



**The Saigon International
University**

KHOA KỸ THUẬT & KHOA HỌC MÁY TÍNH

**ĐỒ ÁN CUỐI KỲ
MÔN TRIỂN KHAI VÀ VẬN HÀNH CÁC
MÔ HÌNH HỌC MÁY**

**ĐỀ TÀI:
ĐÁNH GIÁ CHẤT LƯỢNG HÌNH ẢNH**

SINH VIÊN: ĐẶNG CÔNG THIÊN

MSSV: 31012302899

LỚP: 23DAI

GIẢNG VIÊN: ThS. Phạm Xuân Trí

HK1 NĂM HỌC 2025-2026

I
LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành đến Thầy ThS. Phạm Xuân Trí, giảng viên môn Triển khai và vận hành các mô hình học máy, người đã luôn hỗ trợ và đồng hành cùng em trong suốt hành trình học tập và nghiên cứu.

Thầy đã không chỉ cung cấp cho em những kiến thức lý thuyết vững chắc mà còn truyền đạt cho em những hiểu biết thực tiễn quý giá, giúp em hình dung rõ ràng hơn về cách ứng dụng lý thuyết vào các tình huống thực tế. Những lời hướng dẫn của thầy đã giúp em mở rộng tầm nhìn và tự tin hơn trong quá trình thực hiện dự án.

Em đặc biệt trân trọng thái độ nhiệt huyết và tận tâm của thầy, luôn sẵn sàng giải đáp mọi thắc mắc và hướng dẫn em từng bước để hoàn thành dự án, mặc dù thầy luôn có một công việc rất bận rộn. Thầy đã tạo cho em một môi trường học tập đầy cảm hứng, khuyến khích em không ngừng học hỏi và phát triển bản thân.

Dự án này không chỉ là một sản phẩm học thuật mà còn là minh chứng cho sự tận tâm của thầy trong việc hướng dẫn và truyền đạt kiến thức. Em tin rằng những kiến thức thầy đã chia sẻ sẽ là hành trang quý báu trong sự nghiệp của em trong tương lai.

Một lần nữa, em xin gửi lời cảm ơn sâu sắc đến thầy và kính chúc thầy luôn mạnh khỏe, thành công trong sự nghiệp giảng dạy và tiếp tục truyền cảm hứng cho các thế hệ học trò sau này.

PHIẾU CHẤM

Họ và tên: Đặng Công Thiên.

Lớp: 23DAI.

Tên đồ án: Đánh giá chất lượng hình ảnh.

Đánh giá

TT	Tiêu chí	Thang điểm	Điểm chấm	Ghi chú
1	Hình thức trình bày:			
	- Trình bày đúng quy định hướng dẫn (font, số trang, mục lục, bảng biểu, danh mục tài liệu tham khảo ...).	1,0		
	- Không lỗi chính tả, lỗi đánh máy, lỗi trích dẫn tài liệu tham khảo.	0,5		
	- Trình bày đẹp, văn phong sáng, không tối nghĩa.	0,5		
2	Nội dung			
	Lời mở đầu: Trình bày tóm tắt nội dung và cấu trúc tiểu luận.	1,0		
	Chương 1: Tổng quan đề tài (Cơ sở lý thuyết).	2,5		
	Chương 2: Chuyên sâu, phân tích đề tài.	2,5		
	Chương 3: Kết luận và giải pháp đề tài.	2,0		
3	TỔNG ĐIỂM			

Điểm chữ (Làm tròn đến 1 số thập phân)

Giảng viên

ThS. Phạm Xuân Trí

MỤC LỤC

DANH MỤC CÁC CHỮ VIẾT TẮT.....	V
DANH MỤC CÁC HÌNH ẢNH/BẢNG BIỂU	VI
CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN ĐỀ TÀI.....	1
1.1. Bối cảnh nghiên cứu và thực tiễn.....	1
1.2. Lý do lựa chọn bài toán đánh giá chất lượng hình ảnh	2
1.3. Mục tiêu của đề tài	3
CHƯƠNG 2. MÔ TẢ BÀI TOÁN VÀ DỮ LIỆU	5
2.1. Mô tả bài toán đánh giá chất lượng hình ảnh.....	5
2.2. Các hướng tiếp cận trong bài toán IQA.....	5
2.3. Dataset sử dụng trong đề tài.....	6
2.3.1. Tổng quan về dataset KonIQ-10k	6
2.3.2. Cấu trúc dữ liệu	6
2.4. Giải thích chi tiết các cột trong dataset	7
2.4.1. Cột image_name	7
2.4.2. Các cột c1 đến c5.....	7
2.4.3. Cột c_total	7
2.4.4. Cột MOS (Mean Opinion Score)	7
2.4.5. Cột SD (Standard Deviation)	8
2.4.6. Cột set	8
2.5. Phân chia dữ liệu trong đề tài.....	8
2.6. Ý nghĩa của dataset đối với bài toán và MLOps	8
CHƯƠNG 3. MÔ HÌNH HỌC MÁY VÀ HUẤN LUYỆN.....	9
3.1. Định hướng tiếp cận mô hình trong bài toán đánh giá chất lượng hình ảnh	9
3.2. Mô hình hóa bài toán dưới dạng hồi quy.....	9
3.3. Kiến trúc tổng quát của mô hình	10
3.4. Các kiến trúc mạng nơ-ron tích chập được sử dụng	11
3.4.1. EfficientNet-B0	11
3.4.2. ResNet18.....	11
3.4.3. MobileNetV2.....	11

3.5. Lý do lựa chọn ba kiến trúc để so sánh	12
3.6. Chiến lược huấn luyện và lựa chọn siêu tham số.....	12
3.6.1. Lựa chọn tốc độ học và số vòng huấn luyện.....	12
3.6.2. Hàm mất mát	13
3.7. Thước đo đánh giá mô hình	13
3.7.1. Hệ số tương quan thứ hạng Spearman.....	13
3.7.2. Các thước đo hỗ trợ	13
3.8. Vai trò của chiến lược huấn luyện trong bối cảnh MLOps	13
CHƯƠNG 4. QUẢN LÝ THÍ NGHIỆM, KẾT QUẢ THỰC NGHIỆM VÀ TRIỂN KHAI HỆ THỐNG	15
4.1. Theo dõi thí nghiệm trong vận hành học máy	15
4.2. Phân tích diễn biến huấn luyện trên tập kiểm định	15
4.2.1. Diễn biến hệ số tương quan thứ hạng Spearman	15
4.2.2. So sánh kết quả trên tập kiểm tra	17
4.2.3. Phân tích diễn biến hàm mất mát	18
4.3. Triển khai hệ thống đánh giá chất lượng hình ảnh.....	19
4.3.1. Triển khai dưới dạng dịch vụ	19
4.3.2. Giao diện người dùng	19
4.4. Thảo luận kết quả trong bối cảnh vận hành học máy.....	21
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	22
5.1. Tổng kết kết quả đạt được	22
5.2. Đóng góp của đề tài.....	23
5.3. Hạn chế của đề tài	23
5.4. Hướng phát triển trong tương lai.....	24
5.5. Kết luận chung	24
TÀI LIỆU THAM KHẢO.....	25

DANH MỤC CÁC CHỮ VIẾT TẮT

STT	Chữ viết tắt	Tiếng Anh	Tiếng Việt
1	ML	Machine Learning	Học máy
2	MLOps	Machine Learning Operations	Triển khai và vận hành các mô hình học máy
3	IQA	Image Quality Assessment	Đánh giá chất lượng hình ảnh
4	CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
5	MOS	Mean Opinion Score	Điểm đánh giá trung bình từ người quan sát
6	SROCC	Spearman Rank Order Correlation Coefficient	Hệ số tương quan thứ hạng Spearman
7	PLCC	Pearson Linear Correlation Coefficient	Hệ số tương quan tuyến tính Pearson
8	MSE	Mean Squared Error	Sai số bình phương trung bình
9	RMSE	Root Mean Squared Error	Căn sai số bình phương trung bình
10	API	Application Programming Interface	Giao diện lập trình ứng dụng
11	UI	User Interface	Giao diện người dùng
12	SD	Standard Deviation	Độ lệch chuẩn

