



(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)



1-0033914

(51)⁷ G06N 3/08

(13) B

(21) 1-2019-06904

(22) 06/12/2019

(30) 10201904549Q 21/05/2019 SG

(45) 25/11/2022 416

(43) 25/11/2020 392A

(73) Advanced New Technologies Co., Ltd. (KY)

Cayman Corporate Centre, 27 Hospital Road, George Town, Grand Cayman KY1-9008, Cayman Islands

(72) Jianshu LI (CN).

(74) Công ty TNHH Tâm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) HỆ THỐNG VÀ PHƯƠNG PHÁP HUẤN LUYỆN MẠNG NƠ-RON

(57) Sáng chế đề cập đến phương pháp huấn luyện mạng nơron bao gồm các bước: huấn luyện mạng nơron được huấn luyện trước bao gồm: lớp đầu vào; các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn; và lớp đầu ra. Bước huấn luyện mạng nơron được huấn luyện trước bao gồm các bước: đưa ít nhất một lớp bổ sung vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có các trọng số kết hợp; giữ cho trọng số của các nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi; nhập dữ liệu từ miền đích vào lớp đầu vào; và điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm thu được ở lớp đầu ra.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập rộng rãi đến các phương pháp và các thiết bị để huấn luyện các mạng nơron.

Tình trạng kỹ thuật của sáng chế

Học sâu là phương pháp học máy kết hợp các mạng nơron phân cấp trong các lớp liên tiếp để học từ dữ liệu theo cách lặp đi lặp lại.

Mạng nơron bao gồm một số nút liên kết, được gọi là “các nơron”, mà được tổ chức dưới dạng ba hoặc nhiều hơn ba lớp: lớp đầu vào, một hoặc nhiều lớp ẩn và lớp đầu ra. Thuật ngữ học sâu được sử dụng khi có nhiều lớp ẩn bên trong mạng nơron.

Dữ liệu được đưa vào mạng nơron qua lớp đầu vào, là lớp thứ nhất trong mạng nơron. Lớp đầu vào lấy các tín hiệu/trị số đầu vào và chuyển chúng lên các lớp ẩn. Các lớp ẩn có các nút áp dụng các phép biến đổi khác nhau cho các tín hiệu/trị số đầu vào. Mỗi lớp ẩn có thể có tập hợp các nút được xếp chồng theo phương thẳng đứng. Lớp ẩn cuối cùng chuyển các trị số tới lớp đầu ra. Lớp đầu ra là lớp cuối trong mạng nơron và nhận đầu vào từ lớp ẩn cuối cùng.

Dữ liệu được xử lý trong các lớp ẩn dựa vào các trọng số và/hoặc độ lệch được áp dụng vào các nút này. Cụ thể, việc xử lý xảy ra qua hệ thống các kết nối đặc trưng bởi các trọng số và độ lệch. Tại mỗi nút, đầu vào được tiếp nhận, nút tính tổng trọng số và cũng có thể thêm độ lệch. Theo kết quả và hàm kích hoạt đặt trước, nút quyết định nó có nên được kích hoạt hay không. Nút truyền thông tin này xuống nút kết nối khác.

Các mạng nơron và học sâu thường được sử dụng trong việc nhận diện hình ảnh (ví dụ nhận diện khuôn mặt), giọng nói và các ứng dụng thị giác máy tính. Nhận diện khuôn mặt là kỹ thuật nhận diện được sử dụng để phát hiện/xác định các khuôn mặt của các cá nhân từ các hình ảnh khuôn mặt. Thông thường, các đặc điểm khuôn mặt được trích xuất từ các hình ảnh khuôn mặt để định danh và xác minh người nào đó.

Để triển khai hệ thống nhận diện khuôn mặt sử dụng học sâu, mô hình được huấn luyện trước được xây dựng và được huấn luyện với dữ liệu quy mô lớn. Dữ liệu quy mô lớn có thể thu được từ một quốc gia hoặc khu vực địa lý cụ thể. Theo cách này, mô hình được huấn luyện hoạt động tốt đối với các ứng dụng trong nước (sở tại) hoặc khu vực

địa lý cụ thể (ví dụ Trung Quốc). Tuy nhiên, khi mô hình được huấn luyện được sử dụng trong một quốc gia hoặc khu vực địa lý khác (ví dụ châu Âu), thì hiệu suất có thể giảm do sự thay đổi của các miền từ nơi mô hình được huấn luyện ban đầu. Cụ thể, các khuôn mặt ở các quốc gia hoặc các khu vực khác nhau có các đặc trưng/đặc điểm khác nhau và độ lệch vốn có của chúng. Do đó, hiệu suất ở các quốc gia hoặc các khu vực địa lý khác có thể không tốt như hiệu suất ở nước sở tại.

Do đó, có nhu cầu đề xuất các phương pháp và các thiết bị có thể tối thiểu hóa sự giảm đáng kể về hiệu suất khi mô hình “nước sở tại tạo ra” được triển khai ở nước ngoài.

Bản chất kỹ thuật của sáng chế

Để nâng cao hiệu suất của hệ thống nhận diện khuôn mặt khi mô hình được huấn luyện trước “nước sở tại tạo ra” được triển khai ở nước ngoài, các phương án của sáng chế sử dụng kỹ thuật thích ứng miền. Kỹ thuật thích ứng miền liên quan đến việc sửa đổi hoặc điều chỉnh mô hình được huấn luyện trước được huấn luyện trên miền nguồn (tức là được huấn luyện sử dụng dữ liệu quy mô lớn thu được từ một quốc gia hoặc khu vực địa lý cụ thể, ví dụ Trung Quốc), sao cho hiệu suất trên miền đích (ví dụ khi được triển khai trong một quốc gia hoặc khu vực địa lý khác) có thể được cải thiện.

Theo một phương án, mô hình được huấn luyện trước “nước sở tại tạo ra” ban đầu được làm thích ứng với miền đích mới bằng cách sử dụng một hoặc nhiều lớp bổ sung với các tham số huấn luyện được. Một hoặc nhiều lớp bổ sung có thể được sắp xếp theo cấu hình xếp tầng hoặc song song.

Theo một phương án khác, sáng chế đề xuất phương pháp bao gồm các bước: huấn luyện mạng nơron được huấn luyện trước bao gồm: lớp đầu vào; các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn; và lớp đầu ra. Bước huấn luyện mạng nơron được huấn luyện trước bao gồm: đưa ít nhất một lớp bổ sung vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có trọng số kết hợp; giữ cho trọng số của các nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi; nhập dữ liệu từ miền đích vào lớp đầu vào; và điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm thu được ở lớp đầu ra.

Theo một phương án khác, sáng chế đề xuất hệ thống huấn luyện mạng nơron

bao gồm môđun bộ xử lý và môđun bộ nhớ bao gồm mã chương trình máy tính. Môđun bộ nhớ và mã chương trình máy tính được tạo cấu hình để, với môđun bộ xử lý, khiến cho hệ thống thực hiện ít nhất các bước: truy hồi, từ môđun bộ nhớ, mạng noron được huấn luyện trước bao gồm: (i) lớp đầu vào; (ii) các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn; và (iii) lớp đầu ra. Hệ thống còn thực hiện các bước: đưa ít nhất một lớp bổ sung vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có trọng số kết hợp; giữ cho trọng số của các nút ở các lớp ẩn của mạng noron được huấn luyện trước không thay đổi; nhập dữ liệu từ miền đích vào lớp đầu vào; và điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm thu được ở lớp đầu ra.

Theo một phương án khác, sáng chế đề xuất phương pháp được thực thi bằng máy tính, bao gồm các bước: nhập, thông qua môđun đầu vào, dữ liệu từ miền đích vào lớp đầu vào của mạng noron được huấn luyện trước, trong đó mạng noron được huấn luyện trước bao gồm: các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn; và ít nhất một lớp bổ sung được đưa vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có trọng số kết hợp. Phương pháp này còn bao gồm các bước: thu, từ môđun đầu ra, các đặc điểm từ lớp đầu ra của mạng noron được huấn luyện trước; và tạo cấu hình môđun xử lý để điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm thu được trong khi giữ cho trọng số của các nút ở các lớp ẩn của mạng noron được huấn luyện trước không thay đổi.

Theo một phương án khác, sáng chế đề xuất thiết bị bao gồm: môđun đầu vào được tạo cấu hình để nhận dữ liệu từ miền đích; và môđun xử lý được tạo cấu hình để cấp dữ liệu vào lớp đầu vào của mạng noron được huấn luyện trước, trong đó mạng noron được huấn luyện trước bao gồm: các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn; và ít nhất một lớp bổ sung được đưa vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có trọng số kết hợp. Thiết bị còn bao gồm môđun đầu ra được tạo cấu hình để xuất ra các đặc điểm từ lớp đầu ra của mạng noron được huấn luyện trước. Môđun xử lý còn được tạo cấu hình để

điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm được xuất ra trong khi giữ cho trọng số của các nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi.

