



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

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.

TRƯỜNG ĐẠI HỌC TỰ THỰC QUỐC TẾ SÀI GÒN

KHOA KỸ THUẬT & KHOA HỌC MÁY TÍNH

PHIẾU CHẤM TIỂU LUẬN/ĐỒ ÁN MH

Họ và tên Sinh viên: Đặng Công Thiên

Mã số Sinh viên: 31012302899

Lớp: 23DAI

Tên Tiểu luận/Đồ án MH: Đánh giá chất lượng hình ảnh

Đánh giá

Tiêu chí chung	Tiêu chí cụ thể	Đáp ứng CLO môn học	Thang điểm	Điểm chấm	Ghi chú
Hình thức	Bố cục		0.5		
	Lỗi chính tả, ngữ pháp, diễn đạt, hình thức trình bày		0.5		
Nội dung	Kiến thức tìm hiểu, nghiên cứu		4.0		
	Xây dựng, đánh giá được một giải pháp		4.0		
Kỹ năng nhóm, thuyết trình	Hiệu quả làm việc cá nhân/nhóm		0.5		
	Thuyết trình		0.5		
	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: CƠ SỞ LÝ THUYẾT	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. Các phương pháp đánh giá chất lượng hình ảnh.....	6
2.3.1. Phương pháp dựa trên đặc trưng thủ công	6
2.3.2. Phương pháp học máy truyền thống	6
2.3.3. Phương pháp học sâu	7
2.4. Vai trò của mạng nơ-ron tích chập trong đánh giá chất lượng hình ảnh	7
2.5. Đánh giá chất lượng hình ảnh dưới góc nhìn hồi quy và xếp hạng	8
CHƯƠNG 3. GIẢI PHÁP ĐỀ XUẤT CỦA ĐỒ ÁN.....	9
3.1. Định hướng xây dựng giải pháp.....	9
3.2. Mô hình hóa bài toán trong đồ án	9
3.3. Kiến trúc tổng quát của mô hình đề xuất.....	9
3.4. Các kiến trúc mạng nơ-ron tích chập được lựa chọn	10
3.4.1. EfficientNet-B0	10
3.4.2. ResNet18	10
3.4.3. MobileNetV2.....	11
3.5. Lý do lựa chọn ba mô hình trong đồ án.....	11
3.6. Chiến lược huấn luyện được áp dụng.....	11
3.7. Vai trò của giải pháp đề xuất trong bối cảnh vận hành học máy	12
CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ	13
4.1. Mục tiêu và nguyên tắc thực nghiệm	13
4.2. Bộ dữ liệu sử dụng trong thực nghiệm	13
4.2.1. Giới thiệu bộ dữ liệu KonIQ-10k	13

IV

4.2.3. Cấu trúc dữ liệu	14
4.2.4. Cột image_name	14
4.2.5. Các cột c1 đến c5.....	14
4.2.6. Cột c_total	14
4.2.7. Cột MOS (Mean Opinion Score)	15
4.2.8 Cột SD (Standard Deviation)	15
4.2.9. Cột set	15
4.2.10. Phân chia dữ liệu trong đề tài.....	16
4.2.11. Ý nghĩa của dataset đối với bài toán và MLOps	16
4.3. Các cấu hình thực nghiệm	16
4.3.1. Lựa chọn tốc độ học và số vòng huấn luyện.....	16
4.3.2. Hàm mất mát	17
4.4. Các phương pháp đánh giá mô hình.....	17
4.4.1. Hệ số tương quan thứ hạng Spearman.....	17
4.4.2. Các thước đo hỗ trợ	17
4.6. Kết quả huấn luyện.....	17
4.6.1. Diễn biến hệ số tương quan thứ hạng Spearman	17
4.6.2. So sánh kết quả trên tập kiểm tra	19
4.6.3. Phân tích diễn biến hàm mất mát	20
4.7. Triển khai hệ thống đánh giá chất lượng hình ảnh.....	21
4.7.1. Triển khai dưới dạng dịch vụ	21
4.7.2. Giao diện người dùng	21
4.8. Tổng kết kết quả thực nghiệm và định hướng phát triển.....	23
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 4.1	Minh họa một số dòng dữ liệu trong file <i>koniq10k_distributions_sets.csv</i>	14
2	Bảng 4.1	Phân chia dữ liệu huấn luyện, kiểm định và kiểm tra trong dataset KonIQ-10k	16
3	Bảng 4.2	Các cấu hình siêu tham số huấn luyện được sử dụng trong đề tài	16
4	Hình 4.2	Điễn biến hệ số tương quan thứ hạng Spearman trên tập kiểm định theo số vòng huấn luyện	18
5	Hình 4.3	So sánh hệ số tương quan thứ hạng Spearman trên tập kiểm tra giữa các mô hình	19
6	Hình 4.4	Điễn biến hàm matsu trên tập huấn luyện theo số vòng huấn luyện	20
7	Hình 4.5	Giao diện người dùng ban đầu của hệ thống đánh giá chất lượng hình ảnh	22
8	Hình 4.6	Kết quả đánh giá chất lượng hình ảnh hiển thị trên giao diện người dùng	22

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. Trong bối cảnh bài toán đánh giá chất lượng hình ảnh, khái niệm “chất lượng hình ảnh” không chỉ giới hạn ở độ sắc nét của ảnh. Thay vào đó, chất lượng hình ảnh được hiểu là mức độ hài lòng tổng thể của con người khi quan sát một hình ảnh, bao gồm sự kết hợp của nhiều yếu tố như độ mờ, mức nhiễu, độ tương phản, độ sáng, sai lệch màu sắc và các hiện tượng suy giảm do nén hoặc điều kiện chụp. Các yếu tố này tác động đồng thời đến cảm nhận thị giác và không thể được đánh giá đầy đủ chỉ bằng một chỉ số đơn lẻ.

