**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA CÔNG NGHỆ THÔNG TIN 1**

**---------------------------------------**



MÔN HỌC: CHUYÊN ĐỀ HỆ THỐNG THÔNG TIN

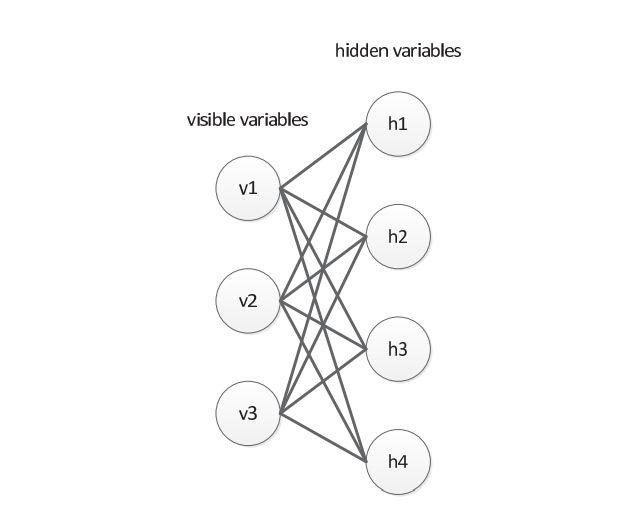
**BÁO CÁO BÀI TẬP CHUYÊN ĐỀ SỐ 3**

|  |
| --- |
| **Giảng viên : N.Q.Hoan**  **Sinh viên thực hiện : Lương Văn Chính – B15DCCN072**  **Nhóm môn học : Nhóm 1** |
|  |

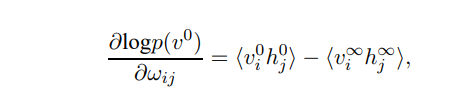
**Hà Nội, tháng 3 -2019**

***Chuyên đề 3: Mạng noron học sâu.***

1. **Nghiên cứu, cho ví dụ (hoặc mô phỏng) mảng noron “Máy Boltzmann”.**
2. **Mạng nở-ron máy Boltzmann (BMR):**



Các nơ ron tạo thành đồ thị 2 cực trong BMR, 1 đơn vị ẩn được kết nối với 1 đơn vị hiển thị. Để tạo ra 1 đồ thị BMR, bắt đầu từ 1 trạng thái ngẫu nhiên và thực hiện lấy mẫu Gibbs. Khi 1 trạng thái mới được đưa ra, tất cả các trạng thái khác sẽ được cập nhật, quá trính này sẽ kết thúc khi đạt cân bằng.



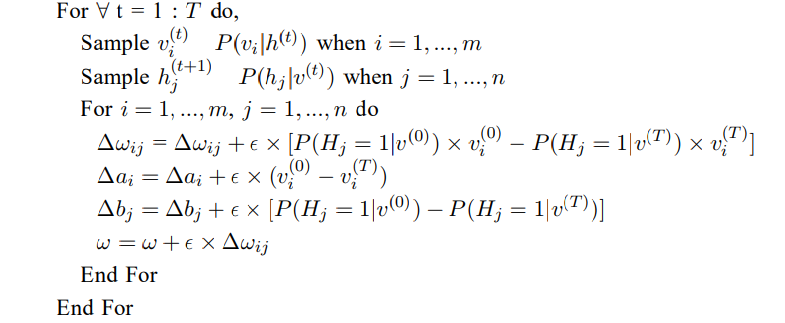
Wij là trọng số giữa đơn vị hiển thị i và đơn vị ẩn j, vi và hj giá trị giữa đơn vị hiển thị và đơn vị ẩn ở lớp thấp nhất và lớp cao nhất tương ứng. Quá trình tính toán sử dụng thuật toán phân kỳ tương phản CD.

Thuật toán CD: k bước cho RBM

Input: RBM( V1,...Vm, H1,...Hn)

Output: ma trận trọng số W, ước lượng Gradient ∆Wij, ∆ ai, , ∆ bj  ; i=1,...m; j=1,....n;

Khởi tạo W với giá trị bất kỳ trong khoảng [0,1] , ∆ Wij=∆ ai= ∆ bj = 0,i =1,...m; j=1....n;



1. **Các biến thể của máy Boltzmann (BMR).**

Hiện tại RBM đang đóng vai trò quan trọng trong nhiều lĩnh vực khác nhau như các mô hình chủ đề, tính chiều giảm, bộ lọc, học máy. Ví dụ, RBM có thể sử dụng để mã hóa dữ liệu sau đó áp dụng để giám sát, phân loại. Ngoài ra RBM còn có thể sử dụng như 1 mô hình phả hệ. Chúng ta có thể tính toán các điểm hiển thị và các điểm bị ẩn P(v,h) với luật Bayesian. Xác suất có điều kiện của 1 đơn vị p(v,h) cũng có thể được tính bằng RBM. Do đó RBM có thể sử dụng như 1 mô hình để phân loại .

Nói chung RBM được sử dụng làm trình trích xuất tính năng cho các mô hình phân loại. RBM trong học không giám sát sẽ không hiệu quả bằng RBM trong có giám sát. Ngoài ra việc lựa chọn các tham số rất quan trọng với hiệu suất của các thuật toán học máy. Do đó máy Boltmanz hạn chế được ra đời bởi Larochelle và Bengio vào năm 2008. Hơn nữa, để học trực tiếp với các bộ dữ liệu lớn, mô hình của DBRM lai (HDRBM) hoạt động tốt do lợi thế kết hợp của cả học tập chung và học từng phần riêng biệt. Tuy nhiên, trong các nhiệm vụ phân loại đa nhãn, hiệu suất của RBM không đạt. Mnih et al đề xuất cái gọi là máy Boltzmann bị hạn chế có điều kiện (CRBM) để cải thiện hiệu suất hơn nữa. Trong khi đó, trong chuỗi thời gian của 1 chu trình, CRBM có thể được sử dụng làm mô hình thế hệ phi tuyến tính. Trong một mô hình vô hướng được thiết lập với các biến có thể nhìn thấy có giá trị thực và các biến nhị phân ẩn. Trong mô hình này, các biến có thể nhìn thấy tại một số bước cuối cùng có thể bị ảnh hưởng trực tiếp bởi các biến ẩn và có thể nhìn thấy ở mỗi bước thời gian. Với kết quả này, suy luận trực tiếp có thể được thực hiện hiệu quả hơn bởi CRBM.

Gần đây, một DRBM độc lập (được gọi là FE-RBM) đã được Elfwing phát triển dựa trên một thuật toán học máy phân biệt. Trong FE-RBM, đầu ra cho bất kỳ vectơ đầu vào và lớp vecto nào được tính theo năng lượng tự do âm của RBM. Mục tiêu học tập đạt được thông qua việc giảm thiểu việc đào tạo bình phương lỗi sử dụng một phương pháp giảm độ dốc ngẫu nhiên. Hơn nữa, lấy kết quả từ nghiên cứu trước đó, năng lượng tự do được chia tỷ lệ theo hằng số dựa trên kích thước mạng để cải thiện độ mạnh của xấp xỉ hàm trong FE-RBM.