Mô tả vắn tắt các hình vẽ

Các phương án chỉ được cung cấp theo cách làm ví dụ và sẽ được hiểu rõ hơn đối với người có hiểu biết trung bình trong lĩnh vực kỹ thuật từ phần mô tả chi tiết sau đây dựa vào các hình vẽ kèm theo, trong đó:

Fig.1 là sơ đồ kiến trúc của mạng nơron được huấn luyện trước.

Fig.2 là sơ đồ kiến trúc của mạng nơron, theo một phương án của sáng chế.

Fig.3 là sơ đồ kiến trúc của mạng nơron, theo một phương án của sáng chế.

Fig.4 là lưu đồ minh họa phương pháp huấn luyện mạng nơron được sử dụng để trích xuất các đặc điểm khuôn mặt từ các hình ảnh khuôn mặt, theo một phương án của sáng chế.

Fig.5 thể hiện sơ đồ giản lược của hệ thống máy tính thích hợp để sử dụng trong việc thực hiện ít nhất một số bước của phương pháp huấn luyện mạng nơron được sử dụng để trích xuất các đặc điểm khuôn mặt từ các hình ảnh khuôn mặt.

Mô tả chi tiết sáng chế

Các phương án sẽ được mô tả, chỉ bằng cách lấy ví dụ, dựa vào các hình vẽ. Các số chỉ dẫn và các ký tự giống nhau trên các hình vẽ chỉ các thành phần tương tự hoặc tương đương.

Một số phần của phần mô tả sau đây trình bày rõ ràng hoặc ẩn ý về các thuật toán và các biểu diễn hàm hoặc ký hiệu của các phép toán trên dữ liệu bên trong bộ nhớ máy tính. Các phần mô tả thuật toán này và các biểu diễn hàm hoặc ký hiệu là phương tiện được sử dụng bởi người có hiểu biết trung bình trong các lĩnh vực xử lý dữ liệu để chuyển tải hiệu quả nhất bản chất công việc của họ tới người có hiểu biết trung bình khác trong lĩnh vực kỹ thuật. Thuật toán ở đây, và nói chung, được hiểu là một trình tự các bước nhất quán dẫn đến kết quả mong muốn. Các bước là các bước yêu cầu các thao tác vật lý của các đại lượng vật lý, như tín hiệu điện, từ hoặc quang có khả năng được lưu trữ, được chuyển, kết hợp, so sánh và được vận hành theo cách khác.

Trừ khi được nêu rõ theo cách khác và như rõ ràng từ phần sau đây, cần hiểu rằng trong toàn bộ bản mô tả sáng chế, các phần mô tả sử dụng các thuật ngữ như “thu”, “quét”, “tính”, “xác định”, “thay thế”, “tạo ra”, “khởi tạo”, “xuất ra” hoặc tương tự, chỉ hoạt động và các quy trình của hệ thống máy tính hoặc thiết bị điện tử tương tự, mà vận hành và biến đổi dữ liệu được biểu diễn dưới dạng các đại lượng vật lý bên trong hệ thống máy tính thành dữ liệu khác được biểu diễn tương tự dưới dạng các đại lượng vật lý bên trong hệ thống máy tính hoặc các thiết bị lưu trữ, truyền hoặc hiển thị thông tin khác.

Bản mô tả sáng chế cũng bộc lộ thiết bị để thực hiện các thao tác của các phương pháp này. Thiết bị như vậy có thể được tạo kết cấu đặc biệt cho các mục đích cần thiết hoặc có thể bao gồm máy tính hoặc thiết bị khác được kích hoạt hoặc được tái cấu hình chọn lọc bởi chương trình máy tính được lưu trong máy tính. Các thuật toán và các hiển thị được biểu diễn ở đây vốn không liên quan đến máy tính cụ thể bất kỳ hoặc thiết bị khác. Các máy khác nhau được sử dụng với các chương trình phù hợp với nội dung được đề cập trong phần mô tả. Theo cách khác, việc xây dựng nhiều thiết bị chuyên dụng để thực hiện các bước của phương pháp cần thiết có thể là phù hợp. Kết cấu của máy tính thích hợp để thực thi các phương pháp/các quy trình khác nhau được mô tả ở đây sẽ xuất hiện từ phần mô tả bên dưới.

Ngoài ra, bản mô tả sáng chế cũng ngầm bộc lộ chương trình máy tính, trong đó người có hiểu biết trung bình trong lĩnh vực kỹ thuật sẽ hiểu rằng các bước riêng lẻ của phương pháp được mô tả ở đây có thể được thực thi bằng mã máy tính. Chương trình máy tính không bị giới hạn ở ngôn ngữ lập trình cụ thể bất kỳ và việc thực thi của nó. Cần hiểu rằng nhiều ngôn ngữ lập trình và việc mã hóa của chúng có thể được sử dụng để triển khai các nội dung trong phần mô tả. Ngoài ra, chương trình máy tính không bị giới hạn ở tiến trình điều khiển cụ thể bất kỳ. Có nhiều biến thể khác của chương trình máy tính có thể sử dụng các tiến trình điều khiển khác nhau mà không vượt ra ngoài phạm vi bảo hộ của sáng chế.

Hơn nữa, một hoặc nhiều bước của chương trình máy tính có thể được thực hiện song song thay vì tuần tự. Chương trình máy tính này có thể được lưu trên vật ghi đọc được bằng máy tính. Vật ghi đọc được bằng máy tính có thể bao gồm các thiết bị lưu trữ như các đĩa từ hoặc quang, các chip bộ nhớ hoặc các thiết bị lưu trữ khác thích hợp để giao tiếp với máy tính. Vật ghi đọc được bằng máy tính cũng có thể bao gồm vật ghi

có dây cứng như trong hệ thống Internet hoặc vật ghi không dây như trong hệ thống điện thoại di động GSM. Chương trình máy tính khi được tải và thực thi trên máy tính như vậy dẫn đến thiết bị thực thi các bước của phương pháp được ưu tiên.

Nhận diện khuôn mặt là công nghệ có khả năng định danh hoặc xác minh con người từ dữ liệu hình ảnh, ví dụ được chụp từ ảnh kỹ thuật số hoặc khung video từ nguồn video. Nhận diện khuôn mặt thông thường liên quan đến kỹ thuật trích xuất các đặc điểm từ các hình ảnh khuôn mặt. Các đặc điểm khuôn mặt được trích xuất được sử dụng để định danh và xác minh con người.

Để cải thiện hiệu suất của hệ thống nhận diện khuôn mặt khi mô hình được huấn luyện trước “nước sở tại tạo ra” được triển khai ở nước ngoài, các phương án của sáng chế sử dụng kỹ thuật thích ứng miền. Kỹ thuật thích ứng miền liên quan đến sự sửa đổi hoặc điều chỉnh mô hình được huấn luyện trước được huấn luyện trên miền nguồn (tức là được huấn luyện sử dụng dữ liệu quy mô lớn thu được từ một quốc gia hoặc khu vực địa lý cụ thể, ví dụ Trung Quốc), sao cho hiệu suất ở miền đích (ví dụ khi được triển khai trong một quốc gia hoặc khu vực địa lý khác) có thể được cải thiện.

Theo một phương án, mô hình được huấn luyện trước “nước sở tại tạo ra” ban đầu được làm thích ứng với miền đích mới bằng cách sử dụng một hoặc nhiều lớp bổ sung với các tham số huấn luyện được. Các tham số chỉ các trọng số và/hoặc các độ lệch của mô hình. Một hoặc nhiều lớp bổ sung có thể được sắp xếp theo cấu hình xếp tầng hoặc song song. Các chi tiết khác về các cấu hình song song và xếp tầng sẽ được cung cấp bên dưới lần lượt dựa vào Fig.2 và Fig.3.

Theo một phương án, sáng chế đề xuất mạng nơron được sử dụng để trích xuất các đặc điểm khuôn mặt từ các hình ảnh khuôn mặt khi được triển khai ở nước ngoài. Mạng nơron gồm có mạng nơron được huấn luyện trước được huấn luyện với dữ liệu hình ảnh khuôn mặt từ miền nguồn. Mạng nơron huấn luyện trước gồm có lớp đầu vào, các lớp ẩn và lớp đầu ra.

Mạng nơron gồm có ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước bên trong các lớp ẩn của mạng nơron được huấn luyện trước. Mạng nơron có thể được huấn luyện theo giao thức sau đây: (i) giữ cho các trọng số được áp dụng cho mỗi nút trong số các nút ở các lớp ẩn của mạng nơron huấn luyện trước đóng băng và không thay đổi; (ii) chọn các mẫu tin cậy nhất (tức là các đặc điểm với ít lỗi

huấn luyện nhất) từ miền đích để huấn luyện mỗi lớp trong số ít nhất một lớp bổ sung; và (iii) trong mỗi lần lặp huấn luyện, điều chỉnh các trọng số được áp dụng cho mỗi nút trong số các nút ở mỗi lớp trong số ít nhất một lớp bổ sung dựa vào các mẫu tin cậy nhất được chọn.