DANH MỤC CÁC HÌNH ẢNH/BẢNG BIỂU

STT	Ký hiệu	Tên bảng / hình ảnh	Trang
1	Hình 2.1	Minh họa một số dòng dữ liệu trong file <i>koniq10k_distributions_sets.csv</i>	6
2	Bảng 2.1	Phân chia dữ liệu huấn luyện, kiểm định và kiểm tra trong dataset KonIQ-10k	8
3	Bảng 3.1	Các cấu hình siêu tham số huấn luyện được sử dụng trong đề tài	12
4	Hình 4.1	Diễn biến hệ số tương quan thứ hạng Spearman trên tập kiểm định theo số vòng huấn luyện	16
5	Hình 4.2	So sánh hệ số tương quan thứ hạng Spearman trên tập kiểm tra giữa các mô hình	17
6	Hình 4.3	Diễn biến hàm mất mát trên tập huấn luyện theo số vòng huấn luyện	18
7	Hình 4.4	Giao diện người dùng ban đầu của hệ thống đánh giá chất lượng hình ảnh	20
8	Hình 4.5	Kết quả đánh giá chất lượng hình ảnh hiển thị trên giao diện người dùng	20

CHƯƠNG 1. GIỚI THIỆU TỔNG QUAN ĐỀ TÀI

1.1. Bối cảnh nghiên cứu và thực tiễn

Trong những năm gần đây, học máy đã trở thành một trong những lĩnh vực nghiên cứu và ứng dụng quan trọng nhất của khoa học máy tính. Nhờ sự phát triển mạnh mẽ của dữ liệu lớn, năng lực tính toán và các mô hình học sâu, học máy đã đạt được nhiều thành tựu nổi bật trong các lĩnh vực như thị giác máy tính, xử lý ngôn ngữ tự nhiên và hệ thống khuyến nghị. Nhiều mô hình học máy hiện đại có khả năng đạt hiệu năng rất cao trong các bài toán cụ thể, thậm chí vượt qua con người trong những điều kiện nhất định.

Tuy nhiên, trong thực tế triển khai, một mô hình học máy không chỉ được đánh giá dựa trên độ chính xác hay giá trị đo lường đạt được trên tập kiểm tra. Việc đưa mô hình vào vận hành trong môi trường thực tế đặt ra nhiều thách thức phức tạp hơn rất nhiều so với giai đoạn nghiên cứu và huấn luyện. Các vấn đề như khả năng tái tạo thí nghiệm, quản lý phiên bản mô hình, triển khai dịch vụ, kiểm thử hệ thống và duy trì mô hình theo thời gian là những yêu cầu không thể thiếu đối với một hệ thống học máy thực tế.

Trong nhiều dự án học máy truyền thống, quy trình thường chỉ dừng lại ở các bước cơ bản:

- Bước 1: Huấn luyện mô hình trên tập dữ liệu đã chuẩn bị
- Bước 2: Báo cáo kết quả trên tập kiểm tra
- Bước 3: Lưu mô hình dưới dạng tệp tin một cách thủ công.

Cách tiếp cận này tuy phù hợp cho nghiên cứu ban đầu hoặc thử nghiệm nhỏ, nhưng bộc lộ nhiều hạn chế nghiêm trọng khi hệ thống cần được sử dụng lâu dài hoặc mở rộng quy mô:

- Khó tái tạo thí nghiệm: thiếu thông tin về dữ liệu, tham số huấn luyện và môi trường chạy.
- Không có cơ chế quản lý và so sánh phiên bản mô hình: khó xác định mô hình nào đang được sử dụng hoặc mô hình nào tốt hơn.
- Khó triển khai và bảo trì: việc cập nhật mô hình thường đòi hỏi thao tác thủ công, dễ gây lỗi.

- Thiếu kiểm thử và giám sát: không đảm bảo hệ thống hoạt động ổn định khi đưa vào vận hành.

Những vấn đề này đã được chỉ ra rõ ràng trong các nghiên cứu về hệ thống học máy quy mô lớn, cho thấy rằng nợ kỹ thuật trong học máy có thể gia tăng nhanh chóng nếu không có quy trình vận hành phù hợp [1]

Để giải quyết các thách thức trên, vận hành học máy (Machine Learning Operations – MLOps) được đề xuất như một hướng tiếp cận nhằm quản lý toàn bộ vòng đời của mô hình học máy, từ dữ liệu, huấn luyện, đánh giá, lưu trữ mô hình cho đến triển khai và vận hành trong môi trường thực tế. MLOps kế thừa các nguyên lý của kỹ nghệ phần mềm và vận hành hệ thống, đồng thời thích ứng với đặc thù của các mô hình học máy [2].

Theo cách tiếp cận MLOps, mô hình học máy không còn là một thực thể độc lập mà trở thành một thành phần trong một hệ thống hoàn chỉnh, có khả năng tái sử dụng, mở rộng và duy trì lâu dài trong thực tế.

1.2. Lý do lựa chọn bài toán đánh giá chất lượng hình ảnh

Đánh giá chất lượng hình ảnh là bài toán ước lượng mức độ chất lượng của một hình ảnh dựa trên cảm nhận thị giác của con người. Khác với các bài toán thị giác máy tính truyền thống, bài toán này không có nhãn tuyệt đối mà phụ thuộc vào đánh giá chủ quan của nhiều người quan sát.

Chất lượng hình ảnh ảnh hưởng trực tiếp đến hiệu quả và độ tin cậy của nhiều hệ thống trong thực tế, bao gồm:

- Trải nghiệm người dùng trong các ứng dụng đa phương tiện
- Chất lượng dữ liệu đầu vào của các hệ thống thị giác máy tính
- Hiển thị và phân phối nội dung trên các nền tảng số
- Các ứng dụng thương mại điện tử và truyền thông

Trong nhiều kịch bản thực tế, không tồn tại ảnh tham chiếu có chất lượng hoàn hảo để so sánh. Do đó, các phương pháp đánh giá chất lượng hình ảnh không tham chiếu được đặc biệt quan tâm. Các phương pháp này chỉ sử dụng thông tin từ ảnh đầu vào để ước lượng chất lượng, phản ánh sát hơn điều kiện sử dụng ngoài thực tế [3].

Một trong những đặc điểm quan trọng của bài toán này là nhãn chất lượng thường được biểu diễn bằng điểm đánh giá trung bình của nhiều người, gọi là Mean Opinion Score (MOS). MOS phản ánh cảm nhận chủ quan của con người và do đó chứa nhiễu, khiến bài toán trở nên thách thức hơn so với các bài toán phân loại truyền thống.

Bài toán đánh giá chất lượng hình ảnh phù hợp để nghiên cứu và triển khai trong bối cảnh MLOps bởi các lý do sau:

- Dữ liệu có quy mô lớn và được công bố công khai
- Nhãn mang tính chủ quan, phản ánh bài toán thực tế
- Tồn tại các thước đo đánh giá chuẩn được chấp nhận rộng rãi trong cộng đồng nghiên cứu
- Có khả năng triển khai trực tiếp thành dịch vụ phục vụ người dùng cuối

1.3. Mục tiêu của đề tài

Xuất phát từ bối cảnh nghiên cứu và bài toán đã trình bày, đề tài này được thực hiện với các mục tiêu chính sau:

- Xây dựng một quy trình MLOps hoàn chỉnh chạy trên môi trường cục bộ cho bài toán đánh giá chất lượng hình ảnh, bao phủ toàn bộ các giai đoạn từ dữ liệu đến triển khai.
- Minh họa đầy đủ vòng đời của mô hình học máy, bao gồm: chuẩn bị và tiền xử lý dữ liệu, huấn luyện và đánh giá mô hình, theo dõi và so sánh các thí nghiệm, quản lý phiên bản mô hình, triển khai dịch vụ và kiểm thử.
- So sánh nhiều kiến trúc mạng nơ-ron tích chập và nhiều cấu hình huấn luyện khác nhau nhằm phân tích ảnh hưởng của kiến trúc và siêu tham số tới hiệu năng đánh giá chất lượng hình ảnh.
- Triển khai mô hình dưới dạng giao diện lập trình ứng dụng và giao diện web, giúp hệ thống dễ trình diễn và phù hợp với người không chuyên về học máy.

