

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

Phạm Thanh Tùng

**NÂNG CAO HIỆU NĂNG MÃ HOÁ VIDEO DÙNG
CHO TRUYỀN THÔNG ĐA PHƯƠNG TIỆN**

Chuyên ngành: Khoa học Máy tính

Mã số: 9480101.01

TÓM TẮT LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

Hà Nội – 2022

Công trình được hoàn thành tại: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

Người hướng dẫn khoa học: **1. PGS.TS Lê Thanh Hà**
2. TS Đinh Triều Dương

Phản biện:

.....

Phản biện:

.....

Phản biện:

.....

L luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia
chấm luận án tiến sĩ họp tại

vào hồi giờ ngày tháng năm

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội

MỤC LỤC

CHƯƠNG 1. MỞ ĐẦU	4
1.1. Đặt vấn đề.....	4
1.2. Đóng góp của luận án	6
1.3. Bố cục luận án	7
CHƯƠNG 2. TỔNG QUAN VỀ NÉN VIDEO.....	7
2.1. Một số khái niệm cơ bản về video	7
2.2. Nén video và hiệu năng nén video.....	8
2.3. Đánh giá chất lượng hình ảnh.....	8
CHƯƠNG 3. XÂY DỰNG CÁC BỘ DỮ LIỆU ĐÁNH GIÁ CHẤT LƯỢNG HÌNH ẢNH CHỦ QUAN.....	9
3.1. Sự cần thiết	10
3.2. Xây dựng bộ dữ liệu đánh giá chất lượng hình ảnh chủ quan theo khối tiêu biểu.....	10
3.3. Xây dựng bộ dữ liệu đánh giá chất lượng hình ảnh chủ quan theo khối ngẫu nhiên	11
CHƯƠNG 4. ĐỀ XUẤT PHƯƠNG PHÁP ĐÁNH GIÁ CHẤT LƯỢNG HÌNH ẢNH THEO THỊ GIÁC NGƯỜI.....	12
4.1. Phương pháp đánh giá chất lượng hình ảnh có tham chiếu sử dụng mạng Neural tích chập.....	13
4.2. Phương pháp đánh giá sự suy giảm chất lượng hình ảnh theo đặc trưng nội dung khối ảnh gốc.....	15
CHƯƠNG 5. ỨNG DỤNG CÁC PHƯƠNG PHÁP ĐỀ XUẤT NÂNG CAO HIỆU NĂNG MÃ HÓA VIDEO.....	17
5.1. Ứng dụng phương pháp đánh giá sự suy giảm chất lượng hình ảnh bằng cách trích chọn đặc trưng cơ bản	18
5.2. Ứng dụng phương pháp đánh giá sự suy giảm chất lượng hình ảnh trích chọn đặc trưng bằng mạng Neural tích chập	19
CHƯƠNG 6. KẾT LUẬN VÀ KIẾN NGHỊ	23
DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ	24

CHƯƠNG 1. MỞ ĐẦU

1.1. Đặt vấn đề

1.1.1. Tính cấp thiết của nghiên cứu

Ngày nay, thông tin hình ảnh chiếm phần lớn bằng thông trên tất cả ứng dụng đa phương tiện như truyền hình, internet và di động. Người dùng có nhu cầu ngày càng tăng cả về số lượng video cũng như chất lượng nội dung video, đồng thời với khả năng truy cập rộng rãi hơn và độ tin cậy tốt hơn. Điều này đang tạo ra áp lực lớn về sự cân bằng giữa dung lượng thiết bị lưu trữ, truyền tải khả dụng của người dùng và tỷ lệ bit cần thiết để truyền nội dung video với chất lượng mong muốn. Do đó, các nhà quản lý mạng, người tạo nội dung và nhà cung cấp dịch vụ đều đang tìm cách tốt hơn để truyền video chất lượng cao nhất ở tỷ lệ bit thấp nhất, điều mà chỉ có thể đạt được thông qua nén video. Nén video (mã hóa video) là một quá trình mã hóa biến đổi một cảnh video thành dữ liệu mới có tổng số bit nhỏ hơn. Các nghiên cứu nâng cao hiệu năng nén video hiện nay tập trung vào bốn hướng chính gồm: (i) cải thiện chất lượng hình ảnh; (ii) giảm độ phức tạp tính toán; (iii) giảm tỷ lệ bit và (iv) phát triển các chuẩn nén thế hệ tiếp theo.

Trong nén video, đánh giá chất lượng hình ảnh có ý nghĩa quan trọng trong việc tính toán hiệu năng và làm căn cứ để lựa chọn cách mã hoá tối ưu. MSE đã được sử dụng rộng rãi trong nén hình ảnh và video do công thức tính toán đơn giản, có thể dễ dàng giải quyết trong toán học. Tuy nhiên, MSE được xác định là tương quan kém với chất lượng cảm nhận thị giác.

Để khắc phục những yếu điểm của MSE đã nói ở trên, một số phương pháp đánh giá chất lượng thay thế đã được phát triển cho các

ứng dụng nén hình ảnh và video. Các phương pháp này cho kết quả gần với thị giác người nhưng tính toán phức tạp dẫn đến việc tích hợp của chúng trong các chương trình nén video thực tế là không khả thi. Gần đây, các nghiên cứu sử dụng học máy để đánh giá chất lượng hình ảnh cho kết quả tương đối khả quan. Từ những nhận định trên, nghiên cứu sinh lựa chọn luận án nghiên cứu “*Nâng cao hiệu năng mã hoá Video dùng cho truyền thông đa phương tiện*” với mong muốn giảm tỷ lệ bit và cải thiện chất lượng theo nhận thức thị giác áp dụng cho từng vùng của khung hình trong nén video.

