BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

LÊ ĐÌNH NGHIỆP

NGHIÊN CỨU CẢI TIẾN KỸ THUẬT PHÁT HIỆN VÀ THAY THẾ ĐỐI TƯỢNG TRONG VIDEO

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2020

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

LÊ ĐÌNH NGHIỆP

NGHIÊN CỨU CẢI TIẾN KỸ THUẬT PHÁT HIỆN VÀ THAY THẾ ĐỐI TƯỢNG TRONG VIDEO

Chuyên ngành: Khoa học máy tính
Mã số: 9 48 01 01

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC:

1. PGS.TS. Phạm Việt Bình

2. PGS.TS. Đỗ Năng Toàn

THÁI NGUYÊN - 2020

LỜI CAM ĐOAN

Tác giả xin cam đoan đây là công trình nghiên cứu của bản thân tác giả. Các kết quả nghiên cứu và các kết luận trong luận án này là trung thực, không sao chép từ bất kỳ một nguồn nào và dưới bất kỳ hình thức nào. Việc tham khảo các nguồn tài liệu đã được thực hiện trích dẫn và ghi nguồn tài liệu tham khảo đúng quy định.

Thái Nguyên, ngày tháng 08 năm 2020

Tác giả luận án

LỜI CẢM ƠN

Luận án được hoàn thành dưới sự hướng dẫn tận tình của PGS. TS. Phạm Việt Bình và PGS.TS. Đỗ Năng Toàn. Lời đầu tiên, tôi xin bày tỏ lòng kính trọng và biết ơn sâu sắc tới hai Thầy.

Tôi xin chân thành cảm ơn TS. Phạm Thế Anh đã đóng góp những ý kiến quý báu cả về học thuật và kinh nghiệm nghiên cứu giúp đỡ tôi trong suốt quá trình thực hiện luận án.

Tôi xin chân thành cảm ơn Ban lãnh đạo trường Đại học Công nghệ thông tin và Truyền thông, Khoa Công nghệ thông tin, Bộ môn Khoa học máy tính, Bộ phận quản lý nghiên cứu sinh – trường Đại học Công nghệ thông tin và Truyền thông – Đại học Thái Nguyên đã tạo mọi điều kiện thuận lợi trong quá trình hoàn thành và bảo vệ luận án.

Xin cảm ơn Ban Giám hiệu Trường Đại học Hồng Đức, các đồng nghiệp tại Phòng Đảm bảo chất lượng và Khảo thí, giảng viên khoa Công nghệ thông tin và Truyền thông – Trường Đại học Hồng Đức, cán bộ Viện Công nghệ thông tin – Đại học Quốc gia Hà Nội đã luôn động viên giúp đỡ tôi trong công tác để tôi có thời gian tập trung nghiên cứu, thực hiện luận án.

Đặc biệt tôi xin bày tỏ lòng biết ơn sâu sắc tới Cha, Mẹ, Vợ, Con và các anh, chị em trong gia đình, những người luôn dành cho tôi những tình cảm nồng ấm và sẻ chia những lúc khó khăn trong cuộc sống, luôn động viên giúp đỡ tôi trong quá trình nghiên cứu. Luận án cũng là món quà tinh thần mà tôi trân trọng gửi tặng đến các thành viên trong Gia đình.

Tôi xin trân trọng cảm ơn!

MỤC LỤC

LỜI CAM ĐOAN i
LỜI CẢM ƠN ii
DANH MỤC CÁC TỪ VIẾT TẮT VÀ KÝ HIỆU vi
DANH MỤC CÁC BẢNG viii
DANH MỤC HÌNH VẼ ix
PHẦN MỞ ĐẦU 1
1. Tính cấp thiết 1
2. Mục tiêu nghiên cứu của luận án 2
3. Đối tượng, phạm vi nghiên cứu của luận án 3
4. Đóng góp chính của luận án 4
5. Phương pháp và nội dung nghiên cứu 5
6. Cấu trúc luận án 5
CHƯƠNG 1. TỔNG QUAN VỀ BÀI TOÁN PHÁT HIỆN VÀ THAY THẾ ĐỐI TƯỢNG TRONG VIDEO 7
1.1. Tổng quan về video và bài toán phát hiện và thay thế đối tượng trong video 7
1.1.1. Khái quát về video 7
1.1.2. Bài toán thay thế đối tượng trong video 12
1.1.3. Một số khái niệm 16
1.1.3.1. Dò tìm đối tượng trong video 16
1.1.3.2. Nhận dạng hình dạng đối tượng trong video 16
1.1.3.3. Phát hiện đối tượng trong video 16
1.1.3.4. Phân vùng đối tượng 17
1.1.3.5. Video inpainting 18
1.1.3.6. Thay thế đối tượng trong video 19
1.1.4. Các thách thức cho bài toán thay thế đối tượng 19
1.2. Tổng quan về các kỹ thuật áp dụng trong hệ thống thay thế đối tượng trong video 22

