



PHÂN LỚP ẢNH DỰA TRÊN TỔ HỢP ĐA ĐẶC TRƯNG

Trần Sơn Hải^{1*}, Lê Hoàng Thái², Nguyễn Thanh Thủy³

¹ Khoa Công nghệ Thông tin – Trường Đại học Sư phạm Thành phố Hồ Chí Minh

² Trường Đại học Khoa học Tự nhiên – ĐHQG TP HCM

³ Trường Đại học Công nghệ Hà Nội

Ngày nhận bài: 22-11-2018, ngày nhận bài sửa: 05-12-2018, ngày duyệt đăng: 21-12-2018

TÓM TẮT

Phân lớp ảnh là một bài toán quan trọng và có nhiều ứng dụng trong lĩnh vực thị giác máy tính. Bài toán này đã và đang được rất nhiều nhà nghiên cứu trong và ngoài nước quan tâm. Nghiên cứu này đề xuất mô hình đa tầng cho phân lớp ảnh tổ hợp nhiều đặc trưng. Mô hình đa tầng đề xuất gồm: tầng phân lớp đơn, module chuyển đổi, và tầng tích hợp. Mô hình này có thể áp dụng cho tổ hợp đặc trưng thủ công và đặc trưng tự động.

Việc so sánh kết quả thực nghiệm của mô hình đề xuất với các mô hình đã có được thực hiện trên bài toán thực tế cho phát hiện quảng cáo và phát hiện bất thường trong ảnh X-ray phổi. Kết quả thử nghiệm cho thấy tính khả thi của mô hình đề xuất.

Từ khóa: phân lớp ảnh, tổ hợp đa đặc trưng, đặc trưng tự động, đặc trưng thủ công.

ABSTRACT

Image Classification Based On Multiple Feature Combination

Image classification is a significant problem with many applications in the field of computer vision. This problem has been receiving much attention from both domestic and foreign researchers. This paper proposes the multi-layered model for image classification based on multiple features. The proposed multi-layered model includes: Single Layer, Transform Modular, and Integral layers. The proposed model can be applied to hand-crafted features and learning features.

Comparison of the experimental results of the proposed model with the existing models was made on the practical problem for detecting advertisements and detecting abnormalities in lung X-ray images. The experimental results show the feasibility of the proposed model.

Keywords: Image classification, multiple features, hand-crafted features, learning features.

1. Giới thiệu

Bài toán phân lớp ảnh là lĩnh vực nghiên cứu được sự quan tâm của các nhà khoa học trong và ngoài nước. Đây là bài toán quan trọng có nhiều ứng dụng trong thị giác máy tính. Các hướng tiếp cận phổ biến hiện nay cho việc giải quyết bài toán phân lớp ảnh là: sử dụng *KNN* và *K-Mean* với các độ đo khác nhau, sử dụng *Support Vector Machine (SVM)* và mạng Nơron nhân tạo (*ANN*), hay dùng các mô hình học sâu như *CNN*, *AlexNet*, *VGG*, *Google Inception*...

Bài toán phân lớp ảnh có thể xem là bài toán tìm một ánh xạ từ không gian biểu diễn ảnh vào một trong L lớp cho trước. Trong giai đoạn phân lớp, *Convolutional Neural*

* Email: haits@hcmue.edu.vn

Network (CNNs) [1] thường dùng của trong các hệ thống xử lý ảnh lớn như Facebook, Google hay Amazon. Ngoài ra, có một số mô hình kết hợp nhiều bộ phân lớp nhằm nâng cao chính xác.

Bài toán phân lớp ảnh vào L lớp cho trước: Cho không gian ảnh $\mathbb{I} = \{I_i\}$, với $I_i \in \mathbb{M}_{W \times H}(\mathbb{R}^D)$. Trong đó, $\mathbb{M}_{W \times H}(\mathbb{R}^D)$ là không gian ma trận với kích thước $W \times H$ của ảnh. Mỗi phần tử của ma trận là giá trị màu thuộc \mathbb{R}^D , nếu dùng không gian RGB thì $D = 3$. Cho không gian đặc trưng $\mathbb{F} = \{F_i\}$ với $F_i \in \mathbb{R}^n$. Trong đó, n là số chiều của vector đặc trưng. Cho không gian kết quả $\mathbb{O} = \{O_i\}$ với $O_i \in [0, 1]^L$.

Ánh xạ trích chọn đặc trưng: e (đầu vào): $\mathbb{I} \rightarrow \mathbb{F}$.

Ánh xạ phân lớp: $h(F_i): \mathbb{F} \rightarrow \mathbb{O}$.

Phân lớp ảnh là việc xác định ảnh đầu vào thuộc lớp nào một cách chính xác nhất, nghĩa là cần xác định ánh xạ e và h thích hợp. Các tham số của e và h được xác định thông qua quá trình huấn luyện nhằm tối ưu hàm lỗi trên tập huấn luyện.

Trong phạm vi nghiên cứu này, nhóm đề xuất mô hình phân lớp đa tầng tổ hợp nhiều đặc trưng của ảnh đầu vào. Mô hình phản ánh được mối quan hệ giữa các đặc trưng thay vì chỉ ghép nhiều vector đặc trưng với nhau. Ngoài ra, mô hình cần có tính uyển chuyển cao để có khả năng áp dụng vào nhiều bài toán phân lớp ảnh trong thực tế. Đồng thời, áp dụng cho cả đặc trưng thủ công cũng như đặc trưng tự động của ảnh đầu vào. Trong đó, khái niệm đặc trưng thủ công và đặc trưng tự động được xác định như sau:

Khái niệm 1: (*Đặc trưng thủ công*) [2], [3]

Đặc trưng thủ công (handcrafted/engineered features) là các đặc trưng được tạo ra dựa trên cơ sở quan sát các đặc thù riêng của ảnh. Có nhiều kỹ thuật xây dựng đặc trưng thủ công như: *LBP, SIFT, BoW, VLAT...*

Khái niệm 2: (*Đặc trưng tự động*) [3], [4]

Đặc trưng tự động (auto-encoders/non-handcrafted/deep/learned features) là đặc trưng được xây dựng từ tập dữ liệu ảnh cho trước. Kỹ thuật học từ tập dữ liệu huấn luyện sẽ tạo ra các đặc trưng này.

2. Phân lớp ảnh dựa trên tổ hợp đa đặc trưng

Vào năm 2011, Mohammadmehdi Bozorgi, Mohd Aizaini Maarof, and Lee Zhi Sam đề xuất bộ phân lớp 2 tầng trên các đặc trưng cấp thấp và thử nghiệm cho phân lớp ảnh người lớn [5]. Tầng phân lớp 1 gồm 2 bộ phân lớp: *SVM* và *AdaBoost*. Các đặc trưng dùng làm đầu vào của tầng này là các đặc trưng *CLD, SCD, EHD*. Kết quả phân lớp của tầng phân lớp 1 sẽ được tổng hợp theo trọng số 0.61. Việc sử dụng mô hình 2 tầng phân lớp giúp cho hệ thống đạt độ chính xác cao đến 91,9%. Kết quả nghiên cứu này cho thấy khi kết hợp nhiều bộ phân lớp sẽ giúp nâng cao độ chính xác.