Khi RBM được áp dụng cho các khu vực như nhận dạng hình ảnh và giọng nói, hiệu suất của chúng có thể bị suy giảm nghiêm trọng bởi những tiếng ồn trong dữ liệu. Năm 2012, Tang et al đã giới thiệu một mô hình hiện đại, máy Boltzmann mạnh (RoBM), có thể được sử dụng để xử lý tiếng ồn và tiếng kêu trong việc nhận dạng. Với RoBM, một sự khái quát hóa tốt hơn có thể đạt được bằng cách loại bỏ các pixel bị hỏng. Được học với nhãn mác dữ liệu với tiếng ồn sử dụng thuật toán nghiêng không giám sát, mô hình RoBM cũng có thể tìm hiểu cấu trúc không gian của các ngoại lệ so với các thuật toán truyền thống, các RoBM đã cho thấy hiệu suất nâng cao ở nhiều mức độ khác nhau các ứng dụng như phân tích hình ảnh và nhận dạng khuôn mặt.

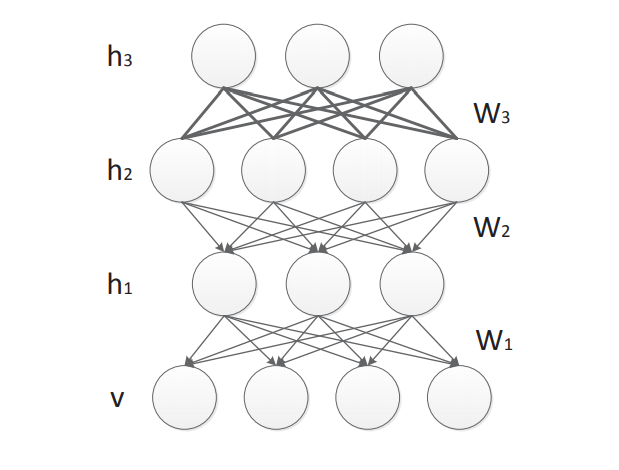
Là một yếu tố chính trong phân phối Boltzmann, lần đầu tiên, nhiệt độ được xem xét trong mô hình đồ họa của DBN của Li et al. Các máy Boltzmann (TRBM) bị hạn chế dựa trên nhiệt độ được đề xuất, trong đó nhiệt độ đóng vai trò là một tham số độc lập được điều chỉnh. Phân tích lý thuyết cho thấy rằng nhiệt độ là yếu tố chính kiểm soát tính chọn lọc của các nơ-ron trong các lớp ẩn. Nó đã được chứng minh hiệu suất của các TRBM được đề xuất có thể được tăng cường bằng cách cài đặt đúng thông số độ sắc nét của chức năng hậu cần. Do mức độ linh hoạt bổ sung được giới thiệu, TRBM có thể thu được kết quả chính xác hơn.

Hơn nữa, nghiên cứu cũng cung cấp một số hiểu biết về RBM từ quan điểm vật lý chỉ ra rằng có thể tồn tại một số mối quan hệ giữa nhiệt độ và một số NN ngoài đời thực.

1. **Phân loại các kiến trúc của mạng noron học sâu.**
2. **Mạng học sâu khả quan (DBN – Deep Belief Network).**
3. **Mạng học sâu khả quan.**

Như phần trước đã đề cập, các biến tường minh và các biến ẩn có mối quan hệ lẫn nhau. Để khám phá sự phụ thuộc giữa các biến, vào năm 2006 Hilton đã xây dựng các DBN bằng cách xếp chồng các lớp RBM.

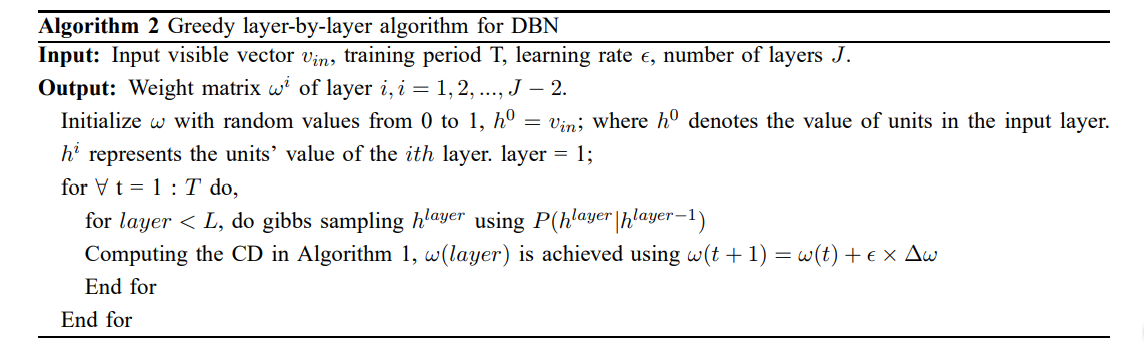
***Cấu trúc và thuật toán:***



Có thể thấy rằng trong một DBN, cứ hai lớp liền kề tạo thành một RBM. Lớp có thể nhìn thấy của mỗi RBM được kết nối với lớp ẩn của RBM trước đó và hai lớp trên cùng là không định hướng, kết nối giữa lớp trên và lớp dưới theo cách từ trên xuống. Các lớp RBM khác nhau trong một DBN được kết nối tuần tự: các RBM thấp hơn được kết nối trước, sau đó là các RBM cao hơn. Sau khi lớp trên cùng được sử dụng, các biến chức năng sẽ được truyền xuống những lớp thấp hơn. So với 1 RBM, xếp chồng sẽ làm tăng khả năng xảy ra của dữ liệu đầu vào.

Quá trình đào tạo của một DBN có thể được chia thành hai giai đoạn: giai đoạn tiền huấn luyện và giai đoạn tinh chỉnh. Trong giai đoạn tiền đào tạo, đào tạo dựa trên học tập không giám sát được thực hiện theo hướng đi xuống cho tính năng khai thác; trong khi ở giai đoạn tinh chỉnh, thuật toán học tập có giám sát được thực hiện để tiếp tục điều chỉnh các tham số trong mạng.

Qui trình: Lớp hiển thị của RBM thấp nhất được đào tạo trước với h(0) làm đầu vào. Các giá trị trong lớp hiển thị sau đó được nhập vào các lớp ẩn, trong đó xác suất kích hoạt P (h | v) của các biến ẩn được tính toán. Các kết quả thu được trong RBM trước đó sẽ được sử dụng làm đào tạo dữ liệu cho RBM tiếp theo và quá trình đào tạo này tiếp tục cho đến khi tất cả các lớp được duyệt qua. Thuật toán đào tạo trước này còn được gọi là thuật toán đào tạo không giám sát từng lớp một.



Trong giai đoạn tinh chỉnh, các DBN được đào tạo với dữ liệu bằng thuật toán lên xuống tương phản. Tóm lại, quy trình đào tạo của một DBN bao gồm một quy trình đào tạo trước từng lớp không được giám sát thực hiện theo cách từ dưới lên và quy trình tinh chỉnh lên xuống có giám sát. Thông qua quá trình đào tạo trước ,có thể đạt được giá trị ban đầu tốt hơn cho các trọng số và thuật toán lên xuống, sau đó được sử dụng để điều chỉnh toàn bộ mạng.

