TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

ĐỒ ÁN CUỐI KÌ

XỬ LÝ NGÔN NGỮ TỰ NHIÊN

Người thực hiện: **HỒ VÕ HOÀNG DUY – 20087541**

VÕ QUỐC HUY – 20081001

Lớp **: 420300138501**

Khoá **: 16**

Người hướng dẫn: **TS BÙI THANH HÙNG**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023

TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

ĐỒ ÁN CUỐI KÌ

XỬ LÝ NGÔN NGỮ TỰ NHIÊN

Người thực hiện: **HỒ VÕ HOÀNG DUY- 20087541**

VÕ QUỐC HUY - 20081001

Lớp **: 420300138501**

Khoá **: 16**

Người hướng dẫn: **TS. BÙI THANH HÙNG**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023

LỜI CẢM ƠN

Chúng tôi xin gửi lời cảm ơn chân thành nhất đế quý thầy TS.Bùi Thanh Hùng – Giảng viên đã tận tình hướng dẫn cho chúng tôi trong suốt quá trình học tập môn Xử lý ngôn ngữ tự nhiên và tạo điều kiện cho chúng tôi làm Đồ án này. Dưới sự hướng dẫn của Thầy chúng tôi được tiếp cận với những kiến thức chuyên môn sâu, rộng trong lĩnh vực xử lý ngôn ngữ tự nhiên. Qua đó có thể hoàn thành Đồ án này một cách thuận lợi.

Một lần nữa chúng tôi xin gửi lời cảm ơn sâu sắc đến Thầy. Chúc Thầy thật nhiều sức khỏe, thành công trong công việc và trong cuộc sống. Xin chân thành cảm ơn Thầy.

ĐỒ ÁN ĐƯỢC HOÀN THÀNH
TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi / chúng tôi và được sự hướng dẫn của TS. Bùi Thanh Hùng. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình. Trường đại học Công nghiệp TP Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày tháng năm

Tác giả

(ký tên và ghi rõ họ tên)

Hồ Võ Hoàng Duy

Võ Quốc Huy

PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

_____

_____

_____

_____

_____

_____

_____

Tp. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

TÓM TẮT

Xác định thông tin văn bản hình ảnh bằng học sâu là một bài toán xử lý hình ảnh và nhận dạng chữ viết tự động sử dụng các mô hình học sâu và xử lý ngôn ngữ tự nhiên để trích xuất các thông tin văn bản từ hình ảnh hoặc file pdf. Bài toán này có thể được sử dụng để tự động trích xuất các thông tin văn bản từ các văn bản của Khoa, Trường, giúp tiết kiệm thời gian và công sức cho nhân viên văn phòng. Ngoài ra, bài toán này cũng có thể được sử dụng để tìm kiếm thông tin trong các văn bản hình ảnh, giúp người dùng dễ dàng tìm thấy thông tin cần thiết, hoặc phân tích các thông tin văn bản trong các văn bản hình ảnh, giúp người dùng hiểu rõ hơn về nội dung của văn bản. Bài toán xác định thông tin văn bản hình ảnh bằng học sâu bao gồm hai bước chính: Phân đoạn hình ảnh: Phân đoạn hình ảnh là quá trình chia hình ảnh thành các vùng có đặc trưng tương đồng. Sau khi phân đoạn hình ảnh, các vùng văn bản sẽ được xác định dựa trên các đặc trưng của chữ viết. Nhận dạng chữ viết: Nhận dạng chữ viết là quá trình chuyển đổi chữ viết trong hình ảnh thành văn bản. Sau khi nhận dạng chữ viết, các thông tin văn bản trong các vùng văn bản sẽ được trích xuất. Bài toán xác định thông tin văn bản hình ảnh bằng học sâu và xử lý ngôn ngữ tự nhiên đang được phát triển mạnh mẽ và có nhiều ứng dụng tiềm năng trong thực tế.

MỤC LỤC

DANH MỤC CHỮ VIẾT TẮT

GNN: Graph Neural Network

GE: Graph Embemdding

GCN: Graph Convolutional Network

GAT: Graph Attention Network

PNAConv: Principal Neighbourhood Aggregation for Graph Nets

SGConv: Simplifying Graph Convolutional Networks

DANH MỤC CÁC HÌNH VẼ

DANH MỤC CÁC BẢNG

XÁC ĐỊNH THÔNG TIN VĂN BẢN HÌNH ẢNH BẰNG CÁC PHƯƠNG PHÁP HỌC SÂU

1.1 Giới thiệu về bài toán

Bài toán xác định thông tin trong văn bản hình ảnh là một thách thức lớn và có ý nghĩa quan trọng trong lĩnh vực xử lý ngôn ngữ tự nhiên và thị giác máy tính. Trong thời đại số hóa ngày nay, việc xác định thông tin từ hình ảnh văn bản trở nên rất cần thiết trong nhiều tình huống. Chẳng hạn như việc nhận diện nơi phát hành văn bản, ngày phát hành, tiêu đề, và nhiều thông tin khác, từ các hình ảnh, file chứa văn bản. Các phương pháp học sâu đã chứng minh hiệu quả trong việc giải quyết bài toán này. Bằng cách sử dụng mô hình học sâu, chúng ta có thể huấn luyện để mô hình tự động nhận diện và trích xuất thông tin từ văn bản trong hình ảnh. Mô hình này có thể học được cấu trúc ngôn ngữ tự nhiên và các đặc điểm hình ảnh, từ đó giúp quá trình xác định thông tin được chính xác và hiệu quả hơn. Điều này không chỉ giúp tự động hóa quy trình, mà còn giảm bớt công sức và thời gian so với việc thủ công. Bài toán này không chỉ làm tăng hiệu suất và độ chính xác trong việc xử lý dữ liệu văn bản từ hình ảnh mà còn mở ra nhiều ứng dụng và hướng nghiên cứu mới trong thực tế, từ tự động hóa công việc văn phòng đến việc quản lý thông tin từ hàng loạt hình ảnh, đồng thời đảm bảo tính minh bạch và chính xác trong việc thu thập và xử lý thông tin văn bản.

1.2 Phân tích yêu cầu của bài toán

1.2.1 Yêu cầu của bài toán

Bài toán xác định thông tin văn bản sử dụng phương pháp học sâu là bài toán quan trọng trong xử lý ngôn ngữ tự nhiên nhằm tìm kiếm và trích xuất thông tin cần thiết từ văn bản. Bài toán này có các yêu cầu sau: Dữ liệu đầu vào của bài toán là văn bản, có thể là hình ảnh hoặc file pdf chứa văn bản (Các văn bản của Khoa, Trường). Bài toán này có thể được giải quyết bằng các phương pháp học sâu kết hợp xử lý hình ảnh, thị giác máy tính và xử lý ngôn ngữ tự nhiên. Sử dụng kỹ thuật xử lý ảnh để có thể làm loại

bỏ nhiễu, tăng cường chất lượng cho dữ liệu đầu vào. Phương pháp học sâu được sử dụng để xây dựng các mô hình phân tích các đặc trưng và cấu trúc của dữ liệu đầu vào. Cuối cùng sử dụng phương pháp xử lý ngôn ngữ tự nhiên để tiến hành trích xuất các vùng văn bản được xác định thành text. Đầu ra mong muốn là text của các vùng được xác định trong văn bản như: nơi phát hành văn bản, ngày phát hành, tiêu đề văn bản,… Bài toán cũng yêu cầu về độ chính xác của mô hình, tốc độ xử lý và khả năng mở rộng.

1.2.2 Các phương pháp giải quyết bài toán

Bài báo "Information Extraction from Receipts using Spectral Graph Convolutional Network" (2021) [1] của tác giả Bui Thanh Hung tập trung vào việc trích xuất thông tin từ hóa đơn đã quét bằng cách sử dụng mạng tích chập đồ thị phổ (Spectral Graph Convolutional Network). Bài báo này cung cấp một cái nhìn tổng quan về các phương pháp trích xuất thông tin từ hóa đơn, bao gồm các phương pháp dựa trên mẫu và dựa trên xử lý ngôn ngữ tự nhiên (NLP). Bài báo cũng giới thiệu về ưu điểm của mạng tích chập đồ thị phổ và cách nó có thể được áp dụng để giải quyết vấn đề trích xuất thông tin từ hóa đơn.

Bài báo nghiên cứu "An Invoice Reading System Using a Graph Convolutional Network" (2019) [2] của D. Lohani, A. Belaïd và Y. Belaïd giới thiệu một hệ thống đọc hóa đơn sử dụng mạng convolutional trên đồ thị (GCN). Hệ thống này có thể đọc hóa đơn được số hóa với độ chính xác cao, ngay cả khi hóa đơn có bố cục khác nhau. Hệ thống sử dụng GCN để học các thông tin cấu trúc và ngữ nghĩa của các thực thể trong hóa đơn. Hệ thống không yêu cầu bất kỳ thông tin định dạng hóa đơn cụ thể nào.

Các bài nghiên cứu đã được giới thiệu trước đây đa phần đều sử dụng Graph Convolution Network để thực hiện trích xuất thông tin, học những biểu diễn của đồ thị tuy nhiên GCN vẫn còn nhiều điểm hạn chế trong việc xử lý đồ thị. GCN phụ thuộc vào cấu trúc đồ thị của dữ liệu đầu vào. Nếu dữ liệu đầu vào không có cấu trúc đồ thị rõ ràng, GCN sẽ gặp khó khăn trong việc học các mối quan hệ giữa các thực thể. GCN có thể bị quá tải bởi dữ liệu. Nếu tập dữ liệu huấn luyện quá lớn, GCN có thể gặp khó khăn trong

việc tìm các mẫu trong dữ liệu. Nếu tập dữ liệu huấn luyện có thiên vị, GCN có thể học các mô hình không chính xác. Các bài nghiên cứu trước đây tập trung vào sử dụng mô hình GCN để trích xuất các thông tin từ hóa đơn.