Theo hướng tiếp cận tổ hợp đa đặc trưng, mỗi ảnh đầu vào sẽ được biểu diễn thành nhiều vector đặc trưng (mỗi đặc trưng biểu diễn cho một góc nhìn về bức ảnh) thay vì chỉ đại diện bởi một vector đặc trưng. Trong đó, khái niệm đơn và đa đặc trưng được hiểu như sau:

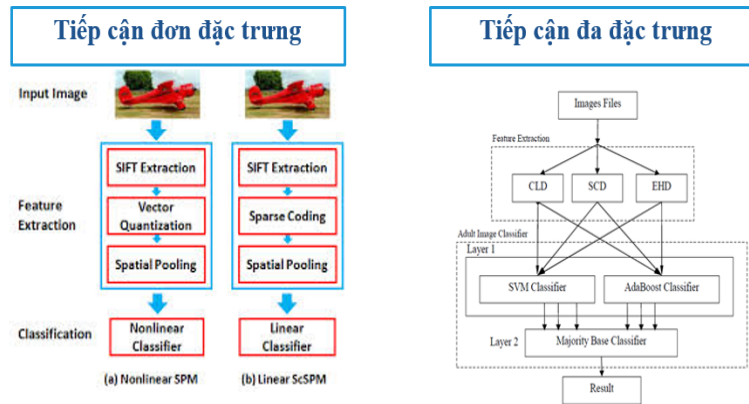
Khái niệm 3: (*Đơn đặc trưng*) [6]

Đơn đặc trưng được hiểu là một vector đặc trưng biểu diễn cho ảnh. Vector đặc trưng này được tạo ra từ một kĩ thuật trích chọn đặc trưng ảnh duy nhất.

Khái niệm 4: (Đa đặc trưng) [7]

Đa đặc trưng (multiple features) là một bộ gồm nhiều vector đặc trưng biểu diễn cho một ảnh hay một số ảnh con khác nhau của một ảnh đầu vào. Mỗi vector đặc trưng của bộ này được tạo ra từ một kĩ thuật trích chọn riêng hoặc là vector đại diện cho một ảnh con của ảnh đầu vào.

Các mô hình phân lớp tổ hợp đa đặc trưng sẽ gồm nhiều bộ phân lớp đơn cho các vector đặc trưng và bộ tích hợp để hợp nhất các kết quả phân lớp đơn.



Hình 1. So sánh tiếp cận đơn và đa đặc trưng trong phân lớp ảnh [5], [8]

Các hướng tiếp cận theo hướng đa đặc trưng gồm kết hợp sớm, kết hợp trễ, mô hình nhiều tầng (Hybrid, Stacking, Ensembling, Fine-Tuning). Các hướng tiếp cận này thường gặp phải các thách thức như: chi phí phân lớp lớn, kết hợp nhiều đặc trưng như thế nào và khai thác mối quan hệ giữa các đặc trưng, và hạn chế trong khả năng ứng dụng thực tiễn.

Bài toán phân lớp ảnh tổ hợp m đặc trưng vào L lớp cho trước:

- Đầu vào: Ma trận biểu diễn ảnh I trong không gian ảnh $\mathbb{I} = \{I_i\}$, với $I_i \in \mathbb{M}_{W \times H}(\mathbb{R}^D)$
- Đầu ra: Vector L chiều $O = (O_1, O_2, \dots, O_L) \in [0,1]^L$.
- Xử lí:
 - Rút trích đặc trưng: $e(I_i): \mathbb{I} \rightarrow \mathbb{F}$ gồm m phép chiếu.
 - pr₁: $F_1 = e_1(I_i)$
 - pr₂: $F_2 = e_2(I_i)$
 - ...
 - pr_m: $F_m = e_m(I_i)$
 - Phân lớp đơn: $h^*(F_i): \mathbb{F} \rightarrow \mathbb{O}$ là hàm vector gồm m hàm thành phần.
 - $h^*_1: OL_1 = h^*_1(F_1)$
 - $h^*_2: OL_2 = h^*_2(F_2)$
 - ...
 - $h^*_m: OL_m = h^*_m(F_m)$
- Hợp nhất kết quả: $h^{**} \circ h^*(F_1, F_2, \dots, F_m)$

Thuật giải 1: Phân lớp ảnh theo m đặc trưng vào L lớp cho trước

Đầu vào: Ma trận biểu diễn ảnh $I_{\text{đầu vào}}$, (ma trận có kích thước $W \times H$, mỗi phần tử ma trận thuộc R^D)

Đầu ra: Kết luận ảnh thuộc 1 trong L lớp cho trước

Xử lí:

1. Rút trích đặc trưng: $e(I_{\text{đầu vào}}): I \rightarrow F$ gồm m phép chiếu

$pr_1: F_1 = e_1(I_{\text{đầu vào}})$

$pr_2: F_2 = e_2(I_{\text{đầu vào}})$

...

$pr_m: F_m = e_m(I_{\text{đầu vào}})$

2. Phân lớp đơn: ánh xạ $h^*: F \rightarrow \mathcal{O}$ là hàm vector gồm m hàm thành phần

$h_1^*: OL_1 = h_1^*(F_1)$

$h_2^*: OL_2 = h_2^*(F_2)$

...

$h_m^*: OL_m = h_m^*(F_m)$

Trong đó, OL_i là vector độ thuộc vào L lớp định nghĩa trước, với $i=1..m$.

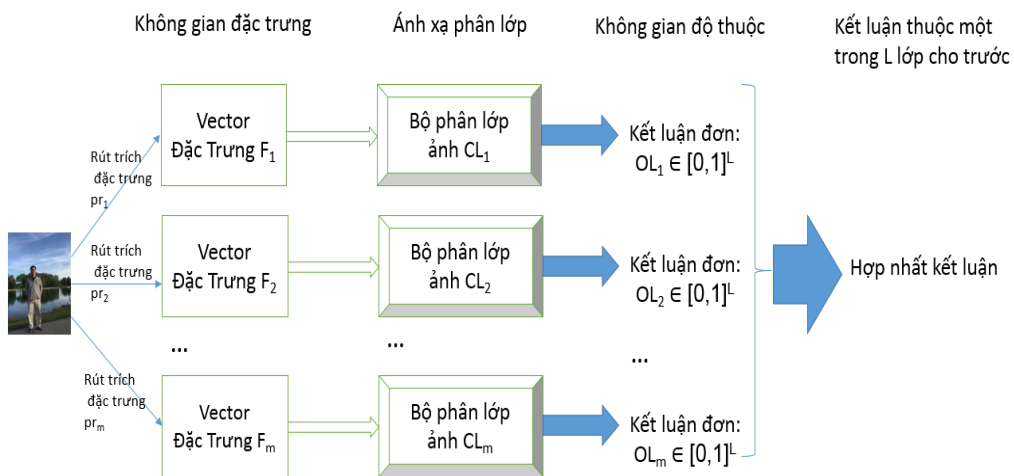
3. Hợp nhất kết quả: ánh xạ hợp $h^{**} \circ h^*(F_1, F_2, \dots, F_m)$. Ta được vector độ thuộc L chiều $O = (O_1, O_2, \dots, O_L) \in [0,1]^L$

