

XÁC ĐỊNH ĐỘ TƯƠNG TỰ CỦA CÁC ẢNH CHỮ KÝ DỰA TRÊN TIẾP CẬN XẾP HẠNG ĐA TẬP

Trần Văn Huy¹, Trần Công Hùng², Ngô Hoàng Huy³, Đào Văn Tuyền²,
Phạm Thị Kim Dung⁴, Nguyễn Văn Quyền⁵, Lê Đình Nghiệp¹

¹Trường Đại học Hồng Đức

²Đại học Quốc tế Sài Gòn

³Đại học CMC, Hà Nội

⁴Swinburne Vietnam, FPT University

⁵Trường Đại học Hải Phòng

tranlehuy@hdu.edu.vn, tranconghung@siu.edu.vn, nhhuu@cmc-u.edu.vn, tuyetdv@gmail.com,

dungptk8@fe.edu.vn, quyenNV@dhhp.edu.vn, ledinhnghiep@hdu.edu.vn

TÓM TẮT: Thuật toán xếp hạng đa tập EMR được sử dụng rộng rãi trong tra cứu ảnh, thông thường, sử dụng các đặc trưng mức thấp như kết cấu và hình dạng để biểu diễn hình ảnh. Đặc trưng như HOG và LBP trong ảnh chữ ký thể hiện khả năng phân biệt ảnh chữ ký rất cao và là một phần bổ sung hiệu quả cho đặc trưng nhúng được tính từ mạng SigNet. Để tăng khả năng phân biệt của thuật toán xếp hạng đa tập EMR trong tra cứu hình ảnh theo nội dung, trong nghiên cứu này, chúng tôi đề xuất kết hợp các đặc trưng mức thấp với các véc tơ nhúng đại diện từ mô hình học sâu. Thực hiện các thử nghiệm trên tập dữ liệu chữ ký tay để chứng minh tính hiệu quả của phương pháp xếp hạng đề xuất trong việc cải thiện chất lượng tra cứu và đo độ tương tự giữa các ảnh chữ ký.

Từ khóa: Tra cứu ảnh chữ ký, Xếp hạng đa tập, EMR, Contrastive loss.

I. GIỚI THIỆU

Trong lĩnh vực tra cứu hình ảnh dựa trên nội dung (CBIR), để đánh giá độ tương tự giữa hai vectơ đặc trưng (đại diện cho hai hình ảnh cần so sánh), thường sử dụng các kỹ thuật đo độ tương tự, đo khoảng cách (như khoảng cách Euclid) và xếp hạng đa tập để so sánh đặc trưng của ảnh truy vấn với các đặc trưng trong cơ sở dữ liệu hình ảnh. Ngay cả khi sử dụng các đặc trưng có sự phân biệt ảnh cao, các độ đo tương tự vẫn đóng vai trò quan trọng trong việc nâng cao độ chính xác của hệ thống.

Các độ đo tương tự truyền thống thường chỉ tập trung vào mức độ tương đồng giữa hai hình ảnh ở mức cục bộ mà không xem xét mối tương tự toàn cục giữa một nhóm các hình ảnh tương tự dẫn đến hiệu quả tìm kiếm độ tương tự không như mong muốn [11]. Để vượt qua những giới hạn này và khám phá cấu trúc bên trong của dữ liệu đặc trưng ảnh, xếp hạng đa tập (Manifold Ranking-MR) đã được nghiên cứu phát triển mạnh từ những năm 2000 và được áp dụng hiệu quả trong nhiều vấn đề về tìm kiếm tương tự.

EMR (Efficient Manifold Ranking) [5, 9] là một thuật toán xếp hạng đa tập được sử dụng rộng rãi trong hệ thống Tra cứu ảnh dựa trên nội dung (Content-Based Image Retrieval - CBIR). Khác với các độ đo tương tự dựa trên khoảng cách Euclidean, EMR đã được chứng minh là rất hiệu quả khi xây dựng được đồ thị quan hệ kề có tính đối xứng giữa các vectơ đặc trưng ảnh, phương pháp giải hàm mục tiêu tính vectơ xếp hạng theo sơ đồ lan truyền trên đồ thị kế phản ánh được cấu trúc nội tại của cơ sở dữ liệu vectơ đặc trưng ảnh.

Chữ ký trên văn bản (loại offline) là một trong những dấu ấn sinh trắc học phổ biến và được chấp nhận rộng rãi liên quan đến con người như: tài liệu, biểu mẫu, séc ngân hàng, cá nhân,... Do đó, việc xác minh chữ ký trở thành một nhiệm vụ quan trọng và đã có nhiều nỗ lực được thực hiện để loại bỏ sự không chắc chắn trong quy trình xác thực thủ công, từ đó biến xác minh chữ ký thành một hướng nghiên cứu quan trọng trong lĩnh vực học máy và nhận dạng mẫu [1-4].

Có rất nhiều thách thức trong việc thiết kế xác minh chữ ký offline khi các đặc trưng thời gian động của chữ ký không có sẵn. Mạng nơron tích chập sâu (Deep Convolutional Neural Networks - DCNNs) có khả năng tuyệt vời trong việc trích xuất các đặc trưng từ hình ảnh chữ ký như SigNet [1]. Các đặc trưng trích rút từ mạng nơron tích chập (CNN), như đặc trưng từ mô hình contrastive [2] hoặc triplet loss [8] thường phải cần một lượng lớn dữ liệu để mạng thích ứng được với sự thay đổi của đối tượng, vì thế khi triển khai các hệ thống thực tế cần kết hợp đặc trưng CNN và đặc trưng mức thấp khi dữ liệu không đủ đa dạng về hướng và kết cấu.

Bên cạnh đặc trưng biểu diễn ảnh chữ ký, độ đo tương tự và phân tương tự cũng là một thành phần quan trọng trong các hệ thống nhận dạng, xác thực ảnh chữ ký.

