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

**KHOA CÔNG NGHỆ THÔNG TIN**

----------



**BÁO CÁO ĐỒ ÁN TỐT NGHIỆP**

***Đề tài***

***Xây dựng ứng dụng chuyển tài liệu viết tay***

***thành văn bản điện tử dựa trên công nghệ***

***Intelligent Character Recognition***

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

**HÀ NỘI, 2024**

**MỤC LỤC**

[LỜI MỞ ĐẦU 1](#_Toc22210)

[Chương I: Bài toán nhận dạng chữ viết tay 2](#_Toc2118)

[1.1 Giới thiệu 2](#_Toc14151)

[1.2 Tầm quan trọng 2](#_Toc12043)

[1.3 Lịch sử phát triển 2](#_Toc2836)

[1.4 Ứng dụng 4](#_Toc30697)

[1.5 Phương pháp và kỹ thuật 5](#_Toc3536)

[1.5.1 CapsNets 5](#_Toc22872)

[1.5.2 Multidimensional Recurrent Neural Networks (MDRNNs) 6](#_Toc10165)

[1.5.3 Connectionist Temporal Classification (CTC) 7](#_Toc24496)

[1.5.4 Transformer models 8](#_Toc27835)

[1.5.5 Encoder - Decoder and Attention Networks 9](#_Toc3097)

[1.5.6 Handwriting Text Generation 12](#_Toc27183)

[1.6 Thách thức và giải pháp 13](#_Toc11343)

[TÀI LIỆU THAM KHẢO 15](#_Toc24057)

**MỤC LỤC HÌNH ẢNH**

[Hình 1 .1: Định tuyến động giữa các capsules [6] 5](#_Toc3360)

[Hình 1.2 : MDRNN 2 chiều [6] 6](#_Toc23364)

[Hình 1.3 : Mô hình hóa trình tự với CTC [6] 7](#_Toc4914)

[Hình 1.4 : Chiến lược giải mã đường dẫn tốt nhất [6] 7](#_Toc4275)

[Hình 1.5 : Tổng quan về kiến trúc của mô hình [6] 8](#_Toc15948)

[Hình 1.6 : Kiến trúc của mô hình mã hóa - giải mã kết hợp với mạng chú ý [6] 9](#_Toc11660)

[Hình 1.7 : Thuật toán Scan As You Read để nhận dạng chữ viết tay, xem kẽ các lớp LSTM theo kiểu đa chiều [6] 10](#_Toc1802)

[Hình 1.8 : Hoạt động của mô hình Listen, Attend and Spell để nhận dạng chữ viết [6] 11](#_Toc14169)

[Hình 1.9 : Generative Adversial networks cho việc tạo dữ liệu văn bản [6] 12](#_Toc10347)

**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.

Công nghệ ICR ra đời đã mở ra một hướng đi mới, giúp tự động hóa quá trình nhận dạng và chuyển đổi các ký tự viết tay thành văn bản điện tử một cách chính xác và nhanh chóng. Đây là bước tiến quan trọng so với công nghệ Optical Character Recognition(OCR) truyền thống, vốn chỉ có khả năng nhận dạng ký tự in. ICR không chỉ giúp tiết kiện thời gian và công sức mà còn nâng cao độ chính xác và hiệu quả trong việc quản lý tài liệu. Trong báo cáo này, sẽ tập trung vào việc ứng dụng công nghệ Intelligent Character Recognition(ICR) để chuyển đổi tài liệu viết tay thành văn bản điện tử.

Báo cáo này gồm 3 chương, cụ thể như sau:

**Chương I: Bài toán nhận dạng chữ viết tay.** Chương này sẽ trình bày từ lịch sử phát triển, tầm quan trọng trong cuộc sống thực tế cho đến các phương pháp và kỹ thuật phổ biến được dùng để nhận dạng chữ viết tay

**Chương II: Công nghệ Optical Character Recognition(OCR) và Intelligent Character Recognition(ICR).** Chương II sẽ giới thiệu về công nghệ OCR và ICR, nguồn gốc, cơ chế hoạt động và sự khác biệt giữa OCR và ICR.