- Tạo nền tảng cho việc mở rộng sang môi trường vận hành thực tế, bao gồm khả năng cập nhật mô hình, giám sát hiệu năng và tích hợp quy trình triển khai liên tục trong tương lai.

Mục tiêu cuối cùng của đề tài không chỉ là đạt được kết quả đánh giá chất lượng hình ảnh tốt, mà quan trọng hơn là xây dựng một hệ thống học máy có khả năng vận hành ổn định, tái sử dụng và mở rộng, đúng theo tinh thần của MLOps.

CHƯƠNG 2. MÔ TẢ BÀI TOÁN VÀ DỮ LIỆU

2.1. Mô tả bài toán đánh giá chất lượng hình ảnh

Đánh giá chất lượng hình ảnh (Image Quality Assessment – IQA) là bài toán ước lượng mức độ chất lượng của một hình ảnh dựa trên cảm nhận thị giác của con người, thay vì dựa trên các tiêu chí toán học thuần túy như độ sắc nét hay mức nhiễu. Trong thực tế, chất lượng hình ảnh mang tính chủ quan, phụ thuộc vào cảm nhận của từng người quan sát, điều kiện hiển thị và bối cảnh sử dụng. Trong đề tài này, bài toán IQA được mô hình hóa dưới dạng bài toán hồi quy (regression) với các đặc điểm sau:

- Đầu vào (input): một ảnh màu ba kênh (RGB)
- Đầu ra (output): một giá trị liên tục biểu diễn mức độ chất lượng của ảnh

Không giống các bài toán phân loại, đầu ra của bài toán IQA không thuộc một tập nhãn rời rạc, mà nằm trên một trục liên tục, phản ánh mức độ chất lượng từ thấp đến cao.

Một điểm quan trọng của IQA là mục tiêu đánh giá không chỉ là sai số tuyệt đối, mà còn là khả năng xếp hạng đúng chất lượng tương đối giữa các ảnh. Do đó, các hệ số tương quan thứ hạng như Spearman Rank Order Correlation Coefficient (SROCC) thường được sử dụng làm thước đo chính trong các nghiên cứu IQA [4].

2.2. Các hướng tiếp cận trong bài toán IQA

Các phương pháp IQA trong nghiên cứu thường được chia thành ba nhóm chính [5]:

- Đánh giá có tham chiếu đầy đủ (Full-Reference IQA): So sánh ảnh cần đánh giá với ảnh gốc chất lượng cao.
- Đánh giá có tham chiếu một phần (Reduced-Reference IQA): Chỉ sử dụng một phần thông tin của ảnh tham chiếu.
- Đánh giá không tham chiếu (No-Reference IQA): Chỉ sử dụng ảnh đầu vào, không có ảnh tham chiếu.

Trong thực tế triển khai, đặc biệt là trong các hệ thống ngoài môi trường phòng thí nghiệm, ảnh tham chiếu thường không tồn tại. Vì vậy, đề tài này tập trung vào bài toán đánh giá chất lượng hình ảnh không tham chiếu, phản ánh sát hơn điều kiện sử dụng thực tế.

2.3. Dataset sử dụng trong đề tài

2.3.1. Tổng quan về dataset KonIQ-10k

Đề tài sử dụng dataset KonIQ-10k (Konstanz Natural Image Quality Database), một dataset đánh giá chất lượng hình ảnh không tham chiếu được công bố rộng rãi và sử dụng phổ biến trong cộng đồng nghiên cứu IQA [3].

KonIQ-10k được xây dựng dựa trên các ảnh tự nhiên thu thập từ môi trường thực tế (in-the-wild), thay vì các ảnh bị làm suy giảm nhân tạo. Điều này giúp dataset phản ánh đa dạng các dạng suy giảm chất lượng như: mờ, nhiễu, ánh sáng kém, nén ảnh, sai lệch màu sắc.

2.3.2. Cấu trúc dữ liệu

Dataset bao gồm hai thành phần chính:

- Tập ảnh: được lưu trong thư mục data/512x384/, mỗi ảnh có kích thước cố định 512×384.
- Tập nhãn và thông tin đánh giá: được lưu trong file koniq10k_distributions_sets.csv.

	image_name	c1	c2	c3	c4	c5	c_total	MOS	SD	set
0	10004473376.jpg	0.0000	0.0000	0.2381	0.6952	0.0667	105	77.3836	0.5273	training
1	10007357496.jpg	0.0000	0.0312	0.4688	0.4896	0.0104	96	68.7286	0.5800	test
2	10007903636.jpg	0.0104	0.0000	0.2083	0.7604	0.0208	96	78.6286	0.5272	training
3	10009096245.jpg	0.0000	0.0000	0.1927	0.6881	0.1193	109	77.2438	0.5562	training
4	100117038.jpg	0.0000	0.0100	0.2100	0.7200	0.0600	100	75.1125	0.5329	training
5	10012398043.jpg	0.0000	0.0680	0.5340	0.3883	0.0097	103	65.1495	0.6192	training
6	10013822223.jpg	0.0000	0.0000	0.4364	0.5455	0.0182	110	68.3063	0.5313	training
7	10014002683.jpg	0.0000	0.0566	0.4340	0.4811	0.0283	106	64.7627	0.6507	training
8	10017119924.jpg	0.0000	0.0270	0.5225	0.4505	0.0000	111	66.2244	0.5485	training
9	10020766793.jpg	0.0000	0.0000	0.1009	0.7156	0.1835	109	81.5062	0.5293	test

Hình 2.1: Minh họa một số dòng dữ liệu trong koniq10k_distributions_sets.csv

2.4. Giải thích chi tiết các cột trong dataset

Một dòng dữ liệu mẫu trong file CSV có dạng như hình 2.1

2.4.1. Cột `image_name`

- Tên file ảnh tương ứng trong thư mục ảnh.
- Dùng để ánh xạ mỗi dòng nhãn với đúng ảnh đầu vào.

2.4.2. Các cột `c1` đến `c5`

Các cột `c1`, `c2`, `c3`, `c4`, `c5` biểu diễn phân bố xác suất đánh giá chất lượng ảnh theo 5 mức chất lượng khác nhau.

- Mỗi mức tương ứng với một mức đánh giá chủ quan (từ rất kém đến rất tốt).
- Giá trị trong mỗi cột là tỉ lệ số người đánh giá ảnh ở mức đó.

2.4.3. Cột `c_total`

- Tổng số lượt đánh giá cho ảnh đó.
- Mỗi ảnh trong KonIQ-10k thường được đánh giá bởi hàng chục đến hàng trăm người.
- Giá trị này phản ánh độ tin cậy thống kê của MOS.

2.4.4. Cột MOS (Mean Opinion Score)

- MOS (điểm đánh giá trung bình) là nhãn chính được sử dụng trong huấn luyện mô hình.
- MOS được tính bằng trung bình có trọng số của các mức đánh giá:

$$MOS = \sum_{i=1}^5 w_i \cdot p_i$$

Trong đó:

w_i : giá trị điểm của mức đánh giá thứ i

p_i : tỉ lệ đánh giá ở mức i (các cột `c1`–`c5`)

MOS trong KonIQ-10k được quy đổi về thang điểm 0–100, giúp biểu diễn trực quan mức độ chất lượng ảnh [3].

Do MOS phản ánh cảm nhận chủ quan của con người nên:

- MOS không phải nhãn tuyệt đối
- Có thể tồn tại nhiều và sai lệch giữa các người đánh giá

2.4.5. Cột SD (Standard Deviation)

- Độ lệch chuẩn của các đánh giá chất lượng
- Phản ánh mức độ không đồng thuận giữa các người đánh giá
- SD cao \rightarrow cảm nhận chất lượng ảnh không thống nhất

Trong đề tài này, SD không được dùng làm nhãn, nhưng cung cấp thông tin quan trọng về độ tin cậy của MOS

2.4.6. Cột set

Chỉ định tập dữ liệu mà ảnh thuộc về:

- training
- validation
- test

Việc phân chia này được thực hiện ngay từ dataset gốc, giúp tránh rò rỉ dữ liệu giữa các tập.