Ít nhất một lớp bổ sung có thể được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước. Phép ghép đầu ra từ ít nhất một lớp bổ sung và đầu ra từ lớp ẩn cuối cùng của mạng nơron được huấn luyện trước được chuyển sang lớp đầu ra. Cấu hình này được gọi là “cấu hình song song”.

Theo cách khác, ít nhất một lớp bổ sung có thể được đưa vào mạng nơron được huấn luyện trước sau lớp ẩn cuối cùng của mạng nơron được huấn luyện trước. Cấu hình này được gọi là “cấu hình xếp tầng”.

Theo cách khác, “cấu hình lai tạp” (tức là tổ hợp của “cấu hình song song” và “cấu hình xếp tầng”) có thể được triển khai, trong đó ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước sau lớp ẩn cuối cùng của mạng nơron được huấn luyện trước và ít nhất một lớp bổ sung khác được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước. “Cấu hình xếp tầng” khiến cho mạng nơron sâu hơn trong khi “cấu hình song song” khiến cho mạng nơron rộng hơn.

Trong bất kỳ một trong các cấu hình trên, có thể bổ sung nhiều hơn một lớp bổ sung. Ví dụ, trong “cấu hình xếp tầng”, việc bổ sung nhiều lớp khiến cho mạng nơron sâu hơn. Trong “cấu hình song song”, việc bổ sung nhiều lớp khiến cho mạng nơron rộng hơn.

Fig.1 là sơ đồ kiến trúc của mạng nơron được huấn luyện trước 100. Mạng nơron 100 bao gồm một số nút liên kết cao, được tổ chức dưới dạng ba hoặc nhiều hơn ba lớp: lớp đầu vào 102, các lớp ẩn 104 và lớp đầu ra 106.

Trong giai đoạn huấn luyện mạng nơron 100, dữ liệu 108 (ở dạng các hình ảnh khuôn mặt từ miền nguồn) được đưa vào mạng nơron 100 qua lớp đầu vào 102. Không có phép toán nào được áp dụng trên dữ liệu đầu vào 108; lớp đầu vào 102 lấy dữ liệu 108 và chuyển chúng lên các lớp ẩn 104.

Các lớp ẩn 104 có các nút (không được thể hiện) áp dụng các biến đổi khác nhau đối với dữ liệu 108. Mỗi lớp ẩn có tập hợp các nút được xếp chồng theo phương thẳng đứng. Dữ liệu được xử lý ở các lớp ẩn 104 dựa vào các trọng số và/hoặc độ lệch được áp dụng cho các nút này. Cụ thể, việc xử lý xảy ra qua hệ thống các kết nối đặc trưng bởi các tham số như các trọng số và độ lệch. Tại mỗi nút, đầu vào được nhận, nút tính tổng trọng số và cũng có thể thêm độ lệch. Theo kết quả và hàm kích hoạt đặt trước, nút quyết định nó có nên được kích hoạt hay không. Nút truyền thông tin xuống các nút kết nối khác. Trọng số biểu diễn cường độ kết nối giữa các nút. Nói cách khác, trọng số quyết định mức độ ảnh hưởng của đầu vào đối với đầu ra .

Lớp ẩn cuối cùng 104a có thể được gọi là lớp đặc điểm khuôn mặt và là lớp gần cuối khi huấn luyện mạng. Lớp đầu ra 106 là lớp cuối trong mạng nơron 100 và nhận đầu vào từ lớp ẩn cuối cùng 104a. Ở lớp đầu ra 106, thu được các đặc điểm khuôn mặt quan tâm. Đầu ra ở lớp đầu ra 106 chứa các đặc điểm khuôn mặt được gắn nhãn. Trong quá trình triển khai, lớp đầu ra 106 (mà được gắn nhãn) được loại bỏ và lớp ẩn cuối cùng 104a xuất ra các đặc điểm mà người sử dụng quan tâm. Theo cách này, mô hình được huấn luyện được xây dựng sử dụng dữ liệu quy mô lớn 108 từ miền nguồn. Thông thường, dữ liệu 108 từ miền nguồn thu được từ khu dân cư cụ thể (ví dụ một quốc gia hoặc khu vực địa lý). Mô hình được huấn luyện hoạt động tốt đối với các ứng dụng trong nước (sở tại) hoặc khu vực địa lý cụ thể (ví dụ Trung Quốc). Tuy nhiên, khi mô hình được huấn luyện trước được sử dụng trong một quốc gia hoặc khu vực khác (ví dụ châu Âu), hiệu suất có thể giảm do sự thay đổi của các miền ở đó mẫu được huấn luyện.

Fig.2 là sơ đồ kiến trúc của mạng nơron 200, theo một phương án của sáng chế. Mạng nơron 200 gồm có một số nút liên kết, được tổ chức dưới dạng ba hoặc nhiều hơn ba lớp: lớp đầu vào 202, các lớp ẩn 204 và lớp đầu ra 206.

Khác biệt chính giữa mạng nơron 100 được mô tả ở trên và mạng nơron 200 là việc đưa lớp mới 216 vào trước lớp đầu ra 206. Nói cách khác, lớp cuối cùng 204a của các lớp ẩn 204 bao gồm lớp ẩn cuối cùng ban đầu 104a và lớp đặc điểm mới thêm 216.

Lớp đầu vào 202 về cơ bản giống với lớp đầu vào 102. Ngoại trừ lớp đặc điểm mới thêm 216, mỗi lớp trong số các lớp ẩn 204 về cơ bản giống với mỗi lớp trong số các lớp ẩn tương ứng 104. Nói cách khác, ngoại trừ lớp đặc điểm mới thêm 216, các trọng số và/hoặc độ lệch áp dụng cho các nút ở các lớp ẩn 204 vẫn không được thay đổi.

Trong giai đoạn làm thích ứng/huấn luyện mạng noron 200, dữ liệu 208 (ở dạng các hình ảnh khuôn mặt từ miền đích) được đưa vào mạng noron 200 qua lớp đầu vào 202. Trái ngược với dữ liệu 108 được mô tả ở trên, dữ liệu 208 có thể là các hình ảnh khuôn mặt từ miền đích, miền đích là quốc gia, khu vực địa lý hoặc vùng khác với “nước sở tại” ban đầu (tức là miền nguồn). Lớp đầu vào 202 lấy dữ liệu 208 và chuyển chúng sang các lớp ẩn 204.

Các lớp ẩn 204 có các nút (không được thể hiện) áp dụng các biến đổi khác nhau đối với dữ liệu 208. Mỗi lớp ẩn có thể có tập hợp các nút được xếp chồng theo phương thẳng đứng. Dữ liệu được xử lý ở các lớp ẩn dựa vào các trọng số và/hoặc độ lệch được áp dụng cho các nút này. Cụ thể, việc xử lý xảy ra qua hệ thống các kết nối đặc trưng bởi các tham số như các trọng số và độ lệch. Tại mỗi nút, đầu vào được nhận, nút tính tổng trọng số và cũng có thể thêm độ lệch. Theo kết quả và hàm kích hoạt đặt sẵn, nút quyết định nó có nên được kích hoạt hay không. Nút truyền thông tin xuống các nút kết nối khác. Trọng số biểu diễn cường độ kết nối giữa các nút. Nói cách khác, trọng số quyết định mức độ ảnh hưởng của đầu vào đối với đầu ra. Như được đề cập, ngoại trừ lớp đặc điểm mới thêm 216, các trọng số được áp dụng cho mỗi nút trong số các nút ở các lớp ẩn 204 vẫn không được thay đổi.

Lớp đầu ra 206 là lớp cuối trong mạng noron 200 và nhận đầu vào từ lớp ẩn cuối cùng 204a. Lớp ẩn cuối cùng 204a bao gồm hai lớp tách biệt – lớp ẩn cuối cùng ban đầu 104a và lớp đặc điểm mới thêm 216.

Việc bổ sung lớp đặc điểm 216 trong mạng noron 200 có thể được coi là “cấu hình song song”. Trong cấu hình song song, các tham số huấn luyện được bổ sung (ví dụ trọng số và độ lệch) được đưa vào ở lớp đặc điểm mới thêm 216. Các tham số bổ sung có thể biến đổi các đặc điểm ban đầu thành các đặc điểm thích ứng mới, mà thực hiện tốt hơn so với các đặc điểm ban đầu. Các tham số bổ sung có thể tạo ra tập các đặc điểm mới, song song với các đặc điểm ban đầu. Phép ghép đặc điểm ban đầu và đặc điểm mới được sử dụng làm đặc điểm cuối cùng. Đặc điểm ban đầu không được thay đổi, trong khi đặc điểm bổ sung thu thập các đặc điểm biến thể miền của miền đích. Quá trình thích ứng có thể được coi là quá trình tinh chỉnh mô hình. Như được mô tả ở trên, trong quá trình tinh chỉnh, các tham số của mô hình ban đầu (ví dụ trọng số và độ lệch) được giữ đóng băng và không thay đổi trong khi các tham số mới thêm (ví dụ trọng số và độ lệch) của lớp đặc điểm mới thêm 216 được huấn luyện.