4. Kết luận ảnh thuộc 1 trong L lớp đã định nghĩa dựa vào giá trị độ thuộc lớn nhất của $O = (O_1, O_2, \dots, O_L)$

3. Mô hình phân lớp ảnh dựa trên tổ hợp nhiều đặc trưng

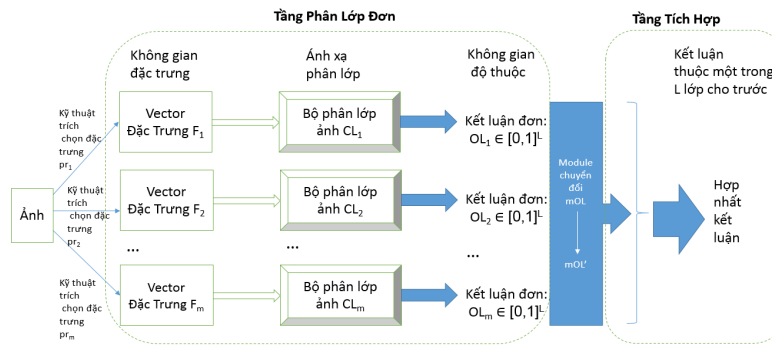
3.1. Kiến trúc của mô hình đa tầng

Quá trình phân lớp ảnh dựa trên m vector đặc trưng (sử dụng m kĩ thuật trích chọn đặc trưng ảnh khác nhau) vào L lớp định nghĩa trước (L tùy vào yêu cầu của các bài toán cụ thể trong thực tế) như sau:



Hình 2. Phân lớp ảnh m vector đặc trưng vào L lớp cho trước

- Kiến trúc mô hình đa tầng đề xuất:
- Tầng phân lớp đơn: Đưa ra kết luận (vector độ thuộc) phân lớp cục bộ dựa trên từng vector đặc trưng của ảnh đầu vào.
 - Module chuyển đổi: Tìm mối tương quan giữa các bộ phân lớp đơn. Module chuyển đổi thực hiện 2 bước tính toán:
 - (i) Xác suất hóa vector độ thuộc;
 - (ii) Áp dụng bộ hệ số tương quan theo Gauss để biến đổi kết quả đầu ra của tầng phân lớp đơn nhằm phản ánh mối tương quan ảnh hưởng qua lại của tất cả các đầu ra của tầng phân lớp đơn với nhau.
 - Tầng tích hợp: Liên kết kết luận của các bộ phân lớp đơn và hợp nhất kết quả.



Hình 3. Kiến trúc mô hình đa tầng cho phân lớp ảnh đa đặc trưng

Thuật giải 2: Phân lớp ảnh tổ hợp nhiều đặc trưng

Đầu vào: Ảnh đầu vào

Đầu ra: Kết luận thuộc một trong L lớp cho trước

Xử lí:

Tiền xử lí: Chuẩn hóa kích thước ảnh và khử nhiễu.

1. Rút trích m vector đặc trưng: (F_1, F_2, \dots, F_m)
2. Tầng phân lớp đơn: Phân lớp theo từng vector đặc trưng bởi.

Duyệt qua m bộ phân lớp $(CL_1, CL_2, \dots, CL_m)$ tính độ thuộc vào L lớp cho trước của ảnh đầu vào.

$CL_i: F_i \rightarrow OL_i, i=1..m.$

Cuối duyệt

3. Module chuyển đổi: Biến đổi m kết quả của tầng phân lớp đơn theo hệ số tương quan
- Biến đổi ma trận độ thuộc của m bộ phân lớp đơn mOL

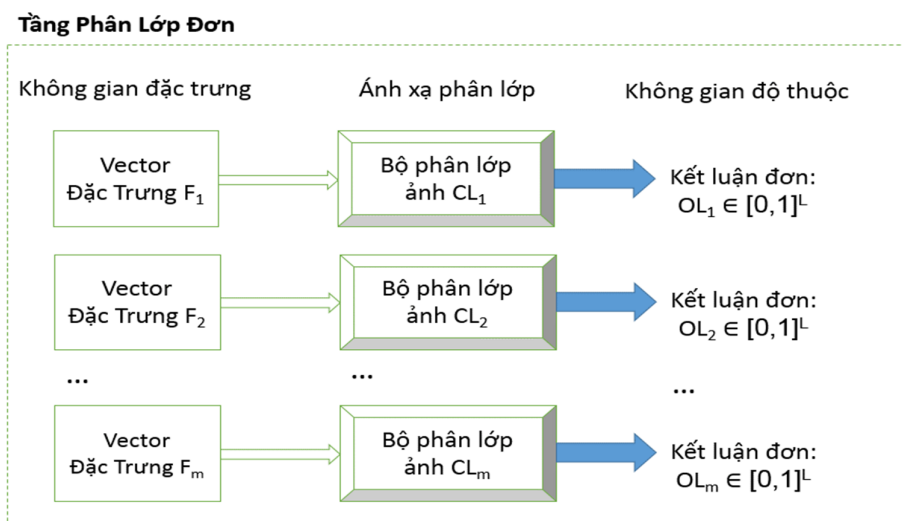
$$mOL = \begin{bmatrix} mOL_1 \\ mOL_2 \\ \dots \\ mOL_m \end{bmatrix}, \text{ với } i = 1..m$$

Trong đó, số dòng m = số vector đặc trưng, số cột L = số lớp
 $mOL \rightarrow mOL'$

4. Tầng tích hợp: Hợp nhất m kết quả của các bộ phân lớp đơn sau khi qua module chuyển đổi
 $mOL' \rightarrow O \in [0, 1]^L$
5. Kết luận thuộc 1 trong L lớp dựa vào độ thuộc lớn nhất

3.2. Tầng phân lớp đơn

Tầng phân lớp đơn gồm m bộ phân lớp đơn $CL_i, i=1..m$. Mỗi bộ phân lớp đơn cho ra một kết luận về độ thuộc vào L lớp của ảnh đầu vào dựa theo vector đặc trưng tương ứng đã được rút trích.



Hình 4. Kiến trúc tầng phân lớp đơn của mô hình đa tầng

Hình trên thể hiện cấu trúc của tầng phân lớp đơn của mô hình phân lớp đa tầng. Tầng phân lớp đơn gồm m bộ phân lớp = m vector đặc trưng của ảnh đầu vào. Các đặc trưng của ảnh đầu vào là kết quả chiếu ảnh đầu vào xuống các không gian khác nhau. Các bộ phân lớp của tầng phân lớp đơn sẽ ánh xạ m vector đặc trưng này cùng một không gian độ thuộc L chiều.