1.1.2. Mục tiêu

Phát triển các bộ dữ liệu thực nghiệm đánh giá chất lượng hình ảnh chủ quan từ đó đề xuất các phương pháp đánh giá lượng hình ảnh theo nhận thức thị giác. Áp dụng các phương pháp đánh giá chất lượng hình ảnh theo nhận thức thị giác trong các bộ nén nhằm nâng cao hiệu năng nén video.

1.1.3. Phương pháp nghiên cứu

- Phương pháp phân tích và tổng hợp lý thuyết;
- Phương pháp thực nghiệm;
- Phương pháp mô hình hóa.

1.1.4. Phạm vi nghiên cứu

Hình ảnh thực nghiệm đánh giá chất lượng chủ quan trích xuất từ các video mẫu chuẩn dưới biến dạng nén thông thường.

Nghiên cứu tập trung vào mô hình hoá phương pháp đánh giá khách quan có tham chiếu dựa trên đặc tính thị giác.

Về ứng dụng nâng cao hiệu năng mã hoá: hai chuẩn mã hoá phổ biến, gần thời điểm nghiên cứu nhất là H264/AVC và H265/HEVC.

1.2. Đóng góp của luận án

Luận án có những đóng góp khoa học tập trung vào ba nhóm vấn đề (1) bộ dữ liệu đánh giá chất lượng ảnh chủ quan, (2) phương pháp đánh giá chất lượng ảnh và (3) bộ nén video cải tiến. Cụ thể các đóng góp như sau:

1.2.1. Đề xuất các bộ dữ liệu đánh giá chất lượng hình ảnh chủ quan

Thứ nhất, luận án đề xuất bộ dữ liệu thực nghiệm đánh giá chất lượng hình ảnh chủ quan theo khối tiêu biểu. *Thứ hai*, trên cơ sở bộ dữ liệu ban đầu, luận án đề xuất bộ dữ liệu thực nghiệm đánh giá chất lượng hình ảnh chủ quan theo khối ngẫu nhiên.

1.2.2. Đề xuất các phương pháp đánh giá chất lượng ảnh khách quan

Thứ ba, phương pháp dự đoán sự thay đổi chất lượng hình ảnh theo nội dung cục bộ bằng cách trích chọn đặc trưng cơ bản. *Thứ tư*, phương pháp đánh giá chất lượng ảnh khách quan có tham chiếu bằng học mạng CNN. *Thứ năm*, phương pháp dự đoán sự thay đổi chất lượng hình ảnh theo nội dung cục bộ bằng trích chọn đặc tính bằng mạng CNN.

1.2.3. Đề xuất các bộ nén video cải tiến

Thứ sáu, bộ nén video H. 264 cải tiến giảm tỷ lệ bit nhưng đảm bảo chất lượng thị giác. *Thứ bảy*, bộ nén H.265/HEVC được cải tiến tương tự như đề xuất *thứ sáu* nhưng sử dụng phương pháp dự đoán chất lượng hình ảnh khi thay đổi hệ số lượng tử trích chọn đặc tính bằng mạng CNN.

1.3. Bố cục luận án

Ngoài chương mở đầu và chương kết luận, luận án gồm 04 chương nội dung.

CHƯƠNG 2. TỔNG QUAN VỀ NÉN VIDEO

Tóm tắt: Trong chương 2, luận án trình bày tổng quan về nén video và các vấn đề có liên quan đến chủ đề nén video. Đồng thời cũng tập trung tìm hiểu về các phương pháp đánh giá chất lượng hình ảnh chủ quan và khách quan. Có nhiều nghiên cứu phát triển phương pháp đánh giá chất lượng hình ảnh khách quan thay thế đánh giá chất lượng chủ quan nhằm ứng dụng trong các bộ nén video. Để nghiên cứu đơn vị đo chất lượng hình ảnh khách quan, bộ dữ liệu thực nghiệm đánh giá chất lượng chủ quan đóng vai trò quan trọng trong việc mô hình hoá, kiểm tra, đánh giá hiệu quả. Những bộ dữ liệu hiện có còn nhiều hạn chế về số lượng và chưa phù hợp với nghiên cứu của luận án.

2.1. Một số khái niệm cơ bản về video

Phần này giới thiệu một số khái niệm cơ bản về video, không gian màu và khung hình video. Thuật ngữ video dùng để chỉ nguồn thông tin hình ảnh trực quan, bao gồm một chuỗi liên tiếp các một khung hình theo thời gian. Không gian màu là một mô hình toán học dùng để mô tả các màu sắc của khung hình video phổ biến là 2 mô hình RGB và YUV. Trong đó mô hình YUV được sử dụng rộng rãi do thành phần màu (UV) ít quan trọng hơn thành phần chói (Y) nên có thể lấy mẫu thấp hơn để làm giảm dung lượng cần để lưu. Cùng với sự phát triển kỹ thuật điện tử, độ phân giải khung hình phổ biến trong truyền thông ngày càng tăng: FHD, 2K, 4K và 8K.

2.2. Nén video và hiệu năng nén video

Phần này giới thiệu khái niệm về nén video. Độ phân giải khung hình, nhu cầu chất lượng hình ảnh ngày càng cao gây áp lực lên quá trình lưu trữ, truyền dẫn nên việc nghiên cứu nâng cao hiệu năng nén video hết sức cần thiết. Lịch sử các chuẩn nén video cho thấy hiệu năng nén ngày càng tăng đặc biệt là hai chuẩn H264/AVC và H265/HEVC gần thời điểm nghiên cứu luận án.