1. **Biến thể của DBNs**

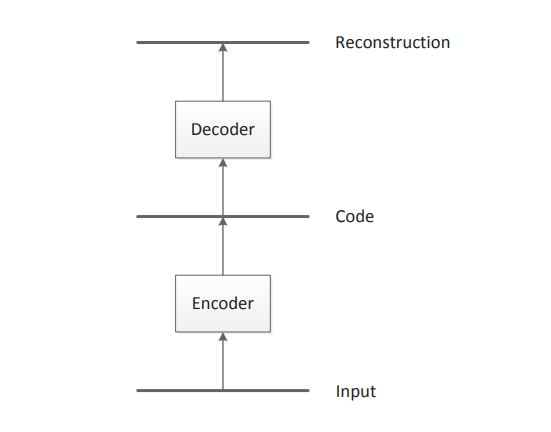
Đặng và Yu đã đề xuất một kiến ​​trúc sâu mới lạ, được gọi là mạng sâu lồi (DCN), để khắc phục những thiếu sót trong khả năng mở rộng học tập. Các DCN bao gồm nhiều mô-đun lớp. Một mô-đun được hình thành với một lớp ẩn duy nhất cũng như hai bộ trọng số trong một đặc biệt mạng lưới thần kinh. Cụ thể hơn, mô-đun thấp nhất bao gồm hai lớp tuyến tính và một lớp phi tuyến tính. Một lớp tuyến tính chứa các biến đầu vào và lớp kia chứa các biến đầu ra. Bên cạnh đó, lớp phi tuyến tính chứa các biến đầu vào phi tuyến. Phương pháp học tập trong DCN là chế độ hàng loạt dựa trên cơ sở dẫn đến song song đào tạo. Ngoài ra, hiệu suất của DCN có thể được cải thiện bằng quy trình tinh chỉnh khai thác cấu trúc.

Để tăng tính linh hoạt của các DBN, một mô hình mới về mạng lưới niềm tin sâu sắc tích chập (CDBN) đã được giới thiệu. Vì các đầu vào phải được vector hóa như một ma trận hình ảnh, thông tin cấu trúc hai chiều (2-D) như một hình ảnh đầu vào không thể được nhập làm đầu vào trực tiếp trong DBN. Tuy nhiên, trong CDBN, chiều cao hình ảnh có thể được trích xuất. Mặc dù thuật toán từng lớp tham lam đóng vai trò quan trọng trong việc đào tạo các DBN, nhiều kỹ thuật học sâu khác cũng đã được nghiên cứu. Năm 2009, Bengio tuyên bố rằng chúng ta có thể quan tâm mỗi cặp lớp của DBN như một bộ mã hóa tự động khử nhiễu (DAE).

1. **Kiến trúc học sâu tự động**
2. **Kiến trúc học sâu tự động.**

Bộ mã hóa tự động (AE), Nó là một thuật toán học giám sát được sử dụng để mã hóa hiệu quả bộ dữ liệu cho mục đích giảm kích thước. Gần đây, các AE đã được sử dụng để tìm hiểu các mô hình dữ liệu tổng quát. Dữ liệu đầu vào được chuyển đổi đầu tiên thành một biểu diễn trừu tượng sau đó được chuyển đổi trở lại định dạng ban đầu bằng chức năng mã hóa.

***Cấu trúc và thuật toán:*** AE là một mạng nơ ron chuyển tiếp dữ liệu ẩn tương tự như MLP. Sự khác biệt giữa một MLP và AE là mục đích của AE là tái cấu trúc đầu vào, trong khi mục đích của MLP là dự đoán các giá trị đích với các đầu vào nhất định. Trong quá trình mã hóa, đầu tiên AE chuyển đổi vectơ đầu vào x thành biểu diễn ẩn h bằng cách sử dụng ma trận trọng số ω; sau đó trong quá trình giải mã, AE ánh xạ trở lại định dạng ban đầu để thu được x˜ với ma trận trọng số khác.

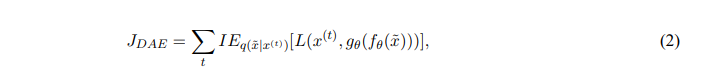


Các lỗi bình phương trung bình (MSE) được sử dụng để đo độ chính xác tái thiết theo phân phối giả định của các tính năng đầu vào.

Tương tự như đối với các DBN, quy trình đào tạo cho một AE cũng có thể được chia thành hai giai đoạn: giai đoạn đầu tiên là để tìm hiểu các tính năng sử dụng học tập không giám sát và thứ hai là tinh chỉnh mạng bằng cách học có giám sát. Cụ thể, trong giai đoạn đầu tiên, việc truyền tiếp tiến trình được thực hiện trước tiên cho mỗi đầu vào để thu được đầu ra giá trị x˜. Sau đó, lỗi bình phương được sử dụng để đo độ lệch của x˜ so với giá trị đầu vào. Cuối cùng, lỗi sẽ được truyền ngược lại qua mạng để cập nhật các trọng số. Trong giai đoạn tinh chỉnh, với mạng có các tính năng phù hợp ở mỗi lớp, chúng ta có thể áp dụng phương pháp học tập có giám sát tiêu chuẩn và thuật toán giảm dần độ dốc để điều chỉnh các tham số ở mỗi lớp.

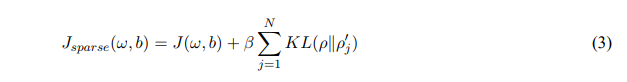
1. **Biến thể của AEs**

Năm 2008, Vincent và cộng sự đề xuất DAEs để phản bác các AE truyền thống. DAE cố ý thêm tiếng ồn vào dữ liệu huấn luyện và huấn luyện các AE với những dữ liệu bị hỏng này. Thông qua quá trình đào tạo, DAE có thể khôi phục phiên bản không có tiếng ồn của dữ liệu đào tạo.DAE sẽ trả về các giá trị bất lợi của tiếng ồn đầu vào bị hỏng một cách ngẫu nhiên. Hàm mục tiêu để tối ưu hóa trong DAE được hiển thị trong Công thức 2:



Một kiến trúc được phát triển trong dựa trên các lớp DAE xếp chồng lên nhau. Với mô hình xếp chồng, việc thực hiện của DAE trở nên dễ dàng hơn vì chúng ta chỉ cần xác định loại và mức độ của tiếng ồn bị hỏng.

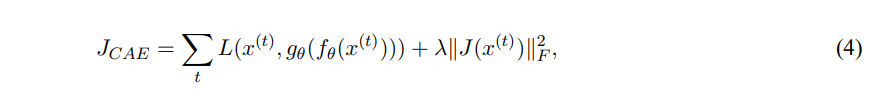
Gần đây, thấy rằng hiệu suất của các nhiệm vụ phân loại sẽ được cải thiện khi biết được độ rời rạc của kết quả. Các biểu diễn rời rạc được sử dụng để tạo ra một cách giải thích đơn giản về dữ liệu đầu vào bằng cách trích xuất cấu trúc ẩn của dữ liệu. Thuật toán học máy cho biểu diễn rời rạc được đề xuất bởi Ranzato vào năm 2006 . Để điều chỉnh một vectơ mã hóa thành một vecto rời rạc nhị phân, độ rời rạc phi tuyến tính được thêm vào giữa một bộ mã hóa tuyến tính và một bộ giải mã tuyến tính. lưu ý rằng đối với đầu vào nhị phân, cần có trọng số lớn để giảm thiểu lỗi tái thiết. Hàm chi phí tổng thể trong AE rời rạc được hiển thị trong công thức 3:



Khi p rời rạc, thường xấp xỉ và bằng 0. N là số lượng nơron, pj là số các hàm, J(w,b) là hàm chi phí, β là trọng số của hàm chi phí.