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: CƠ SỞ LÝ THUYẾT

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

Ví dụ, xét hai ảnh đầu vào khác nhau: ảnh thứ nhất có độ sắc nét cao và nhiễu thấp, trong khi ảnh thứ hai bị mờ và nhiễu rõ rệt. Mô hình có thể dự đoán giá trị chất lượng lần lượt là 0,78 cho ảnh thứ nhất và 0,32 cho ảnh thứ hai. Mặc dù cả hai ảnh đều không thuộc một nhãn rời rạc, giá trị đầu ra liên tục cho phép mô hình xếp hạng đúng chất lượng tương đối giữa các ảnh, đây là yêu cầu quan trọng của bài toán đánh giá chất lượng hình 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.

2.3. Các phương pháp đánh giá chất lượng hình ảnh

Trong quá trình phát triển của lĩnh vực đánh giá chất lượng hình ảnh, nhiều phương pháp khác nhau đã được đề xuất nhằm mô phỏng cách con người cảm nhận chất lượng hình ảnh. Các phương pháp này có thể được phân loại dựa trên cách thức biểu diễn đặc trưng và mức độ sử dụng tri thức thị giác.

2.3.1. Phương pháp dựa trên đặc trưng thủ công

Các phương pháp đánh giá chất lượng hình ảnh truyền thống thường dựa trên việc thiết kế thủ công các đặc trưng mô tả sự suy giảm chất lượng của ảnh, chẳng hạn như độ sắc nét, mức nhiễu, độ tương phản hoặc các thống kê tự nhiên của ảnh. Các đặc trưng này sau đó được đưa vào các mô hình hồi quy để ước lượng điểm chất lượng.

Một hướng tiếp cận phổ biến là khai thác các thống kê của ảnh tự nhiên (Natural Scene Statistics), dựa trên giả định rằng các ảnh tự nhiên chất lượng cao tuân theo những phân bố thống kê nhất định, và sự suy giảm chất lượng sẽ làm phá vỡ các phân bố này. Các phương pháp dựa trên thống kê cảnh tự nhiên đã cho thấy hiệu quả nhất định trong các bài toán đánh giá chất lượng hình ảnh không tham chiếu, đặc biệt khi dữ liệu huấn luyện còn hạn chế [6].

Tuy nhiên, các phương pháp dựa trên đặc trưng thủ công tồn tại nhiều hạn chế, bao gồm khả năng biểu diễn hạn chế, khó mở rộng sang các dạng suy giảm phức tạp và phụ thuộc mạnh vào kinh nghiệm thiết kế đặc trưng của người nghiên cứu.

2.3.2. Phương pháp học máy truyền thống

Để khắc phục một phần hạn chế của các phương pháp thuần thủ công, nhiều nghiên cứu đã kết hợp các đặc trưng thiết kế sẵn với các mô hình học máy như hồi quy tuyến tính, máy vectơ hỗ trợ hoặc rừng ngẫu nhiên. Trong cách tiếp

cận này, vai trò của mô hình học máy là học ánh xạ từ không gian đặc trưng sang không gian chất lượng.

Mặc dù các phương pháp này cải thiện khả năng tổng quát hóa so với các phương pháp thuần quy tắc, chúng vẫn bị giới hạn bởi chất lượng của các đặc trưng đầu vào. Khi gặp các dạng suy giảm mới hoặc dữ liệu có phân bố khác biệt, hiệu năng của các mô hình này thường suy giảm đáng kể [7].

2.3.3. Phương pháp học sâu

Sự phát triển của học sâu đã mở ra một hướng tiếp cận mới cho bài toán đánh giá chất lượng hình ảnh. Thay vì dựa vào các đặc trưng được thiết kế thủ công, các mô hình học sâu có khả năng tự động học các biểu diễn đặc trưng phù hợp trực tiếp từ dữ liệu ảnh.

Các mạng nơ-ron tích chập (Convolutional Neural Network – CNN) đã chứng minh khả năng vượt trội trong việc trích xuất các đặc trưng thị giác đa cấp, từ các đặc trưng mức thấp như cạnh và kết cấu cho đến các đặc trưng mức cao hơn liên quan đến cấu trúc và ngữ nghĩa. Điều này đặc biệt phù hợp với bài toán IQA, nơi chất lượng hình ảnh bị ảnh hưởng đồng thời bởi nhiều yếu tố phức tạp [8].

Trong các nghiên cứu gần đây, các phương pháp dựa trên CNN đã đạt được kết quả vượt trội so với các phương pháp truyền thống, đặc biệt trên các bộ dữ liệu ảnh tự nhiên lớn và đa dạng.

