

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Học tự giám sát cho bài toán khoanh vùng đối tượng

TRƯỜNG TUẤN ANH

anh.tt193984@sis.hust.edu.vn

Ngành Khoa học máy tính

Giảng viên hướng dẫn: TS. Đinh Viết Sang

Chữ kí GVHD

Trường:

Công nghệ thông tin và Truyền thông

HÀ NỘI, 08/2023

LỜI CẢM ƠN

Để hoàn thiện đồ án này, em xin gửi lời cảm ơn sâu sắc đến thầy Đinh Viết Sang. Thầy đã truyền đạt kiến thức và những ý kiến quý báu, giúp em có được định hướng để vượt qua những khó khăn và phát triển kỹ năng của mình. Em rất biết ơn vì tất cả những giờ phút thầy đã dành cho em, dạy bảo và giúp đỡ để giúp em hoàn thành công việc.

Em xin gửi đến cha mẹ của mình lòng biết ơn sâu sắc. Từ tận đáy lòng, em xin chân thành cảm ơn cha mẹ đã luôn ở bên, tin tưởng và tạo điều kiện cho em được học tập và phát triển trong ngôi trường này. Gia đình luôn là chỗ dựa vững chắc nhất để em trưởng thành như ngày hôm nay.

Em muốn gửi lời cảm ơn các thầy cô giáo của Trường Công Nghệ Thông Tin và Truyền Thông, Đại học Bách Khoa Hà Nội đã giúp em xây dựng được một nền tảng kiến thức vững chắc trong quá trình học tập tại trường. Em xin cảm ơn các bạn lớp Khoa học máy tính 06 - K64 và các bạn sinh viên cùng làm việc dưới sự hướng dẫn của thầy Sang. Sự hỗ trợ của các bạn đã giúp em nhiều trong quá trình xây dựng, hoàn thiện ý tưởng và tiến hành đồ án.

Em xin gửi lời cảm ơn đến anh Cường, anh Quang cùng các đồng nghiệp khác trong nhóm nghiên cứu AI ở công ty MISA. Cảm ơn các anh vì đã cho em những kinh nghiệm quý báu, cũng như tạo điều kiện để em hoàn thiện đồ án này tốt hơn.

Do thời gian có hạn, nên đồ án này cũng không thể tránh khỏi những sai sót. Em rất mong nhận được những góp ý từ các thầy cô và các bạn sinh viên khác để đồ án của em được hoàn thiện hơn. Em xin chân thành cảm ơn.

TÓM TẮT NỘI DUNG ĐỒ ÁN

Khoanh vùng đối tượng bất thường từ ảnh chụp là một yêu cầu quan trọng trong lĩnh vực y tế. Tuy nhiên, lượng dữ liệu y tế đã gán nhãn để huấn luyện các mô hình học sâu thường tương đối nhỏ, do dữ liệu liên quan đến thông tin cá nhân của bệnh nhân, hay việc gán nhãn yêu cầu chuyên gia có trình độ cao. Các mô hình trước đây yêu cầu một lượng lớn dữ liệu đã gán nhãn, gây tốn kém thời gian và công sức cho việc tạo dữ liệu. Trong những năm gần đây, mô hình Vision Transformer, với khả năng biểu diễn các đặc trưng ẩn và mối liên hệ giữa chúng, dần chứng minh được tính hữu ích của mình trong các tác vụ liên quan đến thị giác máy tính. Masked Autoencoder là một phương pháp ứng dụng Vision Transformer, giúp mô hình học tự học được các đặc trưng từ ảnh mà không cần sự can thiệp của con người. Ở đồ án này, em đã huấn luyện Masked Autoencoder để nhận diện được các đặc trưng từ ảnh nội soi trực tràng, từ đó tinh chỉnh mô hình trên tác vụ khoanh vùng polyp. Kết quả của đồ án đã chứng minh được tác dụng của phương pháp, đóng góp được một quy trình để huấn luyện mô hình với ít dữ liệu đã gán nhãn, góp phần giảm công sức cần bỏ ra để chuẩn bị dữ liệu.