**Chương III: Áp dụng công nghệ Intelligent Character Recognition để xây dựng ứng dụng chuyển tài liệu viết tay thành văn bản điện tử.** Nội dung của chương này tập trung vào việc làm như thế nào để xây dựng ứng dụng chuyển tài liệu viết tay thành văn bản điện từ sử dụng công nghệ ICR.

# Chương I: Bài toán nhận dạng chữ viết tay

**1.1 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 [2].

Ở mức khải niệm, kỹ thuật nhận dạng chữ viết tay được chia thành hai phương pháp chính [1]:

* + 1. **Online handwriting recognition**

Nhận dạng chữ viết tay trực tuyến (Online handwriting recognition) đề cập đến quá trình nhận dạng chữ viết tay dựa trên thông tin thu được trong thời gian thực ngay lúc người dùng thực hiện hành động viết. Chữ viết tay sẽ được ghi lại và lưu trữ ở dạng kỹ thuật số thông qua các phương tiện khác nhau. Thông thường, một cây bút đặc biệt được sử dụng cùng với bề mặt điện từ. Khi bút di chuyển trên về mặt, tọa độ hai chiều của các điểm liên tiếp được biểu diễn dưới dạng hàm của thời gian và được lưu trữ theo thứ tự. Người ta thường công nhận rằng phương pháp nhận dạng văn bản viết tay trực tuyết đã đạt được kết quả tối hơn so với phương pháp ngoại tuyến. Điều này là do thực tế nhiều thông tin có thể được ghi lại như hướng, tốc độ và thứ tự các nét của chữ viết tay.

* + 1. **Offline handwriting recognition**

Nhận dạng chữ viết tay ngoại tuyến (Offline handwriting recognition) đề cập đến quá trình nhận dạng các từ đã được quét từ một bề mặt (chẳng hạn như một tờ giấy) và được lưu trữ kỹ thuật số ở định dạng ảnh xám. Sau khi được lưu trữ, thông thường phải thực hiện các bước tiền xử lý để cho phép nhận dạng tốt hơn. Nhận dạng ký tự ngoại tuyến có thể được nhóm thành 2 loại chính:

+ Nhận dạng ký tự từ tính (Magnetic Character Recognition - MCR)

+ Nhận dạng ký tự quang học (Optical Character Recognition - OCR)

Trong MCR, các ký tự được in bằng mực từ tính. Thiết bị đọc có thể nhận dạng các ký tự theo từ trường duy nhất của mỗi ký tự. MCR chủ yếu được sử dụng trong các ngân hàng để xác thực. OCR xử lý việc nhận dạng các ký tự thu được bằng phương tiện quang học, thường là máy quét hoặc máy ảnh. Các ký tự ở dạng hình ảnh pixel và có thể được in hoặc viết tay, ở bất kỳ kích thước, hình dạng hoặc hướng nào.

OCR có thể được chia thành nhận dạng ký tự viết tay và nhận dạng ký tự in. Nhận dạng ký tự viết tay khó thực hiện hơn nhận dạng ký tự in do phong cách và phong tục viết tay của con người đa dạng. Trong nhận dạng ký tự in, hình ảnh cần xử lý ở dạng phông chữ tiêu chuẩn như Times New Roman, Arial, Courier,… nên dễ dàng hơn.

Trong báo cào này, khi đề cập đến nhận dạng chữ viết tay sẽ tập trung vào phương pháp nhận dạng chữ viết tay ngoại tuyến.

**1.2 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.

**1.3 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 [4]:

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.

**1.4 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 [3], 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 liệu 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.