Hiệu năng nén của một bộ nén video là khả năng cân bằng giữa chất lượng video sau khi nén, tỉ lệ bit (bitrate) và chi phí tính toán. Để đánh giá hiệu năng của bộ nén người ta so sánh chất lượng hình ảnh video tái tạo với cùng tỷ lệ bit. Do đó, đánh giá chất lượng hình ảnh là một nội dung cơ bản quan trọng trong quá trình xử lý ảnh và video.

2.3. Đánh giá chất lượng hình ảnh

Phần này trình bày về các phương pháp đánh giá chất lượng hình ảnh và các bộ dữ liệu thực nghiệm. Đánh giá chất lượng ảnh có ý nghĩa quan trọng trong các bộ nén: (i) so sánh hiệu năng giữa các bộ nén khác nhau trên các tốc độ bit và nội dung video; (ii) so sánh ảnh hưởng của các tham số và tùy chọn mã hoá nhằm lựa chọn giá trị tối ưu.

Nhận thức thị giác của con người rất phức tạp, khó có thể hiểu đầy đủ và mô hình hoá đầy đủ. Đánh giá chất lượng video còn phức tạp hơn nhiều so với đánh giá chất lượng ảnh do thông tin không chỉ trong không gian hai chiều của khung hình mà còn theo thời gian. Có hai phương pháp chủ yếu là đánh giá chủ quan và khách quan. Đánh giá chất lượng chủ quan bằng con người vẫn là phương pháp đánh giá tốt nhất tuy nhiên không thể áp dụng vào bộ nén video nên được thay thế bằng các phương pháp đánh giá chất lượng khách quan. Bản chất

của các phương pháp đánh giá khách quan là mô phỏng lại phương pháp chủ quan.

Có nhiều phương pháp đánh giá chất lượng khách quan đã được nghiên cứu, phát triển trong đó tập trung thành bốn nhóm chính: (i) các phương pháp dựa trên sai số bình phương tối thiểu tiêu biểu là MSE, PSNR; (ii) các phương pháp đánh giá trên nền tảng thị giác người; (iii) các phương pháp đánh giá chất lượng theo mô hình thống kê tiêu biểu là SSIM và (iv) các phương pháp đánh giá chất lượng dùng học máy. Đồng thời, chúng cũng được phân thành ba loại: (i) các phương pháp có tham chiếu; (ii) các phương pháp không tham chiếu; (iii) Các phương pháp tham chiếu không đầy đủ. Tuy nhiên, kết quả nghiên cứu vẫn còn những nhược điểm và chưa áp dụng được trong các chuẩn mã hoá mới nhất.

Để nghiên cứu đơn vị đo chất lượng hình ảnh khách quan, bộ dữ liệu thực nghiệm đánh giá chất lượng chủ quan đóng vai trò quan trọng trong việc mô hình hoá, kiểm tra, đánh giá hiệu quả. Những bộ dữ liệu hiện có còn nhiều hạn chế về số lượng và chưa phù hợp với nghiên cứu của luận án.

CHƯƠNG 3. XÂY DỰNG CÁC BỘ DỮ LIỆU ĐÁNH GIÁ CHẤT LƯỢNG HÌNH ẢNH CHỦ QUAN

Tóm tắt: Chương 3 làm rõ cơ sở khoa học, phương pháp xây dựng và một số kết quả phân tích, đánh giá ban đầu đối với 2 bộ dữ liệu đánh giá chất lượng hình ảnh chủ quan đề xuất của luận án. Quá trình đánh giá chất lượng hình ảnh tuân thủ quy trình theo tiêu chuẩn ITU-R BT.500-11 và bộ dữ liệu đã loại bỏ các điểm nhiễu.

3.1. Sự cần thiết

Một nhóm các phương pháp đánh giá chất lượng hình ảnh khách xây dựng trên nền tảng mạng CNN được phát triển gần đây cho kết quả khả quan. Các phương pháp học này trích rút đặc trưng thị giác từ dữ liệu thực nghiệm đánh giá chủ quan đã đề cập ở Chương 2 để mô hình hoá biến dạng theo thị giác người. Tuy nhiên, tất cả các dữ liệu nêu trên đều đánh giá chất lượng trên toàn ảnh trong khi chất lượng các vùng trong ảnh rất khác nhau theo nội dung của nó. Do đó, để phát triển phương pháp đánh giá chất lượng ảnh cục bộ, cần thiết phải xây dựng bộ dữ liệu thực nghiệm riêng.

3.2. Xây dựng bộ dữ liệu đánh giá chất lượng hình ảnh chủ quan theo khối tiêu biểu

Thực nghiệm đảm bảo theo tiêu chuẩn ITU-R BT.500-11 về đánh giá chất lượng hình ảnh: số lượng người quan sát mỗi mẫu từ 15 trở lên; thời gian tối thiểu mỗi lần đánh giá là 10 giây; đảm bảo khoảng cách quan sát và chất lượng màn hình.

3.2.1. Tạo bộ dữ liệu ảnh đánh giá

Bộ dữ liệu thực nghiệm đánh giá chất lượng hình ảnh chủ quan được tạo ra từ 20 video gốc nén bởi bộ nén H.264/AVC với 5 mức lượng tử khác nhau. Bộ dữ liệu dùng cho thực nghiệm có tất cả 20 video \times 5 ảnh \times 3 vị trí \times 2 kích thước = 600 cặp khối ảnh. Vị trí chọn khối theo đặc trưng: phẳng, cạnh, nhiều chi tiết.

3.2.2. Thực nghiệm đánh giá chất lượng hình ảnh

Quá trình thực nghiệm tuân theo theo khuyến cáo tiêu chuẩn ITU-R BT.500-11, người quan sát đánh giá chất lượng khối biến dạng có đối chiếu với khối gốc theo thang đo DMOS gồm 5 mức: xuất sắc”, “tốt”, “khá”, “kém” và “xấu”.