Nói chung các độ đo phân tương tự như khoảng cách Euclid là không phù hợp khi so khớp đặc trưng biểu diễn ảnh và học đa tập, học độ đo khoảng cách có mối liên hệ với nhau. Ngoài ra, như bỏ đề tại [Phụ lục A] đã chứng minh về tính tổng quát của dữ liệu đa tập, để nhận dạng và đối sánh ảnh chữ ký, chúng ta có thể sử dụng các thuật toán theo tiếp cận dữ liệu đa tập trong các không gian được giả định là không Euclid. Dữ liệu đa tập, xếp hạng đa tập (MR) là mô hình dựa trên đồ thị đã được áp dụng thành công cho tra cứu hình ảnh dựa trên nội dung (CBIR) bằng các đặc trưng

mức thấp [5-9] và đặc trưng CNN. Trong bài viết này, chúng tôi đề xuất sử dụng xếp hạng đa tập để kết hợp các ưu điểm của đặc trưng mức thấp như HOG, LBP và véc tơ biểu diễn ảnh được trích rút từ mạng SigNet khi nhận dạng và đối sánh ảnh chữ ký offline.

Phần còn lại của bài báo được cấu trúc như sau: Mục II nghiên cứu liên quan và phương pháp đề xuất. Mục III thực nghiệm trên bộ dữ liệu SIGN_HD300. Mục IV là kết luận.

II. NGHIÊN CỨU LIÊN QUAN VÀ PHƯƠNG PHÁP ĐỀ XUẤT

A. Mạng SigNet học độ đo khoảng cách

Deep Metric Learning (DML) là một tiếp cận mạnh mẽ để trích chọn đặc trưng ảnh và là một phương pháp hiệu quả để giảm chiều dữ liệu trong quá trình phân lớp ảnh. DML sử dụng các kiến trúc mạng sâu để khắc phục các hạn chế của dữ liệu gốc bằng cách học các đặc trưng nhúng tương tự thông qua không gian con phi tuyến. Việc này đã được chứng minh là hiệu quả cao trong phân lớp và phân cụm ảnh. DML có khả năng tạo ra các đặc trưng véc tơ từ hình ảnh đầu vào và đã được áp dụng trong nhiều lĩnh vực khác nhau, bao gồm nhận dạng mặt người, xác minh chữ ký [1, 8],...

Trong bài báo này, chúng tôi lựa chọn DML với mạng Siamese làm mô hình tiền huấn luyện để trích xuất các véc tơ nhúng và trích xuất đặc trưng nhúng. Mạng nơron SigNet [1] dạng Siamese là một lớp kiến trúc mạng thường chứa hai phần mạng con đồng nhất. Các mạng nơron tích chập (CNNs) song sinh có cùng cấu hình với các tham số và trọng số chia sẻ giống nhau. Việc cập nhật tham số được thực hiện đối xứng trên cả hai phần mạng con. Những phần mạng con này được kết hợp thông qua một hàm mất mát ở phía trên, tính toán một độ đo tương tự liên quan đến khoảng cách Euclid giữa biểu diễn đặc trưng ở mỗi bên của mạng Siamese. Một hàm mất mát phổ biến được sử dụng chủ yếu trong mạng Siamese là hàm mất mát tương phản, được định nghĩa như sau:

$$L_{Contrastive} = \alpha(1 - y) \left\| f_{w_1}(s_1) - f_{w_2}(s_2) \right\|_{L_2}^2 + \beta \max \left(0, m - \left\| f_{w_1}(s_1) - f_{w_2}(s_2) \right\|_{L_2} \right)^2 \quad (1)$$

trong đó, s_1 và s_2 là hai mẫu (ở đây là hình ảnh chữ ký), y là hàm chỉ số nhị phân đánh dấu xem hai mẫu có thuộc cùng một lớp hay không, α và β là hai hằng số, ở đây m được giới hạn bằng 1 và f là một hàm nhúng chuyển đổi hình ảnh chữ ký thành không gian véc tơ thực thông qua CNN, và w_1, w_2 là các trọng số học được cho một lớp cụ thể của mạng cơ sở. Khác với các phương pháp tiếp cận thông thường gán nhãn tương tự nhị phân cho các cặp, mạng Siamese nhằm đưa các vectơ đặc trưng đầu ra gần nhau cho các cặp đầu vào được gán nhãn là tương tự, và đẩy các vectơ đặc trưng xa nhau nếu các cặp đầu vào không tương tự. Mỗi nhánh của mạng Siamese có thể được xem như một hàm nhúng hình ảnh đầu vào vào một không gian. Sử dụng Contrastive loss (công thức 1) nên không gian này sẽ có tính chất rằng hình ảnh của cùng một lớp (chữ ký thật của một người viết cụ thể) sẽ gần nhau hơn so với hình ảnh của các lớp khác nhau (chữ ký giả mạo hoặc chữ ký của các người viết khác nhau). Cả hai nhánh được kết hợp với nhau thông qua một lớp tính toán khoảng cách Euclid giữa hai điểm trong không gian nhúng. Sau đó, để quyết định hai hình ảnh có thuộc cùng lớp tương tự (thật, thật) hay lớp không tương tự (thật, giả) cần xác định một giá trị ngưỡng dựa trên giá trị khoảng cách.

Bảng 1. Bảng tổng quan về các mạng CNN

Layer	Size	Parameters
Convolution	96×11×11	stride = 1
Local Response Norm	-	$\alpha = 10^{-4}; \beta = 0.75, k = 2; n = 5$
Pooling	96×3×3	stride = 2
Convolution	256×5×5	stride = 1, pad = 2
Local Response Norm	-	$\alpha = 10^{-4}; \beta = 0.75, k = 2; n = 5$
Pooling + Dropout	256×3×3	stride = 2; p = 0:3
Convolution	384×3×3	stride = 1; pad = 1
Convolution	256×3×3	stride = 1; pad = 1
Pooling + Dropout	256×3×3	stride = 2; p = 0:3
Fully Connected + Dropout	1024	p = 0:5
Fully Connected	128	