Trong đồ án này chúng tôi sử dụng nhiều kỹ thuật học sâu để thực hiện giải quyết bài toán trích xuất thông tin từ hình ảnh văn bản hoặc file pdf. Dữ liệu về hình ảnh văn bản hoặc file pdf thường có cấu trúc phức tạp, nhiều thông tin hơn các thông tin chứa trong hình ảnh hóa đơn. Chúng tôi thực hiện xử lý, xác định vùng chứa thông tin văn bản và gán nhãn tương ứng cho mỗi vùng. Sau đó thực hiện huấn luyện các mô hình học sâu như GCN, GAT, GraphSAGE, PNAConv, SGConv. Từ đó có cái nhìn tổng quan và những đánh giá, so sánh giữa các mô hình một cách chi tiết. Chọn ra một mô hình tốt nhất để thực hiện dự đoán và trích xuất thông tin từ các vùng đã được dự đoán từ mô hình.

1.2.3 Phương pháp đề xuất giải quyết bài toán

Bài toán xác định thông tin văn bản từ hình ảnh là một thách thức đầy thú vị, đặc biệt là khi chúng ta đối mặt với các hình ảnh có độ phức tạp cao và nhiều yếu tố khác nhau. Để giải quyết vấn đề này, chúng tôi đề xuất một phương pháp tích hợp các kỹ thuật xử lý dữ liệu, mô hình hóa đồ thị, huấn luyện mô hình học sâu và sử dụng công cụ OCR cho Tiếng Việt để thực hiện nhiệm vụ trích xuất text từ vùng văn bản được xác định.

- Xử lý dữ liệu và gán nhãn: Thu thập một tập dữ liệu là hình ảnh hoặc file pdf chứa các văn bản của Khoa, Trường. Tiến hành xử lý, xác định vùng chữ và gán nhãn cho mỗi vùng chữ. Ở bước này chúng tôi thực hiện gán nhãn và xác định vùng chữ bằng phương pháp thủ công.
- Mô hình hóa đồ thị (Graph Modeling): Sử dụng mô hình đồ thị để biểu diễn mối quan hệ giữa các thành phần trong hình ảnh, chẳng hạn như vị trí, kích thước, và mối quan hệ văn bản với các đối tượng khác.
- Huấn luyện mô hình học sâu: Chia tập dữ liệu thành tập huấn luyện và tập kiểm tra. Áp dụng GNN để học thông tin từ cấu trúc đồ thị, giúp mô hình hiểu được

mối quan hệ phức tạp giữa các thành phần. Sử dụng kỹ thuật Graph Embedding để chuyển đổi đồ thị thành biểu diễn vector, giúp đưa thông tin vào mạng học sâu. Tiến hành huấn luyện mô hình trên tập dữ liệu đã được gán nhãn sử dụng kỹ thuật lan truyền ngược (backpropagation) và tối ưu hóa hàm mất mát. Đánh giá mô hình trên tập kiểm tra để đảm bảo hiệu suất chính xác và tổng quát.

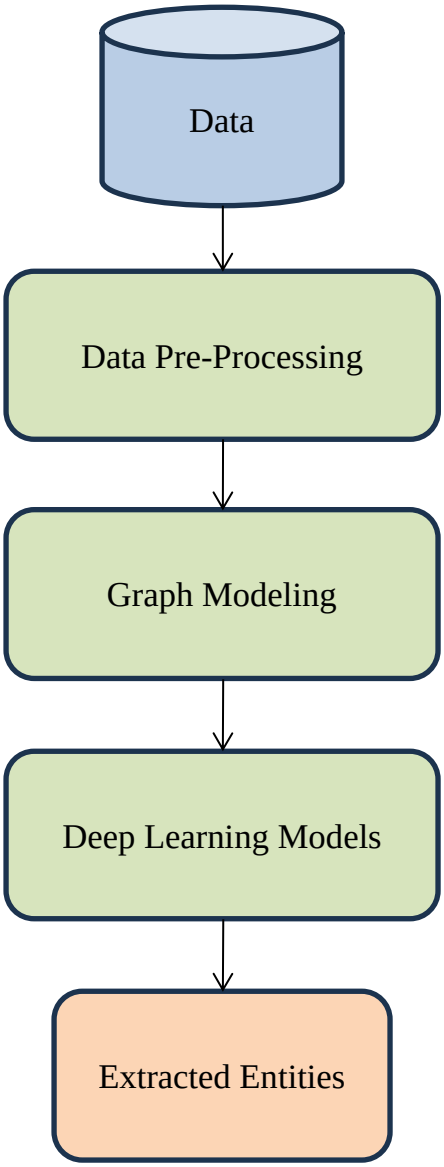
- Sử dụng công cụ OCR: Áp dụng các công cụ OCR để trích xuất văn bản từ vùng chữ văn bản được dự đoán bởi mô hình học sâu. Tối ưu hóa cài đặt OCR để xử lý các trường hợp đặc biệt, chẳng hạn như văn bản nghiêng, kích thước nhỏ, hoặc độ phân giải thấp.

Kết hợp các phương pháp này giúp chúng ta xác định thông tin văn bản từ hình ảnh một cách hiệu quả, đồng thời tận dụng cả sức mạnh của mô hình học sâu và đồ thị để hiểu mối quan hệ phức tạp trong dữ liệu hình ảnh.

1.3 Phương pháp giải quyết bài toán

1.3.1 Mô hình tổng quát

Từ tập dữ dữ liệu ban đầu (các file pdf chứa văn bản) chúng tôi tiến hành tiền xử lý dữ liệu. Chúng tôi thực hiện chuyển từ file pdf sang ảnh. Tiếp theo tiến hành xác định các vùng chứa văn bản cần trích xuất nội dung, ở phần này chúng tôi thực hiện một cách thủ công để xác định vùng chứa nội dung và gán nhãn cho các vùng đó. Sau đó chúng tôi Graph Modeling (mô hình hóa đồ thị) cho dữ liệu đã được xử lí và đưa vào mô hình Deep Learning để tiến hành huấn luyện. Sau đó chúng tôi thực hiện trích xuất text của các vùng được dự đoán bởi mô hình Deep Learning. Mô hình tổng quát được trình bày ở Hình 1.



Hình 1: Mô hình tổng quát

1.3.2 Đặc trưng của mô hình đề xuất

1.3.2.1. Graph Modeling

Trong đồ án này, chúng tôi thực hiện mô hình hóa đồ thị trên các dữ liệu đã được xử lý bao gồm gán nhãn và xác định các vùng chứa thông tin văn bản. Cụ thể, các bước thực hiện như sau:

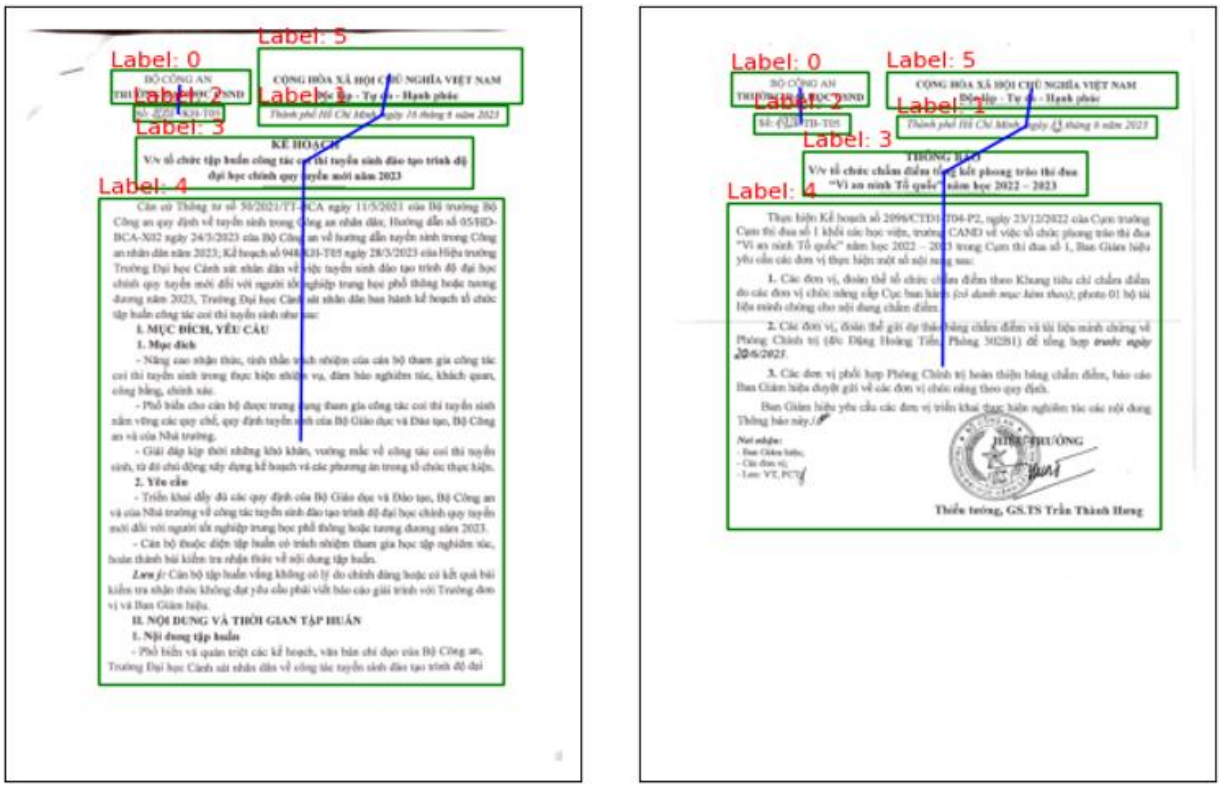
- Khởi tạo các nodes: Ban đầu, chúng tôi khởi tạo các nodes của đồ thị là các vùng dữ liệu đã được gán nhãn. Mỗi nhãn được gán trên vùng được xác định tương ứng với một node trong đồ thị. Ví dụ, chúng ta có một tập dữ liệu hình ảnh chứa các văn bản thì sẽ được gán nhãn như sau (ví dụ: nhãn "Nơi phát hành văn bản", "Ngày phát hành văn bản", "Tiêu đề văn bản",...). Khi đó, mỗi vùng chứa văn bản sẽ tương ứng với một node trong đồ thị, và nhãn của vùng đó sẽ là nhãn của node.
- Kết nối các nodes với nhau: Tiếp đến, chúng tôi thực hiện kết nối các nodes của đồ thị với nhau dựa trên tính toán khoảng cách giữa các bounding box để thêm các cạnh tương ứng vào đồ thị. Cách kết nối đồ thị dựa trên tiêu chí các nodes gần nhau nhất sẽ được kết nối với nhau. Cụ thể, chúng tôi sử dụng công thức Euclidean để tính toán khoảng cách giữa các bounding box. Công thức này được tính như sau:

$$\boldsymbol{d(a,b) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}} \ \boldsymbol{(1)}$$

Trong đó:

- $\boldsymbol{d(a,b)}$ là khoảng cách từ bounding box a đến b
- $\boldsymbol{x_1, y_1}$ là tọa độ của bounding box a
- $\boldsymbol{x_2, y_2}$ là tọa độ của bounding box b

Kết quả của mô hình hóa đồ thị là một đồ thị có các nodes và edges. Các nodes của đồ thị đại diện cho các vùng dữ liệu đã được gán nhãn, và các edges của đồ thị đại diện cho mối quan hệ giữa các nodes. Hình 2 minh họa cách kết nối các nodes của đồ thị.



Hình 2: Văn bản đã được mô hình hóa đồ thị

1.3.2.2. Deep Learning Models

Sau khi có dữ liệu đã được mô hình hóa đồ thị, chúng tôi tiến hành chia tập dữ liệu thành hai tập: tập train và tập test. Tập train là tập dữ liệu được sử dụng để huấn luyện mô hình, còn tập test là tập dữ liệu được sử dụng để đánh giá hiệu quả của mô hình sau khi được huấn luyện.Chúng tôi thực hiện huấn luyện trện các mô hình Graph Neural Network (GNN) và mô hình Graph Embemdding (GE) sau đó chúng tôi thực hiện so sánh và đánh giá kết quả.

- **Mô hình Graph Neural Network**

- *Graph Convolutional Network (GCN) [3]:*Được giới thiệu lần đầu tiên trong bài báo "Semi-Supervised Classification with Graph Convolutional Networks" của Michael Bronstein, Joan Bruna, Yan LeCun, Arthur Szlam và

Pierre Vandergheynst, được xuất bản trên tạp chí IEEE Transactions on Pattern Analysis and Machine Intelligence vào năm 2017. GCN được định nghĩa là một mô hình học sâu có các lớp ẩn là các phép biến đổi tích chập trên đồ thị. Mỗi lớp ẩn của GCN nhận đầu vào là một ma trận đặc trưng của các đỉnh trong đồ thị, và đầu ra là một ma trận đặc trưng mới của các đỉnh trong đồ thị. GCN hoạt động bằng cách tính toán các phép biến đổi tích chập trên ma trận đặc trưng của các đỉnh trong đồ thị. GCN được áp dụng rộng rãi trong các bài toán xử lý dữ liệu đồ thị, chẳng hạn như phân loại, hồi quy, và phát hiện cộng đồng.

Công thức tổng quát của GCN:

$$\boldsymbol{X_l = f(A, X_{l-1}, W_1)\ (2)}$$

Trong đó:

- $\boldsymbol{X_l}$ là ma trận đặc trưng của các đỉnh trong đồ thị sau khi thực hiện phép biến đổi tích chập $\boldsymbol{l}$ lần.
- $\boldsymbol{X_{l-1}}$ là ma trận đặc trưng của các đỉnh trong đồ thị sau khi thực hiện phép biến đổi tích chập $\boldsymbol{l-1}$ lần
- $\boldsymbol{A}$ là ma trận adjacency của đồ thị.
- $\boldsymbol{W_1}$ là ma trận trọng số của lớp ẩn thứ $\boldsymbol{l}$.
- $\boldsymbol{f}$ là hàm biến đổi tích chập.

- *Graph Attention Network (GAT) [4]:* là một loại mạng nơ-ron đồ thị (GNN) sử dụng cơ chế chú ý (attention mechanism) để xác định mức độ quan trọng của các hàng xóm đối với mỗi đỉnh trong đồ thị. Bằng cách trọng số hóa thông tin từ các hàng xóm, GAT tập trung vào các đỉnh quan trọng và xử lý đồ thị một cách linh hoạt. GAT được giới thiệu lần đầu tiên trong bài báo "Graph Attention Networks" của Velickovic et al. (2017). Bài báo này đã chứng minh rằng GAT có thể vượt trội so với các phương pháp GNN truyền thống, chẳng hạn như Graph Convolution Networks (GCNs), trong các bài toán như phân

loại, hồi quy và dự đoán liên kết. GAT được xây dựng dựa trên kiến trúc của GCNs. Tuy nhiên, thay vì sử dụng các phép tích chập thông thường, GAT sử dụng các phép tính chú ý để xác định mức độ quan trọng của các hàng xóm đối với mỗi đỉnh.

Cụ thể, GAT có thể được biểu diễn dưới dạng sau:

$$\boldsymbol{h_i^{l+1}} = \boldsymbol{\sigma}\left(\sum_{\boldsymbol{j \in N_i}} \boldsymbol{\alpha_{ij}^l W_a h_j^l}\right) \textbf{(3)}$$

Trong đó:

- $\boldsymbol{h_j^l}$ là giá trị của đỉnh $\boldsymbol{i}$ ở lớp $\boldsymbol{l}$
- $\boldsymbol{W_a}$ là ma trận trọng số của attention layer
- $\boldsymbol{\alpha_{ij}^l}$ là trọng số chú ý của đỉnh $\boldsymbol{j}$ đối với đỉnh $\boldsymbol{i}$ ở lớp $\boldsymbol{l}$
- $\boldsymbol{N_i}$ là tập hợp các hàng xóm của đỉnh $\boldsymbol{i}$
- $\boldsymbol{\sigma}$ là hàm kích hoạt

GAT là một mô hình GNN mạnh mẽ có thể được áp dụng trong nhiều bài toán trên dữ liệu đồ thị. GAT có thể vượt trội so với các phương pháp GNN truyền thống trong các bài toán đòi hỏi khả năng xử lý thông tin từ các đỉnh lân cận một cách linh hoạt.

- *Principal Neighbourhood Aggregation for Graph Nets (PNAConv) [5]:* là một kiến trúc mạng thần kinh đồ thị (GNN) được giới thiệu trong bài báo "Principal Neighbourhood Aggregation for Graph Nets" vào năm 2020. Mục tiêu chính của PNAConv là giải quyết các vấn đề về hiệu quả và độ phức tạp tính toán của các GNN truyền thống, đồng thời duy trì hiệu quả học tập. PNAConv thay thế phép biến đổi tích chập truyền thống bằng một phương pháp tổng hợp thông tin hiệu quả hơn dựa trên phân tích thành phần chính (PCA). Ý tưởng chính là: Trích xuất các thành phần chính từ ma trận đặc trưng của lân cận của một nút. Các thành phần chính này nắm bắt được thông tin quan trọng nhất từ

các lân cận. Tổng hợp thông tin từ các thành phần chính này để tạo ra một biểu diễn mới cho nút.

PNAConv có một số ưu điểm chính so với các GNN truyền thống:

- Hiệu quả tính toán: PNAConv sử dụng PCA, một phương pháp tính toán hiệu quả, để tổng hợp thông tin từ các lân cận. Điều này giúp PNAConv có thể xử lý được các đồ thị lớn hơn với độ loãng cao.
- Tính chính xác: PNAConv đã đạt được hiệu quả học tập tương đương hoặc thậm chí tốt hơn so với các GNN truyền thống trong nhiều bài toán khác nhau.
- Tính linh hoạt: PNAConv có thể được sử dụng với nhiều loại mạng thần kinh khác nhau, chẳng hạn như mạng CNN và RNN.

PNAConv cũng có một số nhược điểm:

- Giảm độ chi tiết: Do PCA tóm tắt thông tin bằng một số thành phần chính hạn chế, một số chi tiết của thông tin lân cận có thể bị mất.
- Khả năng giải thích kém: PNAConv hoạt động như một hộp đen, khiến việc giải thích các quyết định của nó trở nên khó khăn.

