

NGHIÊN CỨU THUẬT TOÁN HỌC TỪ ĐIỂN PHÂN BIỆT TRONG PHÂN LOẠI HÌNH ẢNH

A PROPOSED DISCRIMINATIVE DICTIONARY PAIR LEARNING ALGORITHM FOR IMAGE CLASSIFICATION

Nguyễn Hoàng Vũ¹, Trần Quốc Cường¹, Trần Thanh Phong¹

¹Trường Đại học Tiền Giang

nguyenhoangvu@tgu.edu.vn; tranquoccuong@tgu.vn; tranthanhphong@tgu.edu.vn

(Nhận bài: 22/10/2020; Chấp nhận đăng: 22/12/2020)

Tóm tắt - Phương pháp học từ điển dựa trên biểu diễn thưa là mô hình đã được áp dụng rộng rãi trong nhiều hệ thống thị giác máy tính với những kết quả đầy hứa hẹn. Trong bài báo này, một giải thuật học cặp từ điển phân biệt (DDPL) mới được đề xuất nhằm cải thiện hiệu quả phân loại hình ảnh. Mô hình đề xuất có khả năng huấn luyện đồng thời một từ điển phân tích và một từ điển tổng hợp với sự kết hợp của ràng buộc không mạch lạc và đại diện hạng thấp. Thuật toán đảm bảo từ điển sau khi được huấn luyện có khả năng phân biệt mạnh và tín hiệu sau mã hóa là tách biệt. So sánh với các phương pháp học từ điển trước đây, DDPL sử dụng ánh xạ mã thưa nên giảm phần lớn gánh nặng tính toán trong huấn luyện và kiểm tra. Kết quả phân loại hình ảnh trên nhiều tập dữ liệu tiêu chuẩn đã chứng minh tính hiệu quả của phương pháp đề xuất.

Từ khóa - Đại diện thưa; học từ điển; từ điển tổng hợp; từ điển phân tích; học từ điển phân biệt; phân loại hình ảnh.

1. Đặt vấn đề

Trong những năm gần đây, phương pháp biểu diễn thưa đã thu hút được nhiều sự chú ý và đã được ứng dụng thành công trong lĩnh vực thị giác máy tính [1], [2]. Biểu diễn thưa đại diện cho một mẫu bằng cách phối hợp tuyến tính với một vài nguyên tử của từ điển được chọn trong tập dữ liệu. Vì vậy từ điển có vai trò rất quan trọng trong quá trình tái cấu trúc mẫu. Trong lĩnh vực phân loại hình ảnh, việc học một từ điển tối ưu từ tập dữ liệu hình ảnh huấn luyện đã đem lại hiệu quả phân loại rất cao, do đó nhiều mô hình học từ điển đã được nghiên cứu và đề xuất [3], [4], [5].

Có hai phương pháp học từ điển là học không giám sát và học có giám sát. Đối với phương pháp học từ điển không giám sát, từ điển được huấn luyện từ một hàm mục tiêu trong đó lỗi tái cấu trúc của dữ liệu huấn luyện được tối thiểu hóa [3]. Mặc dù, từ điển này có khả năng tái cấu trúc chính xác dữ liệu huấn luyện, nhưng nó không có thông tin về lớp nhãn trong từ điển. Trong phương pháp học từ điển có giám sát, thông tin nhãn của lớp được đưa vào trong các giai đoạn huấn luyện vì vậy phương pháp học từ điển có giám sát thường được sử dụng trong phân loại hình ảnh [6], [7], [8]. Tùy thuộc vào cách mã hóa dữ liệu huấn luyện, từ điển được chia thành ba loại: Từ điển tổng hợp, từ điển phân tích và cặp từ điển phân tích - tổng hợp.

- Từ điển tổng hợp đại diện cho dữ liệu huấn luyện bằng cách phối hợp tuyến tính với các nguyên tử của từ điển. Từ điển loại này có thể là một từ điển chung được chia sẻ bởi tất cả các lớp (từ điển đại diện cho tất cả) [9] hoặc từ điển

Abstract - Dictionary learning for sparse coding has been widely applied in the field of computer vision and have achieved promising performance. In this paper, a new method called discriminative dictionary pair learning (DDPL) for image classification was proposed which jointly learned a synthesis dictionary and an analysis dictionary to promote the image classification performance. The DDPL method ensures that the learned dictionary has the powerful discriminative ability and the signals are more separable after coding. Compared with previous dictionary learning methods, DDPL employs projective coding, which largely reduces the computational burden in training and testing. Experimental results on various image classification benchmarks are presented to demonstrate the effectiveness of the proposed method.

Key words - Sparse representation; dictionary learning; synthesis dictionary; analysis dictionary; discriminative dictionary learning; image classification.

dành riêng cho mỗi lớp (mỗi nguyên tử của từ điển liên kết với một lớp) [7], [10]. Mặc dù từ điển tổng hợp đạt được kết quả phân loại rất tốt, thời gian tính toán các hệ số mã thưa lớn đã ảnh hưởng đến thời gian huấn luyện từ điển và kiểm tra dữ liệu vào.

- Từ điển phân tích trực tiếp biến đổi dữ liệu thành không gian đặc điểm thưa bằng cách nhân với từ điển huấn luyện cho việc miêu tả dữ liệu. Đại diện cho cách tiếp cận này là phương pháp học từ điển không giám sát được sử dụng trong khôi phục ảnh [11]. Mặc dù đạt được kết quả hứa hẹn nhưng phương pháp này vẫn còn tính toán phức tạp và không phù hợp cho nhiệm vụ phân loại hình ảnh.