3.2.3. Phân tích kết quả đánh giá

Dữ liệu thu được gồm 12.000 điểm đánh giá cho 600 khối ảnh từ 20 người quan sát được gọi là bộ dữ liệu đánh giá chất lượng hình ảnh chủ quan VP9. Chất lượng hình ảnh giảm khi mức lượng tử (QP) tăng. Khối nhiều chi tiết có điểm chất lượng tốt nhất trong các loại khối và khi khối ảnh quan sát ở khoảng cách xa hơn thì điểm DMOS tốt hơn đồng nghĩa với việc mắt khó nhận biết biến dạng hơn.

3.3. Xây dựng bộ dữ liệu đánh giá chất lượng hình ảnh chủ quan theo khối ngẫu nhiên

3.3.1. Xây dựng bộ dữ liệu bằng thực nghiệm

Dữ liệu đánh giá được trích xuất từ các 40 video gốc và biến dạng nén bởi bộ nén HEVC với mức lượng tử ngẫu nhiên. Các vị trí lựa chọn để trích xuất khối ngẫu nhiên không lặp. Bộ dữ liệu có tất cả 246400 khối gồm 61600 cặp kích thước 64×64 và 61600 cặp kích thước 128×128 .

3.3.2. Xử lý dữ liệu thực nghiệm

Trong tổng 61600 cặp khối ảnh, qua quá trình đánh giá có 40,708 cặp khối ảnh mà mỗi khối từ 15 người đánh giá trở lên với 697,179 điểm đánh giá. Luận án lựa chọn phương pháp Z-score với ngưỡng lọc 2σ với tổng số 422 cặp khối bị loại bỏ trong bộ dữ liệu. Số lượng cặp khối ảnh cuối cùng $N = 40286$ đặt tên gọi là bộ dữ liệu HMII (Human Machine Interaction Image).

Điều này có nghĩa là chất lượng chủ quan không đồng đều trong cùng một ảnh mã hoá. Đồng thời các phép đo chất lượng dựa trên độ trung thực của tín hiệu như RMSE, PSNR không tương đồng như các phép đo cảm nhận chủ quan của con người.

3.3.3. Đánh giá kết quả thực nghiệm với các phương pháp đo chất lượng hình ảnh khách quan

Luận án sử dụng 7 thang đo chất lượng phổ biến để đánh giá mức độ phù hợp với chất lượng chủ quan của bộ dữ liệu đề xuất. Kết quả cho thấy các phương pháp theo yếu tố thị giác SSIM, FSIM cho mức tương quan cao hơn so với các phương pháp sai số bình phương tối thiểu PSNR, MSE. Đồng thời, thang đo trên nền tảng học sâu DIQaM-FR và WaDIQaM-FR cho kết quả tương quan thấp do dữ liệu mô hình hoá thang đo kém phù hợp.

3.3.4. Mô hình hóa đánh giá chất lượng đơn giản

Trong nội dung này, các phương pháp đánh giá tốt nhất ở mục 3.3.3 điều chỉnh trên bộ dữ liệu đề xuất. Đồng thời, mô hình học máy DIQaM-FR và WaDIQaM-FR được dùng để huấn luyện lại trên bộ dữ liệu HMII. Kết quả cho mô hình đánh giá học sâu điều chỉnh với bộ trích chọn đặc trưng VGG đạt được hiệu suất vượt trội so với các hệ số khác. Từ kết quả của thực nghiệm, kích thước dùng cho huấn luyện tốt nhất là 64x64. Đồng thời, học sâu sử dụng mạng CNN là hướng đi khả quan nhằm xây dựng mô hình đánh giá chất lượng hình ảnh.

CHƯƠNG 4. ĐỀ XUẤT PHƯƠNG PHÁP ĐÁNH GIÁ CHẤT LƯỢNG HÌNH ẢNH THEO THỊ GIÁC NGƯỜI

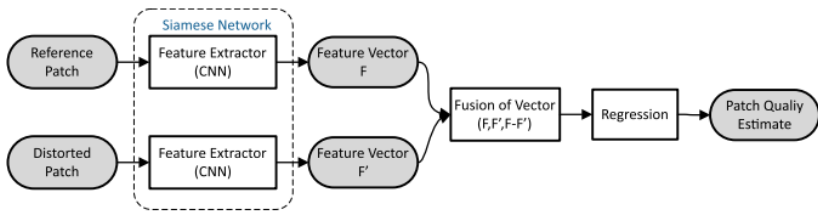
Tóm tắt: Chương 4 là nội dung trọng tâm luận án nhằm đề xuất các phương pháp đánh giá chất lượng hình ảnh theo đặc trưng thị giác. Các phương pháp này được phát triển, kiểm tra và đánh giá dựa trên các bộ dữ liệu thực nghiệm đánh giá chất lượng hình ảnh chủ quan HMII và VP9. Trong đó có 2 phương pháp dự đoán mức độ suy giảm chất lượng hình ảnh biến dạng từ ảnh gốc và hệ số lượng tử. Phương

pháp còn lại dự đoán chất lượng dựa trên ảnh gốc và ảnh biến dạng (phương pháp đánh giá có tham chiếu).

4.1. Phương pháp đánh giá chất lượng hình ảnh có tham chiếu sử dụng mạng Neural tích chập

4.1.1. Kiến trúc của phương pháp đề xuất

Trong nội dung này, luận án trình bày về kiến trúc mạng CNN để dự đoán chất lượng hình ảnh. Kiến trúc của phương pháp IQA đề xuất như Hình 4.1 dưới đây. Dữ liệu đầu vào là 2 khối ảnh gồm khối cần đánh giá chất lượng và khối tham chiếu (khối gốc). Hai khối được xử lý trích chọn đặc trưng thị giác theo 2 nhánh song song nhau. Kết quả vector đặc trưng kết hợp giữa giữa 2 nhánh được dùng cho lớp hồi quy dự báo chất lượng.