Đầu ra ở lớp đầu ra 206 được gọi là “đặc điểm cuối cùng” là phép ghép đặc điểm ban đầu (từ lớp đầu ra ban đầu 106) và đặc điểm mới (từ lớp đặc điểm mới thêm 216). Trong ngữ cảnh này, các đặc điểm này là các đặc điểm trừu tượng từ mạng nơron. Từ quan điểm của mạng nơron, đặc điểm ban đầu là đặc điểm từ lớp đặc điểm ban đầu 204a và đặc điểm mới là đặc điểm từ (các) lớp mới thêm 216. Đặc điểm cuối cùng là phép ghép lớp đặc điểm ban đầu 204a và lớp mới thêm 216. Lớp 206 cũng được gọi là lớp được gắn nhãn, chỉ được sử dụng trong huấn luyện, không triển khai.

Fig.3 là sơ đồ kiến trúc của mạng nơron 300, theo một phương án của sáng chế. Mạng nơron 300 gồm có một số nút liên kết, được tổ chức dưới dạng ba hoặc nhiều hơn ba lớp: lớp đầu vào 302, các lớp ẩn 304 và lớp đầu ra 306.

Khác biệt chính giữa mạng nơron 100 được mô tả ở trên và mạng nơron 300 là việc đưa vào lớp mới 316 làm lớp ẩn cuối cùng.

Lớp đầu vào 302 về cơ bản giống với lớp đầu vào 102. Mỗi lớp trong số các lớp ẩn 304 về cơ bản giống với mỗi lớp trong số các lớp ẩn tương ứng 104, ngoại trừ lớp cuối cùng 316 trong các lớp ẩn 304. Nói cách khác, ngoại trừ lớp cuối cùng 316, các trọng số và/hoặc độ lệch được áp dụng cho các nút ở các lớp ẩn 304 vẫn không được thay đổi.

Trong giai đoạn thích ứng/huấn luyện mạng nơron 300, dữ liệu 308 (ở dạng các hình ảnh khuôn mặt từ miền đích) được đưa vào mạng nơron 300 qua lớp đầu vào 302. Trái ngược với dữ liệu 108 được mô tả ở trên, dữ liệu 308 ở dạng các hình ảnh khuôn mặt từ miền đích, miền đích là quốc gia, khu vực địa lý hoặc vùng khác với “nước sở tại” (tức là miền nguồn). Lớp đầu vào 302 lấy dữ liệu 308 và chuyển chúng sang các lớp ẩn 304.

Các lớp ẩn 304 có các nút (không được thể hiện) áp dụng các biến đổi khác nhau đối với dữ liệu 308. Mỗi lớp ẩn có thể có tập hợp các nút được xếp chồng theo phương thẳng đứng. Dữ liệu được xử lý ở các lớp ẩn dựa vào các trọng số và/hoặc độ lệch được áp dụng cho các nút này. Cụ thể, việc xử lý xảy ra qua hệ thống các kết nối đặc trưng bởi các tham số như các trọng số và độ lệch. Tại mỗi nút, đầu vào được nhận, nút tính tổng trọng số và cũng có thể thêm độ lệch. Theo kết quả và hàm kích hoạt đặt sẵn, nút quyết định nó có nên được kích hoạt hay không. Nút truyền thông tin xuống các nút kết nối khác. Trọng số biểu diễn cường độ kết nối giữa các nút. Nói cách khác, trọng số

quyết định mức độ ảnh hưởng của đầu vào đối với đầu ra. Như đã nêu, ngoại trừ lớp cuối cùng 316, các trọng số được áp dụng cho mỗi nút trong số các nút ở các lớp ẩn 304 vẫn không được thay đổi.

Lớp đầu ra 306 là lớp cuối cùng trong mạng nơron 300 và nhận đầu vào từ lớp ẩn cuối cùng 316.

Việc bổ sung lớp đặc điểm 316 trong mạng nơron 300 có thể được coi là “cấu hình xếp tầng”. Trong cấu hình xếp tầng, các tham số huấn luyện được bổ sung (ví dụ trọng số và độ lệch) được đưa vào ở lớp đặc điểm mới thêm 316. Các tham số bổ sung có thể tạo ra tập các đặc điểm mới. Các đặc điểm bổ sung thu thập các đặc điểm biến thể miền của miền đích. Quá trình thích ứng có thể được coi là quá trình tinh chỉnh mô hình. Như được mô tả ở trên, trong quá trình tinh chỉnh, các tham số của mô hình ban đầu (ví dụ trọng số và độ lệch) được giữ đóng băng và không thay đổi. Các tham số mới thêm (ví dụ trọng số và độ lệch) của lớp đặc điểm mới thêm 316 được huấn luyện. Đầu ra ở lớp đầu ra 306 được gọi là “đặc điểm cuối cùng”.

Giai đoạn thích ứng của các mạng nơron 200 và 300 được mô tả ở trên có thể được coi là “học tự điều chỉnh nhịp độ” liên quan đến quá trình lặp để cải thiện dần hiệu suất của mô hình được huấn luyện trước trên miền đích mới. Đối với mỗi lần lặp, người sử dụng chọn các mẫu tin cậy nhất (tức là các đặc điểm có ít lỗi huấn luyện nhất) từ miền đích để huấn luyện các lớp bổ sung.

Fig.4 là lưu đồ minh họa phương pháp 400 để huấn luyện mạng nơron được sử dụng để trích xuất các đặc điểm khuôn mặt từ các hình ảnh khuôn mặt, theo một phương án của sáng chế. Mỗi bước trong số các bước của phương pháp có thể được thực hiện tuần tự; hoặc khi có thể áp dụng, thực hiện song song hoặc theo thứ tự bất kỳ.

Phương pháp 400 bao gồm bước 402 để tạo ra mạng nơron được huấn luyện trước mà đã được huấn luyện với dữ liệu hình ảnh khuôn mặt từ miền nguồn. Mạng nơron được huấn luyện trước gồm có các lớp ẩn.

Phương pháp 400 còn bao gồm bước 404 để đưa ít nhất một lớp bổ sung vào các lớp ẩn của mạng nơron được huấn luyện trước.

Phương pháp 400 còn bao gồm bước 406 để giữ cho các trọng số mà được áp dụng cho mỗi nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi.

Phương pháp 400 còn bao gồm bước 408 để nhập dữ liệu hình ảnh khuôn mặt từ miền đích vào lớp đầu vào của mạng nơron.

Phương pháp 400 còn bao gồm bước 410 để điều chỉnh lặp đi lặp lại các trọng số của các nút tương ứng ở ít nhất một lớp bổ sung dựa vào các đặc điểm khuôn mặt thu được ở lớp đầu ra của mạng nơron.

Ít nhất một lớp bổ sung có thể được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước. Trong trường hợp này, phương pháp 400 còn bao gồm các bước: (i) thực hiện ghép đầu ra từ ít nhất một lớp bổ sung và đầu ra từ lớp ẩn cuối cùng của mạng nơron được huấn luyện trước; và (ii) chuyển các đầu ra được ghép đến lớp đầu ra của mạng nơron.

Ít nhất một lớp bổ sung có thể được đưa vào mạng nơron được huấn luyện trước sau lớp ẩn cuối cùng của mạng nơron được huấn luyện trước. Theo cách khác, ít nhất một lớp bổ sung có thể được đưa vào mạng nơron được huấn luyện trước sau lớp ẩn cuối cùng của mạng nơron được huấn luyện trước và ít nhất một lớp bổ sung khác có thể được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước.

Liên quan đến bước 410, phương pháp 400 có thể còn bao gồm các bước con sau đây: (i) chọn, từ các đặc điểm khuôn mặt thu được ở lớp đầu ra của mạng nơron, các đặc điểm khuôn mặt tin cậy nhất (tức là các đặc điểm có ít lỗi huấn luyện nhất); và (ii) trong mỗi lần lặp huấn luyện, điều chỉnh các trọng số được áp dụng cho mỗi nút ở mỗi lớp trong số ít nhất là một lớp bổ sung dựa vào các đặc điểm khuôn mặt tin cậy nhất được chọn.

Miền nguồn được liên kết với quốc gia hoặc khu vực địa lý thứ nhất và miền đích được liên kết với quốc gia hoặc khu vực địa lý thứ hai khác quốc gia hoặc khu vực địa lý thứ nhất.

Các kỹ thuật được mô tả trong phần mô tả này tạo ra một hoặc nhiều hiệu quả kỹ thuật. Cụ thể, các phương án của sáng chế giúp cải thiện hiệu quả các đặc điểm ở miền đích mới bằng cách biến đổi đặc điểm và gia tăng đặc điểm. Cấu hình xếp tầng có thể biến đổi các đặc điểm ban đầu thành các đặc điểm thích ứng; trong khi cấu hình song song có thể tạo ra các đặc điểm bổ sung để gia tăng các đặc điểm ban đầu. Cả cấu hình xếp tầng lẫn cấu hình song song có thể nâng cao và cải thiện các đặc điểm ban đầu. Bởi

vậy, bằng cách sử dụng các phương án của sáng chế, mô hình có thể trích xuất nhiều đặc điểm mạnh mẽ hơn và đạt được hiệu suất tốt hơn trong miền đích.