2.5. Phân chia dữ liệu trong đề tài

Dataset được phân chia như sau:

Tập dữ liệu	Số lượng
Huấn luyện	7,058
Kiểm định	1,000
Kiểm tra	2,015

Bảng 2.1: Phân chia dữ liệu huấn luyện

- Tập huấn luyện dùng để tối ưu tham số mô hình.
- Tập kiểm định dùng để lựa chọn mô hình và cấu hình tốt nhất.
- Tập kiểm tra chỉ dùng cho đánh giá cuối cùng, đảm bảo tính khách quan.

2.6. Ý nghĩa của dataset đối với bài toán và MLOps

Việc sử dụng KonIQ-10k mang lại nhiều lợi thế cho nghiên cứu MLOps:

- Dataset lớn \rightarrow phù hợp cho huấn luyện mô hình học sâu.
- Nhãn chủ quan \rightarrow phản ánh bài toán thực tế.
- Có phân bố đánh giá chi tiết \rightarrow cho phép phân tích sâu.
- Phù hợp để đánh giá khả năng tổng quát hóa và độ ổn định của mô hình khi triển khai.

CHƯƠNG 3. MÔ HÌNH HỌC MÁY VÀ HUẤN LUYỆN

3.1. Định hướng tiếp cận mô hình trong bài toán đánh giá chất lượng hình ảnh

Bài toán đánh giá chất lượng hình ảnh không tham chiếu đặt ra nhiều thách thức đối với thiết kế mô hình học máy. Không giống các bài toán phân loại hoặc phát hiện đối tượng, IQA yêu cầu mô hình không chỉ nhận diện nội dung trong ảnh mà còn phải hiểu được mức độ suy giảm chất lượng do nhiều yếu tố khác nhau như nhiễu, mờ, nén ảnh, ánh sáng kém hoặc sai lệch màu sắc.

Trong các nghiên cứu gần đây, các mô hình học sâu dựa trên mạng nơ-ron tích chập (Convolutional Neural Network – CNN) đã cho thấy hiệu quả vượt trội trong việc trích xuất đặc trưng thị giác liên quan đến chất lượng ảnh [6]. Do đó, trong đề tài này, bài toán IQA được tiếp cận bằng cách sử dụng các kiến trúc CNN hiện đại làm mạng trích xuất đặc trưng, kết hợp với một tầng hồi quy để dự đoán điểm chất lượng.

Thay vì thiết kế một kiến trúc hoàn toàn mới, đề tài tập trung vào việc:

- Sử dụng các kiến trúc CNN đã được kiểm chứng
- Khai thác trọng số huấn luyện sẵn (pretrained)
- So sánh các kiến trúc có đặc tính khác nhau dưới cùng một pipeline MLOps

Cách tiếp cận này phản ánh đúng thực tiễn triển khai, nơi việc lựa chọn mô hình không chỉ dựa trên độ chính xác mà còn phụ thuộc vào độ ổn định, chi phí tính toán và khả năng triển khai.

3.2. Mô hình hóa bài toán dưới dạng hồi quy

Trong đề tài này, bài toán đánh giá chất lượng hình ảnh được mô hình hóa dưới dạng bài toán hồi quy (regression) với các đặc điểm sau:

- Đầu vào: một ảnh màu ba kênh (RGB)
- Đầu ra: một giá trị liên tục biểu diễn mức độ chất lượng ảnh

Khác với các bài toán phân loại, đầu ra của mô hình không thuộc một tập nhãn rời rạc, mà nằm trên một miền liên tục. Do đó, kiến trúc mô hình cần được thiết kế sao cho:

- Có khả năng dự đoán giá trị thực
- Đảm bảo đầu ra nằm trong miền hợp lệ
- Ổn định trong quá trình tối ưu.

Trong quá trình huấn luyện, nhãn MOS của tập dữ liệu được chuẩn hóa về khoảng $[0, 1]$, do đó tầng đầu ra của mô hình sử dụng hàm kích hoạt sigmoid.

Hàm sigmoid giúp:

- Giới hạn giá trị đầu ra
- Tránh hiện tượng dự đoán vượt miền
- Ổn định gradient trong quá trình huấn luyện

Khi triển khai mô hình để phục vụ người dùng, giá trị đầu ra được quy đổi về thang điểm $[0, 100]$, phù hợp với cách diễn giải trực quan của con người.

3.3. Kiến trúc tổng quát của mô hình

Tất cả các mô hình trong đề tài đều tuân theo cùng một kiến trúc tổng quát, bao gồm hai thành phần chính:

1. Mạng trích xuất đặc trưng (Backbone CNN): chịu trách nhiệm học các đặc trưng thị giác liên quan đến chất lượng ảnh.
2. Tầng hồi quy (Regression Head): chuyển đổi đặc trưng trích xuất được thành một giá trị chất lượng duy nhất.

Cụ thể:

- Phần backbone sử dụng các kiến trúc CNN huấn luyện sẵn trên tập dữ liệu ImageNet.
- Phần phân loại cuối cùng của backbone được loại bỏ.
- Một tầng kết nối đầy đủ (fully-connected) được thêm vào để dự đoán giá trị chất lượng.
- Hàm kích hoạt sigmoid được áp dụng ở đầu ra cuối cùng.

Thiết kế này giúp:

- Tận dụng tri thức đã học từ dữ liệu lớn
- Giảm thời gian huấn luyện
- Tăng khả năng tổng quát hóa của mô hình [7].

3.4. Các kiến trúc mạng nơ-ron tích chập được sử dụng

Trong đề tài này, ba kiến trúc CNN đại diện cho các nhóm đặc tính khác nhau được lựa chọn để so sánh là EfficientNet-B0, ResNet18, MobileNetV2.

3.4.1. EfficientNet-B0

EfficientNet là họ mô hình CNN được đề xuất với ý tưởng tăng đồng thời độ sâu, độ rộng và độ phân giải ảnh đầu vào một cách có hệ thống [6]. Phiên bản EfficientNet-B0 là cấu hình cơ sở, có số lượng tham số vừa phải nhưng hiệu năng cao.

Ưu điểm của EfficientNet-B0:

- Hiệu năng tốt trên nhiều bài toán thị giác
- Sử dụng tài nguyên tính toán hiệu quả
- Phù hợp cho cả nghiên cứu và triển khai

Trong đề tài, EfficientNet-B0 đóng vai trò là mô hình có kỳ vọng hiệu năng cao nhất.

3.4.2. ResNet18

ResNet là kiến trúc CNN nổi tiếng với cơ chế kết nối tắt (skip connection), giúp khắc phục hiện tượng suy giảm gradient khi mạng trở nên sâu hơn [8].

ResNet18 là phiên bản tương đối nông, có ưu điểm:

- Dễ huấn luyện
- Ổn định
- Thường được sử dụng làm mô hình cơ bản trong nhiều nghiên cứu.

Trong đề tài, ResNet18 được chọn làm mốc so sánh cơ bản, phản ánh hiệu năng của một mô hình CNN truyền thống.

3.4.3. MobileNetV2

MobileNetV2 được thiết kế với mục tiêu tối ưu hóa cho thiết bị hạn chế tài nguyên, sử dụng các khối tích chập tách rời theo chiều sâu (depthwise separable convolution) và cơ chế inverted residual [9].

Ưu điểm của MobileNetV2:

- Số lượng tham số nhỏ
- Tốc độ suy luận nhanh

- Phù hợp cho triển khai trên thiết bị biên hoặc hệ thống yêu cầu độ trễ thấp

Trong đề tài, MobileNetV2 đại diện cho nhóm mô hình ưu tiên khả năng triển khai.

3.5. Lý do lựa chọn ba kiến trúc để so sánh

Việc lựa chọn EfficientNet-B0, ResNet18 và MobileNetV2 không nhằm tìm ra mô hình tốt nhất về mặt học thuật, mà nhằm phản ánh các quyết định thực tế trong MLOps.