2.4. Vai trò của mạng nơ-ron tích chập trong đánh giá chất lượng hình ảnh

Mạng nơ-ron tích chập được thiết kế nhằm mô phỏng cơ chế xử lý thông tin thị giác của hệ thống thị giác con người. Các phép tích chập cho phép mô hình học các bộ lọc không gian có khả năng phát hiện các mẫu thị giác quan trọng liên quan đến chất lượng ảnh.

Trong bài toán đánh giá chất lượng hình ảnh, CNN có các ưu điểm nổi bật sau:

- **Khả năng học đặc trưng tự động từ dữ liệu ảnh thô.**
- **Khả năng biểu diễn các dạng suy giảm chất lượng phức tạp và không tuyến tính.**
- **Khả năng tổng quát hóa tốt khi được huấn luyện trên dữ liệu lớn.**

Ngoài ra, việc sử dụng các mô hình CNN đã được huấn luyện sẵn trên các bộ dữ liệu lớn giúp tận dụng tri thức thị giác đã học, từ đó cải thiện hiệu năng và giảm chi phí huấn luyện cho các bài toán IQA [9].

2.5. Đánh giá chất lượng hình ảnh dưới góc nhìn hồi quy và xếp hạng

Không giống nhiều bài toán thị giác máy tính khác, mục tiêu của đánh giá chất lượng hình ảnh không chỉ là dự đoán giá trị gần đúng với nhãn, mà còn là đảm bảo thứ tự chất lượng tương đối giữa các ảnh được duy trì chính xác. Điều này đặc biệt quan trọng trong các ứng dụng thực tế, nơi việc so sánh chất lượng giữa các ảnh thường có ý nghĩa hơn giá trị tuyệt đối.

Do đó, các nghiên cứu IQA thường kết hợp các thước đo sai số với các thước đo tương quan thứ hạng. Trong đó, hệ số tương quan thứ hạng Spearman được sử dụng rộng rãi để đánh giá mức độ phù hợp giữa thứ hạng dự đoán và thứ hạng thực tế của các ảnh [10].

Việc xem xét đồng thời cả khía cạnh hồi quy và xếp hạng giúp đánh giá mô hình một cách toàn diện hơn, phản ánh đúng yêu cầu của bài toán đánh giá chất lượng hình ảnh.

CHƯƠNG 3. GIẢI PHÁP ĐỀ XUẤT CỦA ĐỒ ÁN

3.1. Định hướng xây dựng giải pháp

Trên cơ sở các phân tích lý thuyết về bài toán đánh giá chất lượng hình ảnh đã trình bày ở Chương 2, đồ án này đề xuất một giải pháp dựa trên học sâu nhằm ước lượng chất lượng hình ảnh không tham chiếu. Giải pháp được xây dựng theo định hướng vừa đảm bảo hiệu năng dự đoán, vừa phù hợp với yêu cầu vận hành thực tế của các hệ thống học máy hiện đại [11].

Thay vì tập trung vào việc thiết kế một mô hình mới hoàn toàn, đồ án lựa chọn cách tiếp cận thực tiễn hơn: khai thác các kiến trúc mạng nơ-ron tích chập đã được kiểm chứng, kết hợp với một cách mô hình hóa bài toán phù hợp và một chiến lược huấn luyện có kiểm soát. Cách tiếp cận này phản ánh đúng bối cảnh triển khai trong thực tế, nơi các yếu tố như độ ổn định, khả năng tái sử dụng và chi phí vận hành đóng vai trò quan trọng không kém so với độ chính xác thuần túy.

3.2. Mô hình hóa bài toán trong đồ án

Trong đồ án này, bài toán đánh giá chất lượng hình ảnh được mô hình hóa dưới dạng bài toán hồi quy, trong đó mô hình học máy học ánh xạ từ không gian ảnh đầu vào sang một giá trị chất lượng liên tục.

Cụ thể:

- Đầu vào là một ảnh màu ba kênh (RGB) có kích thước cố định.
- Đầu ra là một giá trị thực nằm trong khoảng $[0,1]$, biểu diễn mức độ chất lượng tương đối của ảnh.

Việc chuẩn hóa nhãn về khoảng $[0,1]$ giúp quá trình huấn luyện trở nên ổn định hơn, đồng thời phù hợp với việc sử dụng các hàm kích hoạt giới hạn miền giá trị ở tầng đầu ra. Khi triển khai phục vụ người dùng, giá trị dự đoán được quy đổi tuyến tính sang thang điểm $[0,100]$, giúp kết quả dễ diễn giải và phù hợp với cách đánh giá trực quan của con người.

3.3. Kiến trúc tổng quát của mô hình đề xuất

Tất cả các mô hình được sử dụng trong đồ án đều tuân theo một kiến trúc tổng quát thống nhất, bao gồm hai thành phần chính [12]:

1. Mạng trích xuất đặc trưng: thành phần này sử dụng các kiến trúc mạng nơ-ron tích chập đã được huấn luyện sẵn trên các bộ dữ liệu ảnh lớn. Nhiệm vụ của mạng là trích xuất các đặc trưng thị giác liên quan đến chất lượng hình ảnh, bao gồm cả các đặc trưng mức thấp và mức cao.
2. Tầng hồi quy: sau khi trích xuất đặc trưng, một tầng kết nối đầy đủ được sử dụng để ánh xạ các đặc trưng này sang một giá trị chất lượng duy nhất. Hàm kích hoạt sigmoid được áp dụng ở tầng đầu ra nhằm đảm bảo giá trị dự đoán nằm trong miền hợp lệ.