1.2.1. Dò tìm đối tượng 22
1.2.1.1. Dựa trên điểm đặc trưng 22
1.2.1.2. Dựa trên các mô hình từng phần của đối tượng 23
1.2.1.3. Dựa trên mạng nơron nhân chập 24
1.2.1.4. Phát hiện đối tượng quảng cáo 25
1.2.2. Nhận dạng hình dạng đối tượng 26
1.2.2.1. Lượng tử hóa vector 26
1.2.2.2. Lượng tử hóa tích đề các 29
1.2.2.3. Độ đo khoảng cách 30
1.2.2.4. Nhận dạng hình dạng dựa trên tìm kiếm ANN 32
1.2.3. Các kỹ thuật hoàn thiện video 34
1.2.3.1. Video inpainting dựa trên lấy mẫu 35
1.2.3.2. Inpainting ảnh sử dụng DCNN cho không gian 2D 36
1.2.3.3. Video inpainting sử dụng DCNN cho không gian 3D 37
Kết luận chương 1 38
CHƯƠNG 2. PHÁT HIỆN ĐỐI TƯỢNG TRONG VIDEO 39
2.1. Dò tìm đối tượng trong video 39
2.1.1. Khái quát về mô hình dò tìm đối tượng YOLO 40
2.1.2. Mô hình dò tìm đối tượng cải tiến YOLO-Adv 42
2.1.2.1. Cải tiến trong hàm loss 42
2.1.2.2. Cải tiến trong kiến trúc mạng 46
2.1.2.3. Trích chọn đặc trưng 49
2.1.3. Ước lượng, đánh giá mô hình cải tiến 49
2.1.3.1. Dữ liệu kiểm thử 49
2.1.3.2. Độ đo ước lượng 50
2.1.3.3. Môi trường cài đặt 52
2.1.3.4. Ước lượng, đánh giá 52
2.2. Nhận dạng hình dạng đối tượng 59
2.2.1. Mô hình lập chỉ mục PSVQ 60

2.2.2. Tìm kiếm ANN dựa trên cây phân cụm thứ bậc..... 64
2.2.3. Ước lượng, đánh giá 68
2.2.3.1. Dữ liệu và cấu hình hệ thống kiểm thử..... 69
2.2.3.2. Ước lượng, đánh giá chất lượng mã hóa của PSVQ..... 71
2.2.3.3. Ước lượng, đánh giá tốc độ tìm kiếm với PSVQ 73
2.2.3.4. Ước lượng, đánh giá giải thuật tìm kiếm cây phân cụm thứ bậc kết hợp PSVQ..... 75
Kết luận chương 2..... 80
CHƯƠNG 3. THAY THẾ ĐỐI TƯỢNG VÀ HOÀN THIỆN VIDEO 81
3.1. Phân vùng đối tượng..... 81
3.1.1. Các kỹ thuật phân vùng thực thể 82
3.1.2. Mô hình phân vùng thực thể..... 84
3.1.2.1. Phát sinh mặt nạ vùng..... 85
3.1.2.1. Phân vùng thực thể bằng Mask R-CNN 87
3.1.3. Kết quả thực nghiệm mô hình phân vùng..... 90
3.2. Mô hình hoàn thiện video 92
3.2.1. Kiến trúc mô hình V-RBPconv..... 94
3.2.2. Mô hình kiến trúc mạng RBPconv 95
3.2.3. Hàm loss..... 99
3.2.4. Ước lượng, đánh giá mô hình hoàn thiện video 100
3.2.4.1. Môi trường thực nghiệm..... 101
3.2.4.2. Kết quả so sánh định tính..... 103
3.2.4.3. Kết quả so sánh định lượng 104
Kết luận chương 3..... 109
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 110
DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC CÓ LIÊN QUAN ĐẾN LUẬN ÁN 112
TÀI LIỆU THAM KHẢO..... 113
PHỤ LỤC..... 122