Hơn nữa, Makhzani và Frey đã đề xuất một AE rời rạc vào năm 2013. AE rời rạc bao gồm cơ bản kiến trúc của một AE tiêu chuẩn trong khi chỉ giữ các kích hoạt k cao nhất trong các lớp ẩn. Kết quả thu được cho thấy các AE rời rạc hoạt động tốt hơn các DAE và RBM. Họ tuyên bố rằng các AE rời rạc có thể được đào tạo dễ dàng và quy trình mã hóa tiên tiến sẽ góp phần đạt được hiệu suất tốt cho vấn đề quy mô lớn.

Năm 2011, Rifai và cộng sự đề xuất các bộ kết nối tự động (CAE) trong đó lựa chọn tốt được thêm vào hàm chi phí tiêu chuẩn trong giai đoạn tái thiết. Theo cách này, ánh xạ từ vectơ đầu vào sang đại diện sẽ hội tụ với xác suất cao hơn. Kết quả thu được bởi CAE giống hệt hoặc thậm chí tốt hơn những người có được bởi các AE thường xuyên khác như DAEs. Mục tiêu đào tạo của CAE được thể hiện trong Hình 4:



L là hàm chi phí, λ tham số thể hiện tần suất, J(x) là hàm ma trận Jacobian . công thức cho thấy 1 DAE với độ nhiễu ít được coi như một CAE. Năm 2016 Sun et al đề xuất 1 bộ mã hóa tự động tách rời (SDAE) được sử dụng để tính toán với nhiễu không nhìn thấy được. Tổng số nhiễu có thể được giảm bằng cách điều chỉnh các tham số chưa biết của DAE.

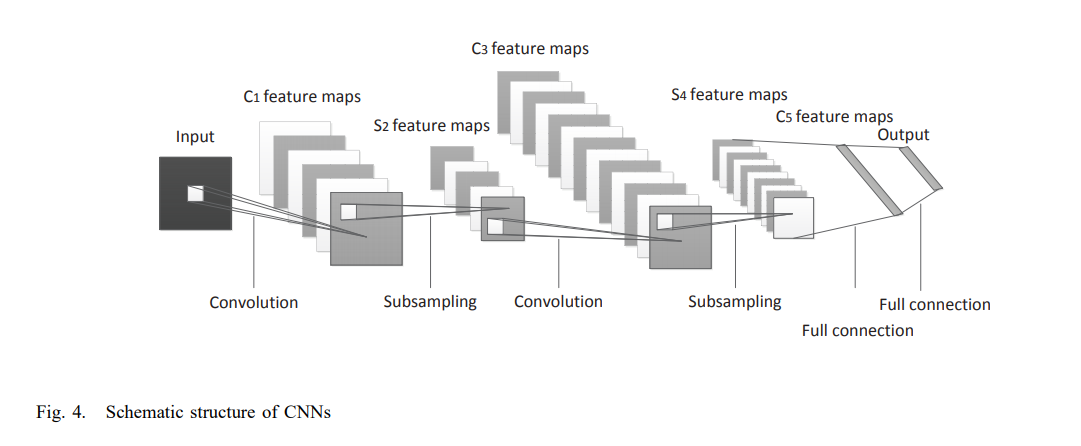
1. **Mạng nơ-ron chuyển đổi.**
2. **Mạng chuyển đổi.**

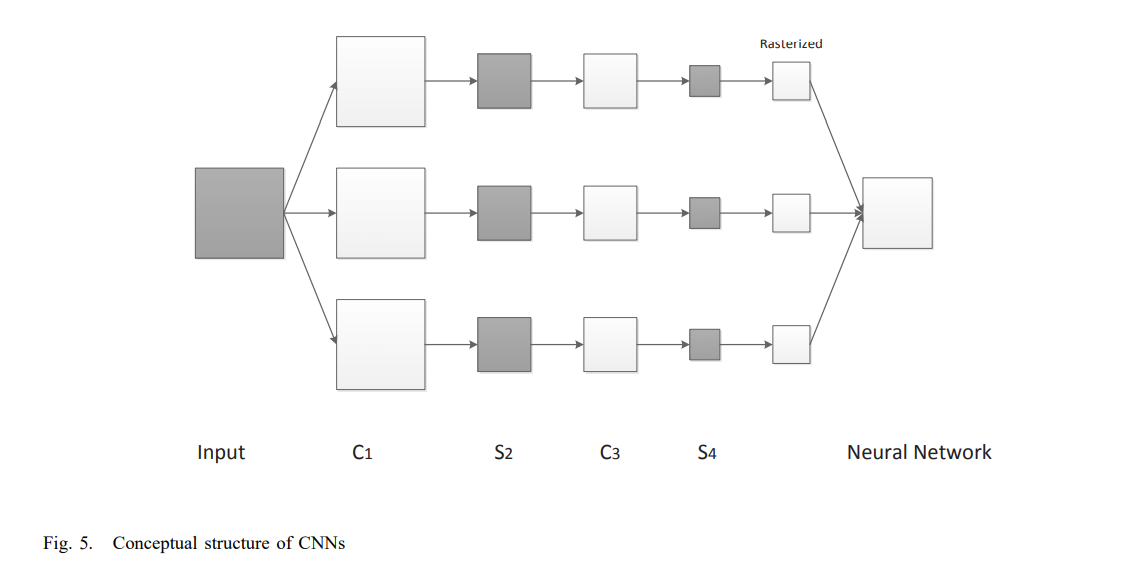
CNN là một kiểu con của kiến ​​trúc mạng rời rạc và đã cho thấy hiệu suất trong xử lý dữ liệu hai chiều với cấu trúc liên kết giống như lưới, chẳng hạn như hình ảnh và video.

Khái niệm về CNN được lấy cảm hứng từ các mạng thần kinh trì hoãn thời gian (TDNN). Trong một TDNN, các trọng số được chia sẻ trong một chiều thời gian, dẫn đến giảm tính toán. Trong CNNs, tích chập đã thay thế nhân ma trận chung trong NN tiêu chuẩn. Theo cách này, số lượng trọng số giảm, do đó giảm sự phức tạp của mạng. Hơn nữa, hình ảnh, như là đầu vào thô, có thể được nhập trực tiếp vào mạng,do đó tránh được quy trình trích xuất tính năng trong các thuật toán học tập tiêu chuẩn. Cần lưu ý rằng CNNs là kiến ​​trúc học sâu thực sự thành công đầu tiên do đào tạo thành công các lớp phân cấp. Các cấu trúc liên kết CNN tận dụng các mối quan hệ không gian để giảm số lượng tham số trong mạng và do đó hiệu suất được cải thiện bằng cách sử dụng các thuật toán lan truyền tiêu chuẩn. Một lợi thế khác của CNN mô hình là nó đòi hỏi tiền xử lý tối thiểu.