Ba mô hình đại diện cho ba nhóm đặc tính quan trọng:

- Hiệu năng dự đoán cao: EfficientNet-B0
- Ổn định và dễ huấn luyện: ResNet18
- Nhẹ và dễ triển khai: MobileNetV2

Sự so sánh này cho phép đánh giá:

- Mức độ đánh đổi giữa độ chính xác và khả năng triển khai
- Sự khác biệt về hành vi huấn luyện
- Tính phù hợp của từng mô hình trong các kịch bản vận hành khác nhau

3.6. Chiến lược huấn luyện và lựa chọn siêu tham số

3.6.1. Lựa chọn tốc độ học và số vòng huấn luyện

Trong đề tài, mỗi kiến trúc được huấn luyện với ba cấu hình siêu tham số khác nhau, bao gồm tốc độ học (learning rate) và số vòng huấn luyện (epoch) như ở Bảng 3.1:

Tốc độ học	Số epoch	Mục đích
1e-3	3	Kiểm tra khả năng hội tụ nhanh
3e-4	5	Cân bằng giữa tốc độ và ổn định
1e-4	6	Tinh chỉnh chậm, giảm overfitting

Bảng 3.1. Các cấu hình siêu tham số huấn luyện

Cách thiết kế này giúp:

- Quan sát hành vi học của mô hình trong các điều kiện khác nhau
- So sánh mức độ ổn định giữa các kiến trúc
- Tránh việc tối ưu siêu tham số quá mức cho một mô hình cụ thể

3.6.2. Hàm mất mát

Do bài toán được mô hình hóa dưới dạng hồi quy, hàm mất mát được sử dụng là sai số bình phương trung bình (Mean Squared Error – MSE). Hàm này đo lường khoảng cách bình phương giữa giá trị dự đoán và giá trị nhãn chuẩn hóa. MSE thường được sử dụng trong các bài toán hồi quy và có tính chất:

- Khả vi
- Dễ tối ưu
- Nhạy với các sai lệch lớn

3.7. Thước đo đánh giá mô hình

3.7.1. Hệ số tương quan thứ hạng Spearman

Hệ số tương quan thứ hạng Spearman (Spearman Rank Order Correlation Coefficient – SROCC) được sử dụng làm thước đo chính trong đề tài.

SROCC đo mức độ tương quan giữa thứ hạng của giá trị dự đoán và thứ hạng của giá trị thực, phản ánh khả năng xếp hạng đúng chất lượng ảnh [4].

SROCC đặc biệt phù hợp với IQA vì:

- MOS mang tính chủ quan
- Thứ tự chất lượng quan trọng hơn sai số tuyệt đối

3.7.2. Các thước đo bổ trợ

Bên cạnh SROCC, các thước đo sau cũng được sử dụng để phân tích mô hình:

- Hệ số tương quan tuyến tính Pearson
- Sai số tuyệt đối trung bình
- Căn sai số bình phương trung bình
- các thước đo quy đổi về thang điểm 0–100

Việc sử dụng nhiều thước đo giúp đánh giá mô hình một cách toàn diện, tránh phụ thuộc vào một chỉ số duy nhất.

3.8. Vai trò của chiến lược huấn luyện trong bối cảnh MLOps

Chiến lược huấn luyện trong đề tài không chỉ nhằm đạt hiệu năng cao nhất, mà còn hướng tới:

- Khả năng tái tạo thí nghiệm
- So sánh công bằng giữa các mô hình
- Tích hợp chặt chẽ với pipeline MLOps

Mỗi lần huấn luyện được xem là một thí nghiệm độc lập, với đầy đủ thông tin về:

- Mô hình
- Siêu tham số
- Kết quả đánh giá

Cách tiếp cận này là nền tảng cho việc theo dõi, lựa chọn và triển khai mô hình trong các chương tiếp theo.

CHƯƠNG 4. QUẢN LÝ THÍ NGHIỆM, KẾT QUẢ THỰC NGHIỆM VÀ TRIỂN KHAI HỆ THỐNG

4.1. Theo dõi thí nghiệm trong vận hành học máy

Trong các hệ thống học máy hiện đại, đặc biệt là trong bối cảnh vận hành học máy, việc theo dõi thí nghiệm huấn luyện không chỉ nhằm ghi nhận kết quả cuối cùng mà còn đóng vai trò then chốt trong việc đảm bảo tính tái tạo, khả năng so sánh và ra quyết định lựa chọn mô hình một cách có cơ sở khoa học.

Mỗi lần huấn luyện mô hình trong đề tài này được xem là một thí nghiệm độc lập, bao gồm đầy đủ các thông tin: kiến trúc mô hình, siêu tham số, diễn biến hàm mất mát, các thước đo đánh giá trên tập kiểm định và kết quả cuối cùng trên tập kiểm tra. Việc ghi nhận có hệ thống các thông tin này cho phép phân tích sâu hành vi học của mô hình, thay vì chỉ dựa vào một giá trị đo lường duy nhất.

Cách tiếp cận này phù hợp với các khuyến nghị trong thực tiễn vận hành học máy, nơi việc theo dõi và so sánh thí nghiệm là điều kiện tiên quyết để giảm nợ kỹ thuật và đảm bảo khả năng mở rộng của hệ thống [10].

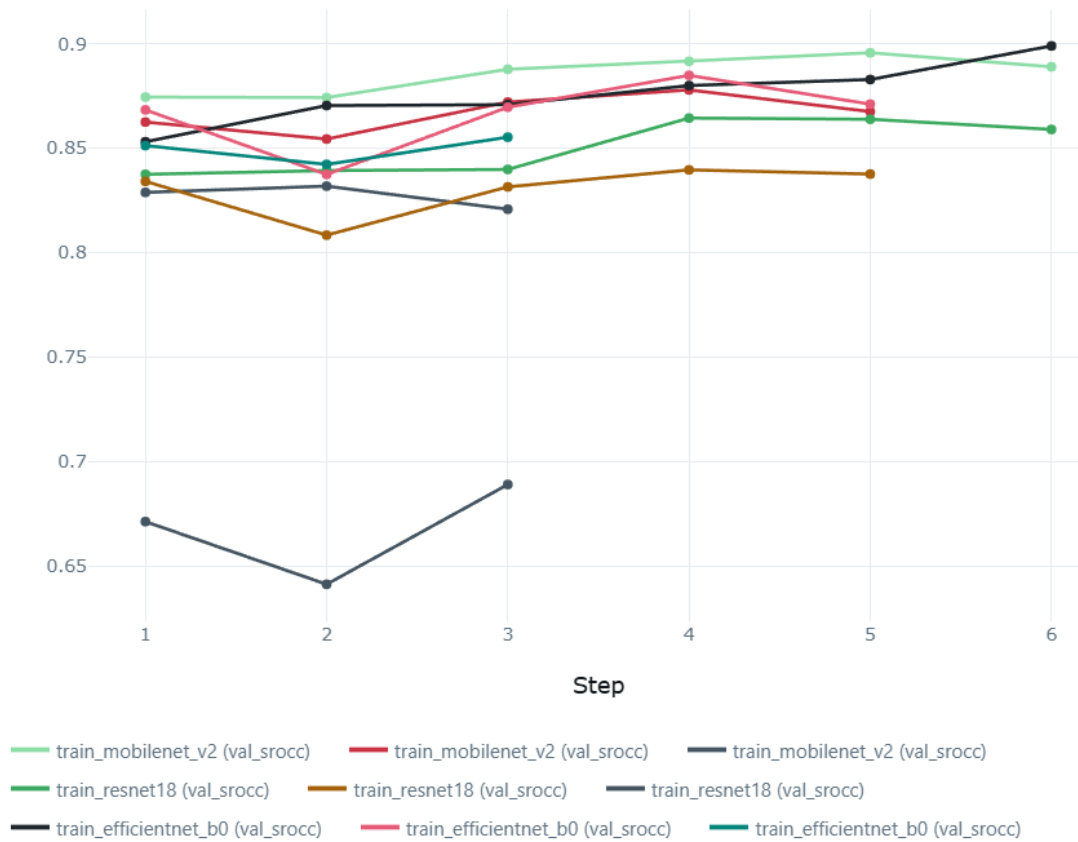
4.2. Phân tích diễn biến huấn luyện trên tập kiểm định

4.2.1. Diễn biến hệ số tương quan thứ hạng Spearman

Hệ số tương quan thứ hạng Spearman (Spearman Rank Order Correlation Coefficient – SROCC) được sử dụng làm thước đo chính trong đề tài do khả năng phản ánh chính xác mức độ tương quan về thứ hạng giữa giá trị dự đoán và giá trị MOS thực tế.

