

ĐẠI HỌC THÁI NGUYÊN  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG  
KHOA CÔNG NGHỆ THÔNG TIN

LÊ THỊ GIANG

NGHIÊN CỨU KỸ THUẬT HỌC SÂU  
NHẬN DẠNG KÝ HIỆU TOÁN HỌC  
HỖ TRỢ CHUYỂN ĐỔI SỐ TÀI LIỆU KHOA HỌC KỸ THUẬT

ĐỀ TÀI LUẬN VĂN THẠC SĨ  
Chuyên ngành: Khoa học máy tính  
Mã số: 8480101

Người hướng dẫn: TS ĐÀM THANH PHƯƠNG

## LỜI CẢM ƠN

Lời đầu tiên em xin gửi lời cảm ơn sâu sắc tới thầy giáo hướng dẫn TS.Đàm Thanh Phương. Thầy đã giao đề tài và tận tình hướng dẫn em trong quá trình hoàn thành đề tài này.

Em xin gửi lời cảm ơn của mình tới các thầy cô giáo trong khoa Công nghệ thông tin đã giảng dạy và giúp đỡ chúng em trong suốt quá trình học tập tại trường.

Tôi xin chân thành cảm ơn bạn bè đồng nghiệp ở trường Cao Đẳng Lào Cai nơi tôi công tác. Đồng nghiệp đã giúp đỡ hỗ trợ tôi rất nhiều trong công việc để tôi có thời gian học tập nghiên cứu tập trung hiệu quả hơn.

Xin cảm ơn Gia đình tôi, bố mẹ, chồng con và những người thân yêu đã ở bên động viên và khích lệ tôi.

Xin chân thành cảm ơn các bạn cùng lớp CHK20 đã đồng hành giúp đỡ tôi trong quá trình học tập.

*Thái Nguyên, tháng 8 năm 2023*  
Học viên

**LÊ THỊ GIANG**

# Mục lục

<b>DANH SÁCH HÌNH VẼ</b> .....	<b>iii</b>
<b>DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT</b> .....	<b>v</b>
<b>MỞ ĐẦU</b> .....	<b>vi</b>
<b>Chương 1. BÀI TOÁN NHẬN DẠNG KÝ TỰ QUANG HỌC</b> .....	<b>2</b>
<b>1.1. Giới thiệu về bài toán nhận dạng ký tự quang học</b> .....	<b>3</b>
1.1.1. Bài toán nhận dạng ký tự quang học .....	3
1.1.2. Hoạt động của hệ thống nhận dạng ký tự quang học .....	3
1.1.3. Các công nghệ nhận dạng ký tự quang học .....	4
1.1.4. Một số lợi ích của nhận dạng ký tự quang học .....	5
<b>1.2. Một số kỹ thuật tiêu biểu trong ORC</b> .....	<b>7</b>
1.2.1. Kỹ thuật xử lý ảnh truyền thống .....	7
1.2.2. Kỹ thuật học sâu .....	7
<b>Chương 2. MỘT SỐ KỸ THUẬT HỌC SÂU</b> .....	<b>12</b>
<b>2.1. CNN và DenseNet</b> .....	<b>13</b>
2.1.1. CNN .....	13
2.1.2. Densenet .....	19
<b>2.2. LSTM và cơ chế chú ý</b> .....	<b>21</b>
2.2.1. LSTM .....	21
2.2.2. Cơ chế chú ý .....	29
<b>Chương 3. ỨNG DỤNG HỌC SÂU TRONG BÀI TOÁN IMAGE2LATEX</b> ..	<b>39</b>
<b>3.1. Cải tiến mô hình</b> .....	<b>40</b>
3.1.1. Thiết lập bộ mã hóa DenseNet cho mô hình cải tiến .....	40
3.1.2. Thiết lập bộ giải mã cho mô hình cải tiến .....	41
3.1.3. Cơ chế chú ý sử dụng trong mô hình cải tiến .....	42
<b>3.2. Thực nghiệm và đánh giá</b> .....	<b>43</b>
3.2.1. Tập dữ liệu và độ đo .....	43
3.2.2. So sánh với mô hình cơ sở [11] .....	45
<b>Kết luận</b> .....	<b>49</b>
<b>Tài liệu tham khảo</b> .....	<b>49</b>
<b>Phụ lục</b> .....	<b>52</b>