Với sự phát triển nhanh chóng của các kỹ thuật tính toán, các kỹ thuật điện toán được tăng tốc GPU đã được khai thác để đào tạo CNN hiệu quả hơn. Ngày nay, CNN đã được áp dụng thành công vào chữ viết tay nhận dạng, nhận diện khuôn mặt, nhận dạng hành vi, nhận dạng giọng nói, hệ thống đề xuất, phân loại hình ảnh,và NLP.

***Cấu trúc và thuật toán:*** Khác với các NN truyền thống thì mối quan hệ giữa đầu vào và đầu ra được lấy bằng cách nhân ma trận, các CNN giảm gánh nặng tính toán với sự tương tác rời rạc trong đó các hạt nhân được làm nhỏ hơn các đầu vào và được sử dụng cho toàn bộ hình ảnh. Tham số trung bình là thay vì học một bộ thông số riêng biệt ở mỗi vị trí, chúng ta chỉ cần học một bộ trong số đó, ngụ ý hiệu suất tốt hơn của CNN. tham số trung bình cũng cho CNN một thuộc tính hấp dẫn được gọi là tương đương, có nghĩa là bất cứ khi nào đầu vào thay đổi, đầu ra cũng thay đổi giống nhau. Do đó, cần ít tham số hơn cho CNN so với các thuật toán NN truyền thống khác, dẫn đến giảm bộ nhớ và cải thiện hiệu quả. Các thành phần của lớp CNN tiêu chuẩn được hiển thị trong hình 4 và sơ đồ nguyên lý khái niệm của một CNN tiêu chuẩn được hiển thị trong hình 5.





Như được hiển thị trong Hình 5, CNN là một mạng lưới thần kinh nhiều lớp bao gồm hai loại lớp khác nhau, tức là, các lớp chập (lớp c) và lớp lấy mẫu phụ (lớp s) .Lớp C và lớp s được kết nối luân phiên và tạo thành phần giữa của mạng. Như Hình 4 cho thấy, hình ảnh đầu vào được tích hợp với khả năng huấn luyện ở tất cả các độ lệch có thể để tạo ra các bản đồ đặc trưng trong lớp c đầu tiên. Một lớp trọng số kết nối là bao gồm trong mỗi bộ lọc. Thông thường, bốn pixel trong bản đồ đặc trưng tạo thành một nhóm. Truyền qua một chức năng sigmoid, những pixel này tạo ra các bản đồ tính năng bổ sung trong lớp s đầu tiên. Thủ tục này tiếp tục và do đó chúng ta có thể có được tính năng bản đồ trong các lớp c và lớp s sau đây. Cuối cùng, các giá trị của các pixel này được hiển thị trong một vectơ duy nhất là đầu vào của mạng .

Do tính đa chiều cao, việc áp dụng phân loại có thể gây ra thừa. Để giải quyết vấn đề này, một quy trình gộp, còn được gọi là lấy mẫu phụ hoặc lấy mẫu xuống, được đưa ra để giảm kích thước tổng thể của tín hiệu. Trong thực tế, lấy mẫu phụ đã được áp dụng thành công cho kích thước dữ liệu giảm nén âm thanh. Trong bộ lọc 2 chiều, lấy mẫu phụ cũng đã được sử dụng để tăng tính bất biến vị trí. Quy trình đào tạo cho một CNN tương tự như quy trình cho một NN tiêu chuẩn sử dụng phương pháp truyền ngược. Cụ thể hơn, Lunun et al đã giới thiệu gradient lỗi để huấn luyện các CNN. Trong giai đoạn đầu tiên, thông tin được truyền bá theo hướng chuyển tiếp qua các lớp khác nhau. Các tính năng nổi bật có được bằng cách áp dụng các bộ lọc kỹ thuật số ở mỗi lớp. Các giá trị của đầu ra sau đó được tính toán. Trong giai đoạn thứ hai, lỗi giữa dự kiến ​​và thực tế các giá trị của đầu ra được tính toán. Ma trận trọng số được điều chỉnh thêm và mạng do đó được tinh chỉnh. Không giống như các thuật toán tiêu chuẩn khác trong phân loại hình ảnh, quá trình tiền xử lý không thực hiện thường xuyên trong CNNs. Thay vì đặt tham số, như trường hợp của NN truyền thống, chúng ta chỉ cần tạo các bộ lọc trong CNNs.

1. **Biến thể của CNNs.**

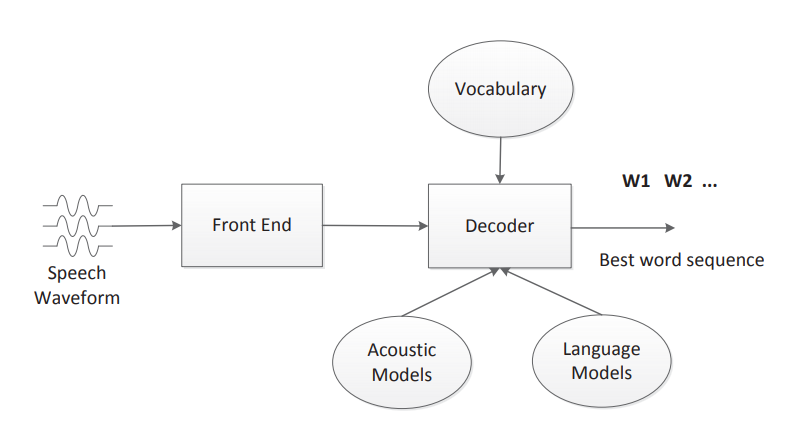
CNN đã trở thành một chủ đề nghiên cứu phổ biến trong vài năm qua. Năm 2013, Eigen và cộng sự đã giới thiệu một cuốn tiểu thuyết mô hình được gọi là mạng chập đệ quy (RCN). Kiến trúc của RCN có thể được xem như là một CNN với số lượng bản đồ đặc trưng giống hệt nhau trong tất cả các lớp và gắn trọng số bộ lọc trên các lớp. Nó chỉ ra rằng, số lượng các lớp ẩn tăng thì gánh nặng tính toán cũng tăng .Điều này rất ít có ý nghĩa để xác định chính xác kích thước của tính năng kích thước bản đồ.

Thành công lớn đã đạt được khi CNN được áp dụng vào nghiên cứu thị giác máy tính. Năm 2008, Desjardins và Bengio đã đề xuất một mô hình mới để sử dụng RBM trong CNN, là giới hạn tích chập máy Boltzmann (CRBM). Trong CRBM, một tích chập được tính toán với RBM bình thường làm hạt nhân. Mặc dù số lượng tham số trong RBM phụ thuộc vào kích thước của hình ảnh đầu vào, độ phức tạp của CRBM chỉ phụ thuộc vào số lượng tính năng được trích xuất và kích thước của lĩnh vực tiếp nhận. thuật toán CD cũng có thể được áp dụng để đào tạo CRBM. Lớp nhìn thấy được khởi tạo với đầu vào hình ảnh. Một đường chuyền lên được thực hiện để tính toán các trạng thái pixel trong lớp ẩn. So với RBM tiêu chuẩn trong các ứng dụng xử lý hình ảnh, CRBM có thể đạt được tốc độ cao hơn với giá trị nhỏ hơn của hàm mẫu. Bên cạnh đó, mạng lưới niềm tin sâu chập (CDBN) cũng đã được phát triển và áp dụng cho các cơ sở không giám sát có thể học mở rộng để biểu diễn phân cấp và học tính năng không giám sát để phân loại âm thanh .

