



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0042157

(51)^{2022.01} G06V 40/16; G06K 9/00; G06N 3/04

(13) B

(21) 1-2021-04219

(22) 09/07/2021

(45) 25/12/2024 441

(43) 25/10/2021 403

(73) TẬP ĐOÀN CÔNG NGHIỆP - VIỄN THÔNG QUÂN ĐỘI (VIETTEL) (VN)

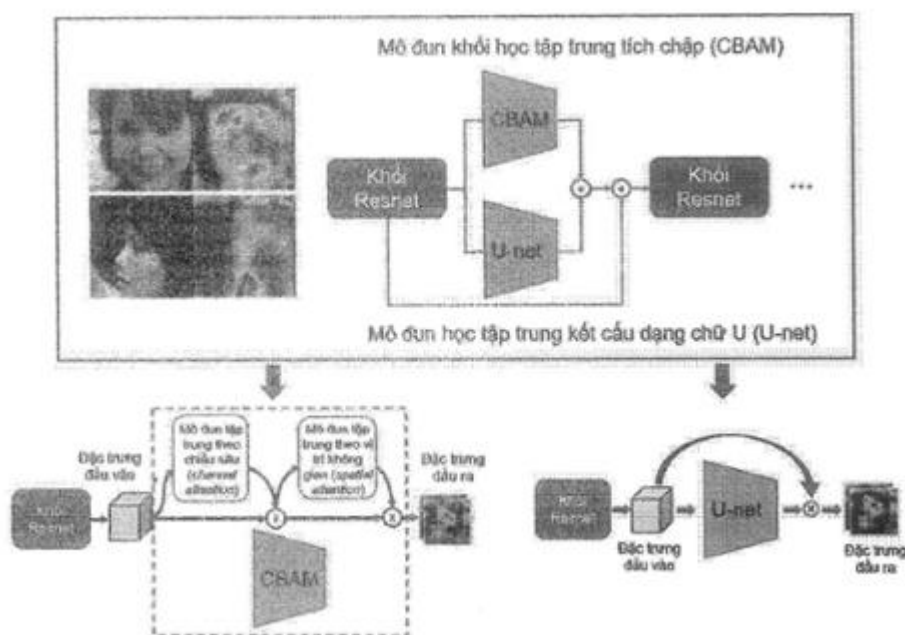
Lô D26 Khu đô thị mới Cầu Giấy, phường Yên Hoà, quận Cầu Giấy, thành phố Hà Nội

(72) VŨ THỊ HẠNH (VN); VÕ QUANG NHẬT (VN); NGUYỄN MẠNH QUÝ (VN);
HOÀNG NGỌC DƯƠNG (VN); NGUYỄN KHẮC DUY NGỌC (VN).

(74) Công ty TNHH Tư vấn Quốc Dân (NACILAW)

(54) PHƯƠNG PHÁP NHẬN DẠNG BIỂU CẢM KHUÔN MẶT

(57) Sáng chế đề xuất phương pháp nhận dạng biểu cảm khuôn mặt. Phương pháp này bao gồm các bước: bước 1: thiết kế và thu thập bộ dữ liệu hình ảnh biểu cảm khuôn mặt nhằm khắc phục vấn đề về thiết hụt dữ liệu, dữ liệu khác biệt và thiên vị, là nguyên nhân gây ra việc học lệch (*overfitting*) cho các mô hình học sâu; bước 2: thiết kế một kiến trúc mạng học sâu mới có khả năng tập trung tốt hơn vào những vùng đặc thù của khuôn mặt để phân biệt biểu cảm, bằng việc tích hợp các mô đun học kết hợp tập trung (*ensemble attention*) vào các kiến trúc mạng học sâu cơ bản; bước 3: huấn luyện mô hình học sâu kết hợp tập trung đã thiết kế ở bước 2 sử dụng bộ dữ liệu đã thu thập được ở bước 1 dựa trên sự kết hợp của hai hàm mất mát (*loss function*) của ArcFace và Softmax để giải quyết vấn đề quá khớp (hay học lệch - *overfitting*).