Hình 4.1. Kiến trúc phương pháp đánh giá chất lượng hình ảnh

4.1.2. Tiêu chí đánh giá hiệu quả

Phương pháp phổ biến để kiểm tra hiệu quả của các thuật toán ước lượng IQA bằng cách sử dụng sai số tuyệt đối trung bình (MAE).

4.1.3. Tối ưu hoá huấn luyện

Phương pháp tối ưu “Adaptive moment estimation optimizer” (ADAM) được sử dụng với các tham số được chọn theo khuyến nghị: $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$ và tốc độ học α ấn định là 5×10^{-4} .

4.1.4. Kỹ thuật trích chọn đặc trưng

Với kiến trúc chung trong *Hình 4.1*, một trong năm mạng CNN gồm: VGGnet, ResNeXt-50, Xception, Inception-v4 và Inception-ResNets lần lượt được sử dụng trích xuất đặc trưng trong kiến trúc mạng CNN dự đoán.

4.1.5. Kết quả huấn luyện

Mô hình sử dụng Resnext-50 có hiệu năng tốt nhất được lựa chọn làm nền tảng kiến trúc trích chọn đặc trưng cho mô hình đánh giá chất lượng ảnh đặt tên là HMI-IQA.

4.1.6. Kiểm nghiệm phương pháp đánh giá chất lượng hình ảnh có tham chiếu sử dụng mạng Neural tích chập

Trong nội dung tiếp theo, bốn bộ dữ liệu đánh giá chất lượng ảnh chủ quan gồm: CSIQ, LIVE, TID2008 và TID2013 được dùng để kiểm nghiệm khả năng dự đoán chất lượng của mô hình HMI-IQA. Chất lượng các khối ảnh được dự đoán bằng mô hình HMI-IQA và tổng hợp thành chất lượng toàn ảnh theo 2 phương pháp: tính trung bình (HMI-IQA-Aver) và tính có trọng số quan tâm (HMI-IQA-Sal).

Kết quả cho thấy chất lượng dự đoán theo phương pháp đề xuất tương đối phù hợp với dữ liệu thực nghiệm. So sánh với các nghiên cứu khác có liên quan, phương pháp đề xuất cho kết quả vượt trội đối với nhóm biến dạng tạo bởi nén (JPEG, JP2K) và tốt hơn hầu hết với các biến dạng còn lại.

4.2. Phương pháp đánh giá sự suy giảm chất lượng hình ảnh theo đặc trưng nội dung khối ảnh gốc

4.2.1. Phương pháp đánh giá sự suy giảm chất lượng hình ảnh bằng cách trích chọn đặc trưng cơ bản

Từ bộ dữ liệu thực nghiệm đánh giá chất lượng hình ảnh chủ quan VP9 thu được, luận án tiếp tục xây dựng phương pháp đánh giá sự suy giảm chất lượng hình ảnh bằng cách trích chọn đặc trưng cơ bản. Phương pháp được dùng để ước lượng chất lượng thị giác Y mỗi của khối nén với mức lượng tử qp theo hàm logistic (skew-symmetric sigmoid) như sau:

$$Y(qp) = \frac{5}{1 + e^{(qp-a)b}}. \quad (1)$$

Hai tham số a , b phụ thuộc nội dung của khối ảnh được nội suy theo các đặc trưng thị giác của khối từ bộ dữ liệu thực nghiệm VP9 như sau:

$$a = 36.2247 - 0.0581f_2 + 0.0153f_3 + 0.0248f_4 + 0.1874f_7 - 0.0127f_8 - 0.0052f_{11}, \quad (2)$$

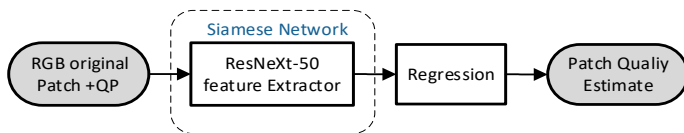
$$b = -0.2003 - 0.0009f_1 + 0.0003f_3 + 0.0002f_5 + 0.0006f_6. \quad (3)$$

Chất lượng trung bình a của khối phụ thuộc chủ yếu vào mật độ cạnh (f_3), phương sai (f_2), độ sáng cao nhất của các điểm ảnh (f_4). Ngoài ra, mật độ cạnh (f_8) và phương sai của vùng lân cận (f_7) cũng có ảnh hưởng đến chất lượng trung bình khối. Độ suy giảm chất lượng b phụ thuộc chủ yếu vào độ sáng và mật độ cạnh. Khối càng sáng (f_1) và nhiều cạnh (f_3) tốc độ suy giảm chất lượng càng nhanh. Phương pháp đề xuất đã phản ánh được tương đối tốt sự ảnh hưởng của các đặc trưng

đến biến dạng dẫn đến thay đổi chất lượng ảnh. Tuy nhiên do số lượng điểm thực nghiệm còn ít, phương pháp mô hình đơn giản nên chưa biểu diễn được nhiều dạng khối. Thêm nữa, việc lựa chọn đặc trưng để xây dựng bộ dữ liệu còn chủ quan, có thể bỏ sót những đặc trưng có ảnh hưởng đến độ nhạy biến dạng hình ảnh.