Thiết kế kiến trúc theo dạng mô-đun này giúp các mô hình trong đồ án dễ dàng thay thế phần mạng trích xuất đặc trưng mà không ảnh hưởng đến các thành phần còn lại của hệ thống.

3.4. Các kiến trúc mạng nơ-ron tích chập được lựa chọn

Đồ án lựa chọn ba kiến trúc mạng nơ-ron tích chập tiêu biểu để triển khai và so sánh, đại diện cho các nhóm đặc tính khác nhau trong thực tế triển khai.

3.4.1. EfficientNet-B0

EfficientNet là họ mô hình được xây dựng dựa trên nguyên lý mở rộng đồng thời độ sâu, độ rộng và độ phân giải một cách có hệ thống. Phiên bản EfficientNet-B0 là cấu hình cơ sở, có số lượng tham số vừa phải nhưng đạt hiệu năng cao trên nhiều bài toán thị giác [13].

Trong đồ án, EfficientNet-B0 được kỳ vọng là mô hình đạt hiệu năng dự đoán cao nhất, đồng thời vẫn đảm bảo hiệu quả tính toán tương đối tốt so với các kiến trúc sâu hơn.

3.4.2. ResNet18

ResNet18 là một trong những kiến trúc mạng nơ-ron tích chập kinh điển, sử dụng cơ chế kết nối tắt nhằm giảm thiểu hiện tượng suy giảm gradient trong quá trình huấn luyện. Với độ sâu vừa phải, ResNet18 thường được sử dụng như một mốc so sánh ổn định trong nhiều nghiên cứu.

Trong đồ án, ResNet18 đóng vai trò là mô hình cơ sở, giúp đánh giá mức độ cải thiện khi sử dụng các kiến trúc hiện đại hơn.

3.4.3. MobileNetV2

MobileNetV2 được thiết kế nhằm tối ưu hóa cho các thiết bị có tài nguyên hạn chế, sử dụng các phép tích chập tách rời theo chiều sâu và cấu trúc dư ngược [14]. Nhờ đó, mô hình có số lượng tham số nhỏ và tốc độ suy luận nhanh.

Trong đồ án, MobileNetV2 đại diện cho nhóm mô hình ưu tiên khả năng triển khai, đặc biệt phù hợp với các kịch bản yêu cầu độ trễ thấp.

3.5. Lý do lựa chọn ba mô hình trong đồ án

Việc lựa chọn ba kiến trúc EfficientNet-B0, ResNet18 và MobileNetV2 không nhằm tìm ra mô hình tốt nhất về mặt học thuật, mà nhằm phản ánh các quyết định thường gặp trong thực tiễn vận hành học máy.

Ba mô hình đại diện cho ba nhóm mục tiêu khác nhau:

- Mô hình ưu tiên hiệu năng dự đoán.
- Mô hình ổn định, dễ huấn luyện và dễ phân tích.
- Mô hình nhẹ, phù hợp cho triển khai thực tế.

Sự so sánh này cho phép đánh giá rõ ràng các đánh đổi giữa chất lượng dự đoán, độ ổn định và khả năng triển khai, từ đó hỗ trợ quá trình lựa chọn mô hình cho vận hành.

3.6. Chiến lược huấn luyện được áp dụng

Mỗi kiến trúc trong đồ án được huấn luyện với nhiều cấu hình tốc độ học và số vòng huấn luyện khác nhau. Cách thiết kế này nhằm quan sát hành vi học của mô hình trong các điều kiện khác nhau, thay vì chỉ tối ưu cho một cấu hình duy nhất.

Hàm mất mát được sử dụng là sai số bình phương trung bình, phù hợp với bài toán hồi quy và có tính chất dễ tối ưu. Trong suốt quá trình huấn luyện, các thước đo đánh giá được theo dõi trên tập kiểm định nhằm hỗ trợ việc lựa chọn mô hình phù hợp.

Chiến lược huấn luyện này giúp đảm bảo rằng các mô hình được so sánh trong cùng một điều kiện, từ đó tăng tính công bằng và độ tin cậy của các kết luận rút ra.

3.7. Vai trò của giải pháp đề xuất trong bối cảnh vận hành học máy

Giải pháp được đề xuất trong đồ án không chỉ tập trung vào việc xây dựng mô hình dự đoán chất lượng hình ảnh, mà còn hướng tới khả năng vận hành lâu dài của hệ thống. Việc chuẩn hóa kiến trúc mô hình, kiểm soát cấu hình huấn luyện và theo dõi kết quả một cách có hệ thống giúp giảm thiểu rủi ro và nợ kỹ thuật trong quá trình mở rộng hoặc cập nhật mô hình.

Giải pháp này tạo tiền đề cho việc triển khai, đánh giá và cải tiến mô hình một cách có kiểm soát, phù hợp với yêu cầu của các hệ thống học máy trong thực tế.