# Danh sách hình vẽ

2.1	Quá trình trượt theo chiều rộng $W_1$ . . . . .	15
2.2	Thêm padding kích thước $P$ vào 2 lề chiều rộng ( $W_1$ ) . . . . .	15
2.3	Cấu trúc các khối nơ ron 3D mạng Alexnet. . . . .	18
2.4	Kết nối cục bộ . . . . .	19
2.5	Sự khác biệt chính giữa ResNet (bên trái) và DenseNet (bên phải) . . .	20
2.6	Các kết nối dày đặc trong DenseNet . . . . .	21
2.7	Mạng nơ ron truy hồi với vòng lặp . . . . .	22
2.8	Cấu trúc trái phẳng của mạng nơ ron truy hồi . . . . .	22
2.9	Sự lặp lại kiến trúc module trong mạng RNN chứa một tầng ẩn . . . .	23
2.10	Sự lặp lại kiến trúc module trong mạng LSTM chứa 4 tầng ẩn (3 sigmoid và 1 tanh) tương tác . . . . .	24
2.11	Diễn giải các kí hiệu trong đồ thị mạng nơ ron . . . . .	24
2.12	Đường đi của ô trạng thái (cell state) trong mạng LSTM . . . . .	24
2.13	Một cổng của hàm sigmoid trong LSTM . . . . .	25
2.14	Tầng cổng quên (forget gate layer) . . . . .	26
2.15	Cập nhật giá trị cho ô trạng thái bằng cách kết hợp 2 kết quả từ tầng cổng vào và tầng ẩn hàm tanh . . . . .	26
2.16	Ô trạng thái mới . . . . .	27
2.17	Điều chỉnh thông tin ở đầu ra thông qua hàm tanh . . . . .	27
2.18	Kết nối ống tiêu (peehole) liên kết trực tiếp ô trạng thái với các cổng . .	27
2.19	Cấu trúc điều chỉnh thêm mới và bỏ qua thông tin đồng thời . . . . .	28
2.20	Một cấu trúc điều chỉnh LSTM . . . . .	28
2.21	Mô hình seq2seq khi chưa có attention layer . . . . .	30
2.22	Mô hình seq2seq khi đã có attention layer . . . . .	31
2.23	Kiến trúc transformer . . . . .	32
2.24	Self Attention . . . . .	33
2.25	Các từ input tương ứng với vector key, query và value . . . . .	34
2.26	Quá trình tính toán trọng số attention và attention vector cho từ I trong câu I study at school. . . . .	35

2.27	Kết quả tính attention vector cho toàn bộ các từ trong câu. . . . .	35
2.28	Sơ đồ cấu trúc Multi-head Attention . . . . .	36
2.29	Sơ đồ của 1 block layer áp dụng multi-head attention layer. . . . .	37
2.30	Sơ đồ quá trình encoder với 6-layer của transformer. . . . .	37
2.31	Quá trình biến đổi giá trị input thành các giá trị output. . . . .	38
3.1	Mạng DenseNet với 3 khối dày đặc . . . . .	41
3.2	Mô hình chú ý C-S . . . . .	42
3.3	Cấu trúc thiết kế hai mô hình Dense Net . . . . .	46
3.4	So sánh 4 mô hình sử dụng các độ đo EM, BLEU và ED . . . . .	46
3.5	Hình ảnh công thức định nghĩa đạo hàm . . . . .	47
3.6	Hình ảnh công thức phân phối chuẩn . . . . .	47
3.7	Hình ảnh công thức phương trình sóng . . . . .	47
3.8	Hình ảnh công thức phương trình Navier-Stokes . . . . .	48

## DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

ORC	Optical Character Recognition - Nhận dạng ký tự quang học.
Image2Latex	Bài toán chuyển đổi hình ảnh công thức toán thành mã latex
CNN	Mạng nơ ron tích chập
RNN	Mạng nơ ron hồi quy
DenseNet	Mạng nơ ron kết nối dày đặc
LSTM	Long sort term memory. Mạng trí nhớ ngắn dài hạn.
$X$	Ma trận.
$x$	Véc tơ.
$x$	Vô hướng.
$\mathbb{R}^n$	Không gian $n$ chiều.
CONV	Tầng tích chập
RELU	Hàm kích hoạt
Pool	Tầng tổng hợp
FC	Fully conection - Kết nối hoàn toàn
C-S Attention	Cơ chế chú ý theo kênh và theo không gian.

## MỞ ĐẦU

Học máy, học sâu, trí tuệ nhân tạo là lĩnh vực đang thu hút được sự quan tâm của xã hội nói chung và những nhà nghiên cứu khoa học máy tính nói riêng trong những năm gần đây. Bằng việc sử dụng lượng dữ liệu huấn luyện lớn, các thuật toán học máy có thể học hỏi từ dữ liệu để từ đó đưa ra được những mô hình giải quyết nhiều bài toán thực tế như hồi quy, phân lớp, dự đoán, phân cụm, giảm chiều và nhiều nhiệm vụ khác[17], [18], [19].

Bài toán Image-to-LaTeX là một trong những bài toán trong lĩnh vực Nhận diện Văn bản Trong Ảnh (OCR - Optical Character Recognition) và Trí tuệ nhân tạo. Nhiệm vụ của bài toán này là chuyển đổi hình ảnh chứa các ký tự và biểu thức toán học thành mã LaTeX tương ứng. [9], [10].

LaTeX là một ngôn ngữ sử dụng để viết các tài liệu khoa học, báo cáo, sách và tài liệu có độ chính xác cao. Nó cho phép tạo ra các công thức toán học phức tạp và định dạng văn bản chuyên nghiệp.

Bài toán Image-to-LaTeX có thể được phân chia thành các bước chính:

1. Nhận dạng ký tự: Đầu tiên, hệ thống cần nhận dạng và phân loại các ký tự trong hình ảnh. Điều này có thể thực hiện thông qua các thuật toán nhận dạng ký tự, bao gồm cả kỹ thuật Học sâu (Deep Learning) như Convolutional Neural Networks (CNNs).
2. Phân tích biểu thức toán học: Sau khi nhận dạng được các ký tự, bước tiếp theo là phân tích biểu thức toán học trong hình ảnh. Điều này có thể là một phần khó khăn vì biểu thức toán học thường có cấu trúc phức tạp và nhiều quy tắc về việc sắp xếp ký tự.
3. Chuyển đổi thành LaTeX: Khi đã nhận dạng được cấu trúc biểu thức toán học, bước cuối cùng là chuyển đổi nó thành mã LaTeX tương ứng. Điều này có thể đòi hỏi sự áp dụng các quy tắc và mẫu biểu thức toán học trong LaTeX.

Có nhiều phương pháp và công cụ có thể được sử dụng để giải quyết bài toán Image-to-LaTeX. Một số phương pháp tiêu biểu bao gồm sử dụng mạng CNN kết hợp với mạng Recurrent Neural Networks (RNNs) [12], [14] hoặc Transformer [3], [4], [13] để xử lý hình ảnh và sinh ra mã LaTeX tương ứng. Tuy nhiên, bài toán này vẫn còn nhiều thách thức, đặc biệt khi xử lý các biểu thức toán học phức tạp và có nhiều biến thể.