Gần đây, Biến đổi Fourier nhanh (FFT) đã được sử dụng trong các CNN. Năm 2014, Mathieu và Henaff đã giới thiệu một quy trình đào tạo nhanh cho các CNN bằng FFT. Vì CNN tìm hiểu các chức năng phức tạp với lượng dữ liệu lớn, ngay cả với các GPU hiện đại, sẽ mất nhiều thời gian, đôi khi vài tuần để đào tạo các CNN tạo ra kết quả khả quan. Khi làm việc với các bộ dữ liệu quy mô web, chi phí sản xuất cao. Hướng tới vấn đề này, một thuật toán đơn giản được phát triển để đẩy nhanh quá trình đào tạo. Phương pháp này được hiện thực hóa bằng cách tính toán các kết quả như các sản phẩm trong miền Fourier. Bản đồ tính năng biến đổi tương tự được sử dụng nhiều lần. Thách thức của việc đào tạo CNN nằm ở sự phức tạp của các cặp ma trận 2 chiều. Với phép biến đổi Fourier, tích chập của các ma trận được chuyển thành vecto.Dựa trên yêu cầu tính toán, bộ xử lý GPU có thể được sử dụng để thực hiện thuật toán.

1. **Tìm kiếm các ứng dụng của mạng noron học sâu.**
2. **Nhận dạng giọng nói.**

Trong vài thập kỷ qua, thuật toán học máy đã được sử dụng rộng rãi trong các lĩnh vực như tự động nhận dạng giọng nói (ASR) và mô hình âm thanh. ASR có thể được coi là một tiêu chuẩn để xác định chuỗi từ từ giọng nói. Trong một số các ứng dụng như phiên âm và chính tả, nhận dạng giọng nói đã được phổ biến rộng rãi. Nhiều vấn đề mà ASR cần phải xử lý, ví dụ, môi trường ồn ào, nhận dạng đa mô hình và nhận dạng đa ngôn ngữ. Thông thường, dữ liệu nên được xử lý trước, sử dụng các kỹ thuật loại bỏ tiếng ồn trước khi các thuật toán nhận dạng giọng nói được áp dụng. Singh và cộng sự đã xem xét một số phương pháp để loại bỏ nhiễu và tăng cường lời nói như trừ phổ, lọc Wiener, cửa sổ, và ước tính biên độ phổ. Các thuật toán học máy truyền thống, chẳng hạn như SVM và NN, đã cung cấp kết quả đầy hứa hẹn trong nhận dạng giọng nói. Ví dụ, các mô hình hỗn hợp Gaussian (GMM) đã được sử dụng để phát triển hệ thống nhận dạng giọng nói bằng cách thể hiện mối quan hệ giữa đầu vào âm thanh và các trạng thái ẩn của mô hình Markov ẩn (HMM).



***Kiến trúc và giải thuật nhận dạng giọng nói:*** Đầu tiên, dạng sóng giọng nói đi qua thính giác nơi tín hiệu được xử lý trước và các sóng quang phổ được tạo ra. Sau đó, các thông tin sẽ được chuyển đến bộ xử lý của mỗi điện thoại. Sau đó, bộ giải mã sẽ giải mã lời nói bằng cách sử dụng n-gram mô hình ngôn ngữ (LM) và HMM. Cuối cùng, đầu ra sẽ được gửi đến trình phân tích, chuyển đổi thành một định dạng có thể đọc được.

Vì tín hiệu giọng nói có thể được coi là tín hiệu đứng yên trong thời gian ngắn, nên trong một khoảng thời gian ngắn, chúng ta có thể giả sử rằng quá trình phát biểu là đứng yên. Do đó, một mô hình Markov có thể mô tả quá trình phát biểu ngẫu nhiên. Ngoài ra, quy trình đào tạo của HMM là tự động và đơn giản để thực hiện. Chúng tôi có thể sử dụng một chuỗi các trạng thái ẩn của HMM để thể hiện âm thanh với các bản phân phối không cố định. HMM sẽ tạo ra một chuỗi các vectơ đại diện cho khả năng của từng trạng thái.

Hiệu suất của HMM có thể bị ảnh hưởng rất nhiều bởi sự không phù hợp giữa lý thuyết và thực tế. Trong trường hợp như vậy, một lượng lớn dữ liệu được yêu cầu. Trong GMM đã được sử dụng để ước tính mật độ đầu ra của các trạng thái HMM. Hơn nữa, GMM đóng vai trò quan trọng trong các nhiệm vụ tạo giọng nói và thường được sử dụng trong ánh xạ theo từng khung hình, đặc biệt là để tăng cường lời nói, ánh xạ phát âm thành âm thanh và chuyển đổi giọng nói. Các hệ thống GMM-HMM đã cải thiện đáng kể độ chính xác của phân loại và cũng có thể được áp dụng để loại bỏ tiếng ồn. Phải thừa nhận rằng, GMM-HMM vẫn còn một số hạn chế. GMM-HMM rất khó thể hiện các mối quan hệ phi tuyến tính hoặc phức tạp hơn giữa tính năng âm thanh và các đầu vào lời nói. Hiệu quả mô hình hóa thường rất thấp đối với dữ liệu gần phi tuyến tính. Hơn nữa, việc mất thông tin thô cũng có thể làm giảm hiệu suất của các hệ thống GMM-HMM.

Các ứng dụng ban đầu của các kỹ thuật học sâu bao gồm nhận dạng giọng nói liên tục(LVCSR) và nhận dạng điện thoại .Trong các ứng dụng này, các DBN được sử dụng để huấn luyện những dữ liệu rời rạc không nhãn. Hơn nữa, phương pháp DBN-HMM, kết hợp HMM với các mô hình học sâu đã đạt được một thành công lớn. Xác suất quan sát được ước tính bằng DBN, trong khi HMM được sử dụng để mô hình hóa các thông tin tuần tự. Phương pháp DBN-HMM tiên tiến đã áp dụng các trường ngẫu nhiên có điều kiện (CRF) để thay thế HMM trong mô hình hóa thông tin tuần tự. Giá trị thông tin lớn nhất (MMI) được sử dụng để đào tạo DBN-CRF. Trong trường hợp này, các trọng số chuyển tiếp, các trọng số của DBN và mô hình ngôn ngữ điện thoại được tối ưu hóa chung bằng cách sử dụng phương pháp học phân biệt rời rạc . So với DBN-HMM với quy trình đào tạo theo khung, DBN-CRF có thể đạt được độ chính xác cao hơn.