CHƯƠNG 4. THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1. Mục tiêu và nguyên tắc thực nghiệm

Mục tiêu của chương này là đánh giá một cách có hệ thống hiệu quả của giải pháp được đề xuất trong Chương 3 thông qua các thí nghiệm thực tế trên bộ dữ liệu chuẩn. Cụ thể, các thí nghiệm được thiết kế nhằm:

- Đánh giá khả năng dự đoán chất lượng hình ảnh của các mô hình đề xuất
- So sánh hiệu năng giữa các kiến trúc mạng nơ-ron tích chập khác nhau
- Phân tích ảnh hưởng của các cấu hình huấn luyện tới độ ổn định và khả năng tổng quát hóa
- Làm cơ sở cho việc lựa chọn mô hình phù hợp để triển khai trong thực tế

Toàn bộ quá trình thực nghiệm tuân thủ các nguyên tắc sau:

- Tách biệt rõ ràng tập huấn luyện, tập kiểm định và tập kiểm tra
- Các mô hình được huấn luyện và đánh giá trong cùng điều kiện dữ liệu
- Kết quả được phân tích dựa trên nhiều thước đo, không phụ thuộc vào một chỉ số duy nhất
- Việc lựa chọn mô hình dựa trên kết quả kiểm định, trong khi tập kiểm tra chỉ dùng để đánh giá cuối cùng

4.2. Bộ dữ liệu sử dụng trong thực nghiệm

4.2.1. Giới thiệu bộ dữ liệu KonIQ-10k

Các thí nghiệm trong đồ án được thực hiện trên bộ dữ liệu KonIQ-10k, một bộ dữ liệu đánh giá chất lượng hình ảnh không tham chiếu được sử dụng rộng rãi trong nghiên cứu học thuật.

KonIQ-10k bao gồm các ảnh tự nhiên thu thập trong điều kiện thực tế, phản ánh nhiều dạng suy giảm chất lượng thường gặp như:

- Mờ do rung tay hoặc lấy nét kém
- Nhiều do điều kiện ánh sáng yếu
- Sai lệch màu sắc
- Suy giảm do nén ảnh

Điểm đặc biệt của bộ dữ liệu này là các ảnh không được tạo suy giảm nhân tạo, giúp phản ánh sát hơn điều kiện sử dụng trong thực tế.

4.2.3. 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.

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

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

4.2.4. Cột image_name

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

4.2.5. 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.

4.2.6. 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.

4.2.7. 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

4.2.8 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 → 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

4.2.9. 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.

4.2.10. 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 4.1: 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.

4.2.11. 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 → phù hợp cho huấn luyện mô hình học sâu.
- Nhãn chủ quan → phản ánh bài toán thực tế.
- Có phân bố đánh giá chi tiết → 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.

4.3. Các cấu hình thực nghiệm

4.3.1. 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 4.2:

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 4.2. 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

4.3.2. 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

4.4. Các phương pháp đánh giá mô hình

4.4.1. 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.

SROCC đặc biệt phù hợp với IQA vì:

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

4.4.2. Các thước đo bô 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.

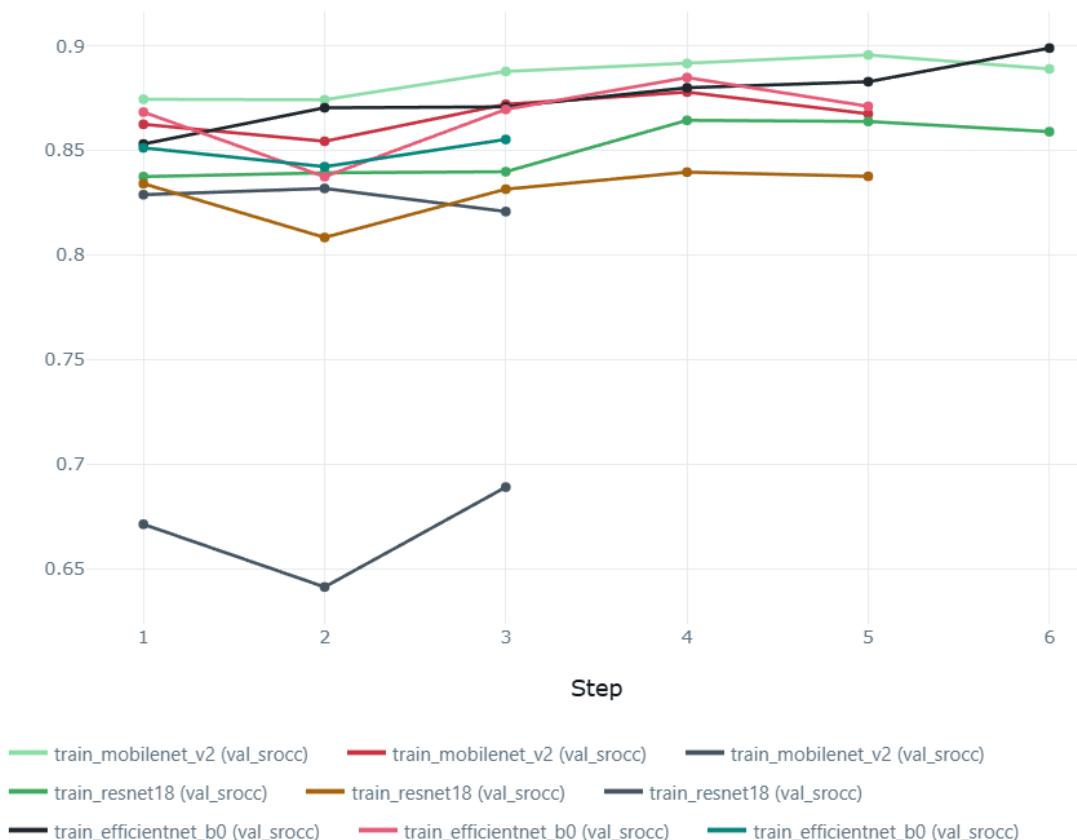
4.6. Kết quả huấn luyện

4.6.1. 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 4.2. 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.