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp nhận dạng biểu cảm khuôn mặt qua hình ảnh. Cụ thể, phương pháp nhận dạng biểu cảm khuôn mặt được đề cập trong sáng chế sử dụng mô hình học sâu (*deep learning*) kết hợp tập trung (*ensemble attention*) và được ứng dụng trong lĩnh vực tìm hiểu tâm lý khách hàng, phân tích tâm lý tội phạm, phát hiện các rối loạn tâm thần, cảm xúc, trị liệu y khoa.

Tình trạng kỹ thuật của sáng chế

Biểu cảm khuôn mặt là một trong những cách thức hiệu quả và phổ biến để con người bày tỏ cảm xúc và suy nghĩ của mình. Nhiều công trình nghiên cứu về phân tích biểu cảm gương mặt tự động đã được thực hiện gần đây do tính ứng dụng cao trong nhiều lĩnh vực phân tích tâm lý khách hàng, như trị liệu y khoa, giao tiếp giữa người và máy. Trong thời gian gần đây, với sự phát triển vượt bậc của công nghệ trí tuệ nhân tạo, nhiều phương pháp nhận dạng biểu cảm khuôn mặt qua hình ảnh đã được đề xuất và đạt được kết quả khá tốt trên những tập dữ liệu phổ biến trong nghiên cứu như FER+, AffectNet. Mặc dù việc áp dụng những mô hình học sâu phức tạp đã đạt được độ chính xác tương đối, nhưng khả năng áp dụng thực tiễn của những mô hình này vẫn còn hạn chế, chủ yếu do những lý do cơ bản sau:

Thứ nhất: số lượng dữ liệu hình ảnh dùng cho việc huấn luyện còn hạn chế, và những dữ liệu này có sự khác biệt với những tình huống thực tế, đặc biệt là hình ảnh khuôn mặt của người châu Á nói chung và người Việt Nam nói riêng. Do dữ liệu còn ít và khác biệt, những mô hình học sâu được huấn luyện trên một tập dữ liệu thường bị học lệch (*overfitting*) và khó đạt kết quả tốt trên tập dữ liệu khác hoặc trên môi trường thực tế.

Thứ hai: những tập dữ liệu được thu thập vẫn khó bao phủ được toàn bộ các trường hợp, ví dụ, ảnh khuôn mặt bị che một phần, khuôn mặt có góc nhìn nghiêng và độ sáng thay

đổi. Do đó, chúng ta cần phát triển những kiến trúc mạng học sâu có khả năng học tập trung tốt hơn vào những vùng đặc thù của khuôn mặt để phân biệt biểu cảm.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề cập một phương pháp mới sử dụng mô hình học sâu kết hợp tập trung cho vấn đề nhận dạng biểu cảm khuôn mặt cải thiện được các khó khăn hiện có để đạt được độ chính xác cao hơn, đặc biệt là áp dụng hiệu quả trên tập dữ liệu gương mặt của người Việt Nam.

Để đạt được mục đích trên, phương pháp nhận dạng biểu cảm khuôn mặt qua hình ảnh sử dụng mô hình học sâu kết hợp tập trung bao gồm các bước:

Bước 1: thiết kế và thu thập dữ liệu hình ảnh biểu cảm khuôn mặt; bước này được thực hiện với mục đích làm giàu có và đa dạng hoá dữ liệu hình ảnh biểu cảm khuôn mặt để huấn luyện mô hình học sâu hiệu quả.

Bước 2: thiết kế một kiến trúc mạng học sâu mới để nhận dạng biểu cảm khuôn mặt; bước này được thực hiện dựa trên việc tích hợp các mô đun học kết hợp tập trung (*ensemble attention*) vào các kiến trúc mạng học sâu cơ bản nhằm tăng cường khả năng học tập trung vào những vùng đặc thù của khuôn mặt để phân biệt biểu cảm.

Bước 3: huấn luyện mô hình học sâu kết hợp tập trung sử dụng hai hàm mất mát (*loss function*) ArcFace và Softmax; bước này được thực hiện dựa trên việc tính tổng của hai hàm mất mát với một tham số α được đặt làm trọng số (công thức (2)), tham số α sẽ được cập nhật tự động dựa vào tốc độ học (*learning rate*) trong quá trình huấn luyện mô hình; hàm mất mát ArcFace được sử dụng nhằm giải quyết vấn đề quá khớp (hay học lệch - *overfitting*) khi huấn luyện dữ liệu ảnh khuôn mặt.