Các lớp tích chập đầu tiên lọc hình ảnh chữ ký đầu vào kích thước 155×220 bằng 96 kernel kích thước 11×11 với bước đi là 1 pixel. Lớp tích chập thứ hai nhận đầu vào từ kết quả (đã được chuẩn hóa và thu gọn) của lớp tích chập đầu tiên và lọc nó bằng 256 kernel kích thước 5×5. Lớp tích chập thứ ba và thứ tư được kết nối với nhau mà không có sự can thiệp của lớp thu gọn hoặc chuẩn hóa. Lớp thứ ba có 384 kernel kích thước 3×3 được kết nối với đầu ra (đã chuẩn hóa, thu gọn và thực hiện dropout) của lớp tích chập thứ hai được đưa vào lớp tích chập thứ tư. Lớp tích chập thứ tư có 256 kernel kích thước 3×3. Điều này dẫn đến mạng nơron học các đặc trưng cấp thấp ít hơn cho các vùng tiếp

nhỏ hơn và nhiều đặc trưng cho các đặc trưng cấp cao hoặc trừu tượng hơn. Lớp kết nối đầy đủ đầu tiên có 1024 neuron, trong khi lớp kết nối đầy đủ thứ hai có 128 neuron. Điều này cho thấy rằng véc tơ đặc trưng đã học cuối cùng của SigNet có kích thước bằng 128.

B. Xếp hạng đa tập hiệu quả

Khác với các thuật toán MR trong CBIR, thuật toán gốc EMR [5, 9] tập trung vào chọn số lượng điểm neo (anchors), sau đó xếp hạng các điểm dữ liệu trong cơ sở dữ liệu hình ảnh. Việc xếp hạng các hình ảnh dựa trên các mối quan hệ lân cận trong biểu đồ EMR được biểu diễn thông qua các điểm neo này và duy trì một cấu trúc đồ thị có khả năng mở rộng. Vì vậy, EMR có khả năng áp dụng xếp hạng một cách hiệu quả trên các cơ sở dữ liệu lớn và do đó nâng cao hiệu suất của các hệ thống CBIR [5-9].

Một véc tơ đặc trưng E_i được gọi là kết nối với E_j nếu $i \neq j$ và tồn tại một điểm neo chung cụ thể A_c ($c = 1, 2, \dots, C$; c là số lượng các điểm neo) sao cho A_c là láng giềng của cả E_i và E_j . Với mỗi ký hiệu E_i , chúng ta sẽ ký hiệu $N_b(i; s)$ là một tập hợp gồm s véc tơ đặc trưng của các điểm neo gần nhất với E_i (s là một tham số thử nghiệm, ví dụ như $s = 5$) và khoảng cách lớn nhất giữa E_i và s được biểu diễn bởi:

$$d_s = \max_{l \in N_b(i, s)} \{d(E_i, A_l)\} \quad (2)$$

Độ đo đa tập là một véc tơ xếp hạng $r_Q = (r_i)$ được xây dựng từ hàm mục tiêu sau:

$$EMR_Q(r) = \frac{1}{2} \left(\sum_{i,j} w_{ij} \left\| \frac{r_i}{\sqrt{D_{ii}}} - \frac{r_j}{\sqrt{D_{jj}}} \right\|^2 + \mu \sum_i \|r_i - r_{0,i}\|^2 \right) \rightarrow \min \quad (3)$$

(i) Ma trận trọng số Z biểu thị khoảng cách giữa các điểm neo được tính bằng:

$$Z = (z_{ci})_{1 \leq c \leq C, 1 \leq i \leq n}, z_{ci} = \frac{K\left(\frac{\|E_i, A_c\|}{d_s}\right)}{\sum_{c' \in N_b(i, s)} K\left(\frac{\|E_i, A_{c'}\|}{d_{c'}}\right)}, \forall c \in N_b(i, s), z_{ci} = 0, \forall c \notin N_b(i, s) \quad (4)$$

ở đây, $K(t) = \frac{3}{4}(1-t^2)$, $-1 \leq t \leq 1$;

(ii) Ma trận kề W được xác định như sau $W \stackrel{\text{def}}{=} Z^T Z$ cụ thể

$$W = (w_{ij})_{1 \leq i, j \leq n}, w_{ij} = \sum_{c \in N_b(i, s) \cap N_b(j, s)} z_{ci} * z_{cj}, 1 \leq i, j \leq n \quad (6)$$

(iii) Ma trận đường chéo $D = (D_{ii})$, với $i = \overline{1, n}$ $D_{ii} \stackrel{\text{def}}{=} \sum_{j=1}^n w_{ij}$ (7)