- Cặp từ điển phân tích tổng hợp mở rộng đầy đủ hơn khả năng học từ điển bằng cách kết hợp giữa đại diện dữ liệu và giảm thời gian tính toán. Rubinstein và Elad [12] đã đề xuất mô hình học cặp từ điển phân tích tổng hợp sử dụng trong xử lý hình ảnh. Phương pháp này có khả năng đại diện dữ liệu tốt và các hệ số mã hóa đảm bảo tính thưa. Gần đây, Gu và các cộng sự [13] đã đề xuất phương pháp học cặp từ điển phân tích tổng hợp (DPL) sử dụng cho phân loại hình ảnh. Trong mô hình DPL, từ điển phân tích tổng hợp dành cho các lớp riêng có khả năng miêu tả rất tốt dữ liệu của lớp và miêu tả kém đối với các lớp khác. Mặc dù mô hình DPL đã đạt được kết quả ấn tượng trong phân loại hình ảnh nhưng DPL chưa khai thác hết khả năng phân biệt của dữ liệu huấn luyện trong quá trình học từ điển.

Trong bài báo này, nhóm tác giả tập trung cải thiện khả năng phân biệt của cặp từ điển phân tích tổng hợp và

¹ Tien Giang University (HoangVu Nguyen, Tran Quoc Cuong, Tran Thanh Phong)

đề xuất thuật toán học cặp cấp từ điển phân tích tổng hợp phân biệt (DDPL) sử dụng cho phân loại hình ảnh. Những đóng góp chính của bài báo như sau:

- Trong mô hình đề xuất, nhóm tác giả tích hợp sự phân biệt của từ điển tổng hợp với sự miêu tả của từ điển phân tích thành một mô hình thống nhất để học cặp từ điển cho mỗi lớp có tính phân biệt mạnh mẽ mà còn giảm thời gian tính toán.

- Để tăng hiệu quả phân loại hình ảnh, nhóm tác giả xem xét ràng buộc không mạch lạc trong các lớp và giữa các lớp trên từ điển tổng hợp nhằm tối thiểu sự miêu tả tương tự của những nguyên tử từ điển của lớp liên quan với các lớp khác.

- Sử dụng ràng buộc quy tắc hạng thấp trên từ điển phân tích để cải thiện sự tương tự của các hệ số mã thừa trên cùng lớp (các nguyên tử trong cùng lớp tương tự nhau).

Thuật toán được thực hiện trên tập dữ liệu hình ảnh tiêu chuẩn và so sánh với các phương pháp học từ điển khác. Kết quả thực hiện đã chứng minh tính hiệu quả của phương pháp đề xuất.

2. Tổng quan về mô hình học cặp từ điển

2.1. Mô hình học từ điển phân biệt

Xét ma trận $X = [X_1, X_2, \dots, X_C]$ là tập hợp các mẫu hình ảnh huấn luyện. Trong đó, $X_i \in \mathbb{R}^{m \times n}$ ($i = 1, \dots, C$) là mẫu của lớp thứ i , m là số chiều, n là số mẫu trong mỗi lớp, C là số lớp. Hầu hết các phương pháp học từ điển phân biệt là huấn luyện một từ điển D để miêu tả X . Trong phân loại hình ảnh, từ điển được học kết hợp với thông tin nhãn của lớp qua hàm mục tiêu:

$$\min_{D,A} \|X - DA\|_F^2 + \lambda \|A\|_p + \Psi(D, A, Y) \quad (1)$$

Trong đó $\lambda \geq 0$ là hằng số, $D = [D_1, D_2, \dots, D_C]$, ($D_i \in \mathbb{R}^{m \times p}$) là từ điển và $A = [A_1, A_2, \dots, A_C]$, ($A_i \in \mathbb{R}^{p \times n}$) là ma trận hệ số mã hóa của X qua D . Trong mô hình huấn luyện (1), ràng buộc $\|X - DA\|_F^2$ đảm bảo khả năng miêu tả của từ điển D ; $\|A\|_p$ là chuẩn p ($p \geq 1$) của ma trận A ; Y là ma trận nhãn của lớp trong X , và $\Psi(D, A, Y)$ là điều kiện ràng buộc phân biệt để đảm bảo khả năng phân biệt của D .

2.2. Mô hình học cặp từ điển phân tích - tổng hợp trong phân loại hình ảnh

Gần đây, Gu và các cộng sự [13] đã mở rộng phương trình (1) thành mô hình học cặp từ điển phân tích - tổng hợp P và D (mô hình DPL), với ma trận hệ số mã hóa $A = PX$. Mô hình DPL được nghĩa như sau:

$$\min_{P,D} \sum_{i=1}^C \|X_i - D_i P_i X_i\|_F^2 + \lambda \|P_i \bar{X}_i\|_F^2 \quad s.t. \|d_j\|_2^2 \leq 1 \quad (2)$$

Trong đó, $\bar{X}_i$ là ký hiệu ma trận bù của X_i trong tập dữ liệu X ; $D = [D_1; D_2; \dots; D_C]$ là từ điển tổng hợp với $D_i \in \mathbb{R}^{m \times p}$ là từ điển phụ thứ i trong D , $P = [P_1; P_2; \dots; P_C]$ là từ điển phân tích với $P_i \in \mathbb{R}^{p \times m}$ là từ điển phụ thứ i trong P . Ma trận D_i và P_i được sử dụng để phân lớp. Với một mẫu kiểm tra y , nhãn của y được xác định bởi:

$$\text{identity}(y) = \arg \min_i \|y - D_i P_i y\|_2 \quad (3)$$

Mô hình DPL đã đạt được kết quả nhận dạng rất tốt trong phân loại hình ảnh với thời gian tính toán thấp nhưng mô hình vẫn chưa khai thác hết khả năng phân biệt của dữ liệu huấn luyện. Nhiều nghiên cứu đã dựa trên mô hình DPL và đề xuất các thuật toán học từ điển bằng cách tích hợp các ràng buộc

