

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Khắc phục thách thức thiếu dữ liệu trong thị giác
máy tính: Ứng dụng của học bán giám sát

CÙ MINH HOÀNG

hoang.cm204975@sis.hust.edu.vn

Ngành: Công nghệ thông tin

Giảng viên hướng dẫn: TS. Đặng Tuấn Linh

Chữ kí GVHD

Khoa: Kỹ thuật máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 06/2024

LỜI CẢM ƠN

Trong đồ án này, tôi xin bày tỏ lòng biết ơn sâu sắc tới TS. Đặng Tuấn Linh - người đã hướng dẫn tôi tận tình trong suốt quá trình học tập và nghiên cứu đồ án này. Nhờ sự chỉ dẫn tận tình của thầy tôi đã cơ hội tiếp cận và làm quen với những kiến thức, kỹ năng mới rồi từ đó phát triển và hoàn thiện đồ án này trong khả năng tốt nhất của mình.

Tôi cũng muốn gửi lời cảm ơn chân thành nhất đến bố mẹ tôi, những người đã sinh thành và nuôi dưỡng tôi để tôi có được như ngày hôm nay. Họ luôn ủng hộ và đồng hành cùng tôi vượt qua mọi thử thách, thành công trong cuộc sống, thể hiện sự quan tâm, lo lắng tối đa dành cho tôi.

Ngoài ra, tôi xin gửi lời cảm ơn tới các bạn sinh viên trường Đại học Bách Khoa Hà Nội, những người đã góp phần tạo nên những khoảnh khắc đẹp đẽ của tôi tuổi trẻ trong thời gian tôi còn là sinh viên.

Cuối cùng tôi xin gửi lời cảm ơn tới Đồng Kiều Việt Anh - người yêu của tôi. Cô ấy là người đã luôn bên cạnh và ủng hộ mọi quyết định của tôi khi đối diện với khó khăn trong đồ án này. Cũng nhờ có cô ấy mà tôi mới có thể hoàn thiện đồ án này tốt nhất.

Một lần nữa tôi xin gửi lời cảm ơn chân thành và sâu sắc nhất tới TS. Đặng Tuấn Linh các bạn sinh viên tại trường đại học cũng như người tôi yêu quý đã luôn ở bên và ủng hộ tôi suốt quá trình hoàn thành đồ án này.

LỜI CAM KẾT

Họ và tên sinh viên: Cù Minh Hoàng.....
Điện thoại liên lạc: 0338586980.....
Email: hoang.cm204975@sis.hust.edu.vn
Lớp: Việt Nhật 01 - K65
Hệ đào tạo: Elitech.....

Tôi – *Cù Minh Hoàng* – cam kết Đồ án Tốt nghiệp (ĐATN) là công trình nghiên cứu của bản thân tôi dưới sự hướng dẫn của *TS. Đặng Tuấn Linh*. Các kết quả nêu trong ĐATN là trung thực, là thành quả của riêng tôi, không sao chép theo bất kỳ công trình nào khác. Tất cả những tham khảo trong ĐATN – bao gồm hình ảnh, bảng biểu, số liệu, và các câu từ trích dẫn – đều được ghi rõ ràng và đầy đủ nguồn gốc trong danh mục tài liệu tham khảo. Tôi xin hoàn toàn chịu trách nhiệm với dù chỉ một sao chép vi phạm quy chế của nhà trường.

Hà Nội, ngày tháng năm

Tác giả ĐATN

Cù Minh Hoàng

TÓM TẮT NỘI DUNG ĐỒ ÁN

Chúng ta đang sống trong thời kì phát triển của công nghệ, thời kì mà trí tuệ nhân tạo nói chung, học máy và học sâu nói riêng đang đi sâu vào từng ngóc ngách của đời sống. Trong công nghệ, đặc biệt phải nói tới trí tuệ nhân tạo, chúng đang phát triển từng ngày từng giờ như vũ bão [1]–[3]. Thế nhưng đi cùng với tốc độ phát triển là lượng dữ liệu sạch ngày càng ít. Dữ liệu sạch là dữ liệu có nhãn, được tinh chỉnh trong một số bài toán theo một cách nhất định. Chính việc thiếu dữ liệu sạch này đã dẫn tới việc khó khăn khi huấn luyện mô hình bằng phương pháp học giám sát [4]–[6].