Hình 4.1 thể hiện sự thay đổi của SROCC trên tập kiểm định theo từng vòng huấn luyện đối với chín thí nghiệm, tương ứng với ba kiến trúc mô hình và ba cấu hình tốc độ học – số vòng huấn luyện khác nhau.

val_srocc



Hình 4.1. Diễn biến hệ số tương quan thứ hạng Spearman trên tập kiểm định theo số vòng huấn luyện

Từ Hình 4.1 có thể rút ra các nhận xét quan trọng:

- Các mô hình sử dụng tốc độ học nhỏ hơn ($1e-4$) và số vòng huấn luyện lớn hơn cho thấy xu hướng cải thiện SROCC rõ rệt và ổn định hơn.
- Kiến trúc EfficientNet-B0 thể hiện quá trình học nhất quán, với SROCC tăng đều theo từng vòng huấn luyện và đạt giá trị cao nhất ở cấu hình huấn luyện dài.
- MobileNetV2 đạt SROCC tương đối cao ngay từ các vòng huấn luyện đầu, phản ánh khả năng hội tụ nhanh của kiến trúc nhẹ.
- ResNet18 cho thấy mức SROCC thấp hơn và dao động nhiều hơn ở các cấu hình huấn luyện ngắn, cho thấy giới hạn của kiến trúc này trong bài toán đánh giá chất lượng hình ảnh không tham chiếu.

Việc phân tích diễn biến SROCC theo thời gian cho phép đánh giá không chỉ mức hiệu năng cuối cùng mà còn cả độ ổn định của mô hình trong quá trình huấn luyện, một yếu tố quan trọng trong vận hành học máy [11].

4.2.2. So sánh kết quả trên tập kiểm tra

Sau khi hoàn tất huấn luyện, các mô hình được đánh giá trên tập kiểm tra, tập dữ liệu hoàn toàn không tham gia vào quá trình huấn luyện hay lựa chọn mô hình. Kết quả này phản ánh khả năng tổng quát hóa thực sự của mô hình.

Hình 4.2 trình bày so sánh giá trị SROCC trên tập kiểm tra của tất cả chín thí nghiệm.

test_srocc



Hình 4.2. So sánh hệ số tương quan thứ hạng Spearman trên tập kiểm tra giữa các mô hình

Từ Hình 4.2 có thể thấy:

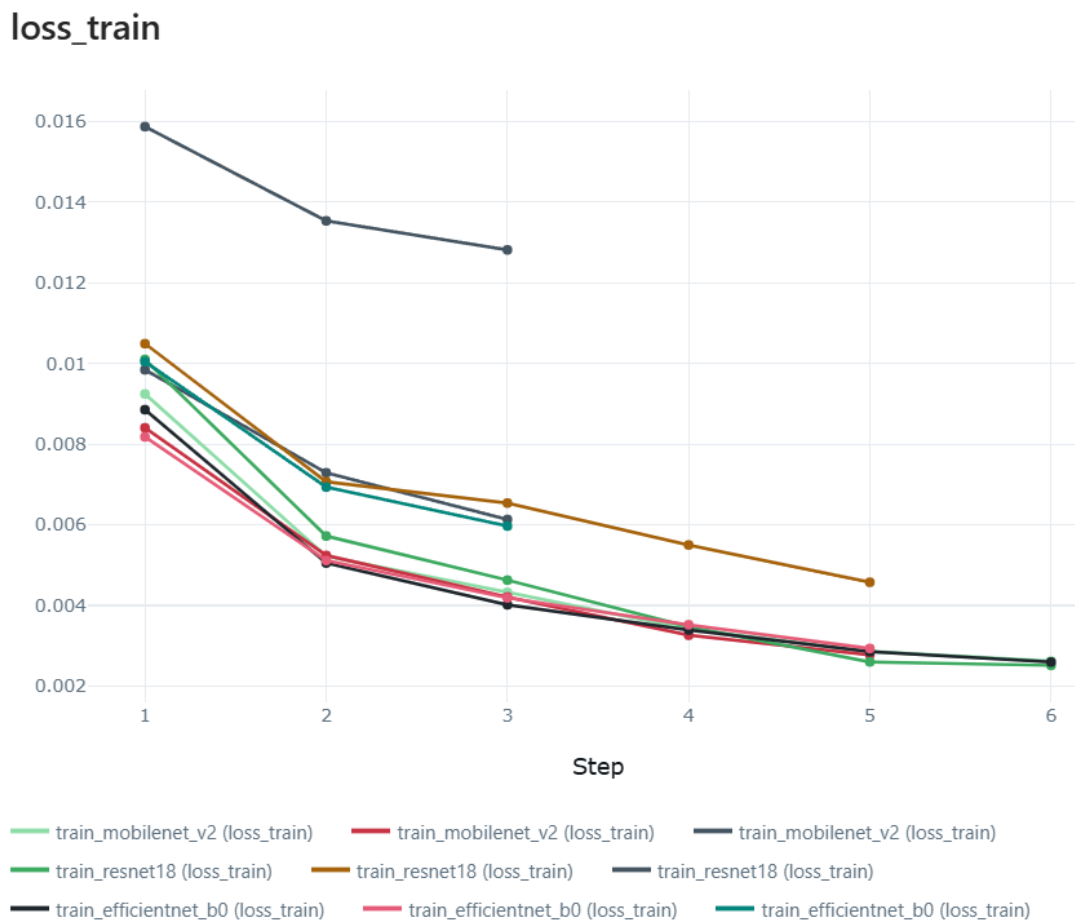
- EfficientNet-B0 đạt giá trị SROCC cao nhất trên tập kiểm tra, xấp xỉ 0.89, cho thấy khả năng xếp hạng chất lượng ảnh rất tốt.

- MobileNetV2 đạt kết quả chỉ thấp hơn EfficientNet-B0 một khoảng rất nhỏ, chứng minh rằng các mô hình nhẹ vẫn có thể đạt hiệu năng cạnh tranh.
- ResNet18 có kết quả thấp hơn rõ rệt, đặc biệt ở cấu hình huấn luyện ngắn, phản ánh hạn chế của kiến trúc này đối với dữ liệu ảnh tự nhiên phức tạp.

Sự nhất quán giữa kết quả trên tập kiểm định và tập kiểm tra cho thấy quy trình lựa chọn mô hình dựa trên tập kiểm định là hợp lý và không gây hiện tượng đánh giá quá lạc quan [12].

4.2.3. Phân tích diễn biến hàm mất mát

Bên cạnh các thước đo tương quan, hàm mất mát trên tập huấn luyện cũng được theo dõi nhằm đánh giá khả năng tối ưu của mô hình. Hình 4.3 minh họa sự thay đổi của hàm mất mát theo số vòng huấn luyện đối với các mô hình.



Hình 4.3. Diễn biến hàm mất mát trên tập huấn luyện theo số vòng huấn luyện

Quan sát Hình 4.3 cho thấy:

- Tất cả các mô hình đều có xu hướng giảm hàm mất mát theo thời gian, chứng tỏ quá trình huấn luyện diễn ra đúng hướng.
- EfficientNet-B0 và MobileNetV2 cho thấy tốc độ giảm hàm mất mát nhanh và ổn định hơn.
- ResNet18 có xu hướng giảm chậm hơn và ít mượt hơn ở một số cấu hình, phù hợp với kết quả SROCC thấp hơn.

Việc kết hợp phân tích hàm mất mát và hệ số tương quan thứ hạng giúp đánh giá mô hình một cách toàn diện, tránh phụ thuộc vào một chỉ số duy nhất [13].