để tăng khả năng phân biệt của từ điển. Chen và các cộng sự [14] đề xuất mô hình học cặp từ điển phân biệt bằng cách bổ sung ràng buộc véc tơ hỗ trợ phân biệt vào mô hình DPL:

$$\begin{aligned} \min_{P,D,A,W,b} \sum_{i=1}^C \|X_i - D_i A_i\|_F^2 + \tau \|A_i - P_i X_i\|_F^2 + \gamma \|P_i\|_F^2 \\ + \lambda_1 \|P_i [\bar{X}_i, X_i - M_i]\|_F^2 \\ + \sum_{i=1}^C 2\lambda_2 L(A, y^c, w_c, b_c) \\ s.t. \|d_j\|_2^2 \leq 1. \end{aligned} \quad (4)$$

Trong đó, M_i là ma trận với các véc tơ cột là trung bình của các véc tơ cột của X_i , $L(A, y^c, w_c, b_c)$ là hàm véc tơ hỗ trợ phân biệt. Nhãn của mẫu thử y được xác định như sau:

$$\begin{aligned} \text{identity}(y) = \arg \min_i \|y - D_i P_i y\|_2 + \eta_1 \|\bar{P}_i y\|_2 \\ - \eta_2 (w_i^T P_i + b_i) \end{aligned} \quad (5)$$

Zhang và các cộng sự [15] đã đề xuất mô hình học từ điển phân tích (ADDL) bằng cách tích hợp từ điển phân tích phân biệt và huấn luyện ma trận phân lớp trên cùng mô hình. Hàm mục tiêu của mô hình ADDL như sau:

$$\begin{aligned} \langle D, S, P, W \rangle = \arg \min_{D,S,P,W} \sum_{l=1}^C \|X_l - D_l S_l\|_F^2 + \alpha f(D_l) \\ + \tau r(P_l, S_l) + \lambda g(H_l, W_l, P_l) \\ s.t. \|d_v\|_2^2 \leq 1, \forall v \in \{1, \dots, K\} \end{aligned} \quad (6)$$

Trong đó, $f(D_l)$ là ràng buộc không mạch lạc, $r(P_l, S_l) = \|P_l X_l - S_l\|_F^2 + \|P_l \bar{X}_l\|_F^2 + \|S_l\|_{2,1}$ là ràng buộc mã thừa trên từ điển phân tích và $g(H_l, W_l, P_l) = \|H_l - W_l P_l X_l\|_F^2 + \|W_l P_l \bar{X}_l\|_F^2$ là hàm huấn luyện để phân loại với H_l là lớp nhãn của X_l , W_l là ma trận phân loại tuyến tính. Sau khi đạt được từ điển phân tích và ma trận phân loại, nếu một mẫu kiểm tra x_{new} thuộc lớp thứ i thì nhãn của x_{new} được xác định bởi:

$$\text{identity}(x_{new}) = \arg \max_i (W P x_{new}) \quad (7)$$

Trong các mô hình học từ điển phân biệt, tính độc lập của các nguyên tử là rất quan trọng, nó góp phần làm gia tăng khả năng phân biệt của từ điển. Một ràng buộc không mạch lạc được định nghĩa để đo lường mối quan hệ giữa các nguyên tử của từ điển [16]:

$$\text{cor}(D) = \|D^T D - I\|_F^2 \quad (8)$$

Trong đó, I là ma trận đơn vị. Từ điển D được gọi là không mạch lạc nếu quan hệ này bằng không. Việc tối thiểu ràng buộc này đảm bảo cho từ điển miêu tả chính xác dữ liệu huấn luyện và đạt được hiệu quả phân loại cao.

3. Mô hình học cặp từ điển phân biệt đề xuất

3.1. Hàm mục tiêu

Cặp từ điển phân tích - tổng hợp được sử dụng để phân loại vì vậy chúng phải có khả năng phân biệt cao. Để tăng khả năng phân biệt của từ điển tổng hợp D , sử dụng ràng buộc không mạch lạc trên các từ điển phụ D_i để tối thiểu mối quan hệ giữa các nguyên tử của D_i . Với ràng buộc này, mỗi từ điển của mỗi lớp chỉ mã hóa tốt nhất mẫu của lớp đó. Ngoài ra, để đảm bảo các hệ số mã hóa của mỗi lớp là tương tự nhau, từ điển phụ P_i được ràng buộc hạng thấp. Từ những phân tích trên, hàm mục tiêu của mô hình đề xuất được thiết kế như sau:

$$\begin{aligned} \min_{P,D} \sum_{i=1}^C \|X_i - D_i P_i X_i\|_F^2 + \lambda \|P_i \bar{X}_i\|_F^2 + \mu \|P_i\|_* + \\ + \eta_1 \sum_{i \neq j}^C \|D_i^T D_j\|_F^2 + \eta_2 \|D_i^T D_i - I\|_F^2 \\ \text{s. t. } \|d_j\|_2^2 \leq 1 \end{aligned} \quad (9)$$

Trong đó, $\mu \geq 0, \eta_1 \geq 0, \eta_2 \geq 0$ là hằng số; $\|P_i\|_*$ là ràng buộc hạng thấp, ký hiệu $\|\cdot\|_*$ chuẩn Schatten 1 (nuclear norm) của ma trận; $\sum_{i \neq j}^C \|D_i^T D_j\|_F^2$ là ràng buộc không mạch lạc để đảm bảo các từ điển phụ trong các lớp độc lập nhau (tức là $D_i^T D_j \approx 0, \forall i \neq j$); ràng buộc $\|D_i^T D_i - I\|_F^2$ làm ổn định từ điển phụ trong mỗi lớp; $I \in \mathbb{R}^{p \times p}$ là ma trận đơn vị.