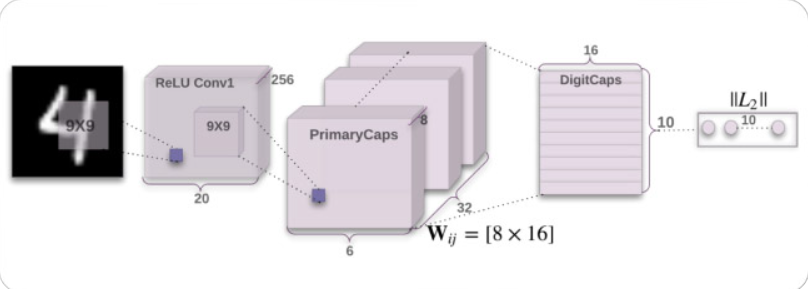
**1.5 Phương pháp và kỹ thuật**

Tham khảo từ [6]:

**1.5.1 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

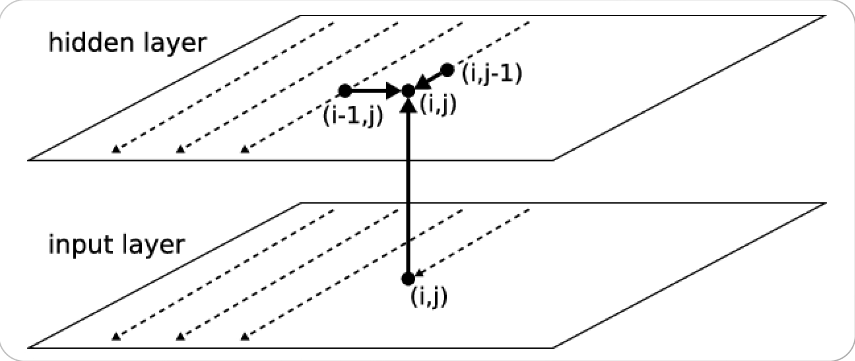


Hình 1.1: Định tuyến động giữa các capsules [6]

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.

**1.5.2 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



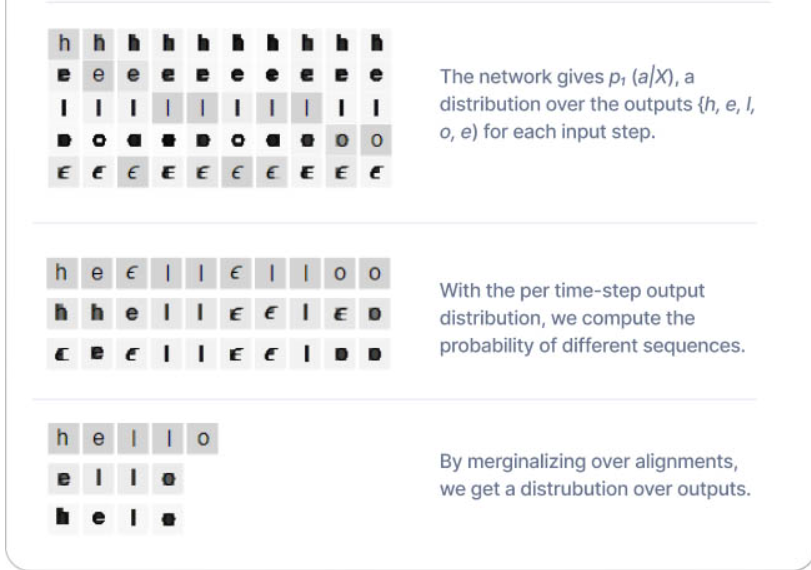
Hình 1.2: MDRNN 2 chiều [6]

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.