Công thức tổng quát:

$$\boldsymbol{X_l = f(X_{l-1}, P_l)}\ \boldsymbol{(4)}$$

Trong đó

- $\boldsymbol{X_l}$ là ma trận đặc trưng của các đỉnh trong đồ thị sau khi thực hiện phép biến đổi PNAConv $\boldsymbol{l}$ lần.
- $\boldsymbol{X_{l-1}}$ là ma trận đặc trưng của các đỉnh trong đồ thị sau khi thực hiện phép biến đổi PNAConv $\boldsymbol{l-1}$ lần.
- $\boldsymbol{P_l}$ là ma trận trọng số của lớp PNAConv thứ $\boldsymbol{l}$.
- $\boldsymbol{f}$ là hàm biến đổi tích chập.

PNAConv là một kiến trúc GNN mới hứa hẹn giải quyết các vấn đề về hiệu quả và độ phức tạp tính toán của các GNN truyền thống. PNAConv đã đạt

được kết quả ấn tượng trong nhiều bài toán khác nhau và có tiềm năng rộng rãi cho các ứng dụng xử lý dữ liệu đồ thị.

- *Simplifying Graph Convolutional Networks (SGConv) [6]:* là một mô hình học sâu đơn giản và hiệu quả cho các bài toán xử lý dữ liệu đồ thị. Được giới thiệu trong bài báo "Simplifying Graph Convolutional Networks" vào năm 2019, SGConv hướng đến việc giảm thiểu độ phức tạp của các GCN truyền thống, đồng thời vẫn duy trì hiệu suất tốt. SGConv giải quyết những vấn đề này bằng cách đơn giản hóa thiết kế của GCN: Loại bỏ các phép biến đổi tích chập: SGConv loại bỏ hoàn toàn các phép biến đổi tích chập. Thay vào đó, nó sử dụng một phương pháp đơn giản hơn để tổng hợp thông tin từ các lân cận. Sử dụng ma trận độ che phủ bậc hai: SGConv sử dụng ma trận độ che phủ bậc hai (power-law degree matrix) để tổng hợp thông tin từ các lân cận. Ma trận này đơn giản hơn và hiệu quả hơn so với ma trận Laplacian thường được sử dụng trong các GCN truyền thống. Sử dụng hồi quy tuyến tính: SGConv sử dụng hồi quy tuyến tính đơn giản để dự đoán nhãn của một nút. Điều này giúp SGConv dễ dàng giải thích hơn các GCN truyền thống.

  SGConv có một số ưu điểm chính so với các GCN truyền thống:

  - Hiệu quả tính toán cao: SGConv đơn giản hơn và hiệu quả hơn về mặt tính toán so với các GCN truyền thống. Điều này cho phép nó xử lý các đồ thị lớn hơn với độ loãng cao.
  - Khả năng giải thích cao: SGConv sử dụng hồi quy tuyến tính đơn giản, giúp cho việc giải thích các quyết định của nó trở nên dễ dàng hơn.
  - Dễ dàng điều chỉnh: SGConv có ít siêu tham số cần điều chỉnh hơn so với các GCN truyền thống.

  SGConv cũng có một số nhược điểm:

- Hiệu suất học tập thấp hơn: SGConv đôi khi có thể có hiệu suất học tập thấp hơn so với các GCN truyền thống trong một số bài toán.
- Không phù hợp cho các bài toán phức tạp: SGConv đơn giản hơn và có thể không phù hợp cho các bài toán phức tạp đòi hỏi các mô hình GCN mạnh mẽ hơn.

Công thức tổng quát:

$$\boldsymbol{X_l = X_{l-1} W_l} \ (\mathbf{5})$$

Trong đó:

- $\boldsymbol{X_l}$ là ma trận đặc trưng của các đỉnh trong đồ thị sau khi thực hiện phép biến đổi SGConv $\boldsymbol{l}$ lần.
- $\boldsymbol{X_{l-1}}$ là ma trận đặc trưng của các đỉnh trong đồ thị sau khi thực hiện phép biến đổi SGConv $\boldsymbol{l-1}$ lần.
- $\boldsymbol{W_l}$ là ma trận trọng số của lớp SGConv thứ $\boldsymbol{l}$.
- SGConv là một mô hình GCN đơn giản và hiệu quả. Nó là một lựa chọn tốt cho các bài toán xử lý dữ liệu đồ thị đòi hỏi hiệu quả tính toán cao và khả năng giải thích tốt. Tuy nhiên, cần lưu ý rằng SGConv có thể có hiệu suất học tập thấp hơn so với các GCN truyền thống trong một số bài toán.

- **Mô hình Graph Embemdding**

- *Graph SAGE [7]:*là một phương pháp học biểu diễn (representation learning) trên đồ thị lớn, được giới thiệu trong bài báo "Inductive Representation Learning on Large Graphs" tại NIPS 2017.

  Phương pháp này có những đặc điểm nổi bật sau:

  - Học biểu diễn theo hướng quy nạp (inductive):
    - Khác với các phương pháp học biểu diễn truyền thống, GraphSAGE không cần phải biết trước toàn bộ đồ thị để học biểu diễn cho các nút.

  - Thay vào đó, phương pháp này chỉ cần biết thông tin về các nút lân cận của từng nút. Điều này cho phép GraphSAGE có khả năng học biểu diễn cho các nút trong các đồ thị lớn và linh động hơn.
- Khả năng tổng hợp thông tin từ các nút lân cận:
  - GraphSAGE sử dụng một phương pháp tổng hợp thông tin (aggregation) hiệu quả để tổng hợp thông tin từ các nút lân cận của một nút.
  - Phương pháp này có thể được áp dụng cho các loại đồ thị khác nhau và có thể được điều chỉnh để phù hợp với các nhiệm vụ cụ thể.
- Hiệu quả tính toán:
  - GraphSAGE có thể được tính toán hiệu quả trên các đồ thị lớn.
  - Phương pháp này sử dụng một kỹ thuật lấy mẫu (sampling) để chỉ cần xử lý một phần nhỏ của đồ thị trong mỗi lần lặp lại của quá trình học.
- Hiệu quả trên các tác vụ phân loại nút:
  - GraphSAGE đã được chứng minh là có hiệu quả trên các tác vụ phân loại nút (node classification) trên các đồ thị lớn.
  - Phương pháp này đã đạt được kết quả tốt hơn so với các phương pháp học biểu diễn truyền thống.

Công thức tổng quát:

$$\boldsymbol{h_i^{l+1} = \sigma\left(\sum_{j \in N(i)} \phi(h_j^l, A_{ij})\right)} \quad \boldsymbol{(6)}$$

Trong đó:

- $\boldsymbol{h_i^l}$ là biểu diễn của nút $\boldsymbol{i}$ ở tầng $\boldsymbol{l}$
- $\boldsymbol{h_j^l}$ là biểu diễn của nút $\boldsymbol{j}$ ở tầng $\boldsymbol{l}$
- $\boldsymbol{A_{ij}}$ là trọng số của cạnh nối nút $\boldsymbol{i}$ và nút $\boldsymbol{j}$
- $\boldsymbol{\phi}$ là hàm tổng hợp thông tin

- $\boldsymbol{\sigma}$ là hàm kích hoạt

GraphSAGE là một phương pháp học biểu diễn tiên tiến cho đồ thị lớn, có khả năng học biểu diễn theo hướng quy nạp, tổng hợp thông tin hiệu quả từ các nút lân cận, tính toán hiệu quả và hiệu quả trên các tác vụ phân loại nút.

Trong đồ án này chúng tôi tiến hành huấn luyện trên 5 mô hình Graph khác nhau. Sau khi có kết quả được huấn luyện chúng tôi tiến hành thực hiện so sánh và đánh giá kết quả giữa các mô hình. Từ đó chọn ra mô hình tối ưu và phù hợp nhất cho bài toán xác định thông tin văn bản hình ảnh dựa trên phương pháp học sâu.

1.3.2.3. Extracted Entities

Sau khi đã huấn luyện, lưu lại các tham số mô hình và thực hiện các bước so sánh và đánh giá kết quả chúng tôi sẽ chọn ra mô hình tốt nhất trong các phương pháp đã đề xuất ở trên để tiến hành việc dự đoán nhãn. Sau khi có được các nhãn dự đoán, chúng tôi tiến hành cắt mỗi bounding box từ hình ảnh dự đoán sau đó thực hiện trích xuất văn bản trong vùng bounding box đó. Chi tiết phần dự đoán và gán nhãn dự đoán xem ở Hình 3.

Hình 3: Văn bản được gán nhãn sau khi thực hiện dự đoán từ mô hình

Sau đó chúng tôi sử dụng các công cụ trích xuất văn bản Tiếng Việt là VietOCR, PaddleOCR để thực hiện trích xuất nội dung trong các vùng bounding chứa nội dung văn bản bao gồm: Nơi phát hành văn bản, Ngày phát hành văn bản, Tiêu đề văn bản,… và các thông tin liên quan khác.

- PaddleOCR [8]: được phát triển bởi đội ngũ nghiên cứu của PaddlePaddle, một nền tảng phát triển AI mã nguồn mở do Baidu AI phát triển. Đội ngũ nghiên cứu bao gồm các chuyên gia từ nhiều lĩnh vực khác nhau, bao gồm nhận dạng hình ảnh, nhận dạng chữ viết tay, xử lý ngôn ngữ tự nhiên, v.v.

  PaddleOCR bao gồm ba giai đoạn chính:

  - Phát hiện văn bản: Sử dụng thuật toán để xác định vị trí của văn bản trong ảnh.
  - Phân loại hướng văn bản: Sử dụng thuật toán để xác định hướng của văn bản (ngang, dọc, nghiêng).

- Nhận dạng văn bản: Sử dụng thuật toán CRNN để nhận dạng các ký tự trong văn bản đã được xác định.