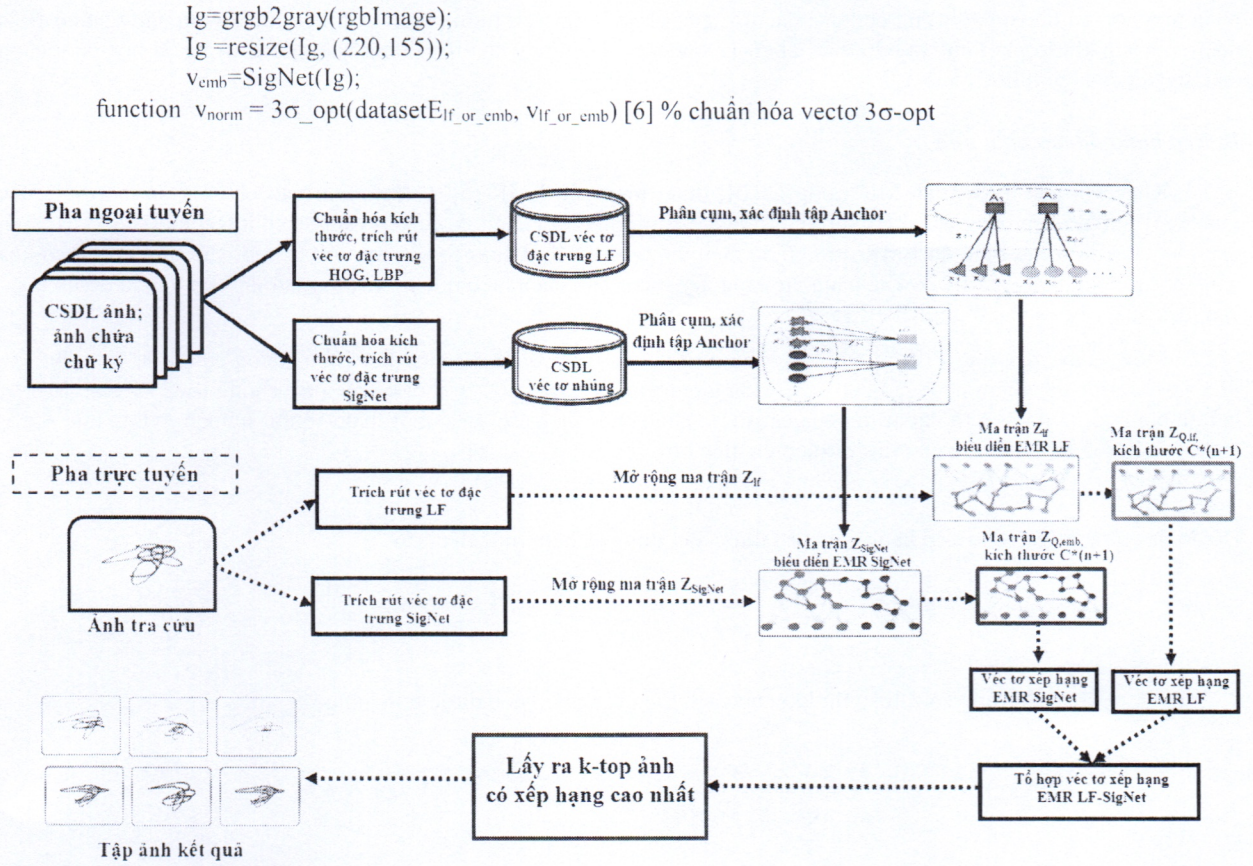
C. Đề xuất thuật toán kết hợp các xếp hạng

Đặc trưng của ảnh chữ ký có thể chia thành hai loại: đặc trưng mức thấp và đặc trưng mức cao. Đặc trưng mức thấp chứa thông tin về cấu trúc và hình dạng của chữ ký, đặc trưng mức cao giúp nâng cao tính phân biệt của chữ ký, tuy nhiên chúng phụ thuộc vào hướng và cấu trúc cụ thể của ảnh chữ ký. Bằng cách sử dụng các đặc trưng mức cao, biểu thị bằng các véc tơ nhúng trích rút từ một mạng học sâu kiểu triplet loss hoặc contrastive loss, tra cứu hình ảnh cho kết quả với độ chính xác tương đối cao [7, 14]. Giả sử rằng tập dữ liệu ảnh chữ ký gốc, mỗi ảnh có một véc tơ nhúng duy nhất tương ứng được xác định bởi DML (cụ thể là mạng SigNet), và với một ảnh chữ ký truy vấn I_Q , hệ

thống lựa chọn xếp hạng theo hình ảnh chữ ký đặc trưng mức thấp $r_{lf, v, Q}^*$ kết hợp với xếp hạng đặc trưng mức cao $r_{emb, Q, i}^*$, trong đó các véc tơ xếp hạng được tính bằng EMR. Hình 1 sau mô tả các bước thực hiện thuật toán tra cứu ảnh chữ ký mà chúng tôi đề xuất:

Thủ tục chuẩn hóa ảnh và đặc trưng mức thấp (HOG, LBP) và đặc trưng mức cao (SigNet) được thực hiện thông qua hai thủ tục sau:

```
function v_lf = get_hog_lbp(rgbImage)
    I = resize(rgbImage, (64, 128));
    v_hog = hog(I);
    v_lbp = lbp(I);
    v_lf = v_hog + v_lbp; % kết nối tự nhiên
function v_emb = get_emb(rgbImage)
```

Hình 1. Sơ đồ mô tả hệ thống (CBIR) tra cứu ảnh chữ ký sử dụng kết hợp xếp hạng đa tập

Chúng tôi đề xuất thuật toán chính EMR-LF-SIGNET kết hợp xếp hạng EMR trên đặc trưng mức thấp và EMR trên đặc trưng mức cao (SigNet) như sau:

Bước 1. Xây dựng đồ thị của xếp hạng EMR_{emb} của CSDL ảnh E với n ảnh trên đặc trưng nhúng (SigNet):

- 1.1. Chuẩn hóa các ảnh của CSDL ảnh về kích thước 155x220, và chuyển về ảnh grayscale
- 1.2. Xác định các vector nhúng chiều deb cho mỗi ảnh khi cho ảnh chạy qua mạng SigNet (thủ tục **get_emb**).
- 1.3. Chuẩn hóa và chuẩn hóa 3σ -opt toàn bộ CSDL vector đặc trưng nhúng (thủ tục **3σ_opt**).
- 1.4. Phân cụm K-means với C cụm tập các vector nhúng thu được ở bước 1.2, nhận được tập tâm cụm và gán cho tập điểm neo của đồ thị EMR.
- 1.5. Xác định ma trận Z_{emb} biểu diễn giá trị khoảng cách giữa từng vector nhúng v_{cb} của CSDL ảnh với r điểm neo gần nhất của v_{eb} .
- 1.6. Chuẩn hóa và chuẩn hóa 3σ -opt toàn bộ CSDL vector đặc trưng mức thấp (thủ tục **3σ_opt**).
- 1.7. Xác định các ma trận của đồ thị EMR_{emb} dựa trên Z_{emb} .

Bước 2. Xây dựng đồ thị của xếp hạng EMR_{lf} trên đặc trưng mức thấp (HOG và LBP kết nối với nhau):

Tương tự các bước nhỏ ở bước 1, ta được ma trận Z_{lf} và các ma trận khác của đồ thị biểu diễn EMR_{lf} cho đặc trưng mức thấp HOG và LBP (thủ tục **get_hog_lbp**), và chuẩn hóa 3σ -opt.