3.2. Giải hàm mục tiêu

Hàm mục tiêu (9) là hàm không lồi, nhóm tác giả sử dụng một biến A để chuyển đổi phương trình (9) thành:

$$\begin{aligned} \{A^*, P^*, D^*\} = \arg \min_{A,D,P} \sum_{i=1}^C \|X_i - D_i A_i\|_F^2 + \tau \|P_i X_i - A_i\|_F^2 + \lambda \|P_i \bar{X}_i\|_F^2 + \mu \|P_i\|_* + \\ + \eta_1 \sum_{i \neq j}^C \|D_i^T D_j\|_F^2 + \eta_2 \|D_i^T D_i - I\|_F^2 \quad \text{s. t. } \|d_j\|_2^2 \leq 1 \end{aligned} \quad (10)$$

Trong đó, $\{A^*, P^*, D^*\}$ được tối ưu bằng cách sử dụng vòng lặp và luân phiên cập nhật A và $\{D, P\}$ theo hai bước sau:

• Cố định D và P , cập nhật A

Khi D và P cố định, hàm mục tiêu liên quan đến biến A được viết lại là:

$$A^* = \arg \min_A \sum_{i=1}^C \|X_i - D_i A_i\|_F^2 + \tau \|P_i X_i - A_i\|_F^2 \quad (11)$$

Đây là dạng bình phương tối thiểu, vì vậy có thể có thể tính được A_i^* :

$$A_i^* = (D_i^T D_i + \tau I)^{-1} (\tau P_i X_i + D_i^T X_i) \quad (12)$$

• Cố định A , cập nhật P và D

+ Cập nhật P

Phương trình (10) được viết lại như sau:

$$\begin{aligned} P^* = \arg \min_P \sum_{i=1}^C \tau \|P_i X_i - A_i\|_F^2 + \lambda \|P_i \bar{X}_i\|_F^2 \\ + \mu \|P_i\|_* \end{aligned} \quad (13)$$

Phương trình (13) được chuyển đổi thành dạng sau bằng cách thêm biến Z :

$$\{P^*, Z^*\} = \arg \min_{P,Z} \sum_{i=1}^C f(P_i) + \mu \|Z\|_* \quad \text{s. t. } P_i = Z \quad (14)$$

Với $f(P_i) = \tau \|P_i X_i - A_i\|_F^2 + \lambda \|P_i \bar{X}_i\|_F^2$.

Sử dụng phương pháp ALM (Augmented Lagrange Multiplier), phương trình (14) được xác định bằng cách giải phương trình sau:

$$\min_{P_i, Z} f(P_i) + \mu \|Z\|_* + \langle T_1, P_i - Z \rangle + \frac{\varepsilon}{2} \|P_i - Z\|_F^2 \quad (15)$$

Trong đó $\varepsilon > 0$ và T_1 là nhân tử Lagrange.

Phương trình (15) có thể giải được bằng thuật toán ADMM (Alternating Direction Method of Multipliers) [17]:

$$\begin{cases} P_i^* = \arg \min_{P_i} \frac{1}{\varepsilon} f(P_i) + \frac{1}{2} \left\| P_i - Z + \frac{T_1}{\varepsilon} \right\|_F^2 \\ Z^* = \arg \min_Z \frac{\mu}{\varepsilon} \|Z\|_* + \frac{1}{2} \left\| Z - P_i + \frac{T_1}{\varepsilon} \right\|_F^2 \\ T_1 = T_1 + \varepsilon (P_i - Z) \end{cases} \quad (16)$$

Từ phương trình (16), có thể tính được P_i^* :

$$P_i^* = \left(\left(\frac{\tau}{\varepsilon} + \lambda \right) X X^T + \frac{I}{2} \right)^{-1} \left(A X^T - Z^* - \frac{T_1}{\varepsilon} \right) \quad (17)$$

Trong đó, Z^* được giải bằng giải thuật SVT (Singular Value Thresholding) [18]:

$$Z^* = US_{\frac{\mu}{\varepsilon}}[\Sigma]V^T \quad (18)$$

Với $(USV^T) = \text{svd}(P - \frac{T_1}{\varepsilon})$ và $S_{\varepsilon}[\cdot]$ là toán tử soft-thresholding (shrink-age), $S_{\varepsilon}[x] = \text{sign}(x)(|x| - \varepsilon)$.