3.3. Module chuyển đổi

Module chuyển đổi: Tìm mối tương quan giữa đầu ra của m bộ phân lớp đơn. Đầu ra của mỗi bộ phân lớp đơn là một vector độ thuộc vào L lớp đã định nghĩa trước. Vì vậy, đầu ra của mỗi bộ phân lớp đơn là vector độ thuộc L chiều. Như vậy, module chuyển đổi sẽ biến đổi ma trận $m \times L$ gồm m dòng L cột (tất cả các đầu ra của tầng phân lớp đơn) thành một ma trận $m \times L$ mới dựa theo mối tương quan của các đầu ra thay vì chuyển thẳng các đầu ra của tầng phân lớp đơn vào tầng tích hợp để hợp nhất kết quả phân lớp.

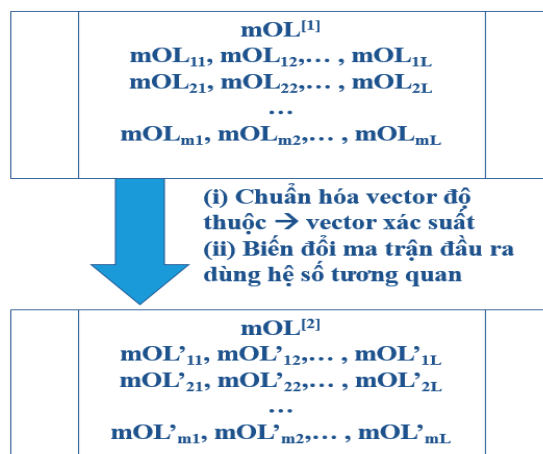
Thuật giải 3: Chuyển đổi đầu ra của m bộ phân lớp đơn

Đầu vào: Ma trận độ thuộc $mOL^{[1]}$ gồm m dòng L cột, mỗi dòng $mOL_i \in [0, 1]^L$ với $i=1..m$

Đầu ra: Ma trận phân ánh sự tương quan của các đầu ra của tầng phân lớp đơn $mOL^{[2]}$ gồm m dòng L cột, mỗi dòng $mOL'_i \in [0, 1]^L$ với $i=1..m$

Xử lí:

1. Chuẩn hóa vector độ thuộc thành vector xác suất
 2. Chuyển đổi ma trận đầu ra bằng hệ số tương quan dựa theo phân phối Gauss
-



Hình 5. Module chuyển đổi của mô hình đa tầng

Chuẩn hóa vector độ thuộc của từng đầu ra thành vector xác suất tương ứng nhằm xác định tầm quan trọng của từng đầu ra của các bộ phân lớp đơn. Vì có m vector độ thuộc đầu ra nên ta sẽ có một ma trận độ thuộc gồm m dòng L cột ứng với m bộ phân lớp đơn của m vector đặc trưng của ảnh đầu vào, và L lớp cần phân loại.

Bước 1. Chuẩn hóa vector độ thuộc thành vector xác suất

Thuật giải 4: Chuẩn hóa vector độ thuộc của từng đầu ra thành vector xác suất

Đầu vào: Ma trận vector độ thuộc

$$mOL^{[1]} = \begin{bmatrix} mOL_1 \\ mOL_2 \\ \dots \\ mOL_m \end{bmatrix}, \text{ với } i = 1..m$$

Trong đó, mOL_i là vector độ thuộc L chiều.

Đầu ra: Cập nhật độ thuộc thành xác suất

$$mOL^{[2]} = \begin{bmatrix} mOL_1 \\ mOL_2 \\ \dots \\ mOL_m \end{bmatrix}, \text{ với } i = 1..m$$

Trong đó, mOL_i là vector xác suất L chiều.

Xử lí:

For $i = 1$ to m do // m bộ phân lớp đơn = số vector đặc trưng

For $j = 1$ to L // L số lớp = số nút đầu ra của vector độ thuộc

$$mOL_{ij}^{[2]} = \frac{mOL_{ij}^{[1]}}{\sum_{j=1}^L mOL_{ij}^{[1]}}$$

End For // j

End For // i

Bước 2. Chuyển đổi ma trận đầu ra bằng hệ số tương quan dựa theo phân phối Gauss. Biến đổi ma trận vector xác suất dùng hệ số tương quan về vị trí:

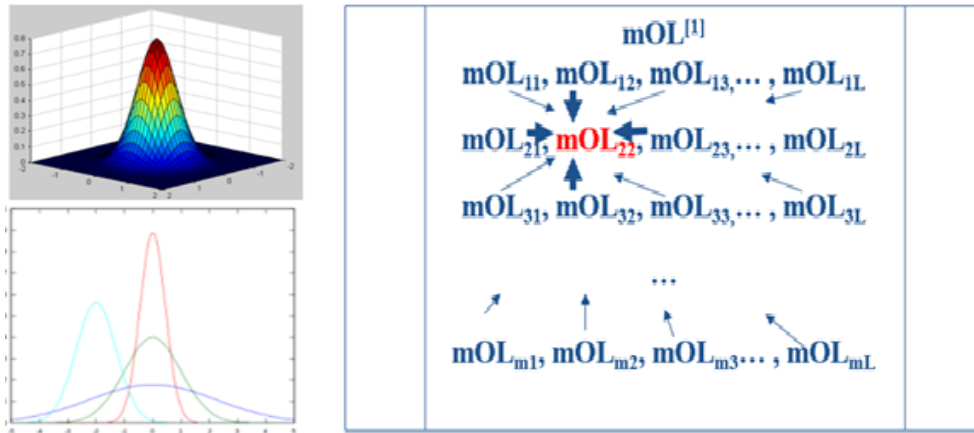
$$mOL_{pq}^{[2]} = \max_{i=1}^m (\max_{j=1}^L (w[p-i, q-j] \cdot mOL_{ij}^{[1]})), \quad (1)$$

trong đó, $p = 1..m$, $q = 1..L$

Trong đó, hệ số $w[p-i, q-j]$ là trọng số liên kết phần tử đầu vào thứ (i, j) $mOL_{ij}^{[1]}$ phần tử đầu ra thứ (p, q) $mOL_{pq}^{[2]}$. Hệ số w được xác định như sau:

$$w[l, k] = \exp(-\beta^2(l^2 + k^2)) \quad (2)$$

Hệ số tương quan giữa các bộ phân lớp đơn và biến đổi ma trận đầu ra w : Tương quan về vị trí theo dạng Gauss.