Trong đồ án này sẽ tập trung nghiên cứu vào vấn đề trên ở hai bài toán chính trong thực tiễn là bài toán phân loại ảnh và đánh giá chất lượng ảnh. Với bài toán phân loại ảnh, ta có rất nhiều bộ dữ liệu [7], [8], thế nhưng với nhu cầu thực tế cần giải quyết trong đồ án này thì các bộ dữ liệu này chưa phù hợp. Đối với bài toán đánh giá chất lượng ảnh, bộ dữ liệu được cung cấp bởi tập đoàn Naver và bộ dữ liệu Koka10k [9]. Chốt lại, trong đồ án sẽ tập trung vào ba bộ dữ liệu: Motor, Naver và Koka10k. Cả ba bộ dữ liệu này đều bao gồm ảnh có nhãn và không có nhãn.

Với bài toán liên quan tới thị giác máy tính, các mô hình mạng nơ ron tích chập [10]–[13] là hướng xử lý phổ biến. Chúng đã có một lịch sử hình thành phát triển lâu đời cùng với hiệu quả to lớn mà chúng đem lại. Thế nhưng trong 5 năm trở lại đây, mô hình Vision Transformer ra đời như là một bước đột phá trong các tác vụ liên quan đến thị giác máy tính. Chính vì vậy mô hình Vision Transformer được sử dụng trong các thử nghiệm của đồ án này.

Khắc phục việc thiếu dữ liệu trong thị giác máy tính có rất nhiều phương pháp khác nhau như học chủ động [14], học không giám sát [15] nhưng nổi trội nhất là học bán giám sát, nên đồ án này sẽ tập trung vào việc huấn luyện mô hình bằng học bán giám sát. Đồ án cũng nghiên cứu và thử nghiệm nhiều thuật toán học bán giám sát [16]–[20]. Rồi từ đó đề xuất ra thuật toán FreeSequenceMatch. Mô hình ViT được huấn luyện trên thuật toán này đã đạt độ chính xác 0,79 trên tập Motor, đạt MAE tổng 0,36 trên tập Naver và MAE tổng 0,28 trên tập Koka10k.

Sinh viên thực hiện
(Ký và ghi rõ họ tên)

ABSTRACT

We are living in a period of technological development, a time when artificial intelligence in general, machine learning and deep learning in particular are penetrating every corner of life. In technology, especially artificial intelligence, they are developing rapidly day by day [1]–[3]. However, along with the speed of development, the amount of clean data is becoming less and less. Clean data is labeled data, refined in some problem in a certain way. This lack of clean data has led to difficulties in training the model using the supervised learning method [4]–[6].

In this project, we will focus on researching the two main problems in practice: image classification and image quality assessment. For the image classification problem, we have many data sets [7], [8], but for the actual needs to be solved in this project, these data sets are not suitable. For the image quality assessment problem, the data set is provided by Naver Corporation and the Koka10k [9] data set. Conclusion The project will focus on three datasets: Motor, Naver and Koka10k. All three datasets include labeled and unlabeled images.

For problems related to computer vision, convolutional neural network models are popular processing methods[10]–[13]. They have a long history of formation and development along with the great effectiveness they bring. But in the past 5 years, the Vision Transformer model was born as a breakthrough in tasks related to computer vision. That is why the Vision Transformer model is used in the tests of this project.

To overcome the lack of data in computer vision, there are many different methods such as active learning [14], unsupervised learning [15] but the most prominent is semi-supervised learning, so the project This session will focus on model training using semi-supervised learning. The project also researches and tests many semi-supervised learning algorithms [16]–[20]. From there, we proposed the FreeSequenceMatch algorithm. The ViT model trained on this algorithm achieved an accuracy of 0.79 on the Motor dataset, a total MAE of 0.36 on the Naver dataset and a total MAE of 0.28 on the Koka10k dataset.

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Các giải pháp hiện tại và hạn chế	3
1.3 Mục tiêu và định hướng giải pháp	3
1.3.1 Mục tiêu.....	3
1.3.2 Định hướng giải pháp	4
1.4 Đóng góp của đề án	4
1.5 Bố cục đề án	4
CHƯƠNG 2. NỀN TẢNG LÝ THUYẾT	6
2.1 Phạm vi bài toán.....	6
2.2 Các kết quả nghiên cứu tương tự	6
2.2.1 Các cách giải quyết vấn đề thiếu dữ liệu trong thị giác máy tính	6
2.2.2 Học bán giám sát	8
2.2.3 Các bộ dữ liệu.....	13
2.2.4 Một số mô hình trong bài toán xử lý ảnh	16
2.3 Công nghệ hỗ trợ	20
2.3.1 Pytorch	20
2.3.2 Thư viện về xử lý hình ảnh trong Python	20
2.3.3 Tensorboard.....	21
CHƯƠNG 3. PHƯƠNG PHÁP ĐỀ XUẤT.....	23
3.1 Tổng quan giải pháp.....	23
3.2 Thu thập dữ liệu	24
3.3 Tạo dữ liệu thực nghiệm: Data30 và Data100	24
3.3.1 Data100	25