Bước 3. Tại thời điểm chạy (runtime), tính vector xếp hạng từ một ảnh truy vấn Q:

- 3.1. Chuẩn hóa ảnh I_Q về kích thước 155x220, chuyển ảnh về ảnh grayscale và chạy qua mô hình SigNet thu nhận được véc tơ nhúng $v_{EQ,emb}$ chiều deb (gọi thủ tục **get_emb** và thủ tục **3σ_opt**).
- 3.2. Mở rộng ma trận Z_{emb} của đồ thị EMR_{emb} [9], thu được ma trận trọng số $Z_{Q,emb}$ mới có kích thước $C \times (n+1)$.
- 3.3. Tính hạng của $v_{EQ,emb}$, chúng ta thu nhận được vector xếp hạng $r^*_{emb.v.Q}$.
- 3.4. Chuẩn hóa ảnh I_Q về kích thước 64x128, tính đặc trưng HOG và LBP (gọi thủ tục **get_hog_lbp** và thủ tục **3σ_opt**).

Kết nối hai đặc trưng, chúng ta thu nhận được vector mức thấp $v_{EQ,lf}$ chiều d_{lf} .

- 3.5. Mở rộng ma trận Z_{lf} của đồ thị EMR_{lf} [9], thu được ma trận trọng số $Z_{Q,lf}$ mới có kích thước $C \times (n+1)$.

3.6. Tính hạng của $v.E_{Q,emb}$, chúng ta thu nhận được vector xếp hạng $r^*_{lf,v,Q}$.

Bước 4. Đầu ra của thuật toán là vector xếp hạng kết hợp: $r^*_Q = RC(r^*_{lf,v,Q}, r^*_{emb,v,Q})$, ở đây $RC(r_{lf}, r_{emb})$ là hàm giá trị vector kết hợp hạng của 2 vector đầu vào để tạo ra vector xếp hạng đầu ra.

Thuật toán EMR-LF-SIGNET chia thành 2 pha offline và online và được thực hiện bằng cách tổ hợp 2 xếp hạng EMR. Đối với pha offline, độ phức tạp là: $O(C*n*d) + O(C^3)$ trong đó C là số cụm, n số mẫu đầu vào, d số chiều véc tơ đặc trưng. Đối với pha online, độ phức tạp là $O(C*n*d)$.

Vậy thuật toán có độ phức tạp là $O(C*n*d) + O(C^3)$ là tương đương với độ phức tạp của EMR gốc [5].

Những đóng góp chính của đề xuất của chúng tôi như sau:

(1) Cung cấp một cách kết hợp giữa đặc trưng HOG, LBP và đặc trưng CNN SigNet một cách hiệu quả, từ đó làm tăng độ chính xác của kết quả tra cứu chữ ký.

(2) Sử dụng bộ dữ liệu SIGN_300HD tiến hành thực nghiệm để chứng minh thuật toán đo độ tương tự sử dụng 2 xếp hạng EMR có hiệu quả hơn khoảng cách Euclid.

D. Xác định độ tương tự của cặp ảnh chữ ký

Trên thực tế chúng ta thường gặp bài toán xác định độ tương tự của cặp ảnh chữ ký, vấn đề này đang đối mặt với hạn chế lớn khi áp dụng EMR cho các ảnh nằm ngoài tập dữ liệu [5]. Để giải quyết vấn đề này, chúng tôi đề xuất tích hợp máy học bằng việc huấn luyện một mô hình Support Vector Machine Regression (SVR) để học kết quả xếp hạng EMR cho mỗi cặp hình ảnh trong bộ dữ liệu huấn luyện.

Tập cặp vector đặc trưng ảnh huấn luyện (I_{v1}, I_{v2}) được chọn như sau: I_{v1} được chọn trong tập 50% ảnh ngẫu nhiên trong CSDL vector đặc trưng ảnh. Với mỗi I_{v1} chọn I_{v2} trong 20 ảnh được chọn gần nhất với I_{v1} về mặt khoảng cách Euclide. Sau khi xây dựng được tập (I_{v1}, I_{v2}) , chúng ta chia tập này thành 80% cặp ngẫu nhiên để huấn luyện mô hình, 20% tập cặp vector đặc trưng ảnh được sử dụng để làm tập kiểm tra huấn luyện. Chú ý là mỗi cặp (I_{v1}, I_{v2}) sẽ cho chúng ta 1 vector $I_{v1}-I_{v2}$. Bằng cách sử dụng xếp hạng EMR như là đầu ra, mô hình SVM có thể học dự đoán độ tương tự hình ảnh một cách chính xác dựa trên các sai lệch của các vector đặc trưng đầu vào của hình ảnh.

Qua đó giảm chi phí tính toán và hiệu quả hơn khi áp dụng cho các tập dữ liệu mới. Kết quả thử nghiệm trên tập dữ liệu huấn luyện (train) độ tương quan là 0,94 và trên tập dữ liệu thử nghiệm (test) độ tương quan là 0,92. Các giá trị hệ số tương quan rất cao đã chứng tỏ tính hiệu quả của mô hình học máy.

III. THỰC NGHIỆM

A. Bộ dữ liệu thực SIGN_300HD

Bộ dữ liệu chữ ký thực nghiệm được thu thập của 300 người với 4800 hình ảnh được chúng tôi sử dụng. Bộ dữ liệu này được tổng hợp từ nhiều nguồn, bao gồm: GPDS 1-150, ICEDAR, Bengali và thu thập từ chữ ký thực tế của 50 người tại Trường Đại học Hồng Đức. Đây là bộ dữ liệu phức tạp và chưa gắn nhãn, có cấu trúc 300 lớp, với 16 ảnh trên một lớp, có dung lượng ~285 MB. Ảnh chữ ký được đánh số từ 01-16 trong mỗi thư mục, chúng tôi đặt tên tạm thời cho bộ dữ liệu này là SIGN_300HD.