+ Cập nhật D

Phương trình (10) được viết lại:

$$\begin{aligned} D^* = \arg \min_D \sum_{i=1}^C \|X_i - D_i A_i\|_F^2 + \eta_1 \sum_{i \neq j}^C \|D_j^T D_i\|_F^2 + \\ \eta_2 \|D_j^T D_i - I\|_F^2 \quad \text{s. t. } \|d_j\|_2^2 \leq 1. \end{aligned} \quad (19)$$

Biến đổi phương trình (19) thành dạng sau bằng cách thêm biến T :

$$\begin{aligned} \min_{D,T} \sum_{i=1}^C \|X_i - D_i A_i\|_F^2 + \eta_1 \sum_{i \neq j}^C \|D_j^T D_i\|_F^2 \\ + \eta_2 \|D_j^T D_i - I\|_F^2 \quad \text{s. t. } D = T, \|t_i\|_2^2 \leq 1 \end{aligned} \quad (20)$$

Phương trình (20) được giải bằng cách sử dụng thuật toán ADMM [17]:

$$\begin{cases} D_i^{k+1} = \arg \min_{D_i} \|X_i - D_i A_i\|_F^2 + \eta_1 \|D_j^T D_i\|_F^2 \\ \quad + \eta_2 \|D_j^T D_i - I\|_F^2 + \rho \|D_i - T_i^k + S_i^k\|_F^2 \\ T_i^{k+1} = \arg \min_{T_i} \rho \|D_i^{k+1} - T_i^k + S_i^k\|_F^2 \quad \text{s. t. } \|t_j\|_2^2 \leq 1 \\ S_i^{k+1} = S_i^k + D_i^{k+1} - T_i^{k+1} \end{cases} \quad (21)$$

Trong đó, k là số vòng lặp và ρ ($0 < \rho < 1$) là đại lượng vô hướng tăng dần để $\rho_{rate} \geq 1$. Từ phương trình (21), có thể tính được D_i^* :

$$D_i^* = (A A^T + I + D_j^T D_j)^{-1} (A^T X + T_i - S_i) \quad (22)$$

Với $T_i = S_i + D_i$.

Thuật toán tổng quát của mô hình DDPL được tóm tắt trong Thuật toán 1. Thuật toán sẽ dừng khi năng lượng của 2 vòng lặp kế cận nhỏ hơn 0.01 hoặc thỏa mãn số vòng lặp giới hạn.

3.3. Phân loại

Sau khi đạt được cập từ điển phân tích - tổng hợp (D^*, P^*), việc xác định nhân của mẫu kiểm tra y được thực hiện như sau: Nếu mẫu kiểm tra y thuộc lớp thứ i thì giá trị $\|y - D_i^* P_i^* y\|_2^2$ là nhỏ nhất.

$$\text{identity}(y) = \arg \min_i \|y - D_i^* P_i^* y\|_2^2 \quad (23)$$

Algorithm1:

Input: Dữ liệu huấn luyện $X = [X_1, X_2, \dots, X_C]$, các thông số $\lambda, \tau, \mu, \eta_1, \eta_2$;

1: Khởi tạo D^0 và P^0 là các ma trận ngẫu nhiên chuẩn Frobenious, $t=0$;

2: while not converge do

3: $t \leftarrow t + 1$;

4: for $i = 1:C$ do

5: cập nhật A_i bởi phương trình (12);

6: cập nhật P_i bởi phương trình (17);

7: cập nhật D_i bởi phương trình (22);

8: end for

9: end while

Output: Từ điển phân tích P , Từ điển tổng hợp D .

4. Kết quả thực nghiệm

Trong phần này, kết quả nhận dạng của thuật toán đề xuất DDPL được đánh giá trên ba tập dữ liệu hình ảnh: Extended YaleB [19], AR [20] và Caltech101 [21]. Thuật toán DDPL được so sánh với một số thuật toán học từ điển bao gồm: Phân loại hình ảnh dựa trên biểu diễn thưa (SRC) [1], học từ điển phân biệt K-SVD (DKSVD) [9], học từ điển phân biệt Fisher cho biểu diễn thưa (FDDL) [22], học từ điển với cấu trúc mạch lạc (DLSI) [16], Lable Consistent K-SVD (LC-KSVD) [7], LC-PDL [23], PCANet [24] và DPL [13]. Đối với các thuật toán SRC, DKSVD và DLSI nhóm tác giả tự thực hiện thí nghiệm dựa theo giải thuật công bố. Đối với các thuật toán còn lại, nhóm tác giả sử dụng trực tiếp các mã nguồn đã được tác giả công bố. Tất cả các thuật toán đều được lập trình trên phần mềm Matlab2014b.

4.1. Các tập dữ liệu tiêu chuẩn

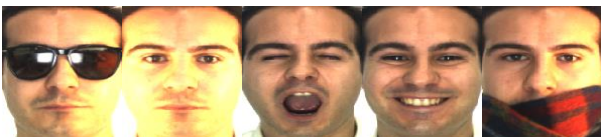
Tập dữ liệu khuôn mặt Extended Yale B [19] chứa 2414 hình ảnh của 38 người, hình ảnh của mỗi người được chụp trong 64 điều kiện ánh sáng được kiểm soát khác nhau. Các hình ảnh đều ở tư thế chính diện và được cắt theo vùng mặt thực tế. Tất cả các hình ảnh có độ phân giải 192×168 pixel (Hình 1a).

Tập dữ liệu khuôn mặt AR [20] bao gồm hơn 4000 hình ảnh trực diện từ 126 cá nhân. Đối với mỗi cá nhân, 26 hình ảnh được chụp trong hai nhóm riêng biệt, gồm nhiều biến thể như thay đổi ánh sáng, biểu cảm và nguy trang trên khuôn mặt (Hình 1b).

Tập dữ liệu hình ảnh Caltech 101 [21] chứa 9144 ảnh gồm 102 lớp. Số lượng hình ảnh trong mỗi lớp thay đổi từ 31 đến 800 (Hình 1c).