Hình 6. Hệ số tương quan dựa trên phân phối Gauss

Như vậy, quá trình xử lý của module chuyển đổi gồm 2 bước:

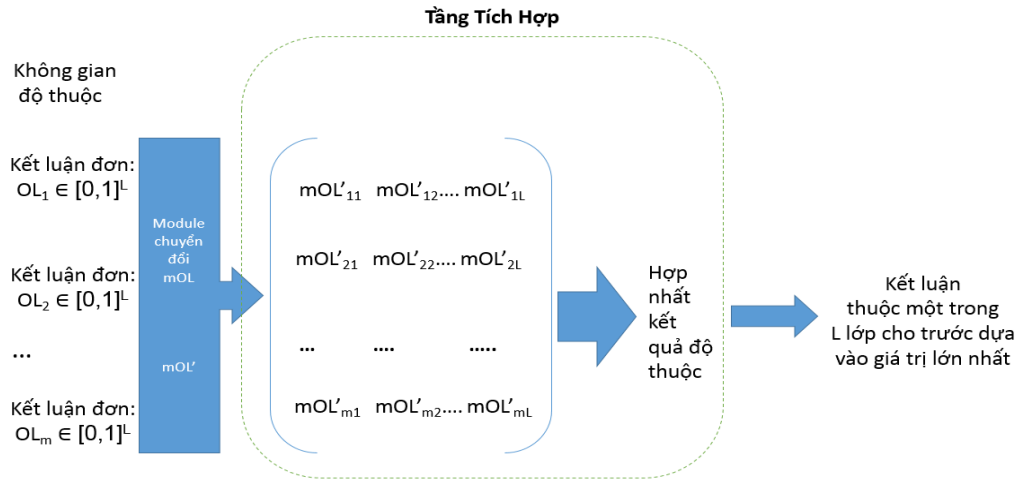
(i) Chuẩn hóa vector đầu ra nhằm phản ánh tầm quan trọng giữa các đầu ra trong cùng một bộ phân lớp đơn.

(ii) Biến đổi ma trận đầu ra dùng hệ số tương quan. Dữ liệu đầu ra sau khi chuẩn hóa theo phân bố tập trung về dạng vector xác suất thuộc $[0, 1]$ nên hệ số tương quan theo dạng Gauss là phù hợp. Bộ hệ số này phản ánh tầm ảnh hưởng của tất cả đầu ra của tất cả các bộ phân lớp đến 1 vị trí đầu ra cụ thể. Năng lượng ảnh hưởng = tương quan vị trí * giá trị, sau đó dùng hàm lấy max để giữ các giá trị lớn và triệt tiêu các giá trị bé.

3.4. Tầng tích hợp

Tầng tích hợp là bộ phân lớp với đầu vào là các kết luận phân lớp của tầng trước sau khi qua module chuyển đổi (để phản ánh mối tương quan của các đặc trưng). Đầu ra của tầng tích hợp là kết quả phân lớp cuối cùng (hợp nhất m kết luận phân lớp đơn). Như vậy, đầu ra của tầng tích hợp sẽ là vector độ thuộc L chiều. Sau đó, dựa trên giá trị lớn nhất về độ thuộc để kết luận ảnh đầu vào thuộc vào 1 trong L lớp nào cho trước.

Mô hình đa tầng sẽ hợp nhất ở mức quyết định (các kết quả phân lớp đơn) thay vì hợp nhất ở mức đặc trưng. Tầng tích hợp hợp nhất kết quả của m bộ phân lớp đơn để cho ra kết luận phân lớp cuối cùng theo cách tiếp cận kết hợp trễ. Tầng tích hợp đóng vai trò xác định tầm quan trọng của từng kết luận của các bộ phân lớp đơn thay vì hợp nhất theo dạng trung bình (đánh đồng tầm quan trọng của các bộ phân lớp đơn) hay hợp nhất theo dạng đa số (kết luận cuối sẽ tuân theo số đông).



Hình 7. Kiến trúc tầng tích hợp của mô hình đa tầng

Bộ tích hợp kết hợp muộn sẽ tăng tính 유연 chuyển và khả năng áp dụng trong nhiều ngữ cảnh thực tế vì tầm quan trọng của các kết luận của các bộ phân lớp đơn sẽ xác định tùy theo dữ liệu.

4. Một số ứng dụng của mô hình phân lớp ảnh đa tầng

4.1. Mô hình đa tầng cho bài toán phát hiện quảng cáo (Multi-CNN)

Việc kiểm tra sự xuất hiện của quảng cáo trên trang web hay video nhằm kiểm tra tính chính xác của thời gian hiển thị quảng cáo có đúng với thỏa thuận của các bên. Cụ thể đây là nhu cầu của 2 bên:

- (i) Nhà cung cấp dịch vụ quảng cáo: Trả tiền để gắn quảng cáo trên các trang web hay video có nhiều người xem. Ví dụ Adword (Google), Inmobi, Facebook.
- (ii) Nhà cung cấp nội dung số cho phép các nhà cung cấp dịch vụ quảng cáo hiển thị quảng cáo. Ví dụ vnexpress.net, hdonline.vn...

Bài toán phát hiện ảnh quảng cáo là bài toán phân lớp ảnh nhị phân hai lớp: Có xuất hiện quảng cáo (lớp *YES*) và không xuất hiện quảng cáo (lớp *NO*). Các thực nghiệm phát hiện ảnh quảng cáo trên hai bộ dữ liệu thực tế cho thấy các kết quả thú vị của mô hình đề xuất cho phân lớp ảnh nhị phân để phát hiện quảng cáo.

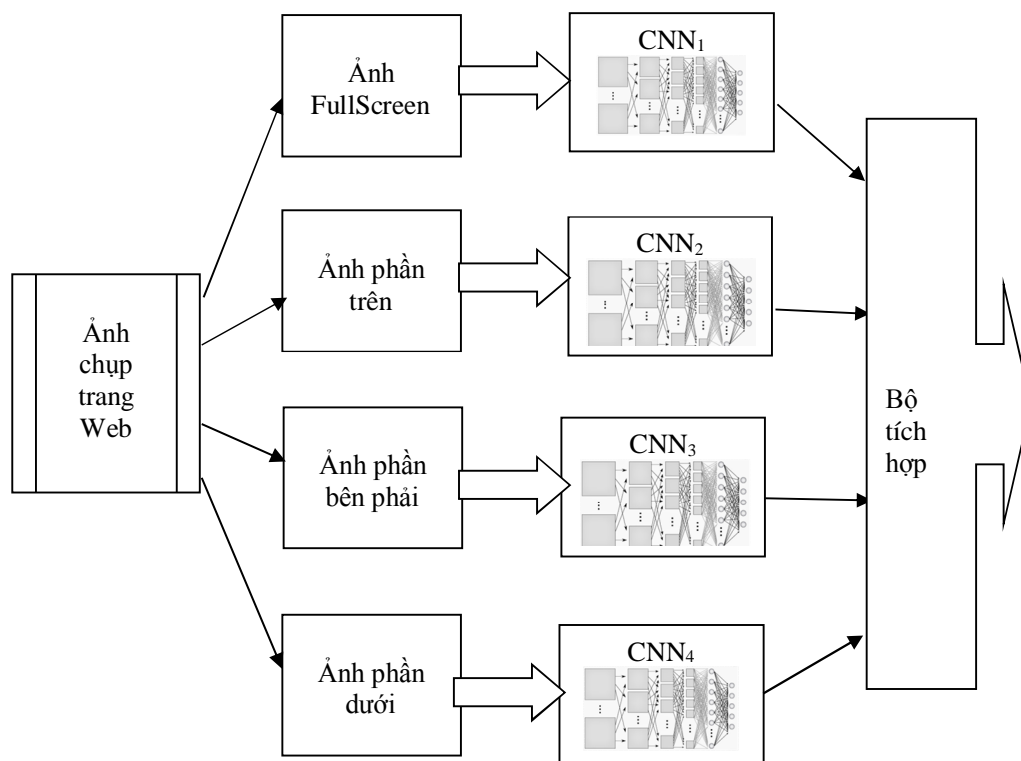
Bộ dữ liệu 1: Thu thập từ trang Web của TWC có trụ sở tại Hoa Kỳ do dự án phần mềm thực tế cung cấp. Dữ liệu ảnh được thu thập từ dự án thực tế của trang TWC News tại Hoa Kỳ dựa trên nhiều trình duyệt khác nhau (Chrome, FireFox, IE) chạy trên nhiều nền tảng khác hệ điều hành khác nhau (Windows, Linux, MacOS) có tổng dung lượng là