Các phương án của sáng chế giúp tạo ra cấu trúc mô hình mới, trong đó các lớp bổ sung với các tham số huấn luyện được áp dụng trên mô hình được huấn luyện trước. Các lớp bổ sung có thể được sắp xếp theo cấu hình xếp tầng hoặc song song.

Trong cấu hình song song, lớp bổ sung được bổ sung song song với lớp ẩn cuối cùng của mô hình được huấn luyện trước. Trong cấu hình xếp tầng, lớp bổ sung được bổ sung sau lớp ẩn cuối cùng của mô hình được huấn luyện trước.

Trong cả hai trường hợp, trong quá trình huấn luyện mô hình mới (tức là, với lớp bổ sung), các trọng số ban đầu được áp dụng cho mỗi nút trong số các nút trong các lớp ẩn của mô hình được huấn luyện trước được giữ đóng băng và vẫn không thay đổi. Đối với mỗi lần lặp huấn luyện, các mẫu tin cậy nhất được chọn từ miền đích mới để huấn luyện lớp bổ sung.

Fig.5 thể hiện sơ đồ giản lược của hệ thống máy tính thích hợp để sử dụng trong việc thực hiện ít nhất một số bước của phương pháp huấn luyện mạng nơron được sử dụng để trích xuất các đặc điểm khuôn mặt từ các hình ảnh khuôn mặt.

Phần mô tả sau đây về hệ thống máy tính/thiết bị điện toán 500 được cung cấp chỉ bằng cách lấy ví dụ và không nhằm mục đích giới hạn.

Như được thể hiện trên Fig.5, thiết bị điện toán ví dụ 500 gồm có bộ xử lý 504 để thực thi các đoạn chương trình phần mềm. Mặc dù một bộ xử lý duy nhất được thể hiện nhằm mục đích rõ ràng, thiết bị điện toán 500 cũng có thể bao gồm hệ thống nhiều bộ xử lý. Bộ xử lý 504 được kết nối với kết cấu hạ tầng truyền thông 506 để truyền thông với các thành phần khác của thiết bị điện toán 500. Kết cấu hạ tầng truyền thông 506 có thể bao gồm, ví dụ, bus truyền thông, thanh ngang hoặc mạng.

Thiết bị điện toán 500 còn bao gồm bộ nhớ chính 508, như bộ nhớ truy cập ngẫu nhiên (RAM) và bộ nhớ phụ 510. Bộ nhớ phụ 510 có thể bao gồm, ví dụ, ổ đĩa cứng 512 và/hoặc ổ đĩa lưu trữ tháo lắp được 514, mà có thể bao gồm ổ băng từ, ổ đĩa quang hoặc tương tự. Ổ đĩa lưu trữ tháo lắp được 514 đọc từ và/hoặc ghi vào bộ phận lưu trữ tháo lắp được 518 theo cách đã biết. Bộ phận lưu trữ tháo lắp được 518 có thể bao gồm băng từ, đĩa quang hoặc tương tự, mà được đọc bởi và ghi bởi ổ đĩa lưu trữ tháo lắp được 514.

Người có hiểu biết trung bình trong lĩnh vực kỹ thuật này cần hiểu rằng bộ phận lưu trữ tháo lắp được 518 gồm có vật ghi đọc được bằng máy tính có các lệnh và/hoặc dữ liệu mã chương trình thực thi được bằng máy tính được lưu trữ trong đó.

Theo một phương án thực hiện thay thế, bộ nhớ phụ 510 ngoài ra hoặc theo cách khác có thể bao gồm phương tiện tương tự khác để cho phép các chương trình máy tính hoặc các lệnh khác được tải vào trong thiết bị điện toán 500. Các phương tiện này có thể bao gồm, ví dụ, bộ phận lưu trữ tháo lắp được 522 và giao diện 520. Các ví dụ về bộ phận lưu trữ tháo lắp được 522 và giao diện 520 bao gồm chip bộ nhớ tháo lắp được (như EPROM hoặc PROM) và ổ cắm liên quan và các bộ phận lưu trữ tháo lắp được khác 522 và các giao diện 520 cho phép phần mềm và dữ liệu sẽ được truyền từ bộ phận lưu trữ tháo lắp được 522 tới hệ thống máy tính 500.

Thiết bị điện toán 500 cũng bao gồm ít nhất một giao diện truyền thông 524. Giao diện truyền thông 524 cho phép phần mềm và dữ liệu sẽ được truyền giữa thiết bị điện toán 500 và các thiết bị bên ngoài thông qua đường truyền thông 526. Theo các phương án khác nhau, giao diện truyền thông 524 cho phép dữ liệu sẽ được truyền giữa thiết bị điện toán 500 và mạng truyền thông dữ liệu, như mạng truyền thông dữ liệu công cộng hoặc riêng tư. Giao diện truyền thông 524 có thể được sử dụng để trao đổi dữ liệu giữa các thiết bị điện toán 500 khác nhau mà các thiết bị điện toán 500 này tạo thành một phần của mạng máy tính liên kết. Các ví dụ về giao diện truyền thông 524 có thể bao gồm modem, giao diện mạng (như thẻ Ethernet), cổng truyền thông, anten có mạch kết hợp và tương tự. Giao diện truyền thông 524 có thể là có dây hoặc không dây. Phần mềm và dữ liệu được truyền thông qua giao diện truyền thông 524 ở dạng các tín hiệu có thể là các tín hiệu điện tử, tín hiệu điện từ, tín hiệu quang hoặc các tín hiệu khác có khả năng thu được bởi giao diện truyền thông 524. Các tín hiệu này được cung cấp cho giao diện truyền thông qua đường truyền thông 526.

Tuỳ ý, thiết bị điện toán 500 còn bao gồm giao diện hiển thị 502 thực hiện các thao tác để kết xuất các hình ảnh tới màn hiển thị đi kèm 530 và giao diện audio 532 để thực hiện các thao tác để phát nội dung audio thông qua các loa đi kèm 534.

Như được sử dụng trong bản mô tả, thuật ngữ “sản phẩm chương trình máy tính” một phần có thể chỉ bộ phận lưu trữ tháo lắp được 518, bộ phận lưu trữ tháo lắp được 522, đĩa cứng được lắp trong ổ đĩa cứng 512 hoặc sóng mang mang phần mềm trên

đường truyền thông 526 (liên kết không dây hoặc dây cáp) tới giao diện truyền thông 524. Phương tiện lưu trữ đọc được bằng máy tính chỉ phương tiện lưu trữ hữu hình lâu dài bất kỳ mà cung cấp các lệnh và/hoặc dữ liệu được ghi tới thiết bị điện toán 500 để thực thi và/hoặc xử lý. Các ví dụ về vật ghi này bao gồm các đĩa mềm, băng từ, CD-ROM, DVD, Blu-ray™ Disc, ổ đĩa cứng, ROM hoặc mạch tích hợp, bộ nhớ USB, đĩa quang từ hoặc thẻ đọc được bằng máy tính như thẻ PCMCIA và tương tự, xem các thiết bị này ở bên trong hay bên ngoài thiết bị điện toán 500. Các ví dụ về phương tiện truyền đọc được bằng máy tính tạm thời hoặc không hữu hình cũng có thể tham gia vào việc cung cấp phần mềm, các chương trình ứng dụng, các lệnh và/hoặc dữ liệu cho thiết bị điện toán 500 bao gồm các kênh truyền radio hoặc hồng ngoại cũng như kết nối mạng với máy tính hoặc thiết bị mạng khác và mạng Internet hoặc mạng Intranet bao gồm chuyển thư điện tử và thông tin được ghi trên trang web và tương tự.

Các chương trình máy tính (cũng được gọi là mã chương trình máy tính) được lưu trong bộ nhớ chính 508 và/hoặc bộ nhớ phụ 510. Các chương trình máy tính cũng có thể thu được thông qua giao diện truyền thông 524. Các chương trình máy tính này, khi được thực thi, cho phép thiết bị điện toán 500 thực hiện một hoặc nhiều dấu hiệu của các phương án được đề cập ở đây. Theo các phương án khác nhau, các chương trình máy tính, khi được thực thi, cho phép bộ xử lý 504 thực hiện các dấu hiệu của các phương án được mô tả ở trên. Theo đó, các chương trình máy tính này biểu diễn các bộ điều khiển của hệ thống máy tính 500.

Phần mềm có thể được lưu trong sản phẩm chương trình máy tính và tải vào trong thiết bị điện toán 500 sử dụng ổ đĩa lưu trữ tháo lắp được 514, ổ đĩa cứng 512 hoặc giao diện 520. Theo cách khác, sản phẩm chương trình máy tính có thể được tải xuống hệ thống máy tính 500 qua đường truyền thông 526. Phần mềm, khi được thực thi bởi bộ xử lý 504, khiến cho thiết bị điện toán 500 thực hiện các chức năng của các phương án được mô tả ở đây.