Ưu điểm:

- Cực kỳ nhẹ và hiệu quả, phù hợp cho các thiết bị edge.
- Hỗ trợ nhiều ngôn ngữ và có độ chính xác cao.
- Dễ dàng triển khai và tích hợp vào các ứng dụng khác.
- Mã nguồn mở và có cộng đồng phát triển lớn.

Nhược điểm

- Hiệu suất có thể không tốt bằng các hệ thống OCR phức tạp hơn.
- Có thể gặp khó khăn với các văn bản phức tạp hoặc chất lượng thấp.

PaddleOCR là một hệ thống OCR thực tế và hiệu quả cho các ứng dụng trên thiết bị edge. Nó cung cấp sự cân bằng tốt giữa kích thước, hiệu suất và độ chính xác.

- VietOCR [9]: là một hệ thống nhận dạng quang học ký tự tiếng Việt được phát triển bởi Trung tâm Nghiên cứu Trí tuệ Nhân tạo và Xử lý Ngôn ngữ (LIRN) thuộc Đại học Quốc gia Hà Nội. VietOCR được xây dựng dựa trên kiến trúc mạng nơ-ron CNN và RNN, được đào tạo trên một tập dữ liệu khổng lồ gồm 10 triệu mẫu ảnh chụp văn bản tiếng Việt. VietOCR có khả năng nhận dạng chính xác các ký tự tiếng Việt, kể cả các ký tự đặc biệt như dấu thanh, dấu mũ, dấu móc. Hệ thống cũng có thể nhận dạng các ký tự tiếng Việt viết tay, viết tắt, viết hoa, viết thường.

  Ưu điểm của VietOCR:

  - Khả năng nhận dạng chính xác cao, kể cả các ký tự đặc biệt
  - Có thể nhận dạng các ký tự tiếng Việt viết tay, viết tắt, viết hoa, viết thường
  - Có thể sử dụng trong nhiều ứng dụng thực tế

  Nhược điểm của VietOCR:

- Hệ thống vẫn còn một số lỗi nhận dạng, đặc biệt là đối với các ký tự viết tay khó đọc
- Hệ thống cần được đào tạo trên một tập dữ liệu lớn để đảm bảo chất lượng nhận dạng

Ứng dụng:

- Nhận dạng văn bản tiếng Việt từ ảnh chụp
- Tự động hóa các quy trình xử lý văn bản tiếng Việt
- Tạo ra các sản phẩm phần mềm hỗ trợ tiếng Việt

Nhìn chung, VietOCR là một hệ thống nhận dạng quang học ký tự tiếng Việt có chất lượng cao, được ứng dụng rộng rãi trong thực tế.

Tuy VietOCR là một công cụ nhận dạng thuần cho văn bản tiếng việt nhưng công cụ này vẫn còn một số hạn chế nhất định như chỉ nhận dạng được chữ trên một hàng, đôi khi nhận dạng chưa chính xác vùng mong muốn.

Vì vậy trong đồ án này chúng tôi sử dụng Nhận dạng ký tự quang học (OCR) trong văn bản tiếng Việt. Nó sử dụng các framework PaddleOCR và VietOCR để đạt được điều này. PaddleOCR là một khung OCR phổ biến cung cấp một loạt các mô hình và công cụ OCR. VietOCR là framework phổ biến cho tác vụ OCR tiếng Việt, dựa trên kiến trúc OCR Transformer OCR. Khi kết hợp giữa PanddleOCR, VietOCR thì việc trích xuất trở nên hiệu quả và chính xác hơn. Đồng thời có thể nhận diện trên một vùng lớn của hình ảnh văn bản không bị giới hạn bởi một hàng duy nhất.

1.4 Thực nghiệm

1.4.1 Dữ liệu

Trong đồ án này, chúng tôi đã sử dụng một tập dữ liệu tự thu thập gồm 30 file PDF chứa các văn bản liên quan đến Khoa và Trường. Việc tự thu thập dữ liệu nhằm mục đích tối ưu hóa nguồn thông tin và đảm bảo rằng dữ liệu thu thập đáp ứng đúng nhu cầu nghiên cứu của chúng tôi.

Chúng tôi cũng đã thực hiện việc kiểm tra và xác nhận tính chính xác của dữ liệu thu thập. Điều này giúp đảm bảo rằng chúng tôi có một nguồn dữ liệu đáng tin cậy để thực hiện các phân tích và mô hình hóa trong quá trình thực hiện đồ án của mình. Hình 4 mô tả một số hình ảnh trong tập dữ liệu.

Hình 4: Một số hình ảnh trong bộ dữ liệu

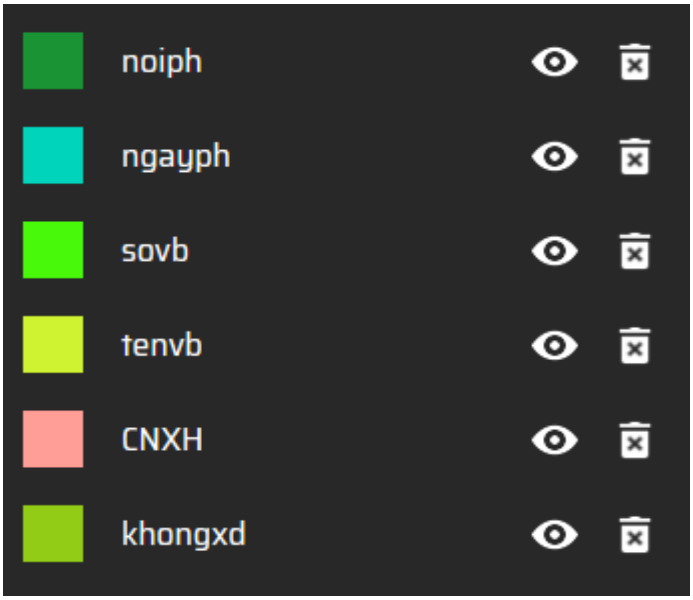
Ngoài ra chúng tôi cần thực hiện các bước tiền xử lý dữ liệu trước khi đưa vào để thực hiện mô hình hóa đồ thị và huấn luyện mô hình học sâu. Chi tiết về tiền xử lý dữ liệu sẽ được trình bày ở phần sau.

1.4.2 Xử lý dữ liệu

Như đã đề cập ở phần trên vì đây là tập dữ liệu thô nên chúng tôi cần thực hiện các bước tiền xử lý dữ liệu trước khi thực hiện và huấn luyện mô hình học sâu.

Đầu tiên chúng tôi chuyển tập dữ liệu ban đầu là các file pdf về thành các hình ảnh tương ứng. Sau đó chúng tôi thực hiện việc xác định các vùng văn bản (bounding box) và gán nhãn tương ứng cho mỗi vùng. Ở bước này chúng tôi thực hiện một cách thủ công trên toàn bộ dữ liệu, sử dụng công cụ Make Sense [10] để thực hiện bounding box và gán nhãn cho dữ liệu văn bản hình ảnh. Hình 5 là một số hình ảnh được gán nhãn và xác định vùng. Hình 6 mô tả nhãn của dữ liệu đối với từng vùng.

Hình 5: Một số hình ảnh dữ liệu đã được gán nhãn



Hình 6: Nhãn của các vùng

Các nhãn của từng vùng được định nghĩa như sau:

- Nơi phát hành văn bản (noiph): label 0
- Ngày phát hành văn bản (ngayph): label 1
- Số văn bản (sovb): label 2
- Tiêu đề văn bản (tenvb): label 3

- Quốc hiệu và tiêu ngữ (CNXH): label 4
- Văn bản chính (khongxd): label 5

Sau khi đã thực hiện xác định từng vùng văn bản và gán nhãn tương ứng chúng tôi tiến hành xây dựng và huấn luyện mô hình học sâu. Sau đó dự đoán kết quả và trích xuất text của các vùng đã được xác định.

1.4.3 Công nghệ sử dụng

Ngôn ngữ	Python
Thư viện	Pytorch [11], Torch Geometric [12]
Công cụ	VietOCR, PaddleOCR

Bảng 1: Công nghệ sử dụng

1.4.4 Cách đánh giá

1.4.4.1 Cross Entropy Loss

Cross Entropy [13] là một khái niệm mở rộng của entropy lý thuyết thông tin bằng cách đo lường sự biến thiên giữa hai phân bố xác suất đối với một biến hoặc một tập hợp các lần xuất hiện ngẫu nhiên nhất định.

Cross-Entropy loss [14] được sử dụng khi điều chỉnh trọng số mô hình trong quá trình huấn luyện. Mục đích là tối thiểu hóa hàm mất mát này–giá trị loss càng nhỏ thì mô hình càng tốt. Một mô hình hoàn hảo thường có giá trị loss bằng 0. Thông thường, nó thường được sử dụng trong các bài toán phân loại đa lớp và đa nhãn.

Cross-Entropy loss đo lường sự khác biệt giữa phân bố xác suất được phát hiện và dự đoán của mô hình phân loại học sâu. Cụ thể Cross-Entropy loss được định nghĩa bằng công thức (1):

$$\boldsymbol{CE} = -\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{y_i} \log \boldsymbol{p}(\boldsymbol{y_i}) \qquad (7)$$

1.4.4.2 Accuracy

Độ chính xác (Accuracy) [15] được đánh giá bằng phần trăm dự đoán chính xác của mẫu kiểm thử trên tổng số mẫu được đem đi kiểm thử được thể hiện dưới công thức (2) dưới đây:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{Correct\ predictions}}{\boldsymbol{All\ predictions}} \qquad (\mathbf{8})$$

1.4.4.3 Recall và Precision

Precision [16] và Recall [17] là hai chỉ số quan trọng được sử dụng để đánh giá hiệu suất của một mô hình phân loại. Chúng thường được sử dụng trong các bài toán phân loại nhị phân, trong đó dữ liệu được chia thành hai lớp: lớp positive và lớp negative.