Mô tả vắn tắt các hình vẽ

Hình 1 là mô tả về kiến trúc mô hình học sâu mới được tích hợp thêm các mô đun học kết hợp tập trung (*ensemble attention*) cho việc nhận dạng biểu cảm khuôn mặt qua hình ảnh.

Hình 2 là sơ đồ huấn luyện mô hình học sâu đã thiết kế sử dụng kết hợp hai hàm mất mát ArcFace và Softmax.

Mô tả chi tiết sáng chế

Sáng chế được mô tả chi tiết dưới đây có dựa vào các hình vẽ kèm theo, các hình vẽ này nhằm mục đích minh họa các phương án của sáng chế mà không làm giới hạn phạm vi bảo hộ sáng chế.

Cũng trong sáng chế này, các thuật ngữ như “RetinaFace”, “ResNet”, “ArcFace”, “Softmax”, “FER+”, “AffectNet” là những danh từ riêng, không có từ tiếng Việt tương ứng và do đó trong sáng chế này sẽ sử dụng những từ chuyên biệt này.

Tham chiếu Hình 1 mô tả về kiến trúc mô hình học sâu mới mà sáng chế đề cập, tích hợp các mô đun học kết hợp tập trung (*ensemble attention*) cho việc nhận dạng biểu cảm khuôn mặt qua hình ảnh. Cụ thể, phương pháp nhận dạng biểu cảm khuôn mặt bao gồm các bước:

Bước 1: thiết kế và thu thập dữ liệu hình ảnh biểu cảm khuôn mặt;

Bước này nhằm giải quyết được khó khăn về vấn đề dữ liệu biểu cảm khuôn mặt còn ít và khác biệt, những mô hình học sâu thường xuyên bị học lệch (*overfitting*) và khó đạt kết quả cao trên môi trường thực tế. Đặc điểm của bộ dữ liệu này là tính giàu có và đa dạng, hình ảnh khuôn mặt bao phủ được tối đa các trường hợp trong thực tế, phân bố dữ liệu theo tỉ lệ phù hợp cho các trường hợp như sau:

- Loại biểu cảm: vui, buồn, giận dữ, ngạc nhiên, khó chịu, sợ hãi, trung tính.
- Đối tượng theo giới tính: nam, nữ.
- Đối tượng theo độ tuổi: trẻ em, thanh thiếu niên, người lớn, người già.
- Đối tượng theo địa lý: người châu Âu, người châu Á, người Việt Nam.
- Tư thế mặt: góc mặt chính diện, góc mặt nghiêng sang trái hoặc sang phải từ 0° đến 90° , góc mặt ngẩng lên hoặc cúi xuống từ 0° đến 45° .

Ảnh đầu vào của mô hình nhận dạng biểu cảm khuôn mặt là các ảnh khuôn mặt đã được chuẩn hóa và căn chỉnh. Việc tìm kiếm và căn chỉnh khuôn mặt từ hình ảnh gốc ban

đầu từ bộ dữ liệu thu thập được thực hiện sử dụng mô hình học sâu RetinaFace. Sau đó, hình ảnh chỉ chứa các khuôn mặt đã cắt ra và căn chỉnh được đưa vào mô hình học sâu kết hợp tập trung đề xuất để xử lý tiếp trong bước 2.

Bước 2: thiết kế một kiến trúc mạng học sâu mới để nhận dạng biểu cảm khuôn mặt;

Tham chiếu Hình 1 mô tả về kiến trúc mô hình này. Mô hình học sâu kết hợp tập trung được thiết kế dựa trên các khối ResNet, trong các khối ResNet này được tích hợp thêm hai mô đun học tập trung bao gồm mô đun khối học tập trung tích chập (*Convolutional block attention module (CBAM)*) và mô đun học tập trung kết cấu dạng chữ U (*U-net*). Các mô đun này phục vụ cho việc trích chọn các đặc trưng tập trung theo chiều sâu (*channel attention*) và đặc trưng tập trung theo vị trí không gian (*spatial attention*) bằng cách học các trọng số quan trọng trong quá trình huấn luyện.