Cần hiểu rằng, phương án trên Fig.5 được thể hiện chỉ nhằm lấy ví dụ. Do đó, theo một số phương án, một hoặc nhiều dấu hiệu của thiết bị điện toán 500 có thể được bỏ qua. Cũng vậy, theo một số phương án, một hoặc nhiều dấu hiệu của thiết bị điện toán 500 có thể được kết hợp với nhau. Ngoài ra, theo một số phương án, một hoặc nhiều dấu hiệu của thiết bị điện toán 500 có thể được chia thành một hoặc nhiều thành phần.

Theo một phương án, sáng chế đề xuất hệ thống huấn luyện mạng nơron được sử dụng để trích xuất các đặc điểm khuôn mặt từ các hình ảnh khuôn mặt. Hệ thống gồm có môđun bộ xử lý (như bộ xử lý 504) và môđun bộ nhớ (như bộ nhớ phụ 510) bao gồm mã chương trình máy tính. Môđun bộ nhớ và mã chương trình máy tính được tạo cấu hình để, với môđun bộ xử lý, khiến cho hệ thống ít nhất là truy hồi, từ môđun bộ nhớ, mạng nơron được huấn luyện trước bao gồm các lớp ẩn được huấn luyện dựa vào dữ liệu hình ảnh khuôn mặt từ miền nguồn.

Hệ thống có thể còn được tạo cấu hình để đưa ít nhất một lớp bổ sung vào các lớp ẩn của mạng nơron được huấn luyện trước và giữ cho trọng số của các nút tương ứng ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi.

Hệ thống có thể còn được tạo cấu hình để nhập dữ liệu hình ảnh khuôn mặt từ miền đích vào lớp đầu vào của mạng nơron và điều chỉnh lặp đi lặp lại trọng số của các nút tương ứng ở ít nhất một lớp bổ sung dựa vào các đặc điểm khuôn mặt thu được ở lớp đầu ra của mạng nơron.

Ít nhất một lớp bổ sung có thể được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước. Hệ thống có thể còn được tạo cấu hình để thực hiện ghép các đầu ra từ ít nhất một lớp bổ sung và đầu ra từ lớp ẩn cuối cùng của mạng nơron được huấn luyện trước; và chuyển các đầu ra được ghép đến lớp đầu ra của mạng nơron.

Ít nhất một lớp bổ sung cũng có thể được đưa vào mạng nơron được huấn luyện trước sau lớp ẩn cuối cùng của mạng nơron được huấn luyện trước.

Hệ thống có thể còn được tạo cấu hình để chọn, từ các đặc điểm khuôn mặt thu được ở lớp đầu ra của mạng nơron, các đặc điểm khuôn mặt có ít lỗi huấn luyện nhất; và điều chỉnh lặp đi lặp lại trọng số của các nút tương ứng ở ít nhất một lớp bổ sung dựa vào các đặc điểm khuôn mặt được chọn có ít lỗi huấn luyện nhất.

Miền nguồn được liên kết với quốc gia hoặc khu vực địa lý thứ nhất và miền đích được liên kết với quốc gia hoặc khu vực địa lý thứ hai khác quốc gia hoặc khu vực địa lý thứ nhất.

Theo một phương án khác, sáng chế đề xuất phương pháp được thực thi bằng máy tính để nhận diện khuôn mặt, bao gồm các bước sau đây. Bước thứ nhất bao gồm

nhập, thông qua môđun đầu vào, dữ liệu hình ảnh khuôn mặt từ miền đích vào lớp đầu vào của mạng nơron. Mạng nơron bao gồm (i) mạng nơron được huấn luyện trước bao gồm các lớp ẩn được huấn luyện dựa vào dữ liệu hình ảnh khuôn mặt từ miền nguồn; và (ii) ít nhất một lớp bổ sung được đưa vào các lớp ẩn của mạng nơron được huấn luyện trước. Dữ liệu hình ảnh khuôn mặt sau đó được chuyển qua các lớp ẩn để xử lý.

Bước tiếp theo bao gồm thu, từ môđun đầu ra, các đặc điểm khuôn mặt từ lớp đầu ra của mạng nơron.

Bước tiếp theo bao gồm tạo cấu hình môđun xử lý để điều chỉnh lặp đi lặp lại trọng số của các nút tương ứng ở ít nhất một lớp bổ sung dựa vào các đặc điểm khuôn mặt thu được trong khi giữ cho trọng số của các nút tương ứng ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi.

Theo một phương án khác, sáng chế đề xuất thiết bị nhận diện khuôn mặt. Thiết bị gồm có môđun đầu vào, môđun đầu ra và môđun xử lý. Môđun đầu vào được tạo cấu hình để thu dữ liệu hình ảnh khuôn mặt từ miền đích. Môđun xử lý được tạo cấu hình để nạp dữ liệu hình ảnh khuôn mặt cho lớp đầu vào của mạng nơron. Mạng nơron bao gồm: (i) mạng nơron được huấn luyện trước bao gồm các lớp ẩn được huấn luyện dựa vào dữ liệu hình ảnh khuôn mặt từ miền nguồn; và (ii) ít nhất một lớp bổ sung được đưa vào các lớp ẩn của mạng nơron được huấn luyện trước. Dữ liệu hình ảnh khuôn mặt sau đó được chuyển qua các lớp ẩn để xử lý. Môđun đầu ra được tạo cấu hình để xuất ra các đặc điểm khuôn mặt từ lớp đầu ra của mạng nơron.

Môđun xử lý còn được tạo cấu hình để điều chỉnh lặp đi lặp lại trọng số của các nút tương ứng ở ít nhất một lớp bổ sung dựa vào các đặc điểm khuôn mặt thu được trong khi giữ cho trọng số của các nút tương ứng ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi.

Một hoặc nhiều thành phần của hệ thống máy tính 500 có thể được sử dụng để nhận biết môđun đầu vào, môđun đầu ra và/hoặc môđun xử lý nêu trên.

Người có hiểu biết trung bình trong lĩnh vực kỹ thuật cần hiểu rằng nhiều biến thể và/hoặc phương án sửa đổi có thể được thực hiện đối với sáng chế như được thể hiện trong các phương án cụ thể mà không vượt ra khỏi phạm vi bảo hộ của sáng chế như được mô tả rộng rãi. Do đó, các phương án của sáng chế được coi là nhằm mục đích minh họa và không làm giới hạn sáng chế.

Ví dụ, một số phương án của sáng chế đề cập đến việc nhận diện khuôn mặt. Tuy nhiên, sáng chế không bị giới hạn ở việc nhận diện khuôn mặt và có thể được sử dụng cho các ứng dụng khác.

YÊU CẦU BẢO HỘ

1. Phương pháp huấn luyện mạng nơron bao gồm các bước:

huấn luyện mạng nơron được huấn luyện trước bao gồm:

lớp đầu vào;

các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn; và

lớp đầu ra, trong đó bước huấn luyện mạng nơron được huấn luyện trước bao gồm các bước:

đưa ít nhất một lớp bổ sung vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có trọng số kết hợp;

giữ cho trọng số của các nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi;

nhập dữ liệu từ miền đích vào lớp đầu vào, trong đó dữ liệu từ miền đích bao gồm dữ liệu có các đặc tính khác với miền nguồn; và

điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm thu được ở lớp đầu ra, trong đó ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước và trong đó phương pháp huấn luyện mạng nơron còn bao gồm các bước:

thực hiện ghép đầu ra từ ít nhất một lớp bổ sung và đầu ra từ lớp ẩn cuối cùng của mạng nơron được huấn luyện trước; và

chuyển các đầu ra được ghép đến lớp đầu ra.

2. Phương pháp theo điểm 1, trong đó ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước sau lớp ẩn cuối cùng của mạng nơron được huấn luyện trước.

3. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó phương pháp này còn bao gồm các bước:

chọn, từ các đặc điểm thu được ở lớp đầu ra, các đặc điểm có ít lỗi huấn luyện nhất; và

điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm có ít lỗi huấn luyện nhất được chọn.

4. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó miền nguồn được liên kết với quốc gia hoặc khu vực địa lý thứ nhất và miền đích được liên kết với quốc gia hoặc khu vực địa lý thứ hai khác quốc gia hoặc khu vực địa lý thứ nhất.

5. Hệ thống huấn luyện mạng nơron bao gồm:

môđun bộ xử lý; và

môđun bộ nhớ bao gồm mã chương trình máy tính;

trong đó môđun bộ nhớ và mã chương trình máy tính được tạo cấu hình để, cùng với môđun bộ xử lý, khiến cho hệ thống này thực hiện ít nhất là bước:

truy hồi, từ môđun bộ nhớ, mạng nơron được huấn luyện trước bao gồm:

lớp đầu vào;

các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn; và

lớp đầu ra;

đưa ít nhất một lớp bổ sung vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có các trọng số kết hợp;

giữ cho trọng số của các nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi;

nhập dữ liệu từ miền đích vào lớp đầu vào, trong đó dữ liệu từ miền đích bao gồm dữ liệu có các đặc tính khác với miền nguồn; và

điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm thu được ở lớp đầu ra, trong đó ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước, và trong đó hệ thống còn được tạo cấu hình để thực hiện các bước:

thực hiện ghép đầu ra từ ít nhất một lớp bổ sung và đầu ra từ lớp ẩn cuối cùng của mạng nơron được huấn luyện trước; và

chuyển các đầu ra được ghép đến lớp đầu ra.