Recall được định nghĩa là tỷ lệ số điểm True Positive trong số những điểm thực sự là Positive (TP + FN):

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}} \qquad (\mathbf{9})$$

Precision được định nghĩa là tỉ lệ số điểm True Positive trong số tất cả các điểm được mô hình dự đoán là Positive (TP + FP):

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}} \qquad (\mathbf{10})$$

Precision và Recall là hai chỉ số bổ sung cho nhau. Một mô hình có Precision cao sẽ có ít điểm False Positive, nhưng có thể bỏ sót một số điểm True Positive. Một mô hình có Recall cao sẽ có ít điểm False Negative, nhưng có thể có nhiều điểm False Positive.

1.4.4.4 F1-Score

F1-Score [18] là một chỉ số đánh giá hiệu suất của một mô hình phân loại nhị phân. Nó là một chỉ số tổng hợp của độ chính xác (precision) và độ nhạy (recall). F1-Score có giá trị từ 0 đến 1, với giá trị càng cao thì mô hình càng tốt. F1-Score được tính bằng công thức (5):

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (11)$$

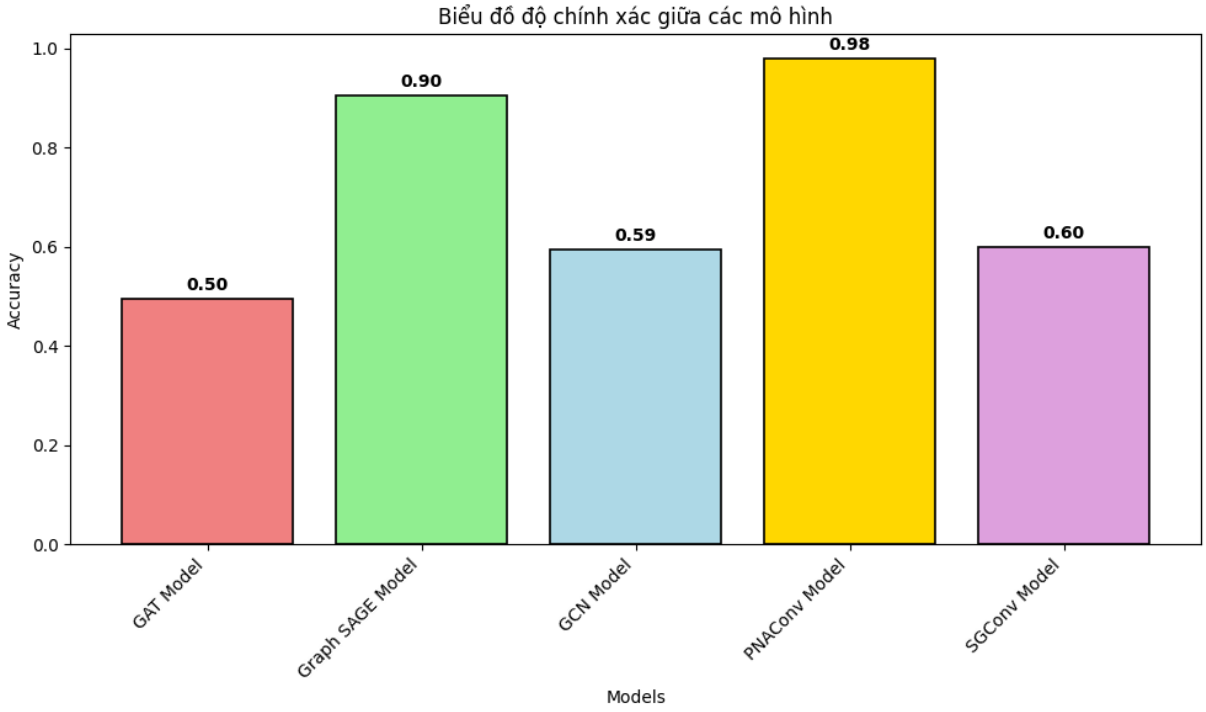
1.5 Kết quả đạt được

1.5.1 Hyperparameter

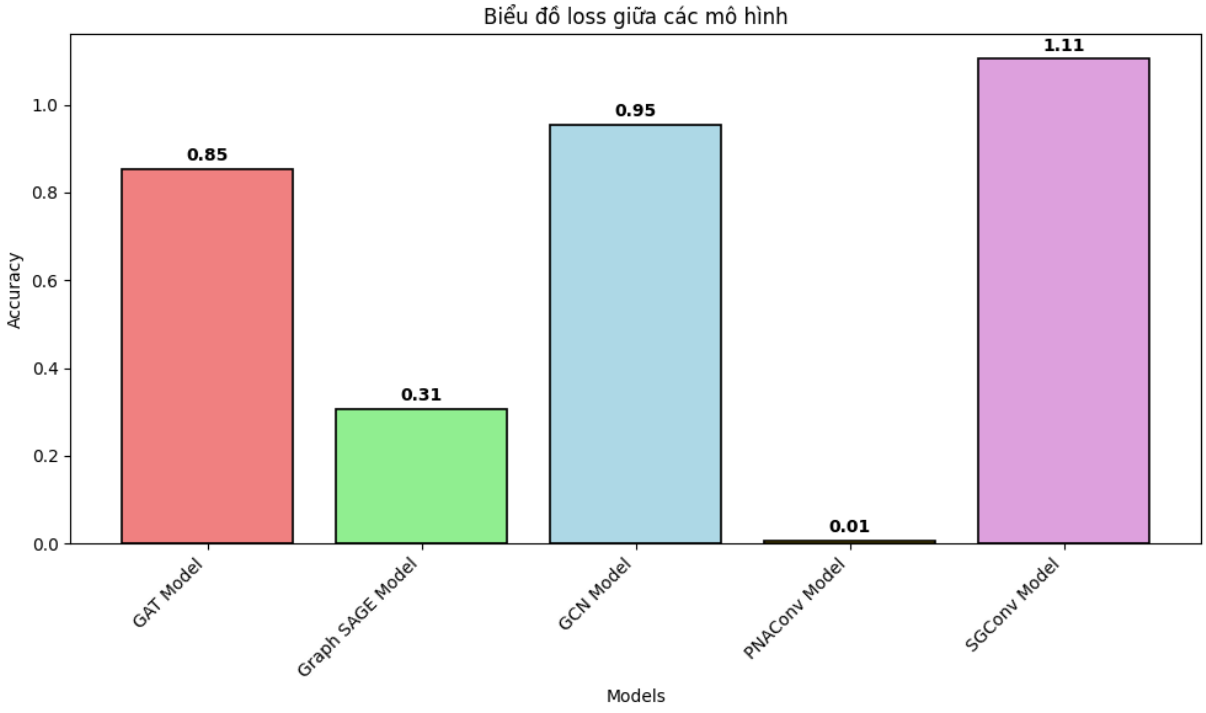
Đồ án này sử dụng bộ tối ưu hóa Adam [19] với learning rate 0.001, ngoài ra chúng tôi sử dụng ReduceLROnPlateau [20] để tự động giảm learning rate của optimizer khi giá trị loss không giảm trong một khoảng thời gian dài. Điều này giúp cải thiện quá trình huấn luyện và ngăn chặn tình trạng quá khớp của mô hình. Đối với mỗi mô hình, chúng tôi xây dựng 2 lớp tích chập để học biểu diễn đặc trưng từ đồ thị và áp dụng hàm kích hoạt Relu giữa các lớp. Các mô hình được huấn luyện với 300 epochs và các tham số như *num_futures* = 2, *hidden_size* = 16, *num_classes* = 6.

1.5.2 Kết quả

Từ những đề xuất ở trên chúng tôi tiến hành thực nghiệm, xử lý dữ liệu, huấn luyện mô hình học sâu. Cuối cùng trích xuất text từ các vùng đã được xác định từ đó đưa ra được những so sánh và đánh giá lựa chọn mô hình tối ưu nhất.



Hình 7: Độ chính xác giữa các mô hình



Hình 8: Loss giữa các mô hình

Từ những hình ảnh trực quan hóa ở trên, ta thấy thấy mô hình PNAConv cho ra kết quả tốt nhất ở tất cả các độ đo. Do đó chúng tôi chọn mô hình PNAConv để thực hiện dự đoán các mẫu.