B. Trích chọn đặc trưng

Ở phần thực nghiệm, chúng tôi lựa chọn 02 đặc trưng HOG với 720 chiều, LBP với 59 là đặc trưng mức thấp (LF) để mô tả mỗi chữ ký trong tập SIGN_300HD. 02 đặc trưng này đã được chuẩn hóa để mỗi thành phần vector của mỗi ảnh nằm trong khoảng $[0, 1]$ và sau đó ghép lại thành một vector với số chiều $d_{lf} = 779$. Đồng thời, kích thước của mỗi hình ảnh trong bộ dữ liệu được chuẩn hóa về kích thước 155×220 và được đưa vào mô hình Ω_2 SigNet, mô hình này sử dụng Contrastive loss và một vector nhúng có kích thước $d_{eb} = 128$.

C. Tham số thực nghiệm

Các tham số được thiết lập chung cho các thực nghiệm, cụ thể tham số $a = 0,99$, số điểm neo $C = 1500$, $r = 3$, n_{best} cho mỗi lần truy vấn $n_b = 16$, 20% số lượng mẫu truy vấn lấy ngẫu nhiên. Hàm giá trị vector RC kết hợp xếp hạng là: $RC(r^*_{lf,v,Q}, r^*_{emb,v,Q}) = \alpha r^*_{lf,v,Q} + \beta r^*_{emb,v,Q}$, ở đây $\alpha = 0,8$, $\beta = 0,2$. Chỉ số đánh giá hiệu quả tra cứu ảnh chữ ký được sử dụng là mAP (xem [9]).

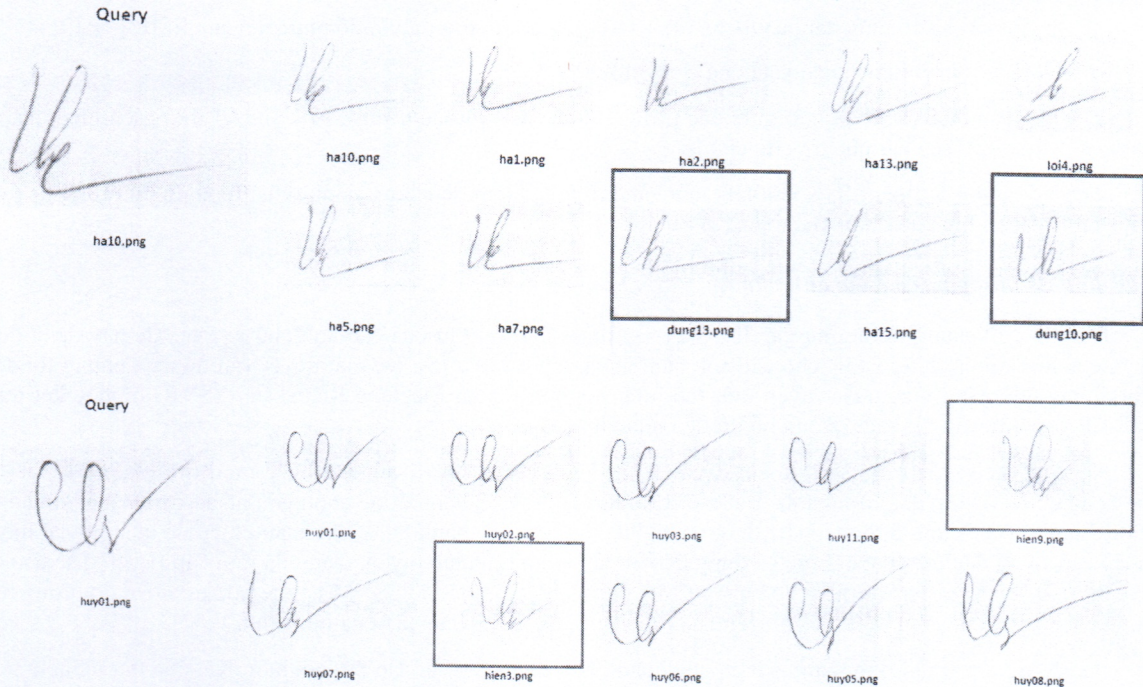
D. Kết quả thực nghiệm và luận giải

Chúng tôi thực hiện các thử nghiệm trên bộ dữ liệu được chọn SIGN_300HD và đánh giá độ chính xác theo mAP khi truy xuất hình ảnh chữ ký với thuật toán EMR-LF-SIGNET. 3 thực nghiệm được tiến hành gồm: 1. Sử dụng khoảng cách Euclid đo độ tương tự của các đặc trưng HOG+LBP, đặc trưng SigNet; 2. Sử dụng EMR tiêu chuẩn cho

xếp hạng đặc trưng HOG+LBP, đặc trưng SigNet; 3. Sử dụng kết hợp xếp hạng EMR theo phương pháp đề xuất, cụ thể:

Thực nghiệm 1: Sử dụng khoảng cách Euclid

Thực nghiệm 1a sử dụng khoảng cách Euclid đo độ tương tự của các đặc trưng HOG+LBP cho kết quả với độ chính xác 49,5%, Thực nghiệm 1b sử dụng khoảng cách Euclid cho đặc trưng SigNet cho kết quả với độ chính xác 52,1%.



Hình 2. Ảnh truy vấn chữ ký

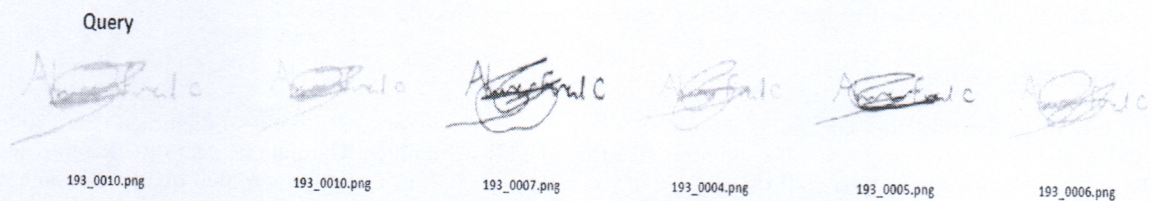
Thực nghiệm 2: Sử dụng EMR

Thực nghiệm 2a sử dụng EMR để xếp hạng cho các véc tơ trích xuất từ bộ đặc trưng HOG+LBP, thực nghiệm 2b sử dụng EMR cho đặc trưng SigNet. Một véc tơ nhúng được chọn tương ứng với hình ảnh truy vấn có kích thước chữ ký lớn nhất. Thực nghiệm 2a đạt độ chính xác 52,1%, thực nghiệm 2b, độ chính xác là 54,1%.