Thứ nhất, mô đun học tập trung CBAM được tạo thành bởi hai mô đun nhỏ hơn nối tiếp nhau: mô đun đặc trưng tập trung theo chiều sâu (*channel attention*) và mô đun đặc trưng tập trung theo vị trí không gian (*spatial attention*). Đầu vào của mô đun tập trung theo chiều sâu là bộ đặc trưng đã được trích xuất từ khối ResNet (*Resnet block*). Khối ResNet này có thể bao gồm hai lớp (dùng trong ResNet 18 và 34) hoặc ba lớp (dùng trong ResNet 50, 101, 152). Đặc trưng đầu vào sẽ được gộp lại (*pooling*) thành hai véc tơ một chiều và sẽ được xử lý bởi một mạng nơ ron. Kết quả xử lý của mô đun này là một véc tơ một chiều và sẽ được nhân với đặc trưng đầu vào và chuyển tiếp đến mô đun tập trung theo vị trí không gian. Tại mô đun tập trung theo vị trí không gian, đặc trưng đầu vào sẽ được gộp thành hai ma trận hai chiều và đưa vào lớp nơ ron tích chập (*convolutional layer*). Tương tự, kết quả của mô đun tập trung này sẽ lại được nhân với đặc trưng đầu vào, và chuyển tiếp cho khối ResNet tiếp theo.

Thứ hai, mô đun học tập trung U-net gồm một bộ mã hóa và một bộ giải mã. Mục đích của mô đun học tập trung U-net cũng tương tự như CBAM, giúp mạng nơ ron tập trung vào thông tin vị trí không gian quan trọng (*spatial attention*) và thực hiện phân loại biểu cảm chính xác hơn.

Tiếp theo, kết quả đầu ra của mô đun học tập trung CBAM và U-net được kết hợp với nhau để sinh ra một bộ đặc trưng được tập trung cuối cùng. Để tránh việc các mô đun tập trung gỡ bỏ những đặc trưng hữu ích, đặc trưng đầu vào từ khối ResNet sẽ được cộng vào bộ đặc trưng tập trung được sinh ra để tạo ra kết quả cuối cùng và đưa sang khối xử lý tiếp theo. Đặc trưng đầu ra của CBAM và U-net sẽ có cùng kích thước với đặc trưng đầu vào. Khối xử lý mô đun kết hợp đặc trưng tập trung và ResNet có thể được nối tiếp nhau lặp lại N lần (kiến nghị $N = 4$ hoặc 5), để tạo thành một kiến trúc mạng tập trung sâu hơn.

Bước 3: huấn luyện mô hình học sâu kết hợp tập trung sử dụng hai hàm mất mát (*loss function*) ArcFace và Softmax;

Tham chiếu Hình 2 mô tả việc huấn luyện mô hình học sâu kết hợp tập trung sử dụng hai hàm mất mát (*loss function*) là ArcFace và Softmax đồng thời, hàm mất mát cuối cùng sẽ là tổng hợp của hai hàm mất mát ArcFace và Softmax.

Đặc điểm của việc huấn luyện mô hình này là sự kết hợp của hai hàm mất mát là ArcFace và Softmax để giải quyết vấn đề quá khớp (hay học lệch - *overfitting*). Hàm mất mát Softmax đã được sử dụng phổ biến trong việc huấn luyện nhiều mô hình học sâu, tuy nhiên hàm này vẫn có nhược điểm chưa cải thiện được vấn đề học lệch. Sáng chế này đề xuất sử dụng kết hợp hàm mất mát ArcFace trong quá trình huấn luyện mô hình. Hàm mất mát ArcFace được thiết kế riêng cho bài toán nhận diện khuôn mặt, nhưng chưa từng được sử dụng cho bài toán nhận dạng biểu cảm khuôn mặt. Việc áp dụng hàm mất mát ArcFace sẽ hạn chế việc học lệch của mô hình, tăng cường khả năng phân loại những đặc trưng được học, và giúp ổn định quá trình huấn luyện. Hàm mất mát ArcFace được định nghĩa như sau (đây là một công thức có sẵn được sử dụng trong lĩnh vực nhận diện khuôn mặt; tuy nhiên, công thức được nêu ở đây để chỉ ra cách áp dụng trong vấn đề của sáng chế):

$$L_{ArcFace} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{yi}+m))}}{e^{s(\cos(\theta_{yi}+m))} + \sum_{j=1, j \neq yi}^n e^{s \cos \theta_j}} \quad (1)$$

