc sức khỏe, nơi dữ liệu chính xác là rất quan trọng.

+ Hiệu quả về chi phí: Công nghệ OCR cũng tiết kiệm chi phí. Chi phí của phần mềm OCR đã giảm đáng kể trong những năm gần đây, giúp nhiều doanh nghiệp và tổ chức có thể tiếp cận phần mềm này. Ngoài ra, công nghệ OCR có thể tiết kiệm tiền cho doanh nghiệp về lâu dài bằng cách giảm nhu cầu nhập và truy xuất dữ liệu thủ công.

+ Số hóa dữ liệu: Công nghệ OCR có thể được sử dụng để số hóa các dữ liệu trên giấy, có thể được lưu trữ và truy cập bằng máy tính điện tử. Điều này đặc biệt hữu ích cho doanh nghiệp và tổ chức phải lưu giữ khối lượng tài liệu lớn.

+ Khả năng tìm kiếm được cải thiện: Công nghệ OCR cũng có thể cải thiện khả năng tìm kiếm của tài liệu kỹ thuật số. Bằng cách trích xuất văn bản từ tài liệu in hoặc viết tay, công nghệ OCR giúp tìm kiếm thông tin cụ thể trong tài liệu. Điều này có thể đặc biệt hữu ích cho các doanh nghiệp và tổ chức phải lưu giữ khối lượng tài liệu lớn.

* 1. **Nhược điểm**

+ Chi phí ban đầu cao: Chi phí ban đầu của công nghệ OCR có thể cao. Các doanh nghiệp và tổ chức phải đầu tư vào phần mềm, phần cứng và huấn luyện OCR để sử dụng công nghệ một cách hiệu quả. Ngoài ra, các doanh nghiệp, tổ chức cũng có thể phải đầu tư thêm nguồn lực để quản lý và bảo trì hệ thống OCR.

+ OCR chỉ hoạt động hiệu quả với văn bản in chứ không phải với văn bản viết tay, máy tính phải học chữ viết tay

+ Giới hạn ở nhận dạng văn bản: Công nghệ OCR bị giới hạn ở khả năng nhận dạng văn bản và không thể nhận dạng hoặc trích xuất các dữ liệu khác như hình ảnh hoặc đồ họa. Điều này có thể gây rắc rối cho các doanh nghiệp và tổ chức phải xử lý tài liệu có chứa sự kết hợp giữa văn bản và các loại dữ liệu khác.

+ Sự phụ thuộc vào chất lượng của tài liệu gốc: Một nhược điểm của OCR là nó phụ thuộc vào chất lượng của tài liệu gốc. Nếu tài liệu gốc có chất lượng kém, phần mềm OCR khó có thể nhận dạng và trích xuất văn bản một cách chính xác.

+ Không chính xác 100%, có thể mắc một số ít sai sót trong quá trình thực hiện

+ Tất cả các tài liệu sau khi nhận dạng phải được kiểm tra cẩn thận và sửa chữa bằng tay nếu cần.

**Chương III: Áp dụng công nghệ icr để xây dựng ừng dụng chuyển tài liệu viết tay thành văn bản điện tử**

1. **Giới thiệu chung về ứng dụng**
2. **Chức năng các thành phần**
3. **Kết quả cài đặt và thử nghiệm**
4. **Đánh giá ứng dụng**

**TÀI LIỆU THAM KHẢO**

1. Joachim M. Buhmann, Annett Faber and Petko Faber. Mustererkennung 1999. Springer ,Berlin, Heidelberg (1999).
2. VR.Plamondon and S.N.Srihari. Online and off-line handwriting recognition: a comprehensive survey.  [IEEE Transactions on Pattern Analysis and Machine Intelligence](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=34) ( Volume: 22, [Issue: 1](https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=17859&punumber=34), January 2000)
3. Venu Govindaraju and H.Xue. Fast handwriting recognition for indexing historical documents.  [First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings.](https://ieeexplore.ieee.org/xpl/conhome/8926/proceeding)
4. https://www.a1-tech.co.uk/handwriting-recognition-with-ocr-applications-and-challenges/
5. https://www.v7labs.com/blog/handwriting-recognition-guide#methods-of-handwriting-recognition
6. https://www.nexdata.ai/news/730
7. Line Eikvil. Optical Character Recognition. December 1993.
8. https://tutorialspoint.com/advantages-and-disadvantages-of-optical-character-reader-ocr