(a) Extended YaleB



(b) AR



(c) Caltech 101

Hình 1. Các tập dữ liệu hình ảnh

4.2. Cài đặt thông số

Bảng 1. Ảnh hưởng của η_1 và η_2 đến độ chính xác phân loại khi $\mu = 0,001$

η_1	0,001	0,01	0,05	0,1	0,15	0,2
η_2	0,00001	0,0001	0,001	0,01	0,1	0,15
Tỉ lệ (%)	97,7	97,9	98,1	95,6	91,5	80,9

Bảng 2. Ảnh hưởng của μ đến độ chính xác phân loại khi $\eta_1 = 0,05$, $\eta_2 = 0,001$

μ	0,00001	0,0001	0,001	0,005	0,01	0,1
Tỉ lệ (%)	97,4	97,8	98,1	96,9	94,6	90,5

Trong hàm mục tiêu (10), có tất cả 5 thông số được xác định như sau: Trong thực nghiệm, với số nguyên tử từ điển bằng số ảnh huấn luyện, giá trị τ và λ không thay đổi nhiều trong tất cả các thí nghiệm, vì vậy τ và λ được cố định là 0,05 và $3e-3$. Đối với tập dữ liệu Extended Yale B, khi cố định $\mu = 0,001$, giá trị η_1 và η_2 tương ứng được thể hiện trên Bảng 1. Dựa theo Bảng 1, độ chính xác phân loại cao nhất đạt được khi $\eta_1 = 0,05$ và $\eta_2 = 0,001$. Bảng 2 thể hiện độ chính xác phân loại khi cố định $\eta_1 = 0,05$ và $\eta_2 = 0,001$. Dựa vào bảng 1 và Bảng 2, các thông số được chọn cho tập dữ liệu Extended Yale B như sau: $\mu = 0,001$, $\eta_1 = 0,05$ và $\eta_2 = 0,001$.

Tương tự, đối với tập dữ liệu AR: $\mu = 0,005$, $\eta_1 = 0,03$, $\eta_2 = 0,001$ và đối với tập dữ liệu Caltech101: $\mu = 0,003$, $\eta_1 = 0,001$, $\eta_2 = 0,01$.

4.3. Nhận dạng khuôn mặt

Đối với tập dữ liệu khuôn mặt Extended Yale B, phân nửa số ảnh trong mỗi lớp (32 ảnh) được chọn ngẫu nhiên dùng cho dữ liệu huấn luyện, nửa còn lại sử dụng cho tập kiểm tra. Mỗi ảnh được ánh xạ thành véc tơ 504 chiều được cung cấp bởi [7]. Từ điển được huấn luyện chứa 570 nguyên tử tương ứng mỗi lớp gồm 15 nguyên tử.

Đối với tập dữ liệu khuôn mặt AR, sử dụng một tập hợp 2600 ảnh của 50 nam và 50 nữ, 20 ảnh được chọn ngẫu nhiên trong mỗi lớp cho tập huấn luyện, số còn lại dùng cho kiểm tra. Số chiều mỗi ảnh là 540. Từ điển được huấn luyện có 500 nguyên tử tương ứng mỗi lớp có 5 nguyên tử.

Kết quả nhận dạng khuôn mặt của mô hình DDPL so sánh với các mô hình khác trên tập dữ liệu Extended Yale B và tập dữ liệu AR được thể hiện trên Bảng 3 và Bảng 4. Từ các kết quả thực nghiệm cho thấy, độ chính xác nhận dạng đạt được của DDPL cao hơn các phương pháp khác từ 0,6%-4% trên tập dữ liệu Extended Yale B, và từ 1%-10,1% trên tập dữ liệu AR. Đặc biệt, mô hình đề xuất DDPL vẫn đạt được tỉ lệ nhận dạng cao hơn khi so sánh với mô hình phân loại hình ảnh sử dụng mạng học sâu PCANet.

Bảng 3. Kết quả nhận dạng (%) trên tập dữ liệu Extended YaleB

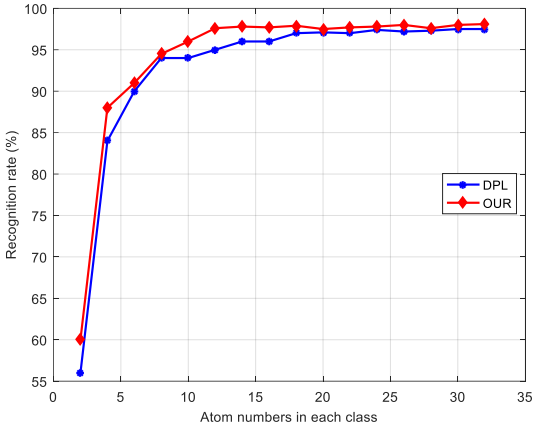
Phương pháp	Độ chính xác	Phương pháp	Độ chính xác
SRC	96,5	DPL	97,5
DKSVD	94,1	LC-DPL	97,8
LC-KSVD	96,7	PCANet	96,9
DLSI	96,5	DDPL	98,1
FDDL	96,7		