4.2.2. Phương pháp đánh giá sự suy giảm chất lượng hình ảnh trích chọn đặc trưng bằng mạng Neural tích chập

Phương pháp đánh giá sự suy giảm chất lượng hình ảnh theo Mục 4.2.1 tương đối đơn giản nhưng cũng đã phần nào cho thấy ảnh hưởng của nội dung khối ảnh đến sự suy giảm chất lượng khi mã hóa video. Từ kết quả trên, luận án đã sử dụng phương pháp học sâu sử dụng mạng CNN trích chọn đặc trưng bộ dữ liệu thực nghiệm quy mô lớn (HMII) nhằm ước lượng chất lượng khối ảnh sau khi mã hóa. Sau khi thực hiện huấn luyện mô hình đánh giá chất lượng hình ảnh sử dụng học sâu như nội dung trình bày trong Mục 4.1, kiến trúc mạng tích chập trích chọn đặc trưng ResNeXt-50 cho thấy hiệu quả tốt hơn so với các kiến trúc khác. Do đó, trong phương pháp đề xuất tiếp tục sử dụng kiến trúc này nhưng có sự điều chỉnh phù hợp với mục đích đề ra như Hình 4.2.



Hình 4.2. Kiến trúc mạng CNN của phương pháp đánh giá sự suy giảm chất lượng hình ảnh

Dữ liệu thực nghiệm đánh giá chất lượng hình ảnh HMII gồm 40286 mẫu. Mỗi mẫu gồm 1 khối ảnh gốc, 1 khối ảnh biến dạng sau mã hoá có cùng kích thước, hệ số lượng tử QP dùng để mã hoá khối gốc và kết quả chất lượng chủ quan DMOS đã được tiền xử lý loại bỏ nhiễu như đề cập ở Chương 3. Mẫu dữ liệu được chuẩn hoá lại thành 3 thành phần khối gốc ghép thêm thành phần hệ số lượng tử QP để thành dữ liệu đầu vào phù hợp kiến trúc trình bày Hình 4.2. Dữ liệu được chia thành 10 phần theo tỷ lệ 8:1:1 dành lần lượt cho việc huấn luyện, hiệu chỉnh và đánh giá phương pháp.

Kết quả kiểm tra trên tập dữ liệu đánh giá gồm 4000 mẫu cho kết quả tương quan tuyến tính Pearson (PLCC) bằng 0.9505 và hệ số tương quan xếp hạng Spearman (SRCC) bằng 0.9088. So sánh với kết quả tương quan giữa duy nhất hệ số lượng tử và chất lượng lần lượt $PLCC = -0,807$ và $SRCC = -0,8438$ cho thấy ảnh hưởng quan trọng của nội dung khối đến chất lượng hình ảnh biến dạng của nó khi nén.

CHƯƠNG 5. ỨNG DỤNG CÁC PHƯƠNG PHÁP ĐỀ XUẤT NÂNG CAO HIỆU NĂNG MÃ HÓA VIDEO

Tóm tắt: Trong Chương 5, hai các phương pháp đánh giá sự suy giảm chất lượng hình ảnh theo đặc trưng nội dung khối ảnh gốc được áp dụng vào bộ nén H264/AVC và H265/HEVC nhằm điều chỉnh hệ số lượng tử theo hai cách: (i) nén theo chất lượng thị giác DMOS cho trước và (ii) giảm tỷ lệ bit đảm bảo chất lượng thị giác.

5.1. Ứng dụng phương pháp đánh giá sự suy giảm chất lượng hình ảnh bằng cách trích chọn đặc trưng cơ bản

5.1.1. Nén video theo chất lượng cho trước trên bộ nén H.264/AVC

Theo mô hình đề xuất, chất lượng của khối suy giảm theo đường cong logistic khi hệ số lượng tử tăng. Thay vì thiết lập hệ số lượng tử, bộ nén được thiết lập mức chất lượng chủ quan cần đạt QC , với thuật toán tìm hệ số lượng tử mỗi khối tương ứng với mức chất lượng QC .

Để đánh giá hiệu quả phương pháp, luận án thực hiện nén bởi bộ nén đề xuất H.254_m một số video mẫu với mức chất lượng $QC=4.0$ và so sánh với bộ nén chuẩn H.264 cùng tỷ lệ bit. Kết quả cho thấy PSNR và hệ số lượng tử trung bình tương đương nhau. Chất lượng theo mô hình đề xuất có một số cải thiện đối với vùng cạnh và vùng nhiều chi tiết. Tuy nhiên những vùng có ít chi tiết trên vùng nền vùng phẳng như vùng đánh dấu màu đỏ, mô hình đề xuất chưa hiệu quả. Thực nghiệm đánh giá chất lượng hình ảnh chủ quan cho thấy chất lượng thị giác video tái tạo sau nén bởi H.254_m tăng 1.62% so với bộ nén gốc.

5.1.2. Nén giảm tỷ lệ bit video đảm bảo chất lượng trên bộ nén H.264/AVC

Mô hình đề xuất áp dụng để thay đổi hệ số lượng tử tăng so với hệ số lượng tử qp thiết lập nhưng chất lượng suy giảm ΔQC không đáng kể. Thuật toán cài đặt trên bộ mã hoá H.264/AVC gốc để có bộ mã hoá H264_m. Thực nghiệm mã hoá các video mẫu trên cả 2 bộ mã với hệ số lượng tử bằng 30, ngưỡng thay đổi chất lượng $\Delta QC = 0.1$

cho kết quả như bảng 6. Các video nén bằng bộ mã hoá H.264_m cho bitrate thấp hơn trung bình 12.45% so với bộ mã hoá gốc do hệ số lượng tử cao hơn. Có một số video giảm đáng kể như *Ducks_take_off*, *Rush_hour*, *Mobcal_ter* tỷ lệ trên 20%. Tuy theo tính chất thị giác, hệ số lượng tử thay đổi theo mỗi video là khác nhau. Một số khối trong khung hình video giữ nguyên hệ số lượng tử thiết lập ban đầu do các đặc trưng khối của khung hình nhạy với các biến dạng dẫn đến tốc độ suy giảm chất lượng quá nhanh.