6. Hệ thống theo điểm 5, trong đó ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước sau lớp ẩn cuối cùng của mạng nơron được huấn luyện trước.

7. Hệ thống theo điểm 5 hoặc 6, trong đó hệ thống còn được tạo cấu hình để thực hiện các bước:

chọn, từ các đặc điểm thu được ở lớp đầu ra, các đặc điểm có ít lỗi huấn luyện nhất; và

điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm có ít lỗi huấn luyện nhất được chọn.

8. Hệ thống theo điểm bất kỳ trong số các điểm từ 5 đến 7, trong đó miền nguồn được liên kết với quốc gia hoặc khu vực địa lý thứ nhất và miền đích được liên kết với quốc gia hoặc khu vực địa lý thứ hai khác quốc gia hoặc khu vực địa lý thứ nhất.

9. Phương pháp được thực thi bằng máy tính, bao gồm các bước:

nhập, thông qua môđun đầu vào, dữ liệu từ miền đích vào lớp đầu vào của mạng nơron được huấn luyện trước, trong đó mạng nơron được huấn luyện trước bao gồm:

các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn, trong đó dữ liệu từ miền nguồn bao gồm dữ liệu có các đặc tính khác với miền đích; và

ít nhất một lớp bổ sung được đưa vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có các trọng số kết hợp, và trong đó ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước;

thu, từ môđun đầu ra, các đặc điểm từ lớp đầu ra của mạng nơron được huấn luyện trước;

tạo cấu hình môđun xử lý để điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm thu được trong khi giữ cho trọng số của các nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi;

thực hiện ghép đầu ra từ ít nhất một lớp bổ sung và đầu ra từ lớp ẩn cuối cùng

của mạng nơron được huấn luyện trước; và

chuyển các đầu ra được ghép đến lớp đầu ra.

10. Thiết bị huấn luyện mạng nơron bao gồm:

môđun đầu vào được tạo cấu hình để nhận dữ liệu từ miền đích;

môđun xử lý được tạo cấu hình để cấp dữ liệu cho lớp đầu vào của mạng nơron được huấn luyện trước, trong đó mạng nơron được huấn luyện trước bao gồm:

các lớp ẩn, trong đó mỗi lớp trong số các lớp ẩn có một hoặc nhiều nút, trong đó mỗi nút trong số một hoặc nhiều nút này có trọng số kết hợp được huấn luyện dựa vào dữ liệu từ miền nguồn, trong đó dữ liệu từ miền nguồn bao gồm dữ liệu có các đặc tính khác với miền đích; và

ít nhất một lớp bổ sung được đưa vào các lớp ẩn, trong đó lớp bổ sung này có một hoặc nhiều nút có các trọng số kết hợp, và trong đó ít nhất một lớp bổ sung được đưa vào mạng nơron được huấn luyện trước song song với lớp ẩn cuối cùng của mạng nơron được huấn luyện trước;

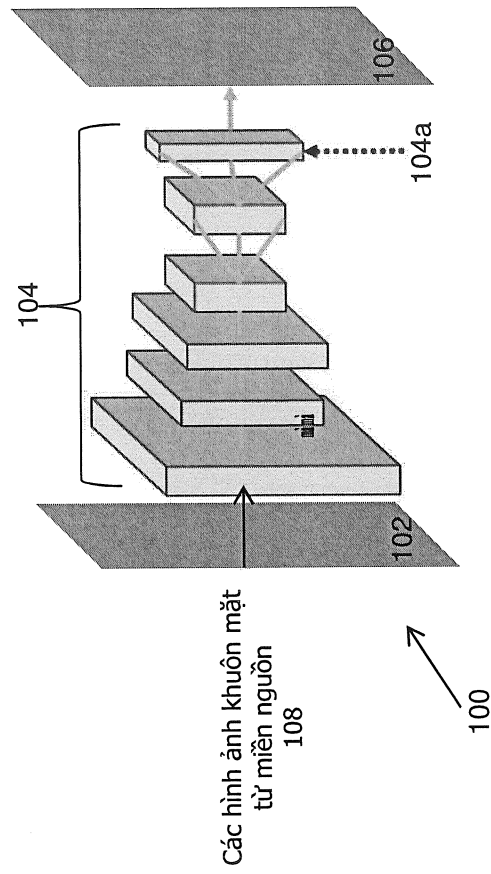
môđun đầu ra được tạo cấu hình để xuất ra các đặc điểm từ lớp đầu ra của mạng nơron được huấn luyện trước, trong đó:

môđun xử lý còn được tạo cấu hình để:

điều chỉnh trọng số của một hoặc nhiều nút ở ít nhất một lớp bổ sung dựa vào các đặc điểm được xuất ra trong khi giữ cho trọng số của các nút ở các lớp ẩn của mạng nơron được huấn luyện trước không thay đổi;

thực hiện ghép đầu ra từ ít nhất một lớp bổ sung và đầu ra từ lớp ẩn cuối cùng của mạng nơron được huấn luyện trước; và

chuyển các đầu ra được ghép đến lớp đầu ra.