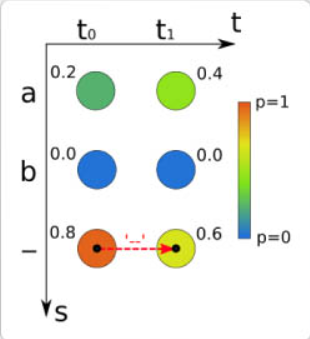
**1.5.3 Connectionist Temporal Classification (CTC)**

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 1.3: Mô hình hóa trình tự với CTC [6]

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 1.4: Chiến lược giải mã đường dẫn tốt nhất [6]

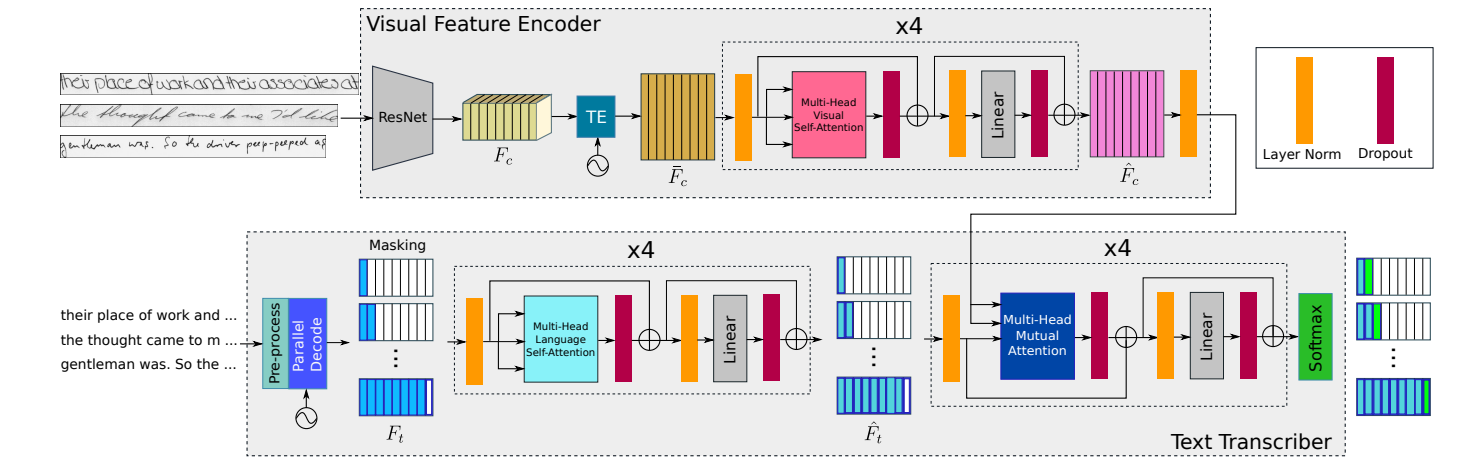
Để giải mã, có thể sử dụng:

+ 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.

**1.5.4 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 1.5: Tổng quan về kiến trúc của mô hình [6]

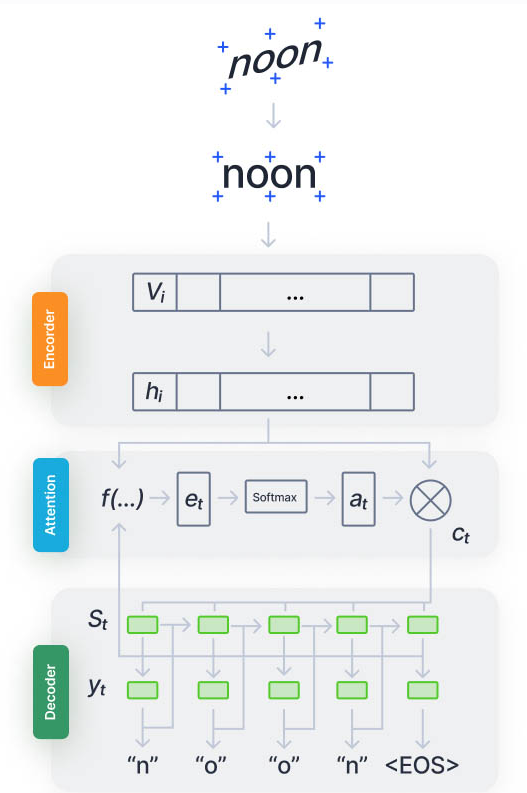
Kiến trúc chia ra làm 2 phần:

+ 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.