5.2. Ứng dụng phương pháp đánh giá sự suy giảm chất lượng hình ảnh trích chọn đặc trưng bằng mạng Neural tích chập

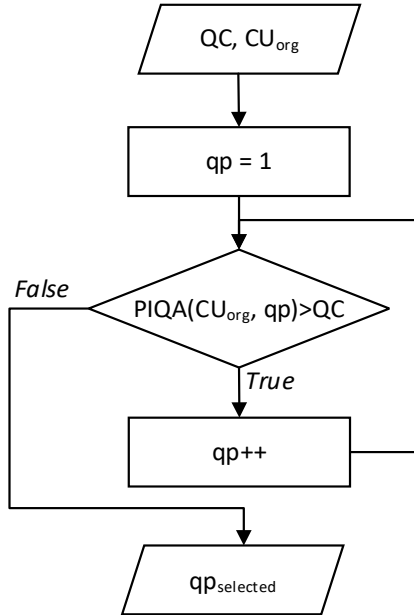
5.2.1. Nén video theo chất lượng cho trước trên bộ nén H.265/HEVC

Tương tự như mục 5.1.1, nội dung này luận án trình bày về thuật toán (PAPS) áp dụng phương pháp dự đoán chất lượng theo hệ số lượng tử trích chọn đặc trưng bằng mạng Neural tích chập đã trình bày ở Chương 4 vào trong bộ nén video H.265/HEVC.

Mục tiêu chính của thuật toán tiết kiệm đáng kể mà không gây ra biến dạng hình ảnh đáng chú ý trong các khung video được tái tạo. Hai yếu tố chính liên quan là giá trị qp_{factor} và qp . Một khối ảnh tham chiếu CU_{org} trong khung hình được mã hoá bởi hệ số lượng tử qp có chất lượng dự đoán DMOS là:

$$y = IQQA(CU_{org}, qp), \quad (5.1)$$

trong đó $IQQA$ là mô hình CNN dự đoán phương pháp đánh giá chất lượng hình ảnh trình bày ở mục **Error! Reference source not found.**



Hình 5.1. Sơ đồ thuật toán PAPS

Thuật toán xác định hệ số lượng tử cho các khối của một khung hình video (thuật toán PAPS) như Hình 5.1. Dữ liệu đầu vào của thuật toán bao gồm khối ảnh gốc CU_{org} và chất lượng cần đạt QC . Kết quả của thuật toán là hệ số lượng tử $qp_{selected}$ ước lượng để chất lượng tối thiểu cần đạt QC . Áp dụng vào bộ mã hoá H.254/HEVC, với yêu cầu chất lượng hình ảnh mã hoá cần đạt QC , mỗi khung hình phân chia thành các khối kích thước 64×64 và đưa vào PAPS. Kết quả, một ma trận hệ số lượng tử được xác định tương ứng với các khối của khung hình làm cơ sở cho quá trình lượng tử hoá. Thực nghiệm đối với các mẫu video chuẩn thuộc 4 nhóm A, B, C, D với 10 khung hình đầu mã hoá

All-intra. Kết quả so sánh chất lượng theo thang đo SSIM bộ nén cải tiến tốt hơn trung bình 5.17% so với bộ nén gốc HM 16.20.

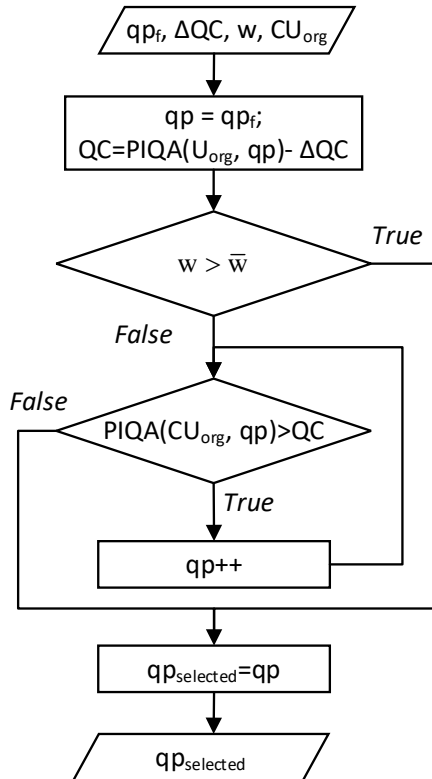
5.2.2. Nén giảm tỷ lệ bit video đảm bảo chất lượng trên bộ nén H.265/HEVC

Luận án đề xuất thuật toán nhằm lựa chọn hệ số lượng tử nhằm giảm tỷ lệ bit nhưng vẫn đảm bảo chất lượng thị giác. Thuật toán xác định hệ số lượng tử cho các khối của một khung hình video (thuật toán APQP) biểu diễn như Hình 5.2. Dữ liệu đầu vào của thuật toán bao gồm khối ảnh gốc CU_{org} , hệ số lượng tử qp và chất lượng điều chỉnh chấp nhận được ΔQC và trọng số chú ý w . Thuật toán đề xuất trong nghiên cứu tăng hệ số lượng tử cho đến khi chất lượng dự đoán còn nhỏ hơn được bằng chất lượng ban đầu cộng thêm ΔQC . Việc thay đổi hệ số lượng tử chỉ áp dụng đối với các khối có trọng số chú ý w thấp hơn trọng số trung bình của toàn khung hình. Kết quả của thuật toán là hệ số lượng tử ước lượng $qp_{selected}$ để chất lượng thị giác thay đổi trong giới hạn.