**Fig.1**

2/5

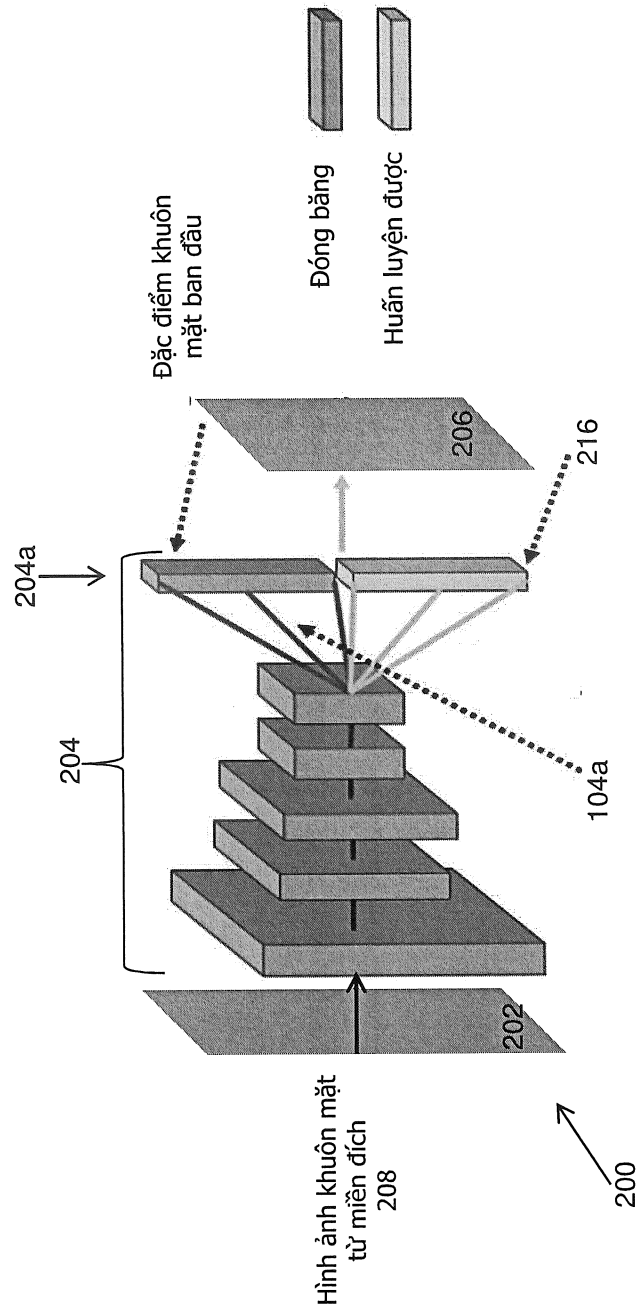


Fig.2

3/5

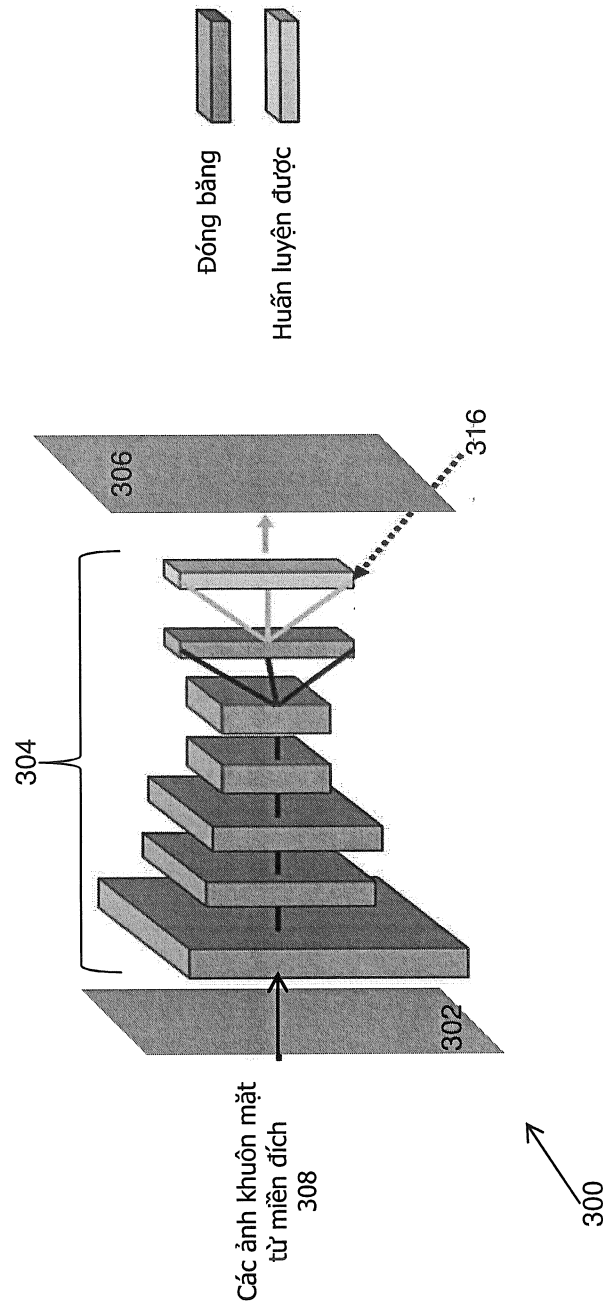
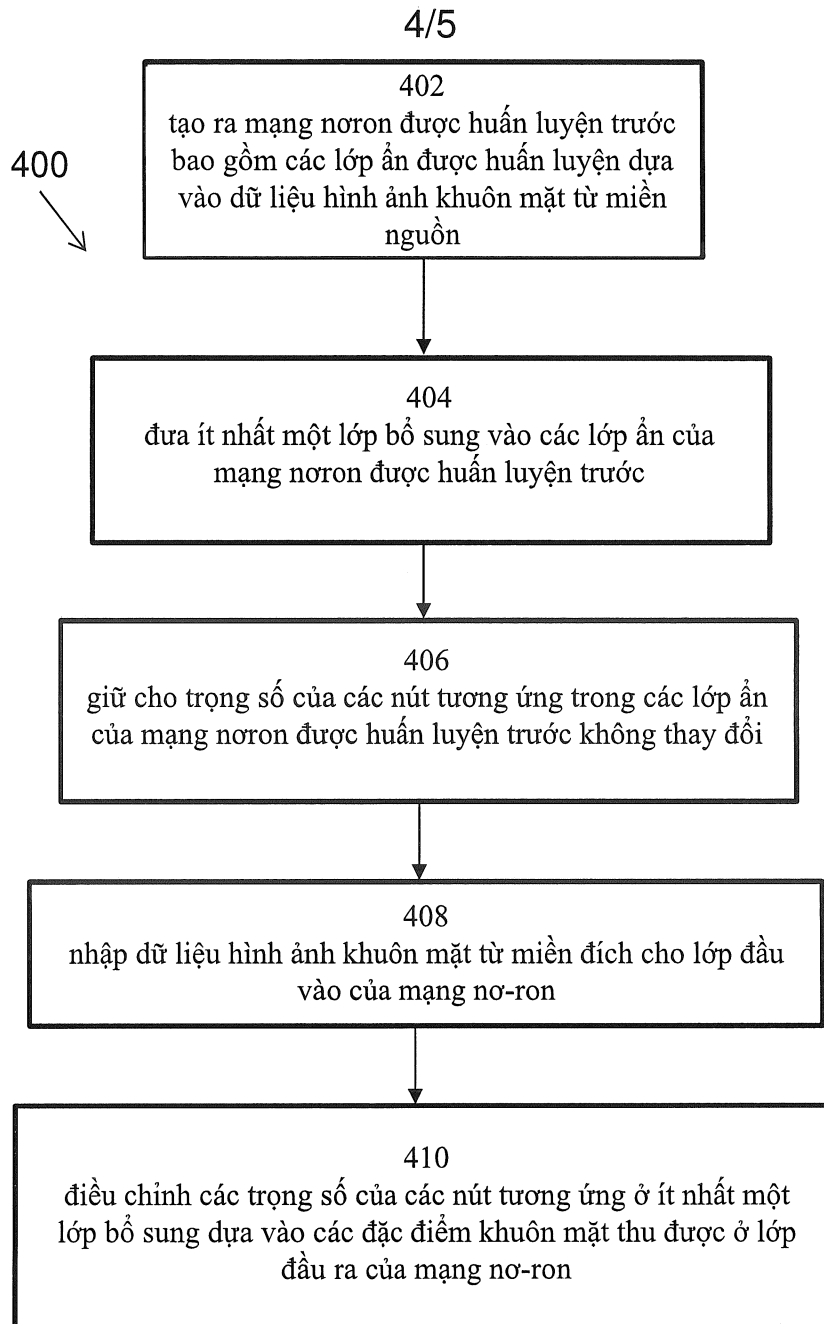


Fig.3

**Fig.4**

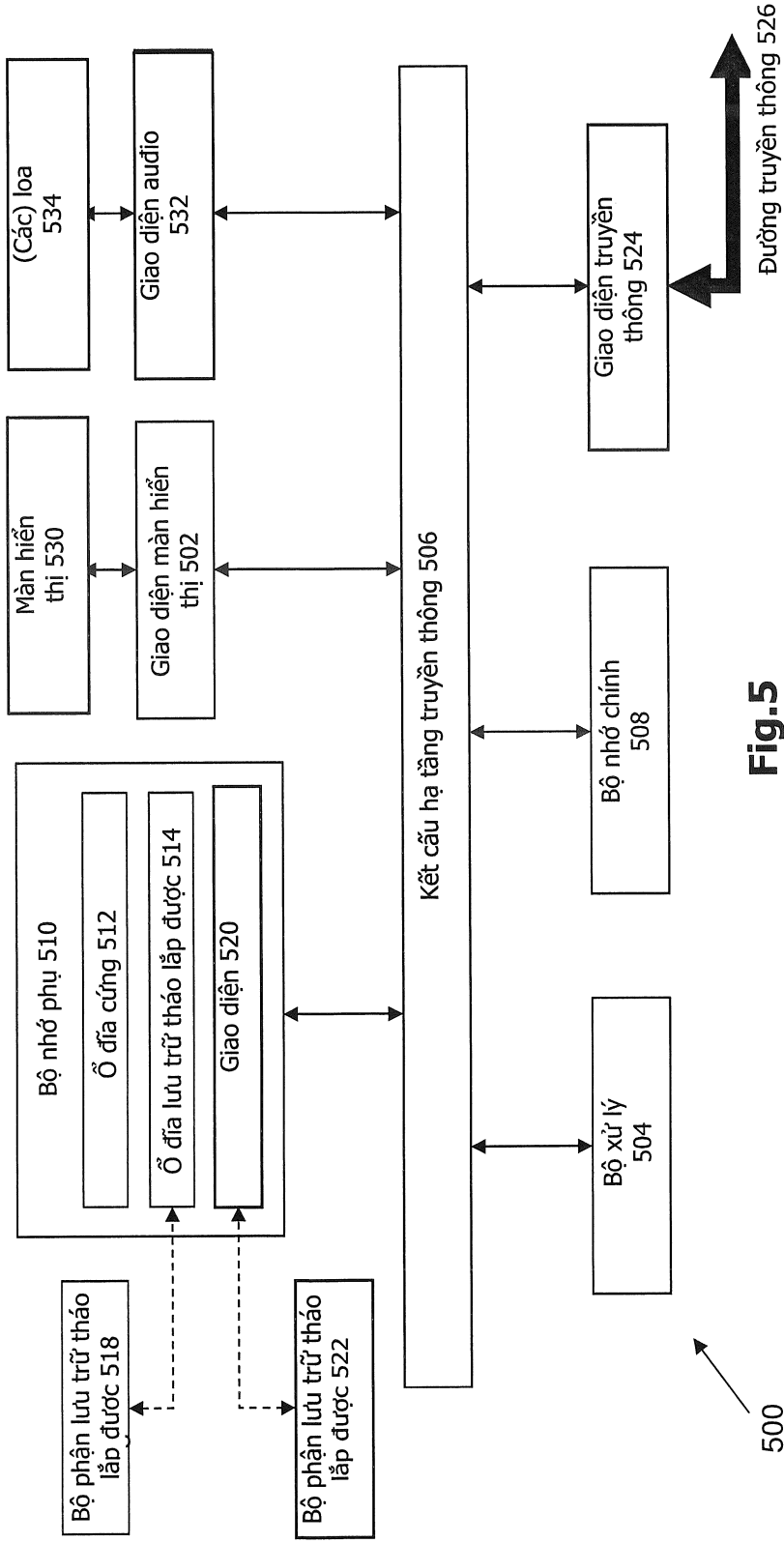


Fig.5