4.3. Triển khai hệ thống đánh giá chất lượng hình ảnh

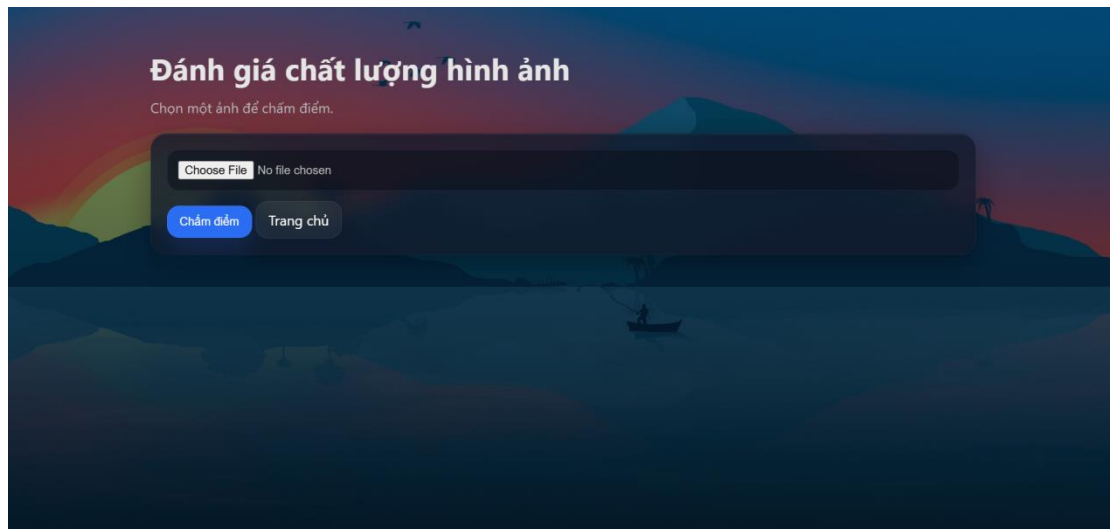
4.3.1. Triển khai dưới dạng dịch vụ

Sau khi lựa chọn mô hình tốt nhất, hệ thống đánh giá chất lượng hình ảnh được triển khai dưới dạng dịch vụ, cho phép người dùng gửi ảnh đầu vào và nhận lại điểm chất lượng tương ứng. Mô hình được tải tự động dựa trên nhãn phiên bản đã được lựa chọn trong kho lưu trữ mô hình, đảm bảo rằng hệ thống luôn sử dụng phiên bản tối ưu nhất tại thời điểm vận hành.

Để giảm độ trễ và tăng hiệu năng, mô hình được lưu trong bộ nhớ tạm sau lần tải đầu tiên, tránh việc khởi tạo lại mô hình cho mỗi yêu cầu. Cách tiếp cận này phù hợp với các thực hành phổ biến trong triển khai mô hình học máy ở môi trường thực tế [14].

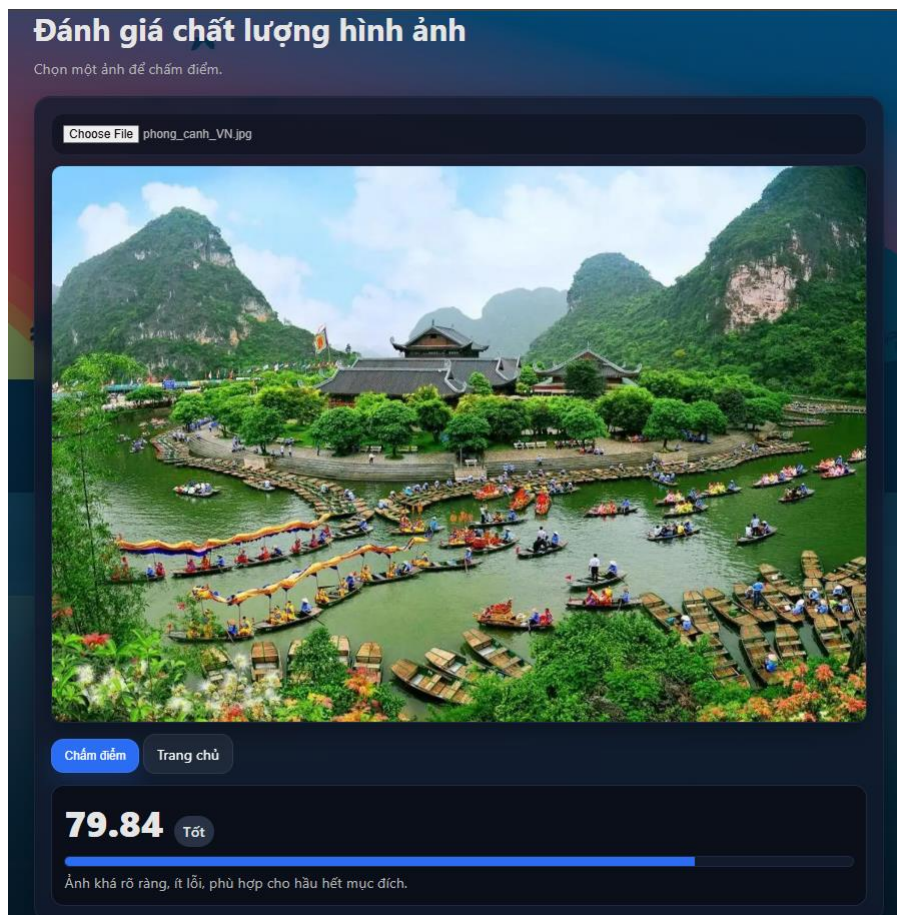
4.3.2. Giao diện người dùng

Hệ thống cung cấp giao diện người dùng trực quan nhằm hỗ trợ việc trình diễn và sử dụng. Hình 4.4 minh họa giao diện ban đầu, nơi người dùng có thể tải ảnh lên để đánh giá.



Hình 4.4. Giao diện người dùng của hệ thống đánh giá chất lượng hình ảnh

Sau khi ảnh được xử lý, kết quả đánh giá được hiển thị trực tiếp trên giao diện, bao gồm điểm chất lượng theo thang 0–100, thanh hiển thị mức độ chất lượng và mô tả ngắn gọn. Hình 4.5 minh họa kết quả đánh giá của một ảnh cụ thể.



Hình 4.5. Kết quả đánh giá chất lượng hình ảnh hiển thị trên giao diện người dùng

Giao diện này giúp hệ thống:

- Dễ dàng trình diễn trong các buổi báo cáo
- Phù hợp với người không chuyên về học máy
- Thể hiện rõ giá trị ứng dụng thực tiễn của mô hình

4.4. Thảo luận kết quả trong bối cảnh vận hành học máy

Kết quả thực nghiệm cho thấy rằng việc lựa chọn mô hình trong vận hành học máy không nên chỉ dựa trên giá trị đo lường cao nhất, mà cần cân nhắc đồng thời nhiều yếu tố như độ ổn định, khả năng tổng quát hóa và chi phí triển khai.

EfficientNet-B0 được lựa chọn để vận hành do đạt hiệu năng cao nhất và ổn định, phù hợp với mục tiêu trình diễn và báo cáo. Tuy nhiên, MobileNetV2 cho thấy tiềm năng lớn trong các kịch bản yêu cầu độ trễ thấp hoặc tài nguyên hạn chế. Những kết quả này minh họa rõ triết lý cốt lõi của vận hành học máy: tối ưu toàn bộ hệ thống thay vì tối ưu một mô hình đơn lẻ.

CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1. Tổng kết kết quả đạt được

Trong đề tài này, một hệ thống đánh giá chất lượng hình ảnh dựa trên học sâu đã được xây dựng và triển khai theo định hướng vận hành học máy, bao phủ đầy đủ các giai đoạn từ dữ liệu, huấn luyện mô hình, theo dõi thí nghiệm, lựa chọn mô hình cho đến triển khai và kiểm thử hệ thống.

Trước hết, bài toán đánh giá chất lượng hình ảnh không tham chiếu đã được mô hình hóa một cách phù hợp dưới dạng bài toán hồi quy, trong đó đầu ra là một giá trị liên tục phản ánh cảm nhận thị giác của con người. Việc sử dụng nhãn Mean Opinion Score (MOS) giúp bài toán gắn chặt với thực tế sử dụng, đồng thời đặt ra những thách thức đặc thù do tính chủ quan và nhiễu của dữ liệu.

Về mặt dữ liệu, đề tài sử dụng bộ dữ liệu KonIQ-10k, một bộ dữ liệu lớn, đa dạng và có tính đại diện cao cho các ảnh tự nhiên trong điều kiện thực tế. Việc tận dụng cấu trúc phân chia tập huấn luyện, kiểm định và kiểm tra có sẵn trong bộ dữ liệu giúp đảm bảo tính khách quan trong đánh giá mô hình và tránh rò rỉ thông tin giữa các tập dữ liệu.