Bài toán Image2Latex có tầm quan trọng đáng kể trong quá trình số hóa tài liệu khoa học kỹ thuật. Điều này thể hiện qua các lý do sau

- Chuyển đổi tài liệu dựa trên hình ảnh sang định dạng mã LaTeX: Tài liệu khoa học kỹ thuật thường chứa nhiều công thức toán học phức tạp. Việc chuyển đổi hình ảnh của các công thức này thành định dạng mã LaTeX có thể giúp tạo ra các tài liệu có thể chỉnh sửa và tái sử dụng dễ dàng hơn. Điều này rất hữu ích trong việc tổ chức, phân tích và chia sẻ tài liệu khoa học.
- Tự động hóa quá trình chuyển đổi: Trước đây, việc chuyển đổi hình ảnh của công thức toán học thành định dạng LaTeX thường phải được thực hiện bằng tay, tốn nhiều thời gian và công sức. Sử dụng các mô hình deep learning trong bài toán Image2Latex giúp tự động hóa quá trình này, giảm bớt công sức và thời gian cần thiết để chuyển đổi tài liệu.
- Khả năng tái sử dụng và tương tác: Khi tài liệu được chuyển đổi thành định dạng mã LaTeX, nó có thể được sử dụng lại và tương tác với các công cụ và môi trường phát triển LaTeX khác. Điều này giúp đơn giản hóa việc chỉnh sửa, biên tập và tương tác với các công thức toán học trong tài liệu.
- Tiện ích trong công việc nghiên cứu: Tài liệu khoa học kỹ thuật thường chứa nhiều công thức toán học quan trọng và phức tạp. Việc có khả năng chuyển đổi hình ảnh công thức thành mã LaTeX giúp nghiên cứu viên nhanh chóng trích xuất và sử dụng lại các công thức này trong công việc nghiên cứu, phân tích và trình bày kết quả.

Bài toán Image2Latex có tầm quan trọng cao trong việc số hóa tài liệu khoa học kỹ thuật, giúp chuyển đổi hình ảnh công thức toán học thành mã LaTeX để tạo ra các tài liệu có thể chỉnh sửa, tái sử dụng và tương tác dễ dàng. Điều này cải thiện hiệu quả và tiện ích trong việc làm việc.

Bài toán nhận dạng ký tự quang học (OCR - Optical Character Recognition) đóng vai trò quan trọng trong bài toán Image2Latex. Mục tiêu của OCR trong Image2Latex là nhận dạng và trích xuất các ký tự và từ từ hình ảnh chứa công thức toán học.[4]

Khi áp dụng OCR trong bài toán Image2Latex, quá trình diễn ra như sau:

**Nhận dạng hình ảnh:** OCR được sử dụng để nhận dạng văn bản từ hình ảnh chứa công thức toán học. Quá trình này có thể bao gồm xử lý tiền xử lý ảnh, như làm sạch nhiễu, tăng cường độ tương phản, chỉnh sửa kích thước và phân đoạn ảnh để tách các ký tự và từ.

**Trích xuất văn bản:** Sau khi hình ảnh được nhận dạng, OCR sẽ trích xuất văn bản từ hình ảnh đó thành các chuỗi ký tự hoặc từ đơn. Các kỹ thuật OCR phổ biến bao gồm phân loại ký tự, nhận dạng từ và xử lý cú pháp để tạo ra đầu ra văn bản dễ đọc được.

**Chuyển đổi thành mã LaTeX:** Các chuỗi văn bản nhận dạng được từ OCR sẽ được chuyển đổi thành mã LaTeX tương ứng với công thức toán học. Quá trình



này có thể sử dụng các kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) để phân tích cú pháp, xác định cấu trúc và ý nghĩa của các biểu thức toán học và tạo ra mã LaTeX phù hợp.

Bài toán nhận dạng ký tự quang học (ORC) đóng vai trò quan trọng trong Image2Latex bằng cách cung cấp khả năng nhận dạng và trích xuất văn bản từ hình ảnh công thức. Điều này là cần thiết để tạo ra dữ liệu đầu vào cho các mô hình xử lý ngôn ngữ tự nhiên hoặc mô hình dự đoán để tạo ra mã LaTeX từ công thức toán học.