3.3.2 Data30	25
3.4 Tiền xử lý ảnh	26
3.5 Học giám sát	27
3.6 Học bán giám sát - Đề xuất thuật toán FreeSequenceMatch	27
3.6.1 Thuật toán FreeSequenceMatch.....	27
3.6.2 Tiền xử lý ảnh.....	30
3.6.3 Hàm mất mát.....	31
CHƯƠNG 4. PHÂN TÍCH LÝ THUYẾT.....	33
4.1 Cơ chế ngưỡng tự thích ứng	33
4.1.1 Ngưỡng toàn cầu tự thích ứng.....	33
4.1.2 Ngưỡng cục bộ tự thích ứng	33
4.2 Phương pháp đánh giá	34
4.2.1 Độ chính xác (Accuracy)	34
4.2.2 Điểm F1 (F1 Score).....	34
CHƯƠNG 5. ĐÁNH GIÁ THỰC NGHIỆM.....	36
5.1 Môi trường thực nghiệm.....	36
5.2 Siêu tham số thực nghiệm.....	36
5.3 Kết quả thực nghiệm	36
5.3.1 Chọn backbone	36
5.3.2 Thử nghiệm các thuật toán bán giám sát trên bộ Data30	37
5.3.3 Thử nghiệm tỉ lệ giữa ảnh có nhãn và không có nhãn trong batchsize trên bộ Data30	37
5.3.4 Thử nghiệm trên bộ Data30 và Data100	38
5.3.5 Thử nghiệm trên bộ Naver và Koka10k	38
5.3.6 Thử nghiệm học giám sát và học bán giám sát	39

CHƯƠNG 6. KẾT LUẬN	41
6.1 Kết luận.....	41
6.2 Hướng phát triển trong tương lai	41
TÀI LIỆU THAM KHẢO.....	44

DANH MỤC HÌNH VẼ

Hình 2.1	Pipeline của active learning	7
Hình 2.2	Luồng của thuật toán FixMatch [16]	9
Hình 2.3	Luồng của RefixMatch [19]	10
Hình 2.4	Luồng của SequenceMatch [20]	11
Hình 2.5	Cơ chế tự điều chỉnh ngưỡng [17]	12
Hình 2.6	Hình minh họa về tập dữ liệu Cifar-100 [7]	14
Hình 2.7	Các nhãn trong bộ Naver2	15
Hình 2.8	Các nhãn trong bộ Motor	16
Hình 2.9	Kiến trúc của mạng Lenet-5 [11]	17
Hình 2.10	Kiến trúc mạng Alexnet [10]	17
Hình 2.11	Kiến trúc mạng VGG-16 [12]	18
Hình 2.12	Kiến trúc của mạng Resnet-50 [13]	18
Hình 2.13	Luồng của ViT [26]	19
Hình 3.1	Luồng của quá trình khắc phục vấn đề thiếu dữ liệu trong thị giác máy tính	23
Hình 3.2	Cách tạo ra bộ dữ liệu Motor30	25
Hình 3.3	Luồng của FreeSequenceMatch	29
Hình 3.4	Ảnh được tăng cường yếu	30
Hình 3.5	Ảnh được tăng cường mạnh	30
Hình 3.6	Ảnh được tăng cường mạnh	31

DANH MỤC BẢNG BIỂU

Bảng 3.1	Phân tích về số lượng của từng nhãn trong tập dữ liệu Motor, có thể thấy nhãn về xe phân khối lớn chiếm tỉ trọng rất nhỏ, điều này làm dữ liệu mất cân bằng có thể ảnh hưởng nhiều tới quá trình huấn luyện.	24
Bảng 3.2	Số liệu thống kê về bộ dữ liệu Motor30	26
Bảng 5.1	Danh sách các siêu tham số thử nghiệm	36
Bảng 5.2	Kết quả thực nghiệm chọn mô hình	37
Bảng 5.3	Kết quả thực nghiệm các thuật toán học bán giám sát	37
Bảng 5.4	Thực nghiệm tỉ lệ ảnh có nhãn và không có nhãn trên thuật toán FreeSequenceMatch	38
Bảng 5.5	Kết quả chạy FreeSequenceMatch trên bộ dữ liệu Motor30 và Motor100	38
Bảng 5.6	Cách tăng cường với bài toán đánh giá chất lượng ảnh	39
Bảng 5.7	Kết quả thực nghiệm mô hình FreeSequenceMatch trong bài toán đánh giá ảnh	39
Bảng 5.8	Kết quả thực nghiệm học giám sát và bán giám sát	40