Trong đó N là số lượng hình ảnh được huấn luyện, s và m là hai hằng số dùng để thay đổi độ lớn giá trị của đặc trưng và tăng khả năng phân loại đặc trưng, θ_{yi} là góc giữa

đặc trưng được trích xuất và trọng số của mạng học sâu. Mục tiêu học là tối đa khoảng cách góc θ để phân hóa đặc trưng của những loại biểu cảm khuôn mặt khác nhau.

Hàm mất mát cuối cùng sẽ là tổng hợp của hai hàm mất mát theo công thức (2), đây là công thức mới lần đầu được sử dụng trong sáng chế này :

$$L_{\text{cuối cùng}} = \alpha * L_{\text{ArcFace}} + (1 - \alpha) * L_{\text{Softmax}} \quad (2)$$

Tham số α sẽ được cập nhật tự động dựa vào tốc độ học (*learning rate*). Vào giai đoạn đầu của việc huấn luyện, khi tốc độ học còn cao (tốc độ học kiến nghị ban đầu là 0,01), α sẽ được đặt ở giá trị cao (ví dụ, $\alpha = 0,9$) để ưu tiên hàm mất mát ArcFace và làm giảm vấn đề học lệch. Sau khi mô hình đã ổn định (tốc độ học giảm dần xuống 0,0001), giá trị α cũng sẽ được giảm dần để huấn luyện phân lớp biểu cảm dựa vào hàm mất mát Softmax. Việc giảm tốc độ học được quyết định dựa vào độ chính xác trên tập thử nghiệm (*validation*). Nếu sau mười chu kỳ huấn luyện (*10 epoch*) độ chính xác trên tập thử nghiệm không tăng thêm, thì tốc độ học sẽ được giảm xuống còn 1/10 tốc độ học cũ. Tỷ lệ giảm tương ứng của α sẽ được quyết định dựa vào thực nghiệm huấn luyện, và tùy thuộc vào bộ dữ liệu được huấn luyện.

Kết thúc bước 3, mô hình học sâu kết hợp tập trung đã được hoàn thành huấn luyện và được sử dụng để dự đoán biểu cảm khuôn mặt qua hình ảnh. Mô hình này được ứng dụng trong các phần mềm hoặc các chương trình máy tính để xử lý hình ảnh, với đầu vào là hình ảnh khuôn mặt người từ camera hoặc video có sẵn, và đầu ra là kết quả phân tích biểu cảm của các khuôn mặt xuất hiện trong đó (ví dụ, khuôn mặt người A có biểu cảm vui, khuôn mặt người B có biểu cảm tức giận, v.v.).

Mặc dù các mô tả nêu trên chứa nhiều chi tiết cụ thể, chúng không được coi là giới hạn về phương án thực thi sáng chế, mà chỉ nhằm mục đích minh họa một số phương án thực thi được ưu tiên.

Yêu cầu bảo hộ

1. Phương pháp nhận dạng biểu cảm khuôn mặt, bao gồm các bước sau:

bước 1: thiết kế và thu thập dữ liệu hình ảnh biểu cảm khuôn mặt;

tại bước này, một bộ dữ liệu lớn được thiết kế và thu thập với mục đích làm giàu có và đa dạng hoá dữ liệu hình ảnh biểu cảm khuôn mặt để huấn luyện mô hình học sâu hiệu quả; bộ dữ liệu được thiết kế sao cho hình ảnh khuôn mặt bao phủ được tối đa các trường hợp trong thực tế, phân bố dữ liệu theo tỉ lệ phù hợp; cụ thể, bộ dữ liệu được thiết kế và thu thập có đặc điểm về phân bố dữ liệu với tỉ lệ hợp lý cho tất cả các trường hợp sau:

loại biểu cảm: vui, buồn, giận dữ, ngạc nhiên, khó chịu, sợ hãi, trung tính;