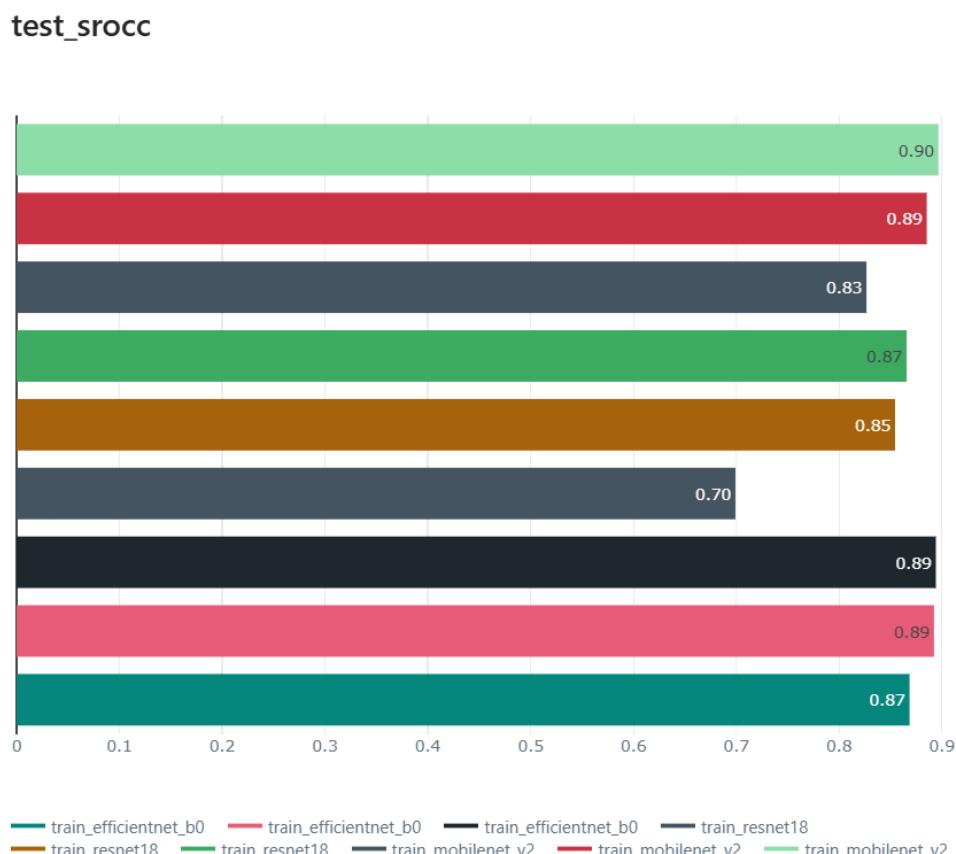
- MobileNetV2 đạt SROCC tương đối cao ngay từ các vòng huấn luyện đầu, phản ánh khả năng hội tụ nhanh của kiến trúc nhẹ.
- ResNet18 cho thấy mức SROCC thấp hơn và dao động nhiều hơn ở các cấu hình huấn luyện ngắn, cho thấy giới hạn của kiến trúc này trong bài toán đánh giá chất lượng hình ảnh không tham 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.

4.6.2. 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.



Hình 4.3. So sánh hệ số tương quan thứ hạng Spearman trên tập kiểm tra giữa các mô hình

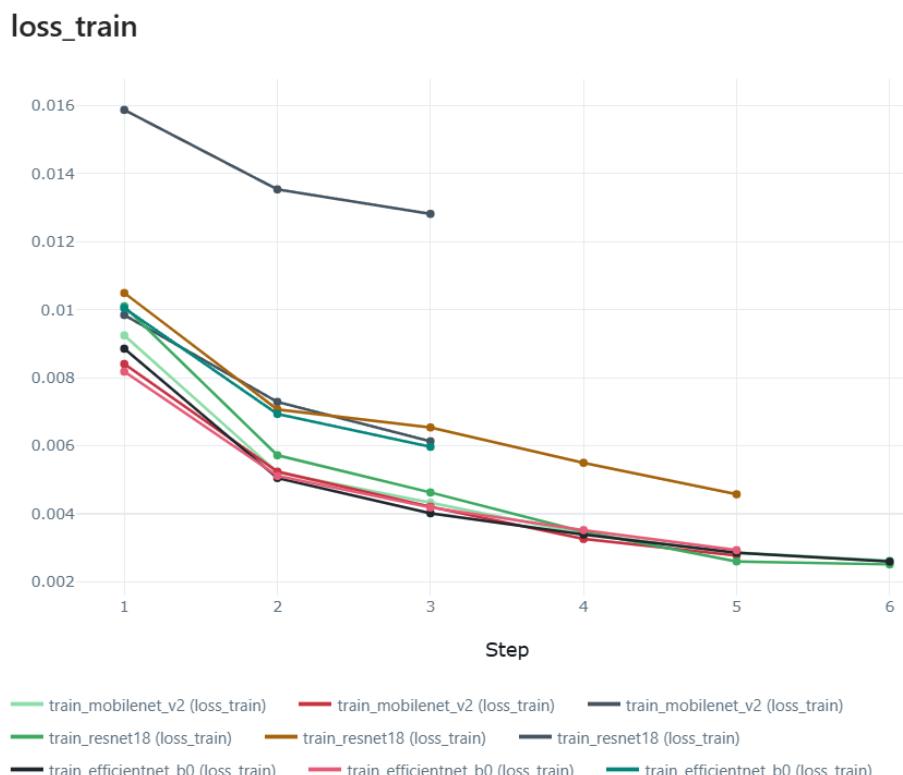
Từ Hình 4.2 có thể thấy:

- 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 nhát 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.

4.6.3. 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 4.4. 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 măt măt theo thời gian, chứng tỏ quá trình huấn luyện diễn ra đúng hướng.
- EfficientNet-B0 và MobileNetV2 cho thấy tốc độ giảm hàm măt măt 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 măt măt 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.

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

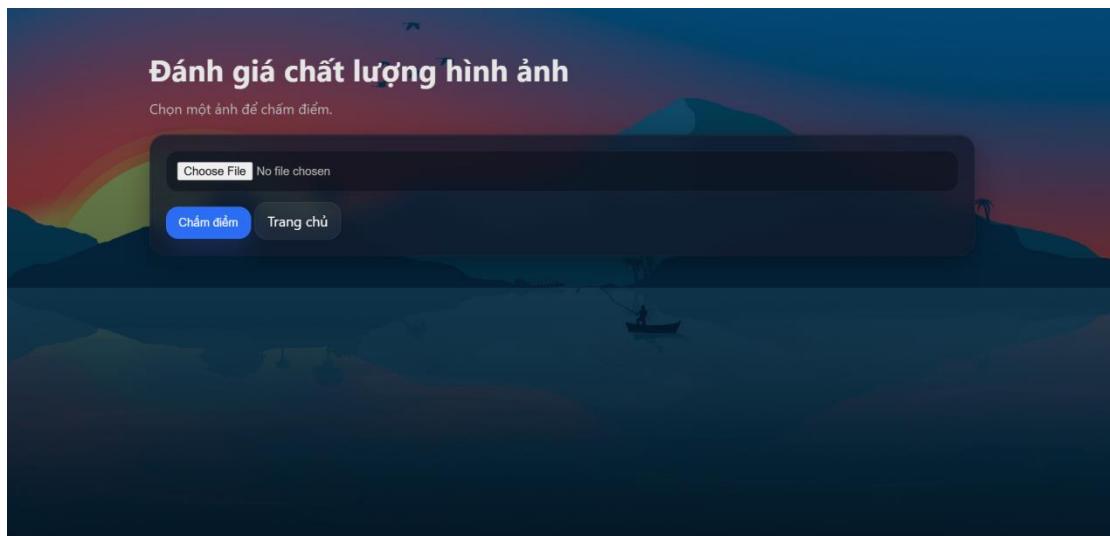
4.7.1. Triển khai dưới dạng dịch vụ

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.