**1.5.5 Encoder - Decoder and Attention Networks**



Hình 1.6: Kiến trúc của mô hình mã hóa - giải mã kết hợp với mạng chú ý [6]

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.

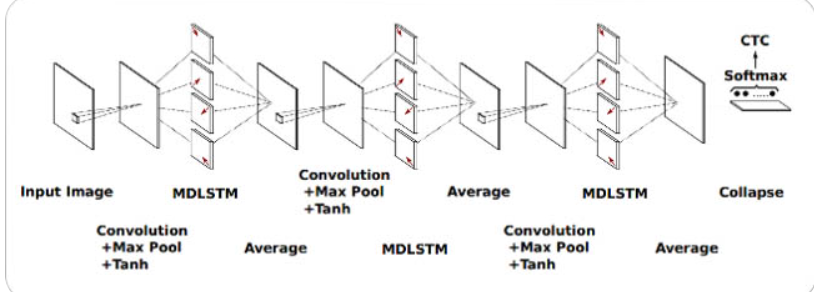
+ ***Trích xuất đặc trưng***: Sử dụng một kiến trúc tương tự ResNet để mã hóa hình ảnh đầu vào thành một bản đồ đặc trưng hình ảnh 2D.

+ ***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.

**1.5.5.a) 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 1.7: Thuật toán Scan As You Read để nhận dạng chữ viết tay, xem kẽ các lớp LSTM theo kiểu đa chiều [6]

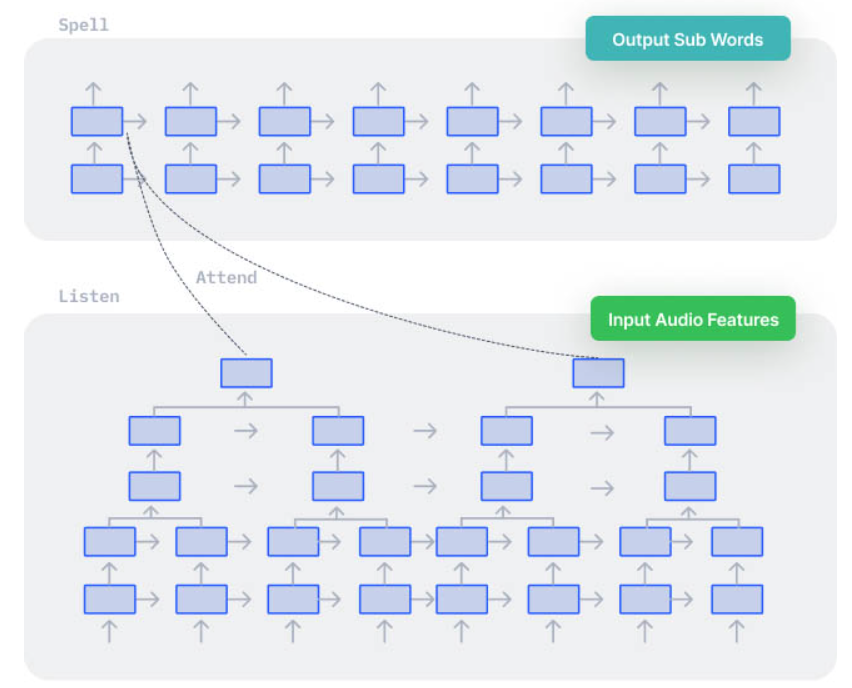
**1.5.5.b) Convolve, Attend and Spell**

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 1.8: Hoạt động của mô hình Listen, Attend and Spell để nhận dạng chữ viết [6]