DANH MỤC CÁC TỪ VIẾT TẮT VÀ KÝ HIỆU

Từ viết tắt	Tên đầy đủ (và tạm dịch)
ANN	Approximate Nearest Neighbor (Lân cận xấp xỉ gần nhất)
ADC	Asymmetric distance computation (Tính khoảng cách bất đối xứng)
AVI	Audio Video Interleave (tệp tin đa phương tiện chứa cả âm thanh và hình ảnh bên trong)
CAM	Class Activation Map (Bản đồ kích hoạt lớp)
CPU	Central processing unit (Bộ vi xử lý trung tâm)
CNN	Convolution Neural Network (Mạng nơron tích chập)
DCNN	Deep Convolution Neural Network (Mạng nơron tích chập sâu)
FID	Frechet Inception Distance (khoảng cách Frechet)
FVI	Free-form video inpainting (hoàn thiện/tái tạo video với mặt nạ bất kỳ)
FCN	Fully Convolutional Network (Mạng tích chập đầy đủ)
GAN	Generative Adversarial Networks (Mạng sinh đối kháng)
GPU	Graphics processing unit (Bộ xử lý đồ họa)
HD	High Definition (chuẩn độ nét cao)
HOG	Histogram of oriented gradients (Biểu đồ hướng gradient)
IoU	Intersection over Union (Tỷ lệ trùng khớp giữa hai hộp bao)
IVFADC	Inverted file index Asymmetric distance computation (Chỉ mục danh sách ngược của ADC)
LPIPS	Learned Perceptual Image Patch Similarity (Chỉ số đo sự tượng đồng các mẫu ảnh)
MSE	Mean square error (Lỗi bình phương trung bình)
MPEG	Moving Picture Experts Group (Nhóm chuyên gia hình ảnh động)
NMS	Non-Maxima Suppression (Loại bỏ các điểm không cực trị)

Từ viết tắt	Tên đầy đủ (và tạm dịch)
NTSC	National Television System Committee (Ủy ban quốc gia về các hệ thống truyền hình)
PRM	Peak Response Mapping (Ánh xạ độ nhạy tối đa)
PSNR	Peak signal-to-noise ratio (Tỉ số tín hiệu cực đại trên nhiễu)
PAL	Phase Alternation Line (Hệ truyền hình màu xoay pha)
PQ	Product quantization (Lượng tử hóa tích đề các)
PSL	Peak Simulation Layer (Tầng kích hoạt cực đại)
PSVQ	Product sub-vector quantization (Lượng tử hóa tích đề các cụm vector)
RGB	Red, Green, Blue (Hệ màu RGB)
RoI	Region of Interest (Vùng chứa đối tượng)
R-CNN	Region-based Convolutional Neural Networks (Mạng nơron tích chập dựa trên đề xuất vùng)
SIFT	Scale-Invariant Feature Transform (Biến đổi đặc trưng bất biến tỷ lệ)
SSD	Single Shot Detector (Bộ dò điểm đặc trưng SSD)
SURF	Speeded up robust features (Đặc trưng SURF)
SD	Standard Denfinition (Độ nét tiêu chuẩn)
SSIM	Structural Similarity Index (Chỉ số đồng nhất có cấu trúc)
VGG	Visual Geometry Group (Nhóm hình học trực quan)
YOLO	You only look once (Mạng nhìn đối tượng chỉ một lần)

DANH MỤC CÁC BẢNG

Số hiệu bảng	Tên bảng	Trang
2.1	Thông số phần cứng thực nghiệm mô hình YOLO-Adv	52
2.2	Hiệu năng thực thi trên tập dữ liệu Flickrlogos-47	55
2.3	So sánh mAP của mô hình dò tìm đối tượng trên tập dữ liệu Flickrlogos-32	58
2.4	Các tập dữ liệu đặc trưng	69
2.5	Các tham số dùng để xây dựng các bộ lượng tử	75
3.1	So sánh kết quả của mô hình sử dụng với các phương pháp khác sử dụng nhiều phương pháp tạo mặt nạ huấn luyện khác nhau.	90
3.2	Kết quả định lượng trên tập dữ liệu Places2 của các mô hình: CA, Pconv, EC và RBPConv.	106
3.3	Kết quả định lượng trên tập dữ liệu FVI với các mô hình: EC, CombCN, 3Dgated và V- RBPConv	107
A.1	Số lượng đối tượng cho các tập huấn luyện và kiểm thử của tập dữ liệu FlickrLogo-47.	124