Bảng 4. Kết quả nhận dạng (%) trên tập dữ liệu AR

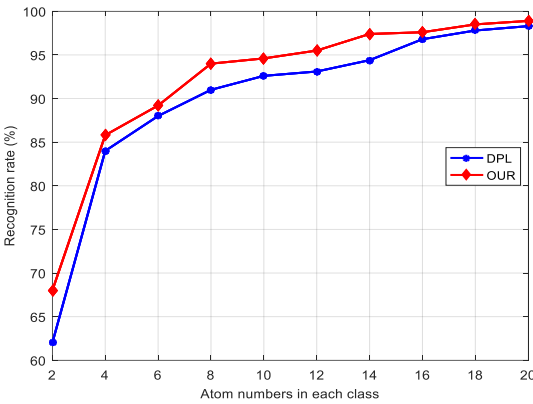
Phương pháp	Độ chính xác	Phương pháp	Độ chính xác
SRC	97,5	DPL	98,3
DKSVD	88,8	LC-DPL	98,6
LC-KSVD	97,8	PCANet	98,0
DLSI	97,5	DDPL	98,9
FDDL	97,5		

Để đánh giá hiệu quả nhận dạng dựa trên kích thước từ điển, thuật toán DDPL được so sánh với DPL. Với số mẫu huấn luyện cố định, giá trị kích thước từ điển thay đổi từ 2 đến 32 đối với tập dữ liệu Extended Yale B và kích thước

từ điển thay đổi từ 2 đến 20 đối với tập dữ liệu AR. Kết quả so sánh được thể hiện trên Hình 2 cho thấy, mô hình DDPL vẫn đạt được tỉ lệ nhận dạng cao hơn DPL.



(a) Extended YaleB



(b) AR

Hình 2. Tỉ lệ nhận dạng (%) với số nguyên tử từ điển

4.4. Nhận dạng đối tượng

Theo cài đặt thử nghiệm từ [7], đối với tập dữ liệu hình ảnh Caltech101, 30 mẫu trong mỗi lớp được sử dụng để huấn luyện và phần còn lại được sử dụng để kiểm tra. Kích thước từ điển là 1020, tương ứng với mỗi lớp từ điển chứa 10 phần tử.

Bảng 5. Độ chính xác nhận dạng trên tập dữ liệu Caltech101

Phương pháp	Độ chính xác
SRC	70,7
DKSVD	71,2
LC-KSVD	73,6
DLSI	73,1
FDDL	73,2
DPL	73,9
LC-PDL	74,1
PCANet	74,8
DDPL	75,6

Kết quả nhận dạng đối tượng sử dụng tập dữ liệu Caltech101 được thể hiện trên Bảng 5. Một lần nữa, có thể thấy tỉ lệ nhận dạng của mô hình DDPL cao hơn các mô hình khác.

4.5. So sánh thời gian tính toán

Thời gian huấn luyện và thử nghiệm của các phương pháp FDDL, DPL, LC-PDL và DDPL trên các tập dữ liệu hình ảnh sử dụng phần mềm Matlab2014b được trình bày trên Bảng 6.

Bảng 6. Thời gian tính toán (s) trên tập huấn luyện (Train) và kiểm tra (Test)

Phương pháp	YaleB		AR		Caltech101	
	Train	Test	Train	Test	Train	Test
FDDL	4501	120	15000	574	63111	2615
DPL	3,0	0,15	7,5	0,29	90,8	9,75
LC-PDL	2,5	0,1	4,8	0,16	56,5	6,8
DDPL	8,03	0,15	31,2	0,29	405	9,75

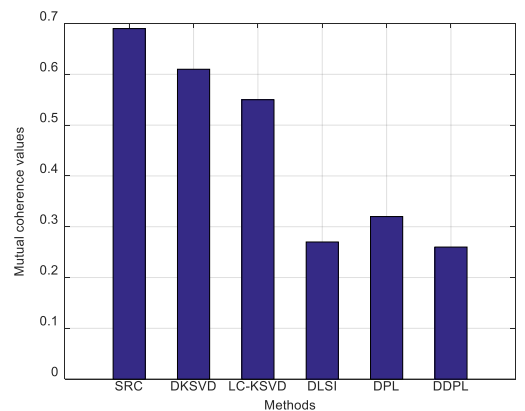
Từ Bảng 6 có thể thấy, các phương pháp học cặp từ điển (DPL, LC-PDL và DDPL) có thời gian nhận dạng trung bình thấp hơn rất nhiều so với phương pháp học từ điển tổng hợp FDDL. Nguyên nhân là do phương pháp học từ điển tổng hợp FDDL mất thời gian tính toán các hệ số mã thừa trong khi học cặp từ điển phân tích - tổng hợp có thể tái cấu trúc tín hiệu với hệ số mã thừa bằng cách ánh xạ tuyến tính. Mặc dù thời gian huấn luyện của phương pháp đề xuất DDPL cao hơn một ít so với phương pháp DPL và LC-PDL do có thêm vòng lặp trong lúc cập nhật từ điển phân tích. Tuy nhiên, kết quả nhận dạng cho thấy tỉ lệ nhận dạng của phương pháp đề xuất vẫn cao hơn các phương pháp khác, điều này chứng tỏ phương pháp đề xuất là rất hiệu quả.

4.6. Sự tương quan giữa các lớp trong từ điển

Giá trị tương quan giữa các lớp trong từ điển được mô tả là mức độ giống nhau tối đa của các nguyên tử trong từ điển hoàn chỉnh được đo bằng biểu thức:

$$\mu(D) = \max_{d_i \in D_i, d_j \in D_j, i \neq j} \left\| \frac{d_i}{\|d_i\|_2}, \frac{d_j}{\|d_j\|_2} \right\| \quad (24)$$

Trong đó, d_i và d_j là từ điển phụ của lớp thứ i và lớp thứ j .