9.49 GB, được thu thập từ tháng 6/2016 đến tháng 10/2016.

Bộ dữ liệu 2: Bộ dữ liệu nội địa trong thị trường Việt Nam. Bộ dữ liệu được thu thập từ trang Web Báo Tiền Phong điện tử với số lượng 2000 hình có diện tích toàn trang Web, quảng cáo có thể xuất hiện ở bên góc phải của trang báo điện tử do đơn vị thứ 3 cung cấp như Google Adword, Facebook, rubiconproject.

Mô hình đa tầng dùng cho bài toán phát hiện quảng cáo gồm bốn bộ phân lớp CNN và một bộ tích hợp (xem Hình 8).

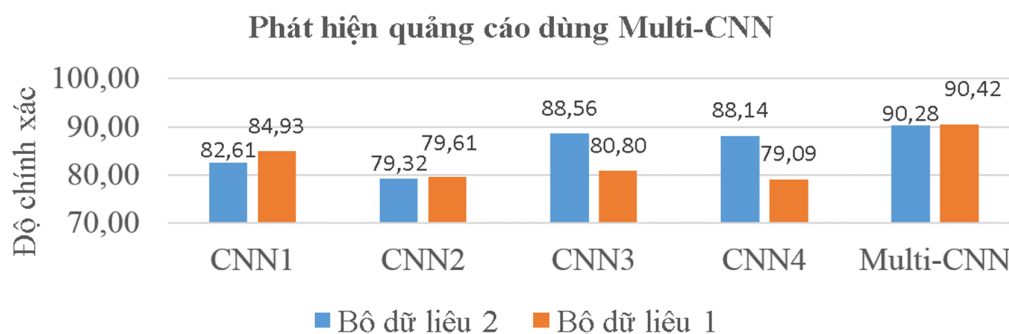
Tầng phân lớp đơn dùng bốn bộ phân lớp CNN_1 , CNN_2 , CNN_3 , CNN_4 , để xử lí các đầu vào tương ứng cho ảnh toàn bộ, ảnh phần trên, ảnh phần bên phải và ảnh phần dưới của trang web. Bốn bộ phân lớp đơn CNN dùng các tầng tích chập với tầng *Pooling*... để tự động xây dựng đặc trưng của phần ảnh đầu vào tương ứng. Tầng softmax của các bộ phân lớp đơn CNN cho ra bốn kết luận phân lớp của tầng phân lớp đơn (kết luận độ thuộc vào lớp có quảng cáo, lớp *YES*, hay lớp không có quảng cáo, lớp *NO*). Sau đó, tầng tích hợp hợp nhất bốn kết luận của tầng phân lớp đơn cho ra kết luận phân lớp cuối cùng của mô hình *Multi-CNN*.



Hình 8. Mô hình *Multi-CNN* cho phát hiện quảng cáo

Kết quả thử nghiệm mô hình *Multi-CNN* trên hai bộ dữ liệu: Bộ dữ liệu một thu thập từ trang WEB của TWC và bộ dữ liệu hai thu thập từ trang Web của trong nước cho thấy tổ hợp nhiều đặc trưng tự động sẽ nâng cao độ chính xác phân lớp.

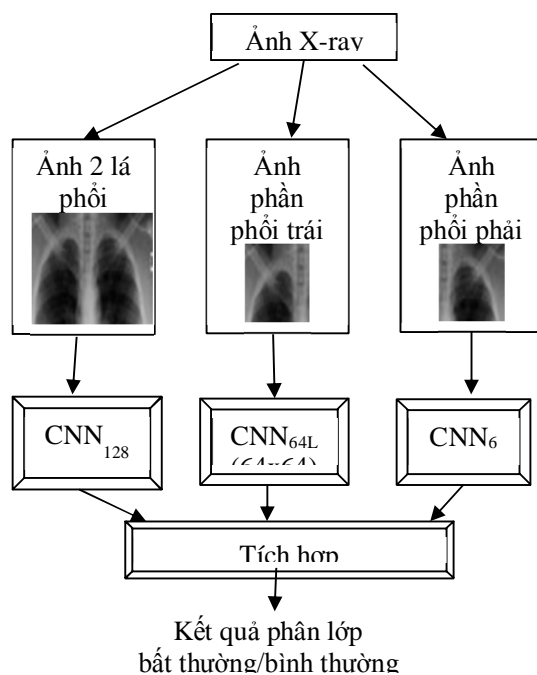
Ngoài ra, nếu dùng một CNN kết quả sẽ cao thấp tùy bộ dữ liệu. Ví dụ, bộ phân lớp CNN4 cho kết quả cao với bộ dữ liệu 2 nhưng cho kết quả thấp với bộ dữ liệu 1. Mô hình đa tầng Multi-CNN tổ hợp các đặc trưng tự động cho kết quả cao hơn các mô hình chỉ sử dụng một CNN. Hơn nữa, mô hình đề xuất tổ hợp nhiều đặc trưng nên cho kết quả ít thay đổi với cả hai bộ dữ liệu 90,28% cho bộ dữ liệu 2 và 90,42% cho bộ dữ liệu 1 (xem Hình 9).