Đánh giá: Kết quả thực nghiệm rõ ràng cho thấy ưu điểm của việc sử dụng các đặc trưng nhúng và đo độ tương tự bằng EMR so với khoảng cách Euclid. Tuy nhiên, cần lưu ý rằng có những trường hợp cụ thể, kết quả truy xuất có thể không như mong đợi, mặc dù véc tơ nhúng của hình ảnh truy vấn là rất tốt. Nguyên nhân chính là sự khác biệt giữa dữ liệu tiền huấn luyện (pretrain) và bộ dữ liệu SIGN_300HD. Điều này thường gặp khi áp dụng Deep Learning trong CBIR, do thường sử dụng mô hình tiền huấn luyện.

Thực nghiệm 3: Kết hợp xếp hạng EMR

Trong phần này, chúng tôi dùng hai EMR để xếp hạng tương ứng cho các véc tơ trích xuất từ bộ đặc trưng SigNet và véc tơ đặc trưng HOG+LBP, từ đó kết hợp các xếp hạng này theo thuật toán đề xuất (EMR-LF-SIGNET) cho kết quả đạt độ chính xác 64,4%. Hình sau mô tả kết quả truy vấn của ảnh 193.0010.png khi sử dụng kết hợp xếp hạng EMR với $n_b = 5$.



Hình 3. Ảnh truy vấn chữ ký sử dụng phương pháp kết hợp EMR

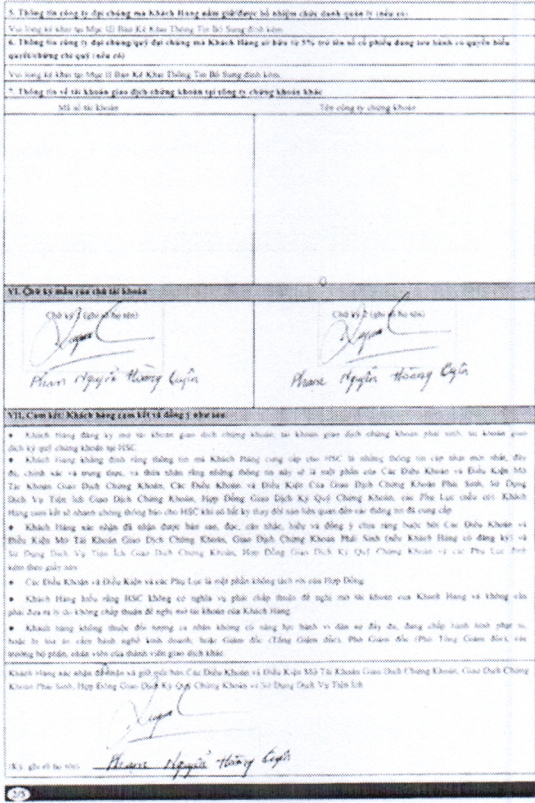
Kết quả thực nghiệm được tổng hợp vào Bảng 2. Các kết quả thu được thể hiện EMR cho kết quả tốt khi chạy riêng rẽ và tốt hơn hẳn khi được kết hợp, độ chính xác được tính bằng mAP, cụ thể như sau: Khoảng cách Euclid với đặc trưng HOG+LBP, Khoảng cách Euclid với đặc trưng SigNet, EMR với đặc trưng HOG+LBP, EMR với đặc trưng SigNet và cuối cùng là kết quả của kết hợp xếp hạng EMR.

Bảng 2. Bảng kết quả thực nghiệm

STT	Phương pháp	mAP
1	Khoảng cách Euclid với đặc trưng HOG, LBP	49,5%
2	Khoảng cách Euclid với đặc trưng SigNet	52,1%
3	EMR cho đặc trưng HOG, LBP	54,8%
4	EMR cho đặc trưng SigNet	54,1%
5	Kết hợp xếp hạng EMR (phương pháp đề xuất)	64,4%

E. Nhận dạng chữ ký trên văn bản thực tế

Quá trình định vị vùng chữ ký trên ảnh chụp hợp đồng, văn bản, và công văn sử dụng thuật toán YOLOv5 diễn ra qua một loạt các bước chính. Đầu tiên, ảnh đầu vào được đưa vào mạng YOLOv5 để xác định các vùng có khả năng chứa chữ ký. Thuật toán sẽ phân loại và dự đoán các bounding box xung quanh chữ ký có trong ảnh. Sau khi đã xác định được vùng chứa chữ ký, chúng ta tiến hành cắt và trích xuất những phần ảnh này để chuẩn bị cho bước tiếp theo.



Hình 4. Vùng chữ ký được định vị bằng thuật toán YOLOv5

Tiếp theo, các phần ảnh chứa chữ ký được chuẩn hóa về kích thước 155x220 để đảm bảo tính nhất quán cho việc trích xuất đặc trưng. Sau khi chuẩn hóa, các ảnh này được đưa vào mạng SigNet để trích xuất các véc tơ đặc trưng có chiều dài 128. Mạng SigNet được đào tạo đặc biệt để nhận diện và trích xuất thông tin quan trọng từ chữ ký trên các ảnh nhỏ kích thước này.

Quá trình này sử dụng một kết hợp giữa YOLOv5 và mạng SigNet giúp xác định và trích xuất thông tin chữ ký một cách hiệu quả từ các văn bản hợp đồng, văn bản và công văn. Các véc tơ đặc trưng 128 chiều từ mạng SigNet có thể được sử dụng để so sánh và nhận biết chữ ký giữa các tài liệu khác nhau, đồng thời cung cấp thông tin quan trọng cho các ứng dụng liên quan đến xác thực và nhận dạng chữ ký trong lĩnh vực số hóa tài liệu. Hình 4 minh họa quá trình crop chữ ký.