Về mặt mô hình, ba kiến trúc mạng nơ-ron tích chập tiêu biểu đã được lựa chọn và so sánh, bao gồm một mô hình hướng tới hiệu năng cao, một mô hình đóng vai trò mốc so sánh ổn định và một mô hình ưu tiên khả năng triển khai. Mỗi kiến trúc được huấn luyện với ba cấu hình siêu tham số khác nhau, cho phép phân tích ảnh hưởng của tốc độ học và số vòng huấn luyện tới hiệu năng và độ ổn định của mô hình.

Kết quả thực nghiệm cho thấy rằng kiến trúc EfficientNet-B0 đạt giá trị hệ số tương quan thứ hạng Spearman cao nhất trên cả tập kiểm định và tập kiểm tra, đồng thời thể hiện quá trình học ổn định và nhất quán. MobileNetV2 đạt hiệu năng rất sát với mô hình tốt nhất trong khi có cấu trúc gọn nhẹ hơn đáng kể, cho thấy tiềm năng lớn trong các kịch bản yêu cầu tài nguyên hạn chế. ResNet18, mặc dù ổn định và dễ huấn luyện, thể hiện hiệu năng thấp hơn trong bài toán đánh giá chất lượng hình ảnh tự nhiên phức tạp.

Quan trọng hơn, đề tài không chỉ dừng lại ở việc so sánh các mô hình dựa trên giá trị đo lường, mà còn minh họa rõ cách thức lựa chọn mô hình phù hợp cho

vận hành dựa trên nhiều tiêu chí tổng hợp, bao gồm hiệu năng, độ ổn định và khả năng triển khai.

5.2. Đóng góp của đề tài

Đề tài này mang lại các đóng góp chính sau:

Thứ nhất, đề tài xây dựng thành công một hệ thống đánh giá chất lượng hình ảnh hoàn chỉnh theo định hướng vận hành học máy, cho phép quản lý và theo dõi toàn bộ vòng đời của mô hình thay vì chỉ tập trung vào giai đoạn huấn luyện.

Thứ hai, đề tài cung cấp một nghiên cứu so sánh có kiểm soát giữa nhiều kiến trúc mạng nơ-ron tích chập và nhiều cấu hình huấn luyện khác nhau trong cùng một điều kiện dữ liệu và đánh giá, giúp làm rõ các đánh đổi giữa hiệu năng và khả năng triển khai trong thực tế.

Thứ ba, hệ thống được triển khai dưới dạng dịch vụ kèm giao diện người dùng trực quan, giúp kết nối trực tiếp giữa nghiên cứu học máy và ứng dụng thực tiễn. Điều này đặc biệt có ý nghĩa trong bối cảnh các dự án học máy ngày càng cần được trình diễn, đánh giá và sử dụng bởi các đối tượng không chuyên về kỹ thuật.

Thứ tư, đề tài tạo ra một nền tảng có cấu trúc rõ ràng, dễ tái sử dụng và dễ mở rộng, phù hợp cho việc học tập, trình diễn, cũng như làm cơ sở cho các nghiên cứu và hệ thống vận hành học máy phức tạp hơn trong tương lai.

5.3. Hạn chế của đề tài

Bên cạnh các kết quả đạt được, đề tài vẫn tồn tại một số hạn chế nhất định.

Thứ nhất, hệ thống hiện tại mới được triển khai và vận hành trong môi trường cục bộ, chưa được kiểm chứng trong các môi trường phân tán hoặc có tải truy cập lớn. Do đó, các vấn đề liên quan đến khả năng mở rộng và chịu tải chưa được đánh giá đầy đủ.

Thứ hai, việc huấn luyện mô hình mới chỉ tập trung vào một bộ dữ liệu duy nhất. Mặc dù KonIQ-10k có tính đại diện cao, việc đánh giá thêm trên các bộ dữ liệu khác có thể giúp kiểm chứng tốt hơn khả năng tổng quát hóa của mô hình.

Thứ ba, các thước đo đánh giá trong đề tài chủ yếu tập trung vào chất lượng dự đoán, trong khi các yếu tố vận hành khác như độ trễ suy luận, mức tiêu thụ tài nguyên và chi phí triển khai chưa được phân tích chi tiết một cách định lượng.

5.4. Hướng phát triển trong tương lai

Dựa trên những kết quả và hạn chế đã nêu, đề tài có thể được mở rộng theo nhiều hướng khác nhau.

Trong ngắn hạn, hệ thống có thể được mở rộng để hỗ trợ nhiều phiên bản mô hình cùng tồn tại, cho phép thử nghiệm và so sánh mô hình trực tiếp trong môi trường vận hành. Việc bổ sung cơ chế chuyển đổi mô hình linh hoạt sẽ giúp quá trình cập nhật và cải tiến mô hình trở nên an toàn và hiệu quả hơn.

Trong trung hạn, việc tích hợp các cơ chế giám sát chất lượng mô hình theo thời gian là một hướng phát triển quan trọng. Điều này bao gồm theo dõi sự thay đổi phân bố dữ liệu đầu vào, sự suy giảm hiệu năng dự đoán và các dấu hiệu cho thấy mô hình không còn phù hợp với dữ liệu mới.

Trong dài hạn, hệ thống có thể được triển khai trên các nền tảng điện toán đám mây hoặc thiết bị biên, tùy theo yêu cầu ứng dụng. Khi đó, việc lựa chọn mô hình không chỉ dựa trên hiệu năng dự đoán mà còn cần cân nhắc kỹ lưỡng các yếu tố như độ trễ, chi phí và khả năng mở rộng.

Cuối cùng, đề tài có thể được mở rộng sang các bài toán đánh giá chất lượng đa phương tiện khác như video hoặc âm thanh, nơi các thách thức về dữ liệu, mô hình và vận hành trở nên phức tạp hơn nhưng cũng mang lại giá trị ứng dụng cao hơn.

5.5. Kết luận chung

Tóm lại, đề tài đã chứng minh rằng việc xây dựng một hệ thống học máy hiệu quả không chỉ phụ thuộc vào việc lựa chọn mô hình tốt nhất, mà còn đòi hỏi một cách tiếp cận toàn diện, có hệ thống và hướng tới vận hành lâu dài. Thông qua bài toán đánh giá chất lượng hình ảnh, đề tài đã minh họa rõ ràng cách kết hợp giữa nghiên cứu học máy và thực tiễn triển khai, góp phần thu hẹp khoảng cách giữa mô hình trong phòng thí nghiệm và hệ thống trong thế giới thực.

TÀI LIỆU THAM KHẢO

- [1] D. Sculley et al., “Hidden Technical Debt in Machine Learning Systems,” *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2503–2511, 2015.
- [2] E. Breck et al., “The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction,” *Proceedings of the IEEE International Conference on Big Data*, pp. 1123–1132, 2017.
- [3] V. Hosu, H. Lin, T. Szirányi, and D. Saupe, “KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment,” *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.
- [4] Z. Wang and A. C. Bovik, “Mean squared error: Love it or leave it? A new look at signal fidelity measures,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [5] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-Reference Image Quality Assessment in the Spatial Domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [6] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *Proceedings of the International Conference on Machine Learning*, pp. 6105–6114, 2019.
- [7] J. Yosinski et al., “How transferable are features in deep neural networks?” *Advances in Neural Information Processing Systems*, pp. 3320–3328, 2014.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- [9] M. Sandler et al., “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.
- [10] T. Zaharia et al., “Accelerating the Machine Learning Lifecycle with MLflow,” *IEEE Data Engineering Bulletin*, vol. 41, no. 4, pp. 39–45, 2018.

- [11] C. Spearman, “The proof and measurement of association between two things,” *American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.
- [12] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 1137–1143, 1995.
- [13] Z. Wang and A. C. Bovik, “Modern Image Quality Assessment,” *Synthesis Lectures on Image, Video, and Multimedia Processing*, Morgan & Claypool, 2006.
- [14] M. Amershi et al., “Software Engineering for Machine Learning: A Case Study,” *Proceedings of the International Conference on Software Engineering*, pp. 291–300, 2019.