đối tượng theo giới tính: nam, nữ;

đối tượng theo độ tuổi: trẻ em, thanh thiếu niên, người lớn, người già;

đối tượng theo địa lý: người châu Âu, người châu Á, người Việt Nam;

tư thế mặt: góc mặt chính diện, góc mặt nghiêng sang trái hoặc sang phải từ 0° đến 90° , góc mặt ngẩng lên hoặc cúi xuống từ 0° đến 45° ;

bước 2: thiết kế một kiến trúc mạng học sâu mới để nhận dạng biểu cảm khuôn mặt;

tại bước này, một mô hình học sâu mới được xây dựng bằng cách tích hợp các mô-đun học kết hợp tập trung (*ensemble attention*) vào các kiến trúc mạng học sâu cơ bản (như *resnet*), mục đích là để tăng cường khả năng học tập trung của mô hình vào các vùng quan trọng của khuôn mặt và các đặc trưng quan trọng của biểu cảm khuôn mặt;

bước 3: huấn luyện mô hình học sâu kết hợp tập trung sử dụng hai hàm mất mát (*loss function*) *arcface* và *softmax*;

tại bước này, hàm mất mát sử dụng để huấn luyện mô hình là tổng của hai hàm mất mát *arcface* và *softmax* theo công thức dưới đây:

$$L_{\text{cuối cùng}} = \alpha * L_{\text{ArcFace}} + (1 - \alpha) * L_{\text{Softmax}}$$

trong đó, tham số α được cập nhật tự động dựa vào tốc độ học (*learning rate*); vào giai đoạn đầu của việc huấn luyện, α sẽ được đặt ở giá trị cao để ưu tiên hàm mất mát *arcface* và giảm việc học lệch; sau khi mô hình đã ổn định, giá trị

α được giảm dần để huấn luyện mô hình phân lớp biểu cảm khuôn mặt dựa vào hàm mất mát softmax.

2. Phương pháp nhận dạng biểu cảm khuôn mặt theo điểm 1, trong đó:

ở bước 2, mô hình học sâu kết hợp tập trung được thiết kế dựa trên các khối resnet, trong các khối resnet này được tích hợp thêm hai mô đun học tập trung bao gồm mô đun khối học tập trung tích chập (*convolutional block attention module (CBAM)*) và mô đun học tập trung kết cấu dạng chữ U (*U-net*); các mô đun này phục vụ cho việc trích chọn các đặc trưng tập trung theo chiều sâu (*channel attention*) và đặc trưng tập trung theo vị trí không gian (*spatial attention*) bằng cách học các trọng số quan trọng trong quá trình huấn luyện; trong đó:

mô đun học tập trung tích chập (CBAM) được tạo thành bởi hai mô đun nhỏ hơn nối tiếp nhau: mô đun đặc trưng tập trung theo chiều sâu (*channel attention*) và mô đun đặc trưng tập trung theo vị trí không gian (*spatial attention*); trong đó:

đầu vào của mô đun tập trung theo chiều sâu là bộ đặc trưng đã được trích xuất từ khối resnet (*resnet block*); khối resnet này có thể bao gồm hai lớp (dùng trong resnet 18 và 34) hoặc ba lớp (dùng trong resnet 50, 101, 152); đặc trưng đầu vào sẽ được gộp lại (*pooling*) thành hai véc tơ một chiều và sẽ được xử lý bởi một mạng nơ ron; kết quả xử lý của mô đun này là một véc tơ một chiều và sẽ được nhân với đặc trưng đầu vào và chuyển tiếp đến mô đun tập trung theo vị trí không gian;

tại mô đun tập trung theo vị trí không gian, đặc trưng đầu vào sẽ được gộp thành hai ma trận hai chiều và đưa vào lớp nơ ron tích chập (*convolutional layer*); tương tự, kết quả của mô đun tập trung này sẽ lại được nhân với đặc trưng đầu vào, và chuyển tiếp cho khối resnet tiếp theo;

mô đun học tập trung U-net gồm một bộ mã hóa và một bộ giải mã; mục đích của mô đun học tập trung U-net cũng tương tự như CBAM, giúp mạng nơ ron tập trung vào thông tin vị trí không gian quan trọng (*spatial attention*) và thực hiện phân loại biểu cảm chính xác hơn;