Tiếp theo là những tiến bộ gần đây trong nhận dạng giọng nói trong vài năm qua. Năm 2015, các DNN đã được sử dụng trong nhận dạng ngôn ngữ tự động (LID). Các thí nghiệm đã được thực hiện trong bài kiểm tra cách phát âm ngắn từ hai bộ dữ liệu: 5 triệu cách nói LID của Google và ngôn ngữ NIST công khai. Nhiều thành tựu khác cũng có thu được trong ASR, đặc biệt là trong các tình huống nói xa, hệ thống nhận dạng giọng nói bằng âm thanh (AVSR). Tiến bộ lớn cũng đã được thực hiện bằng cách sử dụng RBM để biểu diễn sóng âm tăng cường. Hơn nữa, các DNN đã được được sử dụng để theo dõi trạng thái hộp thoại , chuyển kiến ​​thức ngôn ngữ chéo, các ngân hàng bộ lọc, thiết kế hệ thống trích xuất tính năng tự động từ âm thanh và đào tạo thích ứng loa (SAT) cho âm thanh mô hình hóa.

1. **Thị giác máy tính và nhận dạng mẫu.**

Thị giác máy tính nhằm mục đích làm cho máy tính hiểu chính xác và xử lý hiệu quả dữ liệu trực quan như video và hình ảnh. Máy được yêu cầu nhận biết dữ liệu chiều cao trong thế giới thực và tạo ra biểu tượng hoặc thông tin số cho phù hợp. Mục tiêu cuối cùng của thị giác máy tính là ban cho các máy tính khả năng nhận thức khả năng của con người. Về mặt khái niệm, thị giác máy tính đề cập đến các ngành khoa học điều tra làm thế nào để trích xuất thông tin từ hình ảnh trong các hệ thống nhân tạo. Các lĩnh vực sau được bao gồm dưới dạng tên miền phụ của máy tính tầm nhìn: phát hiện sự kiện, tái tạo cảnh, phát hiện và nhận dạng đối tượng, ước tính tư thế đối tượng, hình ảnh phục hồi, học thống kê, chỉnh sửa hình ảnh và nâng cao video.

Nhận dạng mẫu là một môn khoa học nhằm xác định mẫu của một giá trị đầu vào nhất định. Nó là một khái niệm chung chung bao gồm một số tên miền phụ như phân loại, hồi quy, ghi nhãn chuỗi và gắn thẻ bài phát biểu. Do sự phát triển công nghiệp nhanh chóng, ngày càng có nhiều yêu cầu về khả năng thu hồi và xử lý thông tin, điều này đã mang lại những thách thức mới cho việc nhận dạng mẫu. Gần đây, phát triển trong kiến ​​trúc học tập sâu đã cung cấp các cách tiếp cận mới cho vấn đề nhận dạng mẫu, mà sẽ được thảo luận trong những gì sau đây.

1. Công nhận

Trong vài năm qua, các kỹ thuật học sâu đã đạt được những tiến bộ to lớn trong các lĩnh vực thị giác máy tính và nhận dạng mẫu, đặc biệt là trong các lĩnh vực như nhận dạng đối tượng. Trong các ứng dụng phân loại, lựa chọn tính năng là một vấn đề quan trọng. Thông thường, các tính năng được chỉ định thủ công trong các thuật toán phân loại truyền thống, có hạn chế chung chung. Một số kiến ​​trúc học sâu điển hình, chẳng hạn như CNN, có thể tự động chọn các tính năng và đạt được hiệu suất vượt trội dựa trên tài nguyên tính toán được tăng tốc GPU. Lưu ý rằng thị giác của con người các hệ thống khác với các hệ thống thị giác máy tính và đã chứng minh rằng các DNN có thể dễ dàng bị đánh lừa bởi hình ảnh không thể nhận ra . Tuy nhiên, điều này không có nghĩa là các kỹ thuật học sâu không phù hợp với nhiệm vụ phân loại. Các nghiên cứu gần đây đã chỉ ra rằng trong các nhiệm vụ phân loại, các kỹ thuật học sâu có thể đạt được kết quả đầy hứa hẹn.

Trong nhận dạng đối tượng, còn được gọi là phân loại đối tượng, phương pháp học sâu đã đạt được sự vượt trội hiệu suất so với các thuật toán phân loại thông thường . Ở đây chúng tôi xem xét một số tiến bộ gần đây về nhiệm vụ phân loại. Đối với nhận dạng biển báo giao thông Đức, DNN nhiều cột đã được đề xuất . Đến nghiên cứu các điều kiện tâm thần kinh dựa trên các mẫu kết nối chức năng (FC), các phân loại tiêu chuẩn như SVM đã được sử dụng rộng rãi. Gần đây, các DNN đã được sử dụng để phân loại các mẫu FC trạng thái nghỉ ngơi toàn bộ não tâm thần phân liệt (SZ) . Để cải thiện hiệu suất của phân loại, một mạng lưới thần kinh sâu đa phương thức lề tối đa mới (3mDNN) đã được đề xuất để tận dụng lợi thế của nhiều mô tả cục bộ của một hình ảnh . So với các thuật toán tiêu chuẩn, phương pháp này, xem xét thông tin của nhiều mô tả, có thể đạt được khả năng phân biệt.

Các kỹ thuật học sâu cũng có thể được áp dụng để nhận dạng tư thế tay (HPR). Kể từ khicác tính năng được tạo bởi các thuật toán truyền thống bị hạn chế và rất khó để phát hiện và theo dõi tay với bình thườngmáy ảnh, DNN được sử dụng để tạo ra các tính năng nâng cao .Dựa trên chức năng quang phổ hồng ngoại (FNIRS), các kỹ thuật học sâu đã đạt được kết quả đầy hứa hẹn trong việc phân loại các mô hình kích hoạt não cho BCI.

1. Khám phá.

Chúng ta có thể sử dụng phát hiện hình ảnh để tìm ra các mô hoặc tế bào bất thường trong hình ảnh y tế. Mô hình dựa trên bộ phận biến dạng (DPM) được đề xuất bởi Felzenzwalb là một trong những phương pháp phổ biến nhất. DNN đã được sử dụng rộng rãi để phát hiện và đã cho thấy hiệu suất vượt trội.

Như đã đề cập , các DBN được sử dụng trong các hệ thống chẩn đoán hỗ trợ máy tính (CAD) để phát hiện sớm ung thư vú. Trong trường hợp này, độ chính xác của bộ phân loại là yếu tố quan trọng nhất đối với hệ thống CAD. So với các thuật toán phân loại tiêu chuẩn như phương pháp cây quyết định C4.5, phân cụm mờ được giám sát (SFC) kỹ thuật, phương pháp Fuzzy-GA, phương pháp mạng thần kinh chức năng cơ sở (RBFNN) và mạng nơ ron tối ưu hóa (PSOWNN), các DBN có thể đạt được hiệu suất tốt hơn cho CAD hệ thống. Các DBN cũng có thể được sử dụng để giảm kích thước phi tuyến tính của các tính năng đầu vào. Nghiên cứu về phát hiện khối u não và nhiệm vụ phân đoạn đã nhận được sự chú ý ngày càng tăng trong vài năm qua . Do hiệu quả tính toán của nó, phương pháp chụp cộng hưởng từ (MRI) cho khối u não lâm sàng phát hiện đã được giới thiệu bằng cách sử dụng các kỹ thuật học tập sâu. Tuy nhiên, phương pháp phát hiện khối u não dựa trên MRI chịu sự khác biệt giữa kích thước và hình dạng dự đoán của khối u não. Các CNN được sử dụng để giải quyết vấn đề này cho khả năng học tập mạnh mẽ của nó.