Hình 3. Giá trị tương quan giữa các lớp trong từ điển

Giá trị tương quan giữa các lớp trong từ điển tổng hợp so sánh với các phương pháp khác được thể hiện trên Hình 3. Có thể thấy rằng, giá trị của mô hình DDPL thấp hơn các mô hình khác, điều này chứng tỏ các từ điển phụ trong mô hình DDPL là độc lập và có tính phân biệt cao. Đây là đặc điểm chứng tỏ mô hình DDPL rất phù hợp trong phân loại hình ảnh.

5. Kết luận

Bài báo đã giới thiệu một thuật toán mới Học cặp từ điển phân biệt sử dụng cho phân loại hình ảnh. Với thiết kế các điều kiện ràng buộc mạch lạc và quy tắc hạng thấp, thuật toán đã cải thiện khả năng miêu tả và độ phân biệt của từ điển so với các mô hình truyền thống. Thuật toán đã học được cặp từ điển sử dụng cho phân loại hình ảnh rất hiệu quả với từ điển tổng hợp có độ tương quan giữa các lớp thấp và từ điển phân tích có hệ số mã hóa trên cùng lớp là như nhau. Kết quả thực nghiệm cho thấy tính vượt trội của phương pháp đề xuất.

TÀI LIỆU THAM KHẢO

- [1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, và Y. Ma, “Robust face recognition via sparse representation”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009, doi: 10.1109/TPAMI.2008.79.
- [2] Y. Xu, D. Zhang, J. Yang, và J. Y. Yang, “A two-phase test sample sparse representation method for use with face recognition”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1255–1262, Sep. 2011, doi: 10.1109/TCSVT.2011.2138790.
- [3] M. Aharon, M. Elad, và A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation”, *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006, doi: 10.1109/TSP.2006.881199.
- [4] M. Elad và M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries”, *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006, doi: 10.1109/TIP.2006.881969.
- [5] X. Y. Jing, F. Wu, X. Zhu, X. Dong, F. Ma, và Z. Li, “Multi-spectral low-rank structured dictionary learning for face recognition”, *Pattern Recognit.*, vol. 59, pp. 14–25, Nov. 2016, doi: 10.1016/j.patcog.2016.01.023.
- [6] Q. Zhang và B. Li, “Discriminative K-SVD for dictionary learning in face recognition”, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2691–2698, 2010, doi: 10.1109/CVPR.2010.5539989.
- [7] Z. Jiang, Z. Lin, và L. S. Davis, “Label consistent K-SVD: Learning a discriminative dictionary for recognition”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, 2013, doi: 10.1109/TPAMI.2013.88.
- [8] I. Ramirez, P. Sprechmann, và G. Sapiro, “Classification and clustering via dictionary learning with structured incoherence and shared features”, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3501–3508, doi: 10.1109/CVPR.2010.5539964.
- [9] Q. Zhang và B. Li, “Discriminative K-SVD for dictionary learning in face recognition”, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2691–2698, doi: 10.1109/CVPR.2010.5539989.
- [10] H. Nguyen, W. Yang, B. Sheng, và C. Sun, “Discriminative low-rank dictionary learning for face recognition”, *Neurocomputing*, vol. 173, pp. 541–551, Jan. 2016, doi: 10.1016/j.neucom.2015.07.031.
- [11] R. Rubinstein, T. Peleg, M. E.-I. T. on Signal, and undefined 2012, “Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model”, *ieeexplore.ieee.org*.
- [12] R. Rubinstein và M. Elad, “Dictionary Learning for Analysis-Synthesis Thresholding”, *IEEE Trans. SIGNAL Process.*, vol. 62, no. 22, 2014, doi: 10.1109/TSP.2014.2360157.
- [13] S. Gu, L. Zhang, W. Zuo, và X. Feng, “Projective dictionary pair learning for pattern classification”, *Adv. Neural Inf. Process. Syst.*, vol. 1, no. January, pp. 793–801, 2014.
- [14] B. Chen, J. Li, B. Ma, và G. Wei, “Discriminative dictionary pair learning based on differentiable support vector function for visual recognition”, *Neurocomputing*, vol. 272, pp. 306–313, 2018, doi: 10.1016/j.neucom.2017.07.003.
- [15] Z. Zhang *et al.*, “Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 8, pp. 3798–3814, Aug. 2018, doi: 10.1109/TNNLS.2017.2740224.
- [16] T. Lin, S. Liu, and H. Zha, “Incoherent dictionary learning for sparse representation,” in *Proceedings - International Conference on Pattern Recognition*, 2012, pp. 1237–1240.
- [17] S. Boyd *et al.*, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” *Found. Trends R Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010, doi: 10.1561/22000000016.
- [18] J. F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010, doi: 10.1137/080738970.
- [19] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and pose,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001, doi: 10.1109/34.927464.
- [20] A. Mart Nez and R. Benavente, “The AR Face Database,” *CVC Tech. Rep.*, 1998.
- [21] L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories,” *Comput. Vis. Image Underst.*, vol. 106, no. 1, pp. 59–70, Apr. 2007, doi: 10.1016/j.cviu.2005.09.012.
- [22] M. Yang, L. Zhang, X. Feng, and D. Zhang, “Fisher Discrimination Dictionary Learning for sparse representation,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 543–550, 2011, doi: 10.1109/ICCV.2011.6126286.
- [23] Z. Zhang, W. Jiang, Z. Zhang, S. Li, G. Liu, and J. Qin, “Scalable Block-Diagonal Locality-Constrained Projective Dictionary Learning,” *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2019-August, pp. 4376–4382, May 2019.
- [24] T. H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, “PCANet: A Simple Deep Learning Baseline for Image Classification?,” *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015, doi: 10.1109/TIP.2015.2475625.