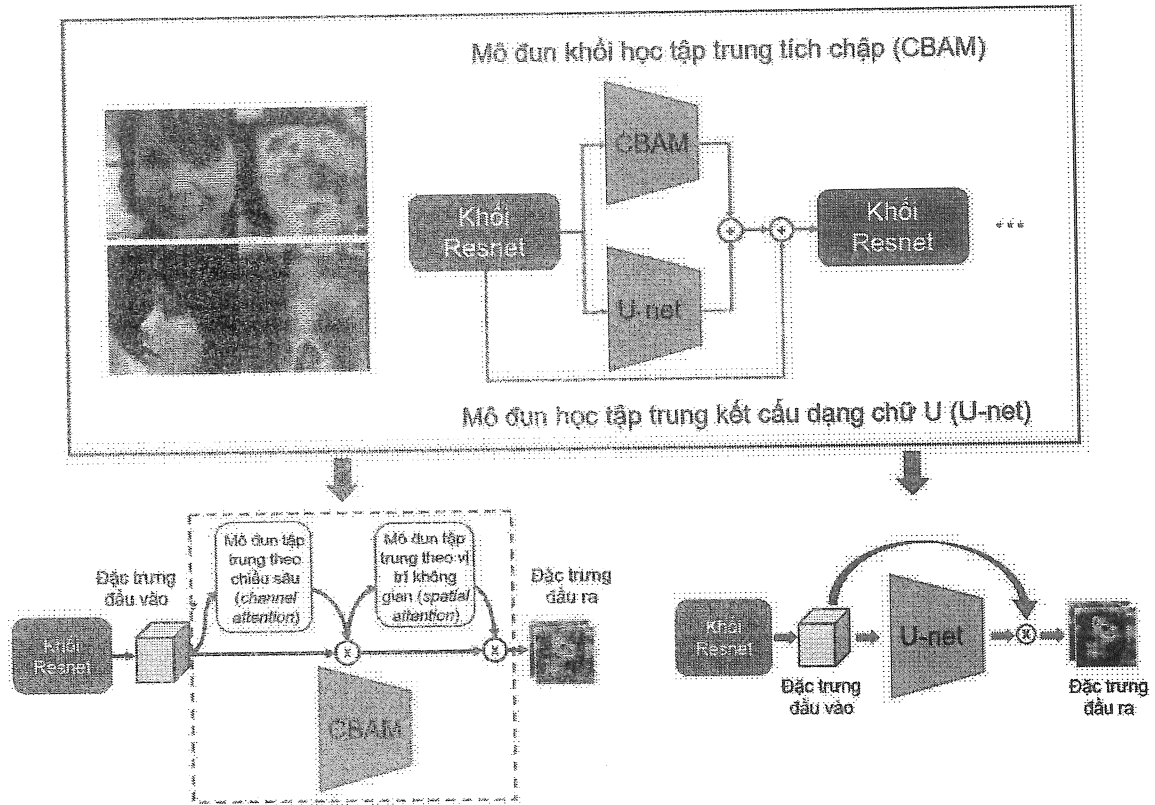
kết quả đầu ra của mô đun học tập trung CBAM và U-net được kết hợp với nhau để sinh ra một bộ đặc trưng được tập trung cuối cùng; để tránh việc các mô đun tập trung gỡ bỏ những đặc trưng hữu ích, đặc trưng đầu vào từ khối resnet sẽ được cộng vào bộ đặc trưng tập trung được sinh ra để tạo ra kết quả cuối cùng và đưa sang khối xử lý tiếp theo; đặc trưng đầu ra của CBAM và U-net sẽ có cùng kích thước với đặc trưng đầu vào; khối xử lý mô đun kết hợp đặc trưng tập trung và resnet có thể được nối tiếp nhau lặp lại N lần (kiến nghị $N = 4$ hoặc 5), để tạo thành một kiến trúc mạng tập trung sâu hơn.