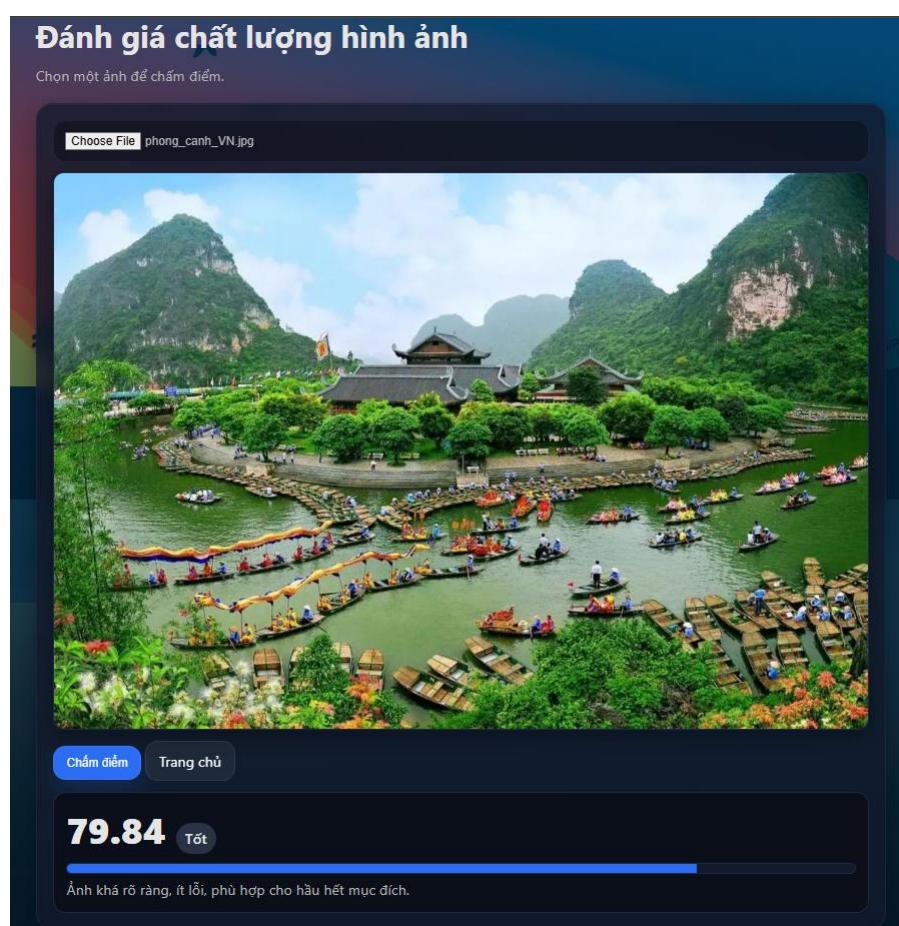
4.7.2. Giao diện người dùng

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 4.5. 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 4.6. 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

4.8. Tổng kết kết quả thực nghiệm và định hướng phát triển

Thông qua các thí nghiệm được trình bày trong chương này, có thể khẳng định rằng giải pháp đánh giá chất lượng hình ảnh không tham chiếu dựa trên học sâu được đề xuất đã đạt được các mục tiêu đặt ra ban đầu, cả về mặt học thuật lẫn khả năng triển khai trong thực tế.

Trước hết, bài toán đánh giá chất lượng hình ảnh đã được mô hình hóa hiệu quả 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 thông qua nhãn Mean Opinion Score. Việc sử dụng bộ dữ liệu KonIQ-10k với quy mô lớn và phân bố đa dạng giúp mô hình học được các đặc trưng thị giác phù hợp với điều kiện ảnh tự nhiên trong thực tế, đồng thời đảm bảo tính khách quan trong quá trình đánh giá nhờ cấu trúc phân chia tập huấn luyện, kiểm định và kiểm tra rõ ràng.

Về mặt mô hình, ba kiến trúc mạng nơ-ron tích chập tiêu biểu đã được triển khai và so sánh trong cùng điều kiện thực nghiệm. Kết quả cho thấy

EfficientNet-B0 đạt hệ số tương quan thứ hạng Spearman cao nhất và thể hiện quá trình huấn luyện ổn định, phù hợp với các kịch bản ưu tiên hiệu năng dự đoán. 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 hệ thống yêu cầu tối ưu tài nguyên và độ trễ thấp. ResNet18, mặc dù ổn định và dễ huấn luyện, thể hiện hiệu năng thấp hơn trong bối cảnh dữ liệu ảnh tự nhiên phức tạp.

Việc theo dõi diễn biến huấn luyện thông qua nhiều thước đo khác nhau, bao gồm hệ số tương quan thứ hạng và hàm mất mát, cho phép đánh giá mô hình một cách toàn diện, không chỉ dựa trên kết quả cuối cùng mà còn xét đến độ ổn định và khả năng hội tụ. Cách tiếp cận này phản ánh đúng tinh thần vận hành học máy, nơi quyết định lựa chọn mô hình cần dựa trên tập hợp nhiều tiêu chí thay vì một chỉ số đơn lẻ.

Bên cạnh các kết quả đạt được, hệ thống vẫn tồn tại một số hạn chế, bao gồm việc mới chỉ được triển khai trong môi trường cục bộ, chưa được đánh giá trong điều kiện tải lớn hoặc môi trường phân tán, cũng như chưa phân tích định lượng sâu các yếu tố vận hành như độ trễ suy luận và mức tiêu thụ tài nguyên. Trong tương lai, hệ thống có thể được mở rộng theo nhiều hướng, bao gồm hỗ trợ nhiều phiên bản mô hình cùng tồn tại, tích hợp cơ chế giám sát chất lượng mô hình theo thời gian, triển khai trên các nền tảng điện toán đám mây hoặc thiết bị biên, và mở rộng sang các bài toán đánh giá chất lượng đa phương tiện khác như video hoặc âm thanh.

Tổng thể, các kết quả thực nghiệm trong chương này đã chứng minh tính hợp lý và hiệu quả của giải pháp đề xuất, đồng thời minh họa rõ cách tiếp cận toàn diện trong việc xây dựng và vận hành một hệ thống học máy phục vụ bài toán đánh giá chất lượng hình ảnh trong thực tế.

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] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a completely blind image quality analyzer,” IEEE Signal Processing Letters, vol. 20, no. 3, pp. 209–212, 2013.
- [7] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” IEEE Transactions on Image Processing, vol. 24, no. 8, pp. 2579–2591, 2015.
- [8] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” IEEE Transactions on Image Processing, vol. 23, no. 2, pp. 684–695, 2014.
- [9] J. Kim and S. Lee, “Fully deep blind image quality predictor,” IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 1, pp. 206–220, 2017.
- [10] C. Spearman, “The proof and measurement of association between two things,” American Journal of Psychology, vol. 15, no. 1, pp. 72–101, 1904.

- [11] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [12] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *International Conference on Learning Representations*, 2015.
- [13] 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.
- [14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.