Hình 9. Kết quả phát hiện quảng cáo dùng Multi-CNN

4.2. Mô hình đa tầng cho bài toán phát hiện bất thường trong ảnh X-ray phổi (Multi-CNN)

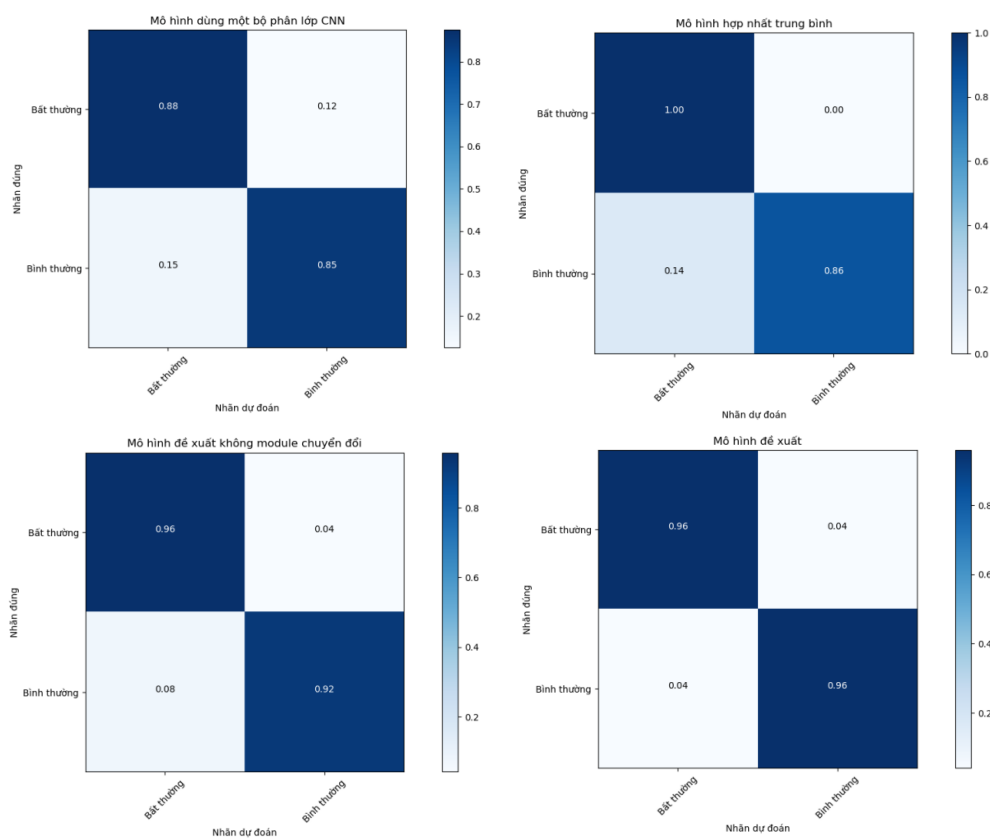
Tương tự, mô hình đa tầng với ba bộ phân lớp CNN cho phát hiện bất thường đậm độ (hỗ trợ chẩn đoán ung thư) cho ảnh y khoa phổi [29] gồm ba bộ phân lớp CNN (CNN_{128F} , CNN_{64L} , và CNN_{64R}) và một bộ tích hợp để hợp ba kết luận của ba bộ phân lớp CNN.



Hình 10. Mô hình đa tầng cho phân lớp ảnh X-ray phổi

Bộ phân lớp CNN_{128F} xác định có hay không bất thường trong ảnh X-ray (kích thước 128×128) dựa trên toàn bộ ảnh X-ray phổi. Bộ phân lớp CNN_{64L} xác định có hay không bất thường trong ảnh X-ray dựa trên vùng ảnh X-ray lá phổi trái. CNN_{64R} xác định có hay không bất thường trong ảnh X-ray dựa trên vùng ảnh X-ray lá phổi phải. Bộ tích hợp nhất ba kết luận cho ra kết luận cuối cùng: Ảnh X-ray đầu vào thuộc lớp nào trong hai lớp, có bất thường trong ảnh X-ray phổi (lớp *YES*) hay không có bất thường trong ảnh X-ray phổi (lớp *NO*) (xem Hình 10).

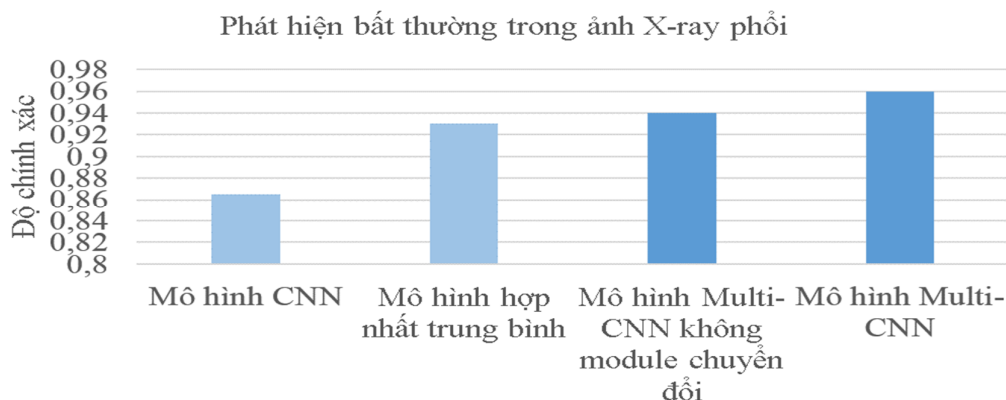
Mô hình đề xuất, mô hình dùng một bộ phân lớp *CNN* truyền thống và mô hình hợp nhất theo trung bình dùng ba bộ phân lớp *CNN* cùng được thử nghiệm trên tập 400 ảnh X-ray của Bệnh viện An Bình, Thành phố Hồ Chí Minh. Trong đó, 200 ảnh dùng để huấn luyện và 200 ảnh để kiểm tra. Để chỉ ra cụ thể kết quả phân lớp, lớp nào được phân lớp chính xác, và dữ liệu thuộc lớp nào thường bị phân lớp sai, nghiên cứu sử dụng kết quả ma trận *confusion* để so sánh. Các giá trị lớn trên các ô đường chéo chính của ma trận *confusion* cho khả năng tách lớp tốt của cả bốn mô hình: mô hình đề xuất *Multi-CNN*, mô hình dùng một bộ phân lớp *CNN* truyền thống, mô hình hợp nhất trung bình (xem Hình 11).



Hình 11. So sánh kết quả phân lớp giữa các mô hình khác nhau

Thống kê độ chính xác trong phát hiện bất thường ảnh X-ray phổi nhằm hỗ trợ chẩn đoán ung thư (độ chính xác phân lớp nhị phân: Có bất thường hay bình thường) giữa các mô hình cho thấy hiệu quả của mô hình đề xuất. Việc hợp nhất nhiều đặc trưng theo trung bình đã giúp nâng cao độ chính xác trong phân lớp ảnh X-ray phổi so với chỉ dựa vào một đặc trưng. Mô hình đa tầng tổ hợp nhiều đặc trưng cũng cho kết quả tốt hơn so với các hợp nhất trung bình (xem Hình 12).

Nhìn chung, mô hình dùng nhiều bộ phân lớp (mô hình đề xuất có/không module chuyển đổi và mô hình hợp nhất theo trung bình) giúp nâng cao độ chính xác so với chỉ dùng một bộ phân lớp. Mô hình hợp nhất trung bình gặp nhiều lỗi khi nhận sai từ bất thường thành bình thường, trong khi các chuyên gia y khoa mong muốn tỉ lệ nhầm lẫn này là thấp nhất có thể và có thể chấp nhận tỉ lệ sai sót nhỏ từ bình thường thành bất thường. Mô hình đề xuất sử dụng bộ tích hợp học máy giúp đánh trọng số tầm quan trọng của từng bộ phân lớp đơn phù hợp với tập dữ liệu theo từng ngữ cảnh cụ thể. Vì vậy, giúp tăng độ chính xác phân lớp so với việc đánh đồng tầm quan trọng của các bộ phân lớp đơn theo cách hợp nhất trung bình.