Mô hình PNAConv cho ra kết quả vượt trội hơn so với các mô hình khác như GCN, GAT, GraphSAGE, SGConv là do một số lý do sau:

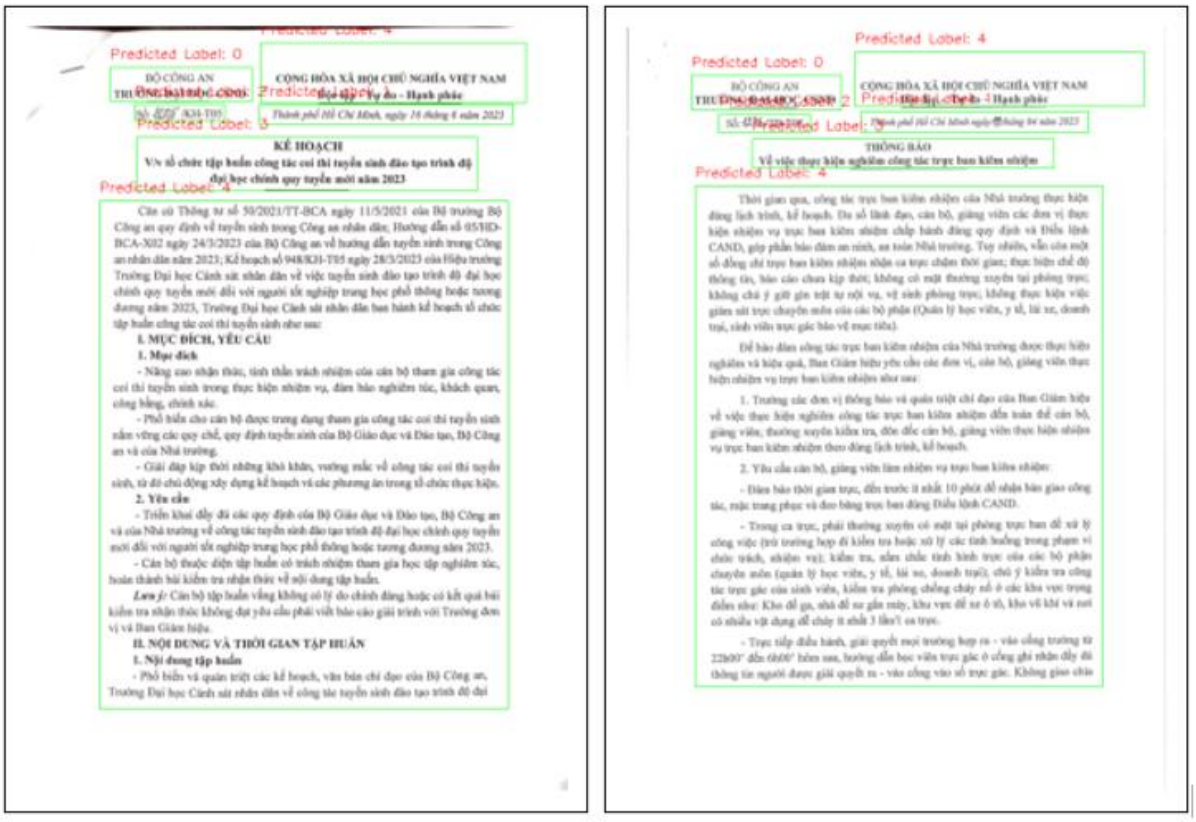
- PNAConv sử dụng cấu trúc Attention để học trọng số cho từng kết nối cạnh. Điều này giúp mô hình tập trung vào các kết nối quan trọng hơn, dẫn đến kết quả chính xác hơn. Ví dụ, trong một đồ thị mạng xã hội, các kết nối giữa các cá nhân có ảnh hưởng lớn hơn có thể được coi là quan trọng hơn. PNAConv sẽ học trọng số cao hơn cho các kết nối này, giúp mô hình dự đoán chính xác hơn các mối quan hệ giữa các cá nhân.
- PNAConv sử dụng một lớp tuyến tính sau khi tính toán thông tin từ các nút và cạnh. Điều này giúp mô hình học được các mối quan hệ phức tạp hơn giữa các nút, dẫn đến kết quả tốt hơn. Ví dụ, trong một đồ thị phân loại văn bản, các nút có thể đại diện cho các từ trong một câu. PNAConv có thể học được các mối quan hệ phức tạp giữa các từ, chẳng hạn như từ đồng nghĩa, từ trái nghĩa, từ đồng nghĩa, v.v., giúp mô hình phân loại văn bản chính xác hơn.
- PNAConv có thể được áp dụng cho các đồ thị có kích thước lớn hơn. Điều này là do PNAConv sử dụng một phương pháp lan truyền hiệu quả hơn, giúp mô hình học được các thông tin từ các nút xa hơn. Ví dụ, trong một đồ thị mạng xã hội, các nút có thể đại diện cho các cá nhân. PNAConv có thể học được các thông tin từ các cá nhân ở xa hơn, giúp mô hình dự đoán chính xác hơn các mối quan hệ giữa các cá nhân.

Nhìn chung, PNAConv là một mô hình học sâu trên đồ thị mạnh mẽ có thể đạt được kết quả vượt trội hơn so với các mô hình khác. Chi tiết trong Bảng 2.

	Accuracy	Recall	Precision	F1-score	Loss
GCN	0.59	0.59	0.62	0.56	0.95
GAT	0.5	0.5	0.51	0.48	0.85
GraphSAGE	0.9	0.9	0.92	0.9	0.31
SGConv	0.6	0.6	0.59	0.59	1.11
PNAConv	**0.98**	**0.98**	**0.98**	**0.98**	**0.01**

Bảng 1: Kết quả so sánh giữa các mô hình

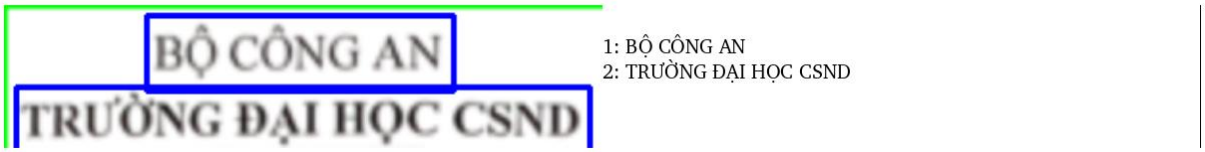
Một số kết quả thực hiện dự đoán trên mô hình PNAConv.



Hình 9: Một số kết quả sau khi thực hiện dự đoán trên mô hình PNAConv

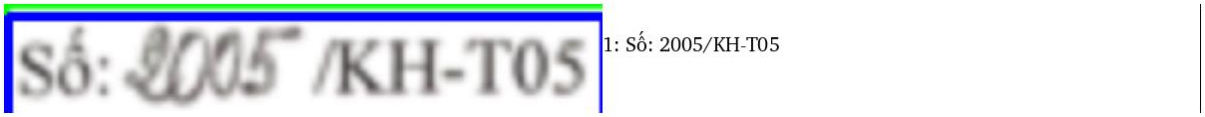
Sau khi thực hiện dự đoán và xác định được vùng chứa văn bản và nhãn của mỗi vùng chúng tôi thực hiện trích xuất nội dung trong các vùng được xác định. Nội dung trích xuất bao gồm: Nơi phát hành văn bản, ngày phát hành văn bản, số văn bản và tiêu đề văn bản. Dưới đây là một số kết quả đạt được khi thực nghiệm.

- Nơi phát hành văn bản:

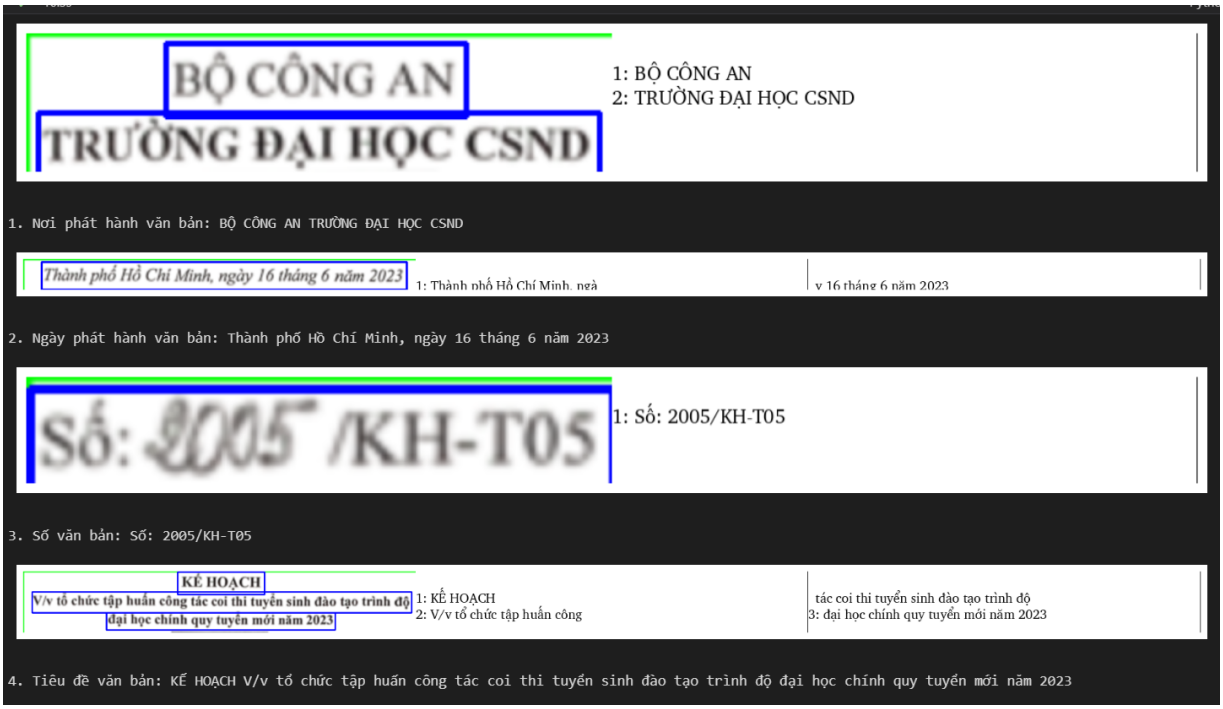


- Ngày phát hành văn bản:

- Số văn bản:



- Tiêu đề văn bản:



Hình 10: Kết quả thực nghiệm

1.6 Kết luận

1.6.1. Kết quả đạt được

Đồ án đã hoàn thành một nhiệm vụ quan trọng, đó là giải quyết bài toán xác định thông tin từ văn bản trong hình ảnh bằng cách sử dụng phương pháp học sâu. Trong quá trình nghiên cứu, đồ án cũng đề xuất và thảo luận về một số phương pháp mới để hiệu quả hóa quá trình giải quyết bài toán nói trên. Điều này không chỉ giúp mở rộng kiến thức về lĩnh vực này mà còn đóng góp vào sự phát triển của các phương pháp xử lý hình ảnh và xác định thông tin.