Để đánh giá hiệu quả của phương pháp lựa chọn hệ số lượng tử nhằm giảm tốc độ bit, đặt thuật toán của phương pháp được cài đặt vào bộ nén H.265/HEVC phiên bản HM 16.0. Bộ nén có điều chỉnh trong nghiên cứu sử dụng thiết lập nén liên khung RA và LDP.

Kết quả áp dụng với thiết lập nén liên khung RA và LDP trên H.265/HEVC cho thấy hiệu năng tăng 6.89% theo thang đo SSIM và 7.47% theo thang đo chủ quan MOS. Đánh giá chất lượng chủ quan cho thấy rằng phương pháp được đề xuất có thể tạo ra chất lượng hình ảnh gần tương đương về nhận thức thị giác so với bộ nén gốc với cùng tỷ lệ bit giảm trung bình 18.83%. Đối với các video thử nghiệm có

nhiều vùng phức tạp, nhiều chi tiết, thuật toán được đề xuất cho hiệu quả nén tốt hơn.



Hình 5.2. Sơ đồ thuật toán APQP

Để làm rõ thêm hiệu quả nén của phương pháp đề xuất, luận án tiến hành so sánh hiệu năng với 3 nghiên cứu liên quan. Kết quả cho thấy hiệu năng nén trung bình với phương pháp đề xuất tốt hơn so với 2 nghiên cứu nhưng kém hơn so với 1 nghiên cứu. Tuy nhiên, phương pháp đề xuất cho hiệu quả hơn đối với các video có độ phân giải cao (nhóm A) và các video có tỷ lệ khối nhiều chi tiết chiếm đa số.

CHƯƠNG 6. KẾT LUẬN VÀ KIẾN NGHỊ

Luận án “*Nâng cao hiệu năng mã hóa video dùng cho truyền thông đa phương tiện*” tập trung nghiên cứu nội dung đánh giá chất lượng hình ảnh phỏng theo thị giác người. Đánh giá chất lượng hình ảnh có ý nghĩa quan trọng trong đánh giá hiệu năng và làm căn cứ để lựa chọn cách mã hoá tối ưu nâng cao hiệu năng nén video.

Những đóng góp khoa học chính của luận án bao gồm hai cơ sở dữ liệu thực nghiệm, ba phương pháp đánh giá chất lượng hình ảnh theo đặc trưng thị giác, hai bộ nén video cải tiến.

Những kết quả thực nghiệm, đánh giá với các phương pháp đề xuất cho thấy: Bộ dữ liệu thực nghiệm được xây dựng đảm bảo tin cậy phù hợp để mô hình hóa, kiểm tra các phương pháp đánh giá chất lượng ảnh khách quan theo đặc trưng thị giác. Các phương pháp đánh giá chất lượng ảnh đã mô hình hóa được đặc trưng thị giác từ dữ liệu. Bộ mã hóa cải tiến có hiệu năng tốt hơn so với bộ mã hóa gốc về các đơn vị đo thị giác khách quan cũng như theo cách đánh giá chủ quan.

Một số hướng nghiên cứu tiếp theo là: Tiếp tục thực nghiệm bổ sung dữ liệu HMII với nhiều loại biến dạng, nhiều khung hình video có nội dung khác nhau. Nghiên cứu áp dụng phương pháp chất lượng hình ảnh có tham chiếu vào bộ lựa chọn tối ưu vòng lặp để cải tiến bộ mã hóa video toàn diện hơn.

DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ

- 1) T. T. Pham, X. V. Hoang, N. T. Nguyen, D. T. Dinh and L. T. Ha, "End-to-End Image Patch Quality Assessment for Image/Video With Compression Artifacts," in *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3040416.
- 2) T. P. Thanh, C. Ma Thi, T. N. Manh, L. Le Dinh and H. Le Thanh, "Compression Artifacts Image Patch database for Perceptual Quality Assessment," *2020 12th International Conference on Knowledge and Systems Engineering (KSE)*, 2020, pp. 55-60, doi: 10.1109/KSE50997.2020.9287704.
- 3) Pham, Thanh Tung and Dinh, Trieu Duong and Hoang, Van Xiem and Vu Huu, Tien and Le, Thanh Ha (2019), Distortion Model based on Perceptual of Local Image Content. In: <http://www.icce-asia2019.org/>, Bangkok, Thailand.
- 4) Thanh Tùng Phạm, Triều Dương Đình, Văn Trọng Đặng and Thanh Hà Lê, Phương pháp lựa chọn hệ số lượng tử trên bộ nén H265/HEVC theo đặc trưng thị giác bằng mạng Neural tích chập, *Hội nghị Quốc gia lần thứ 24 về Điện tử, Truyền thông và Công nghệ Thông tin (REV-ECIT 2021)* (đã chấp nhận).
- 5) M. Le Dinh, L. V. Tung, X. H. Van, D. Dinh Trieu, T. P. Thanh and H. Le Thanh, "Improving 3D-TV view synthesis using motion compensated temporal interpolation," *2016 International Conference on Advanced Technologies for Communications (ATC)*, 2016, pp. 312-317, doi: 10.1109/ATC.2016.7764796.
- 6) Nguyễn Đình Nam, Phạm Thanh Tùng, Lê Thanh Hà, Vũ Hữu Tiến, Hoàng Văn Xiêm, Sáng chế: Phương pháp nén Video dựa trên đặc trưng thị giác ảnh, Số đơn: 1-2018-00173.