Luận văn của em nghiên cứu vấn đề sử dụng các kỹ thuật học máy học sâu để giải quyết bài toán Image2Latex. Vì vậy luận văn được bố cục gồm 3 chương, bao quát một số kiến thức nền tảng và cần thiết cũng như giải quyết bài toán này bằng mạng nơ ron học sâu

**Chương 1:** Trình bày tổng quan kỹ thuật nhận dạng ký tự quang học (ORC).

**Chương 2:** Trình bày các kiến thức sát nhất về các kỹ thuật Deeplearning sử dụng trong việc cải tiến mô hình nhận dạng trong chương 3.

**Chương 3:** Trình bày quá trình cải tiến một mô hình sử dụng mạng nơ ron học sâu để giải quyết bài toán Image2Latex. Chương 3 cũng đưa ra các đánh giá, thực nghiệm cũng như đưa ra phần demo nhận dạng được công thức toán bất kỳ.

Tuy đã có nhiều cố gắng nhưng do thời gian và khả năng có hạn nên các vấn đề trong khóa luận vẫn chưa được trình bày sâu sắc và không thể tránh khỏi có những sai sót trong cách trình bày. Mong được sự góp ý xây dựng của thầy cô và các bạn. Em xin chân thành cảm ơn!

*Thái Nguyên, ngày 10 tháng 8 năm 2023*

Học viên

**Lê Thị Giang**

# Chương 1

## BÀI TOÁN NHẬN DẠNG KÝ TỰ QUANG HỌC

Bài toán nhận dạng ký tự quang học (OCR - Optical Character Recognition) và bài toán chuyển đổi hình ảnh công thức thành mã latex (Image -to -Latex, Image2Latex) có mối quan hệ chặt chẽ với nhau. [3], [4].

Nhận dạng ký tự quang học là một kỹ thuật sử dụng trong xử lý hình ảnh để nhận dạng và trích xuất văn bản từ hình ảnh. Nó được sử dụng để nhận diện và chuyển đổi các ký tự và từ, câu, văn bản từ hình ảnh sang định dạng văn bản có thể xử lý được (như định dạng word hay latex).[1], [20].

Trong bài toán Image2Latex, OCR có thể được sử dụng để nhận dạng các ký tự và từ hình ảnh chứa công thức toán học. Quá trình OCR sẽ chuyển đổi hình ảnh thành các chuỗi văn bản, có thể là các ký tự hoặc từ đơn. Sau đó, các chuỗi văn bản này có thể được đưa vào mô hình Image2Latex để tạo ra mã LaTeX tương ứng với công thức toán học.[10], [11].

Tuy nhiên, trong bài toán Image2Latex, OCR chỉ là một phần trong quy trình chuyển đổi từ hình ảnh sang LaTeX. OCR giúp chuyển đổi hình ảnh thành văn bản có thể xử lý được, nhưng nó không cung cấp thông tin về cấu trúc và ý nghĩa của công thức toán học. Vì vậy, sau khi sử dụng OCR để nhận dạng văn bản từ hình ảnh, cần sử dụng các mô hình học sâu khác (như CNN, RNN, transformer) để xử lý và tạo ra mã LaTeX dựa trên văn bản nhận dạng được từ OCR.[2], [5], [6].

Tóm lại, OCR là một công cụ hữu ích để chuyển đổi hình ảnh thành văn bản, trong khi Image2Latex là quá trình xử lý tiếp theo sử dụng các mô hình học máy, học sâu, mạng nơ ron để chuyển đổi văn bản thành mã LaTeX tương ứng với công thức toán học.

Nội dung chương này sẽ trình bày một số kiến thức về ORC, được tham khảo chủ yếu trong tài liệu [20].