Từ khóa: *Phân vùng ngữ nghĩa, Phân vùng polyp.*

ABSTRACT

Abnormal object segmentation from captured images is an essential requirement in healthcare. However, labeled medical data for training deep learning models is often relatively scarce due to concerns related to patient privacy and the expertise required for accurate labeling. Previous models demanded a large amount of labeled data, which was time-consuming and labor-intensive for data creation. In recent years, Vision Transformers have effectively shown their usefulness in computer vision tasks, as they can effectively represent hidden feature relationships. The Masked Autoencoder is one such application of the Vision Transformer, enabling the model to learn features from images without human intervention. In this thesis, the model Masked Autoencoder is trained to recognize features from colonoscopy images and then fine-tune the model for the polyp segmentation task. This thesis's results demonstrated the method's effectiveness, contributing to training models with limited labeled data and reducing the effort required for data preparation.

Keywords: *Semantic Segmentation, Polyp Segmentation*

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề	1
1.2 Phạm vi đề án và mô tả bài toán	2
1.2.1 Phạm vi đề án	2
1.2.2 Mô tả bài toán	3
1.3 Đóng góp của đề án.....	3
1.4 Bố cục đề án.....	4
CHƯƠNG 2. NỀN TẢNG LÝ THUYẾT	5
2.1 Các khái niệm cơ bản trong học máy	5
2.1.1 Khái niệm học máy	5
2.1.2 Phân loại các thuật toán học máy.....	5
2.1.3 Quy trình học máy	7
2.1.4 Hạn chế và thách thức trong học máy.....	9
2.1.5 Gradient descent và một số biến thể.....	10
2.2 Giới thiệu về mạng nơ-ron (Neural Network)	11
2.2.1 Mạng nơ-ron nhân tạo	11
2.2.2 Mạng Perceptron nhiều lớp.....	12
2.2.3 Mạng nơ-ron tích chập	13
2.3 Bài toán phân vùng ảnh (Image Segmentation).....	17
2.3.1 Tổng quan về bài toán	17
2.3.2 Các dạng phân vùng ảnh khác nhau	18
CHƯƠNG 3. PHƯƠNG PHÁP SỬ DỤNG	20
3.1 Các nghiên cứu liên quan	20
3.1.1 Transformer	20

3.1.2	Vision Transformer (ViT)	21
3.1.3	Masked Autoencoder (MAE)	22
3.1.4	UNet	23
3.1.5	UNet Transformer (UNETR)	24
3.2	Tổng quan về phương pháp.....	25
3.2.1	Huấn luyện bộ mã hóa.....	25
3.2.2	Huấn luyện bộ giải mã	27
CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ		28
4.1	Tập dữ liệu.....	28
4.2	Độ đo đánh giá.....	28
4.3	Ngôn ngữ lập trình và thư viện	29
4.4	Môi trường triển khai.....	29
4.5	Tham số huấn luyện.....	30
4.6	Kết quả và đánh giá	31
4.6.1	Huấn luyện mô hình tự mã hóa.....	31
4.6.2	Huấn luyện mô hình phân vùng ảnh.....	32
4.6.3	Triển khai mô hình.....	36
CHƯƠNG 5. KẾT LUẬN		38
5.1	Kết luận.....	38
5.2	Hướng phát triển trong tương lai	38
TÀI LIỆU THAM KHẢO.....		39