IV. KẾT LUẬN

Kết quả thực nghiệm của nghiên cứu đề xuất rõ ràng cho thấy tính hiệu quả của phương pháp trong việc tra cứu ảnh chữ ký. Sự kết hợp hiệu quả giữa các véc tơ đặc trưng HOG, LBP và véc tơ đặc trưng tính từ mạng SigNet trong hệ thống CBIR đã được chứng minh là hữu ích và mang lại kết quả đáng kể. Điều này thể hiện qua việc tận dụng cả đặc điểm hình dạng, kết cấu và đặc điểm phân biệt ngữ nghĩa sau quá trình đào tạo để giải quyết một phần vấn đề về khoảng cách ngữ nghĩa giữa các đặc trưng mức thấp và mức cao.

Phương pháp kết hợp các xếp hạng theo đặc trưng mức thấp và xếp hạng theo véc tơ nhúng đã được trình bày trong nghiên cứu, với việc tính hai giá trị xếp hạng tương ứng bằng thuật toán EMR và sau đó tổng hợp thành một xếp hạng duy nhất. Thông qua sự kết hợp này, nghiên cứu đã nâng cao được chất lượng xếp hạng EMR của các véc tơ nhúng. Tất cả các thử nghiệm đã được thực hiện và kết quả đã chứng minh rằng phương pháp đề xuất là hiệu quả và có tiềm năng trong lĩnh vực này.

Trong thời gian tới, chúng tôi dự kiến áp dụng cách xếp hạng EMR trong xây dựng độ đo tương tự cho ảnh chữ ký, với mục tiêu tối ưu hóa chất lượng so khớp ảnh chữ ký trong các trường hợp chữ ký có đặc điểm đường nét phức tạp.

TÀI LIỆU THAM KHẢO

- [1] Dey, Sounak & Dutta, Anjan & Toledo, J. & Ghosh, Suman & Lladós, Josep & Pal, Umapada. (2017). "SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification".
- [2] Ebrahim Parcham, Mahdi Ilbeygi, Mohammad Amini, "CBCapsNet: A novel writer-independent offline signature verification model using a CNN-based architecture and capsule neural networks," *Expert Systems with Applications*, vol. 185, 2021, 115649.
- [3] M. Taşkiran and Z. G. Çam, "Offline signature identification via HOG features and artificial neural networks," *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Herl'any, Slovakia, 2017, pp. 000083-000086, DOI: 10.1109/SAMI.2017.7880280.
- [4] A. Soleimani, B. N. Araabi and K. Fouladi, "Deep multitask metric learning for offline signature verification," *Pattern Recognition Letters*, vol. 80, pp. 84-90, 2016.
- [5] B. Xu, J. Bu, C. Chen, C. Wang, D. Cai and X. He, EMR, "A scalable graph-based ranking model for content-based image retrieval," *IEEE Transactions on knowledge and data engineering*, vol. 27, no. 1, pp. 102-114, 2013.
- [6] H. X. Trung, D. V. Tuyet, N. H. Huy, S. Ablameyko, N. Q. Cuong and H. V. Quy, "A Novel Non-Gaussian Feature Normalization Method and its Application in Content Based Image Retrieval," *Nonlinear Phenomena in Complex Systems*, vol. 22, no. 1, pp. 1-17, 2019.
- [7] Q. H. Van, H. N. Hoang, T. D. Van, S. Ablameyko, H. T. Van, D. P. T. Kim and M. Ablameyko, "A modified Efficient Manifold Ranking Algorithm for Large Database Image Retrieval," *Nonlinear Phenomena in Complex Systems*, 2020.
- [8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815-823, 2015.
- [9] Huy Tran Van, Dzong Pham Thi Kim, Huy Ngo Hoang and Quy Hoang Van, "Enhancing the performance of manifold ranking in image retrieval using combined rank on low-level features and embedded vectors," *J. Inf. Hiding Multim. Signal Process*, 11(4), 172-186 (2020).
- [10] Rassoul Hajizadeh, A. Aghagolzadeh, M. Ezoji, "Local distances preserving based manifold learning," *Expert Systems with Applications*, vol. 139, 2020.
- [11] D. e. a. Zhou, "Learning with local and global consistency," *Advances in neural information processing systems*, 16, pp. 595-602, 2003.

Phụ lục

Bổ đề 1: Mọi tập hữu hạn n véc tơ m chiều, mọi số nguyên dương $d: 1 \leq d \leq \left\lfloor \frac{m}{2} \right\rfloor$ và $M = \{x_i = (x_{i,j})_{j=1}^n\} \subset R^m (m > 1)$ thỏa mãn

$x_{i,k} \neq x_{j,k} \forall i \neq j, \forall 1 \leq k \leq d$ (*) thì tồn tại một đa tập d chiều $D \subset R^m: x_i \in D \forall i = \overline{1, n}$.

Chúng tôi xin dẫn bổ đề về tính tổng quát này do ông Hoàng Văn Quý - Khoa CNTT&TT, Trường Đại học Hồng Đức cung cấp, chúng tôi xin trân trọng cảm ơn ông Hoàng Văn Quý đã cho phép trích dẫn kết quả nghiên cứu chưa được công bố về dữ liệu đa tập.

EXPLORING SIMILARITY AMONG SIGNATURE IMAGES BASED ON THE MANIFOLD RANKING APPROACH

Tran Van Huy, Tran Cong Hung, Ngo Hoang Huy Dao Van Tuyet,
Pham Thi Kim Dzong, Nguyen Van Quyen, Le Dinh Nghiep

ABSTRACT: The Efficient Manifold Ranking (EMR) algorithm is widely used in image retrieval, typically employing low-level features such as texture and shape to represent images. Features like HOG and LBP in signature images exhibit a high discriminative capability, serving as an effective complement to the embedded features computed from the SigNet network. To enhance the discriminative power of the EMR algorithm in content-based image retrieval, this study proposes combining low-level features with embedded feature vectors from a deep learning model. Experimental tests conducted on a dataset of handwritten signatures demonstrate the effectiveness of the proposed ranking method in improving retrieval quality and measuring image similarity.

Keywords: Signature image retrieval, Manifold ranking, EMR, Contrastive loss.