Hình 12. Kết quả thử nghiệm phát hiện bất thường trong ảnh X-ray phổi

Ngoài ra, thử nghiệm mô hình đề xuất với việc có hay không sử dụng module chuyển đổi cho thấy rằng module chuyển đổi giúp nâng độ chính xác phân lớp vì phản ánh mối quan hệ giữa các đầu ra của các bộ phân lớp đơn.

5. Kết luận

Đối với từng lớp bài toán cụ thể cần phải chọn loại đặc trưng thích hợp vì mỗi loại đặc trưng có ưu và khuyết điểm riêng. Nếu như các đặc trưng toàn cục giúp đánh giá tổng thể về bức ảnh, mối tương quan giữa các đối tượng trong ảnh, thì các đặc trưng cục bộ cung cấp các đặc trưng cụ thể của từng vùng hay từng đối tượng trong ảnh. Ngoài thách thức về việc lựa chọn và trích chọn đặc trưng phù hợp với loại ảnh thì việc quyết định bộ phân lớp phù hợp cũng hết sức quan trọng. Vì mỗi bộ phân lớp có những điểm mạnh điểm yếu khác nhau, có bộ phân lớp thì đạt độ chính xác cao nhưng đòi hỏi thời gian và bộ dữ

liệu huấn luyện lớn. Do đó, tùy vào yêu cầu thực tế của ứng dụng cần lựa chọn bộ phân lớp và đặc trưng phù hợp.

Nghiên cứu đề xuất mô hình phân lớp đa tầng nhằm giải quyết thách thức của hệ thống phân lớp ảnh theo đa đặc trưng có kiến trúc mô hình uyển chuyển, bao gồm:

- Tầng phân lớp đơn (tầng 1);
- Module chuyển đổi: Tìm mối tương quan (liên hệ) giữa các bộ phân lớp đơn;
- Tầng tích hợp: Liên kết các bộ phân lớp đơn và hợp nhất kết quả.

Mô hình phân lớp theo đa đặc trưng phản ánh nhiều góc nhìn, nhìn cách phân tích về cùng một đối tượng nên giúp tăng hiệu quả phân lớp. Mỗi ảnh đầu vào sẽ được biểu diễn bởi nhiều đặc trưng ứng với nhiều hệ quy chiếu khác nhau nên phản ánh đầy đủ đa dạng thông tin về đối tượng trong ảnh. Ví dụ, cùng là đối tượng giảng viên A khi trên bục giảng sẽ cần một vector biểu diễn đặc trưng khác (một hệ quy chiếu khác) khi cũng đối tượng ấy ngồi trong quán cafe. Tương tự, phân tích bằng đặc trưng SIFT một ảnh đầu vào sẽ cho các giá trị bất biến với các phép biến đổi tỉ lệ, tịnh tiến, phép quay, hay ít thay đổi với độ sáng. Trong khi, nếu phân tích bằng PCA sẽ cho ta các đặc trưng thành phần chính của ảnh đầu vào dựa trên thông kế tập dữ liệu huấn luyện.

Ngoài ra, việc sử dụng nhiều vector đặc trưng sẽ giảm ảnh hưởng của dữ liệu nhiễu. Nếu một vector đặc trưng trích chọn không đạt yêu cầu thì các vector đặc trưng còn lại sẽ giúp cung cấp thông tin cho việc phân lớp chính xác và tin cậy.

❖ **Tuyên bố về quyền lợi:** Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.

TÀI LIỆU THAM KHẢO

- [1] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE transactions on pattern analysis and machine intelligence*, 40(4), pp. 834-848, 2018.
- [2] M. Budnik, E. L. Gutierrez-Gomez, B. Safadi, D. Pellerin, and G. Quénot, “Learned features versus engineered features for multimedia indexing,” *Multimedia Tools and Applications*, 76(9), pp. 11941-11958, 2017.
- [3] Nanni, Loris, Stefano Ghidoni and Sheryl Brahmam, “Handcrafted vs. non-handcrafted features for computer vision classification,” *Pattern Recognition*, 71, pp. 158-172, 2017.
- [4] Hinton, E. Geoffrey, Alex Krizhevsky and D. Wang Sida, “Transforming auto-encoders,” *International Conference on Artificial Neural Networks*, Springer, Berlin, Heidelberg, 2011.
- [5] Bozorgi, Mohammadmehdi, Mohd Aizaini Maarof and Lee Zhi Sam. “Multi-classifier Scheme with Low-Level Visual Feature for Adult Image Classification,” *International Conference on Software Engineering and Computer Systems*, Springer, Berlin, Heidelberg, 2011.

- [6] Y. Di, G. Jiang, L. Yan, H. Liu and S. Zheng, "Multi-scale Segmentation Of High Resolution Remote Sensing Images By Integrating Multiple Features," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42, 2017.
- [7] T. Hu, Z. Li, W. Su, X. Mu and J. Tang, "Unsupervised video summaries using multiple features and image quality," *Multimedia Big Data (BigMM)*, IEEE Third International Conference on. IEEE, 2017.
- [8] J. Yang, K. Yu, Y. Gong and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," *CVPR Computer Vision and Pattern Recognition, IEEE Conference on. IEEE*, 2009.
- [9] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare and P. M. Atkinson, "A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, pp. 133-144, 2018.
- [10] S. Zheng, F. Nie, C. Ding and H. Huang, "A Harmonic Mean Linear Discriminant Analysis for Robust Image Classification," In *Tools with Artificial Intelligence (ICTAI), IEEE 28th International Conference*, pp.402-409, 2016.
- [11] B. A. Maxwell and A. Shafer. Steven, "Physics-based segmentation of complex objects using multiple hypotheses of image formation," pp. 269-295, 1997.
- [12] Tran Son Hai, Le Hoang Thai and Nguyen Thanh Thuy, "Phân Lớp ảnh đa đặc trưng và ứng dụng," *Hội nghị Quốc gia lần thứ 20 về Điện tử, Truyền thông và Công nghệ Thông tin (REV-ECIT)*, Thành phố Hồ Chí Minh, Việt Nam, 2017.
- [13] A. Kumar, J. Kim, D. Lyndon, M. Fulham, and D. Feng, "An ensemble of fine-tuned convolutional neural networks for medical image classification," *IEEE journal of biomedical and health informatics*, 21(1), pp. 31-40, 2017.
- [14] Le Hoang Thai, Nguyen Do Thai Nguyen and Tran Son Hai, "A Facial Expression Classification System Integrating Canny, Principal Component Analysis and Artificial Neural Network," *International Journal of Machine Learning and Computing*, 1(4), 2011.
- [15] Le Hoang Thai, Tran Son Hai and Thanh Nguyen Thuy, "Applying Multi Support Vector Machine for Flower Image Classification," *International Conference on Context-Aware Systems and Applications. Springer*, Berlin, Heidelberg, 2012.