DANH MỤC HÌNH VẼ

Hình 1.1	Các ứng dụng của AI trong lĩnh vực y tế	1
Hình 1.2	Ảnh nội soi polyp đại tràng	2
Hình 1.3	Các loại polyp thường gặp	3
Hình 1.4	Minh họa kết quả của học tự giám sát trên dữ liệu thực tế (polyp)	3
Hình 2.1	Học có giám sát	5
Hình 2.2	Học không giám sát	6
Hình 2.3	Học tăng cường	7
Hình 2.4	Ví dụ về Độ chệch (Bias) và Phương sai (Variance)	9
Hình 2.5	Ví dụ về Chưa khớp (Underfitting) và Quá khớp (Overfitting)	10
Hình 2.6	Biểu diễn Gradient Descent	11
Hình 2.7	Minh họa nơ-ron thần kinh và mô hình tuyến tính	12
Hình 2.8	Kiến trúc mạng nơ-ron	13
Hình 2.9	Minh họa quá trình tính toán lan truyền thuận và ngược	14
Hình 2.10	Kiến trúc mạng nơ-ron tích chập (CNN)	15
Hình 2.11	Minh họa lớp tích chập	16
Hình 2.12	Minh họa lớp tổng hợp	16
Hình 2.13	Minh họa trực quan phân loại ảnh trong mạng nơ-ron tích chập	17
Hình 2.14	Phân biệt giữa thuật toán Object Detection và Instance Segmentation	18
Hình 2.15	Hai kiểu phân vùng hình ảnh chính là phân vùng ngữ nghĩa và phân vùng cá thể	18
Hình 3.1	Kiến trúc Transformer	20
Hình 3.2	Kiến trúc mô hình Vision Transformer	21
Hình 3.3	Kiến trúc mô hình Masked Autoencoder	22
Hình 3.4	Kiến trúc mô hình UNet	23
Hình 3.5	Kiến trúc mô hình UNETR	24
Hình 3.6	Phương pháp huấn luyện mô hình và chuyển tiếp	25
Hình 4.1	Minh họa polyp trong bộ dữ liệu	28
Hình 4.2	Kết quả huấn luyện mô hình tự mã hóa	31
Hình 4.3	Tái tạo hình ảnh kiểm thử PolypGen2021	32
Hình 4.4	Kết quả huấn luyện mô hình phân vùng ảnh 1	33
Hình 4.5	Kết quả huấn luyện mô hình phân vùng ảnh 2	33
Hình 4.6	Kết quả huấn luyện mô hình phân vùng ảnh 3	34

Hình 4.7	Kết quả huấn luyện mô hình phân vùng ảnh 4	34
Hình 4.8	Kết quả huấn luyện mô hình phân vùng ảnh 5	34
Hình 4.9	Một số hình ảnh biểu diễn kết quả phân vùng polyp trên tập kiểm thử	35
Hình 4.10	Giao diện sử dụng mô hình tự mã hóa	36
Hình 4.11	Giao diện sử dụng mô hình phân vùng ảnh	37

DANH MỤC BẢNG BIỂU

Bảng 4.1	So sánh thông số môi trường GPU miễn phí trên Google Colab và Kaggle	30
Bảng 4.2	Lượng dữ liệu trong bộ huấn luyện và kiểm thử mô hình phân vùng ảnh	31

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Ý nghĩa
AI	Artificial Intelligent (Trí tuệ nhân tạo)
ANN	Artificial Neural Network (Mạng nơ-ron nhân tạo)
BCE	Binary Cross Entropy
BERT	Bidirectional Encoder Representations from Transformers (Biểu diễn Thể hiện Mã hóa Hai chiều từ Transformer)
BGD	Batch Gradient Descent
CNN	Convolutional Neural Network (Mạng nơ-ron tích chập)
CUDA	Compute Unified Device Architecture
DL	Deep Learning (Học sâu)
GD	Gradient Descent
GPU	Graphics Processing Unit (Bộ xử lý đồ họa)
IoU	Intersection over Union (Giao trên hợp)
LSTM	Long Short Term Memory (Bộ nhớ dài-ngắn hạn)
MAE	Masked Autoencoder
MGD	Mini-batch Gradient Descent
ML	Machine Learning (Học máy)
MLP	Multilayer Perceptron (Perceptron nhiều lớp)
MSE	Mean squared error (Sai số toàn phương trung bình)
RNN	Recurrent Neural Network (Mạng nơ-ron hồi quy)
SGD	Stochastic Gradient Descent
SSL	Self-supervised Learning (Học tự giám sát)
UNETR	Mô hình UNet Transformer
ViT	Mô hình Vision Transformer