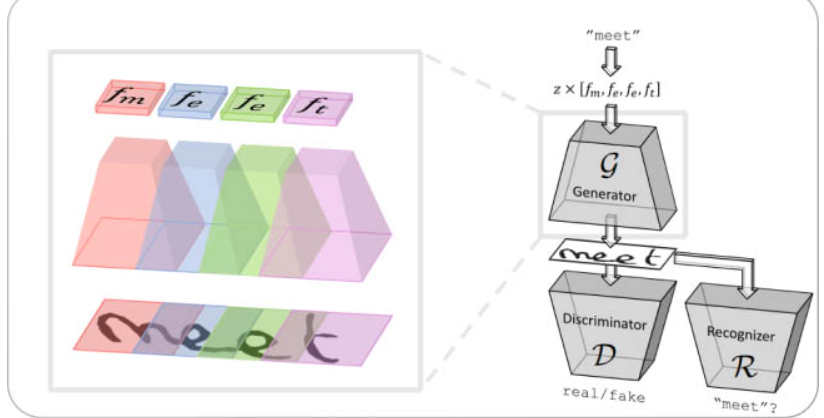
Bộ mã hóa bao gồm một mạng nơ-ron tích chập (CNN) để trích xuất các đặc trưng hình ảnh từ văn bản viết, được mã hóa tuần tự bởi một mạng nơ-ron tái lặp (RNN). Bộ giải mã là một RNN khác nhau giải mã từng ký tự một, từ đó xây dựng từ hoàn chỉnh và viết ra. Một cơ chế chú ý kết nối bộ mã hóa và bộ giải 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.

**1.5.6 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 1.9: Generative Adversial networks cho việc tạo dữ liệu văn bản [6]

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.

**1.6 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 [5]:

+ 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.

**1.7 Kết luận**

Trong chương này, báo cáo đã trình bày một cái nhìn tổng quan về bài toán nhận dạng chữ viết tay, nhấn mạnh tầm quan trọng, lịch sử phát triển và ứng dụng của công nghệ này. Các phương pháp và kỹ thuật dùng để cải thiện hiệu suất nhận dạng chữ viết tay như CapsNets, MDRNNs, CTC, các mô hình Transformer, mạng Encoder-Decoder, Attention Networks và Handwriting Text Generation. Báo cáo cũng đã nêu ra các thách thức trong việc nhận dạng chữ viết tay như việc xử lý độ đa dạng của chữ viết tay và cải thiện độ chính xác của các mô hình.

Đề giải quyết những vấn đề trên, báo cáo này sẽ áp dụng công nghệ Intelligent Character Recognition (ICR), một sự cái tiến đáng kể so với Optical Character Recognition (OCR). ICR không chỉ kế thừa các khả năng cơ bản của OCR mà còn sử dụng các kỹ thuật học sâu và trí tuệ nhân tạo để xử lý các biến thể phức tạp và không đồng nhất của chữ viết tay. Chương tiếp theo sẽ đi vào chi tiết về công nghệ OCR và ICR, bao gồm các kỹ thuật cụ thể và cách chúng được áp dụng để giải quyết các vấn đề nhận dạng chữ viết tay.

**TÀI LIỆU THAM KHẢO**

1. Dalbir and Sanjiv Kumar Singh. Review of Online & Offline Character Recognition. International Journal Of Engineering And Computer Science (Volume 4, Issue 5, May 2015), pp 11729-11730
2. R. Seiler, M. Schenke, F. Eggiman. Off-Line Cursive Handwriting Recognition Compared with On-Line Recognition. Proc. ICPR 96, Vienna, Vol. 4(1996), pp.505-509
3. https://www.a1-tech.co.uk/handwriting-recognition-with-ocr-applications-and-challenges/
4. https://www.docsumo.com/blog/optical-character-recognition-history
5. https://www.nexdata.ai/news/730
6. https://www.v7labs.com/blog/handwriting-recognition-guide#methods-of-handwriting-recognition