3. Phương pháp nhận dạng biểu cảm khuôn mặt theo điểm 1, trong đó:

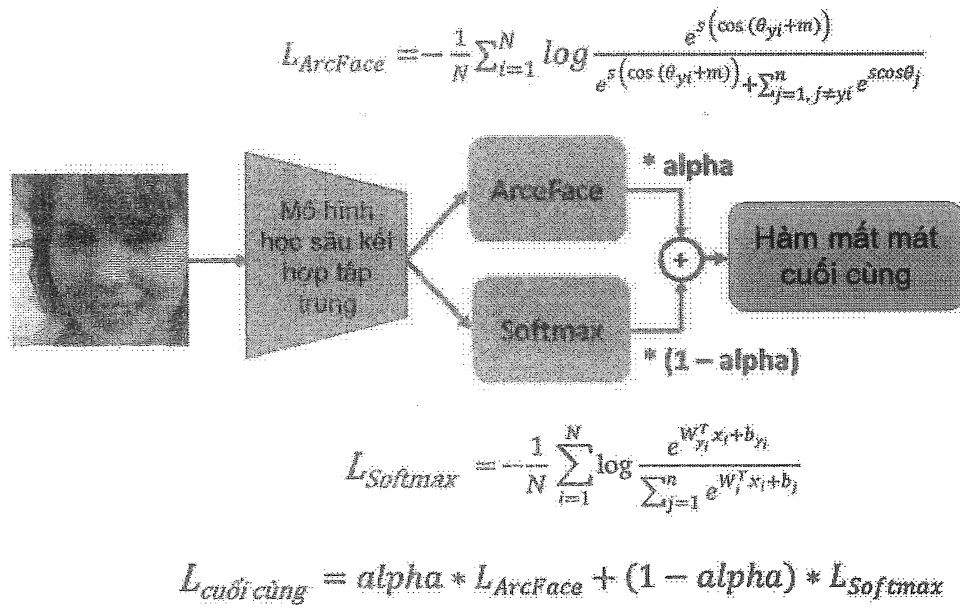
ở bước 3, sử dụng kết hợp hàm mất mát arcface và softmax trong quá trình huấn luyện mô hình; hàm mất mát arcface được thiết kế riêng cho bài toán nhận diện khuôn mặt, nhưng chưa từng được sử dụng cho bài toán nhận dạng biểu cảm khuôn mặt, việc áp dụng hàm mất mát arcface sẽ hạn chế việc học lệch của mô hình, tăng cường khả năng phân loại những đặc trưng được học, và giúp ổn định quá trình huấn luyện; hàm mất mát cuối cùng sẽ là tổng hợp của hai hàm mất mát theo công thức dưới đây:

$$L_{\text{cuối cùng}} = \alpha * L_{\text{ArcFace}} + (1 - \alpha) * L_{\text{Softmax}}$$

trong đó, tham số α sẽ được cập nhật tự động dựa vào tốc độ học (*learning rate*); vào giai đoạn đầu của việc huấn luyện, khi tốc độ học còn cao (tốc độ học kiến nghị ban đầu là 0,01), α sẽ được đặt ở giá trị cao (ví dụ, $\alpha = 0,9$) để ưu tiên hàm mất mát arcface và giảm vấn đề học lệch; sau khi mô hình đã ổn định (tốc độ học giảm dần xuống 0,0001), giá trị α cũng sẽ được giảm dần để huấn luyện phân lớp biểu cảm dựa vào hàm mất mát softmax; việc giảm tốc độ học được quyết định dựa vào độ chính xác trên tập thử nghiệm (*validation*); nếu sau mười chu kỳ huấn luyện (10 epoch) độ chính xác trên tập thử nghiệm không tăng thêm, thì tốc độ học sẽ được giảm xuống còn 1/10 tốc độ học cũ; tỷ lệ giảm tương ứng của α sẽ được quyết định dựa vào thực nghiệm huấn luyện, và tùy thuộc vào bộ dữ liệu được huấn luyện.



Hình 1



Hình 2