Để đảm bảo tính chất thực tế và tính hiệu quả của những phương pháp đề xuất, đồ án đã tiến hành một loạt thực nghiệm chi tiết. Quá trình này bao gồm việc thu thập dữ liệu, xây dựng mô hình, và tiến hành các thử nghiệm so sánh giữa các phương pháp khác nhau. Những kết quả thu được không chỉ giúp đánh giá hiệu suất của mỗi phương

pháp mà còn đưa ra những thông tin cần thiết để hiểu rõ hơn về sự khác biệt và ưu nhược điểm giữa chúng.

Đồ án không chỉ là một công trình nghiên cứu mà còn là một đóng góp quan trọng trong việc áp dụng và phát triển các phương pháp học sâu vào việc xử lý văn bản trong hình ảnh. Những kết quả và phương pháp đề xuất có thể đóng vai trò quan trọng trong việc giải quyết các thách thức thực tế liên quan đến xác định thông tin từ dữ liệu hình ảnh, mở ra những cơ hội mới trong lĩnh vực này.

1.6.2 Hạn chế

Mặc dù đồ án đã đạt được những thành tựu đáng kể trong việc giải quyết bài toán xác định thông tin văn bản từ hình ảnh bằng phương pháp học sâu, nhưng cũng không tránh khỏi những hạn chế nhất định. Một số điểm yếu và hạn chế của đồ án có thể được đề cập như sau:

- Dữ liệu hạn chế: Một số đề xuất và phương pháp mới có thể gặp khó khăn khi áp dụng trên dữ liệu hạn chế. Nếu dữ liệu không đủ đa dạng hoặc đại diện cho các tình huống thực tế, hiệu suất của mô hình có thể bị ảnh hưởng.
- Tính tổng quát hóa: Mô hình có thể gặp khó khăn trong việc tổng quát hóa kết quả trên các tập dữ liệu mới không được sử dụng trong quá trình huấn luyện. Điều này đặt ra thách thức trong việc ứng dụng mô hình vào các bối cảnh thực tế và đa dạng.
- Hiệu suất tương đối: So với một số phương pháp truyền thống, hiệu suất của phương pháp học sâu có thể bị ảnh hưởng bởi sự phức tạp và đòi hỏi nhiều dữ liệu lớn. Điều này có thể là một hạn chế đặc biệt nếu tài nguyên hoặc dữ liệu là hạn chế.

Tất cả những hạn chế này cung cấp cơ hội cho các nghiên cứu và phát triển tương lai để cải thiện và mở rộng ứng dụng của phương pháp học sâu trong lĩnh vực xác định thông tin từ văn bản hình ảnh.

1.6.3. Hướng phát triển trong tương lai

Để nâng cao chất lượng và khả năng ứng dụng của đồ án trong tương lai, có một số hướng phát triển có thể được đề xuất:

- Mở rộng dữ liệu: Tăng cường và mở rộng tập dữ liệu đào tạo có thể giúp cải thiện khả năng tổng quát hóa và hiệu suất của mô hình trên các tình huống thực tế đa dạng. Điều này có thể đòi hỏi sự hợp tác với cộng đồng để thu thập và chia sẻ dữ liệu.
- Tối ưu hóa mô hình: Nghiên cứu và thử nghiệm các kiến trúc mô hình mới hoặc cải tiến để tối ưu hóa hiệu suất và giảm độ phức tạp tính toán. Việc này có thể bao gồm sự tận dụng các kỹ thuật học sâu tiên tiến, kiến trúc mô hình mới, hoặc việc sử dụng các phương pháp transfer learning.
- Xử lý đa nhiệm: Nghiên cứu cách tích hợp khả năng xử lý đa nhiệm của mô hình, giúp nó đồng thời giải quyết nhiều mục tiêu hoặc loại dữ liệu khác nhau trong một hình ảnh. Điều này có thể làm tăng tính linh hoạt và đa dạng của ứng dụng.
- Tích hợp đối tượng mở rộng: Nghiên cứu và phát triển khả năng xác định và xử lý đối tượng mới và không biết trước (open-set recognition), giúp mô hình chấp nhận và xử lý những đối tượng chưa được biết đến trong quá trình huấn luyện.
- Tăng cường tính di động: Xem xét cách triển khai mô hình trên các nền tảng di động hoặc tài nguyên có hạn. Điều này có thể bao gồm việc tối ưu hóa kích thước mô hình, giảm tải tính toán, và tối ưu hóa về tài nguyên để phù hợp với môi trường ứng dụng.
- Tăng cường thông tin giải thích: Nghiên cứu cách làm cho mô hình trở nên có thể giải thích hơn, giúp hiểu rõ quyết định của mô hình và tăng cường độ tin cậy. Điều này có thể giúp người dùng và nhà nghiên cứu hiểu rõ hơn về cách mô hình thực hiện các dự đoán.

Bằng cách thúc đẩy những hướng phát triển này, đồ án có thể không chỉ giữ vững mà còn nâng cao sự đóng góp và tính ứng dụng của nó trong lĩnh vực xác định thông tin từ văn bản hình ảnh.

LÀM VIỆC NHÓM

Trình bày tóm tắt cách thức làm việc nhóm
Phân chia công việc của các thành viên trong nhóm
Tổng số lần gặp nhau (tính theo buổi)
Tổng thời gian gặp nhau (tính theo giờ)

TÀI LIỆU THAM KHẢO

1. Hung, B. T. (2022). Information Extraction from Receipts Using Spectral Graph Convolutional Network. In Intelligent Computing & Optimization: Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021 (ICO2021) 3 (pp. 602-612). Springer International Publishing.
2. Lohani, D., Belaïd, A., & Belaïd, Y. (2019). An invoice reading system using a graph convolutional network. In Asian Conference on Computer Vision (pp. 144-158). Springer, Cham.
3. Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
4. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.
5. Corso, G., Cavalleri, L., Beaini, D., Liò, P., & Veličković, P. (2020). Principal neighbourhood aggregation for graph nets. Advances in Neural Information Processing Systems, 33, 13260-13271.
6. Wu, F., Souza, A., Zhang, T., Fifty, C., Yu, T., & Weinberger, K. (2019, May). Simplifying graph convolutional networks. In International conference on machine learning (pp. 6861-6871). PMLR.
7. Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in neural information processing systems, 30.
8. PaddleOCR: https://github.com/PaddlePaddle/PaddleOCR
9. VietOCR: https://github.com/pbcquoc/vietocr
10. Make Sense: https://www.makesense.ai/
11. Pytorch: https://pytorch.org/
12. Torch Geometric: https://github.com/pyg-team/pytorch_geometric
13. Cross Entropy: https://en.wikipedia.org/wiki/Cross-entropy

14. Mao, A., Mohri, M., & Zhong, Y. (2023). Cross-entropy loss functions: Theoretical analysis and applications. arXiv preprint arXiv:2304.07288.
15. Šimundić, A. M. (2009). Measures of diagnostic accuracy: basic definitions. *ejifcc*, *19*(4), 203.
16. Michaud, E. J., Liu, Z., & Tegmark, M. (2023). Precision machine learning. *Entropy*, *25*(1), 175.
17. Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.
18. Humphrey, A., Kuberski, W., Bialek, J., Perrakis, N., Cools, W., Nuyttens, N., ... & Cunha, P. A. C. (2022). Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth. *Monthly Notices of the Royal Astronomical Society: Letters*, *517*(1), L116-L120.
19. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
20. Al-Kababji, A., Bensaali, F., & Dakua, S. P. (2022, March). Scheduling techniques for liver segmentation: Reducelronplateau vs onecyclelr. In *International Conference on Intelligent Systems and Pattern Recognition* (pp. 204-212). Cham: Springer International Publishing.

PHỤ LỤC

Phần này bao gồm những nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung đồ án như số liệu, biểu mẫu, tranh ảnh. . . . nếu sử dụng những câu trả lời cho một *bảng câu hỏi thì bảng câu hỏi mẫu này phải được đưa vào phần Phụ lục ở dạng nguyên bản* đã dùng để điều tra, thăm dò ý kiến; **không được tóm tắt hoặc sửa đổi**. Các tính toán mẫu trình bày tóm tắt trong các biểu mẫu cũng cần nêu trong Phụ lục của luận văn. Phụ lục không được dày hơn phần chính của đồ án

TỰ ĐÁNH GIÁ

STT	Nội dung	Điểm chuẩn	Tự chấm	Ghi chú
1 (8.5)	**1.1 Giới thiệu về bài toán**	0.5		
	1.2 Phân tích yêu cầu của bài toán	1		
	1.3 Phương pháp giải quyết bài toán	1.5		
	1.4 Thực nghiệm	4		
	1.5 Kết quả đạt được	1		
	1.6 Kết luận	0.5		
2 (1)	**Báo cáo (**chú ý các chú ý 2,3,4,6 ở trang trước, nếu sai sẽ bị trừ điểm nặng**)**	1đ		
3 (0.5)	**Điểm nhóm** (chú ý trả lời các câu hỏi trong mục làm việc nhóm)	0.5đ		
Tổng điểm				