Tương tự, các phương pháp học sâu cũng có thể được áp dụng để chú thích các biến thể di truyền để xác định các biến thể gây bệnh. Thông thường, thuật toán suy giảm kết hợp (CADD) được sử dụng rộng rãi nhất để chú thích biến thể mã hóa và không mã hóa . các DANN được sử dụng thay cho trình phân loại SVM. DANNs phù hợp với số lượng lớn mẫu và tính năng. Cụ thể hơn, để dự đoán các vùng trật tự / rối loạn protein. Các kỹ thuật học sâu chỉ yêu cầu dữ liệu ảnh thô. Ngoài ra, các kỹ thuật học sâu cũng đã được áp dụng để phát hiện Glaucoma và hệ thống tương tác giữa người và robot.

Là một ứng dụng quan trọng khác của thị giác máy tính, phát hiện thay đổi hình ảnh đóng một vai trò quan trọng trong chỉ dân sự mà còn lĩnh vực quân sự. Mục tiêu của phát hiện thay đổi hình ảnh là để phân loại sự khác biệt giữa hai hình ảnh được chụp tại thời điểm khác nhau cho cùng một cảnh. Phát hiện hình ảnh đã được sử dụng rộng rãi trong điều khiển từ xa cảm biến, chẩn đoán y tế, đánh giá thảm họa và giám sát video. Đặc biệt, radar khẩu độ tổng hợp (SAR) xử lý ảnh là một ứng dụng được sử dụng rộng rãi trong phát hiện thay đổi. Trong các phương pháp hiện đại, một hình ảnh khác biệt (DI) được tạo ra giữa các hình ảnh SAR đa thời gian để phát hiện thay đổi. Tuy nhiên, DI có thể có ảnh hưởng xấu đến hiệu suất phát hiện thay đổi. Để tránh điều này, các kỹ thuật học sâu đã bỏ qua quá trình tạo DI. Để kiểm soát giao thông và giám sát an ninh hàng hải, phát hiện tàu trên hình ảnh chụp đã được sử dụng rộng rãi. Tuy nhiên, so với hồng ngoại và hình ảnh radar khẩu độ tổng hợp, hình ảnh tàu vũ trụ dễ bị ảnh hưởng bởi điều kiện thời tiết. Ngoài ra, độ khó của xử lý hình ảnh tăng lên khi cơ sở dữ liệu lớn hơn được xử lý cho độ phân giải cao hơn. Để khắc phục những hai thiếu sót, các DNN được kết hợp với các máy học cực đoan (ELM) trong . So sánh với các phương pháp tiên tiến khác, phương pháp này đã đạt được độ chính xác cao hơn với thời gian phát hiện ít hơn.

1. **Ứng dụng khác.**

Căn chỉnh khuôn mặt đóng một vai trò quan trọng trong các ứng dụng hình ảnh khác nhau như nhận dạng khuôn mặt. Tuy nhiên, đối với tình huống phức tạp, căn chỉnh khuôn mặt có thể dẫn đến khó khăn trong quá trình phân tích quá trình. Do đó, các mô hình khác nhau cho sự thay đổi hình dạng và ngoại hình đã được xem xét để giải quyết vấn đề này.

Ước tính Pose là một ứng dụng quan trọng khác trong thị giác máy tính. Mục tiêu chung của Pose là ước tính vị trí tương đối của một đối tượng cụ thể đối với máy ảnh. Kỹ thuật này đóng một vai trò không thể thay thế trong các nhiệm vụ khác nhau như xây dựng một cánh tay robot. Xem xét nhiều tư thế khác nhau và bối cảnh lộn xộn, các nhà nghiên cứu phát hiện ra rằng nhận dạng tự động sử dụng các phương pháp truyền thống bị ảnh hưởng bởi xác định sai và khai thác tính năng không đầy đủ. Do đó, một kiến ​​trúc học tập sâu đã được giới thiệu cho hình ảnh mẫu .Đặc biệt, Li et al đề xuất một thuật toán học tập đa tác vụ không đồng nhất khuôn khổ trong đó một DCNN đã được thông qua cho ước tính tư thế con người.

Ngoài ra, phương pháp học sâu đã được sử dụng thành công trong ước tính chuyển động, đặc biệt là cho nhiệm vụ theo dõi video. Theo dõi đối tượng đã thu hút nhiều sự chú ý do giá trị lý thuyết của nó và triển vọng ứng dụng trong các lĩnh vực như xe tự lái, robot và giám sát video tình báo. Thiết kế một mô hình xuất hiện đối tượng mạnh mẽ, xây dựng phân loại và biểu diễn tính năng là hai vấn đề chính. Phân loại thông thường đã đạt được mức độ thành công nhất định, cấu trúc, tuy nhiên, có hiệu suất hạn chế, đặc biệt đối với sự xuất hiện của đối tượng phi tuyến tính hoặc biến thể thay đổi theo thời gian các. Tương tự, biểu diễn tính năng truyền thống bao gồm nhiều tính năng nổi tiếng khác nhau như SFIT, HoG, ma trận hiệp phương sai, các tính năng dựa trên không gian con và biểu đồ màu, v.v ... Hầu hết các tính năng thủ công không thể phản ánh tính chất thay đổi thời gian. Do đó, trình theo dõi CNN đã được giới thiệu để giải quyết các hạn chế của cấu trúc phân loại nông và các tính năng thủ công trong các nhiệm vụ theo dõi đối tượng. Một hạn chế của kiến ​​trúc học sâu truyền thống là từ việc sử dụng một mô hình quan sát duy nhất. Trong trường hợp này, các trình theo dõi phải đối phó với các tính năng bị ảnh hưởng do nhiễu. Do đó, Wu et al đã giới thiệu một trình theo dõi học tập sâu trong khu vực có chứa nhiều mô hình sâu, và mỗi trong số đó chịu trách nhiệm theo dõi một vùng.

Trong thị giác máy tính, khử nhiễu là một vấn đề quan trọng vì hình ảnh kỹ thuật số bị hỏng do nhiễu thông qua mua lại và truyền tải. Mặc dù có nhiều thuật toán khử nhiễu, hầu hết chúng được thiết kế đặc biệt trường hợp và thiếu tính tổng quát. Bộ lọc Wiener hoạt động tốt để loại bỏ tiếng ồn Gaussian. Tuy nhiên, nó đòi hỏi kiến thức về các chức năng tự tương quan của đầu vào. Đối với việc loại bỏ hình ảnh nhiễu, lọc trung bình xử lý tiếng ồn muối và hạt tiêu hiệu quả . Do đó, xếp chồng tự động khử nhiễu thưa thớt (SSDAEs) đã được đề xuất với hiệu suất loại bỏ nhiễu đầy hứa hẹn. Hơn nữa, đa cột thích ứng SSDAEs (AMC-SSDAE) đã được giới thiệu để cải thiện tính hiệu quả của bộ lọc .