ÔN TẬP Được truyền đạt bởi Vincent Vanhoucke

Mạng nơ-ron tích chập sâu cho hình ảnh Phân loại: Một đánh giá toàn diện

Waseem Rawat
wrawat10@gmail.com
Zenghui Wang

wangz@unisa.ac.za Khoa Kỹ thuật Điện và Khai thác, Đại học Nam Phi, Florida 1710, Nam Phi

Mạng nơ-ron tích chập (CNN) đã được áp dụng cho các tác vụ thị giác từ cuối những năm 1980. Tuy nhiên, mặc dù có một vài ứng dụng rải rác, chúng vẫn nằm im cho đến giữa những năm 2000 khi sự phát triển về sức mạnh tính toán và sự ra đời của một lượng lớn dữ liệu được gắn nhãn, được bổ sung bởi các thuật toán được cải tiến, đã góp phần vào sự tiến bộ của chúng và đưa chúng lên hàng đầu trong thời kỳ phục hưng của mạng nơ-ron đã chứng kiến sự tiến triển nhanh chóng kể từ năm 2012. Trong bài đánh giá này, tập trung vào ứng dụng của CNN vào các tác vụ phân loại hình ảnh, chúng tôi sẽ đề cập đến sự phát triển của chúng, từ những ngươi tiền nhiệm cho đến các hệ thống học sâu hiện đại gần đây.
Trên đương đi, chúng tôi phân tích (1) những thành công ban đầu của họ, (2) vai trò của họ trong thời kỳ phục hưng học sâu, (3) các tác phẩm tượng trưng được chọn đã góp phần vào sự phổ biến gần đây của họ và (4) một số nỗ lực cải thiện bằng cách xem xét các đóng góp và thách thức của hơn 300 ấn phẩm. Chúng tôi cũng giới thiệu một số xu hướng hiện tại và những thách thức còn lại của họ.

1 Giới thiệu

Phân loại hình ảnh, có thể được định nghĩa là nhiệm vụ phân loại hình ảnh thành một trong một số lớp được xác định trước, là một vấn đề cơ bản trong thị giác máy tính. Nó tạo thành cơ sở cho các nhiệm vụ thị giác máy tính khác như định vị, phát hiện và phân đoạn (Karpathy, 2016). Mặc dù nhiệm vụ này có thể được coi là bản chất thứ hai đối với con người, nhưng nó khó khăn hơn nhiều đối với một hệ thống tự động. Một số biến chứng gặp phải bao gồm tính biến thiên của đối tượng phụ thuộc vào quan điểm và tính biến thiên cao trong lớp khi có nhiều loại đối tượng (Ciresan, Meier, Masci, Gambardella & Schmidhuber, 2011). Theo truyền thống, phương pháp tiếp cận hai giai đoạn được sử dụng để giải quyết vấn đề phân loại. Các tính năng thủ công đầu tiên được trích xuất từ hình ảnh bằng cách sử dụng các mô tả tính năng và chúng được dùng làm đầu vào cho một bộ phân loại có thể đào tạo được. Trở ngại chính của phương pháp tiếp cận này là độ chính xác của nhiệm vụ phân loại phụ thuộc rất nhiều vào thiết kế của tính năng

Tính toán thần kinh 29, 1–98 (2017) doi:10.1162/NECO_a_00990

giai đoạn chiết xuất, và điều này thư ờng đư ợc chứng minh là một nhiệm vụ khó khăn (LeCun, Bottou, Bengio và Haffner, 1998).

Trong những năm gần đây, các mô hình học sâu khai thác nhiều lớp xử lý thông tin phi tuyến tính, để trích xuất và chuyển đổi đặc điểm cũng như để phân tích và phân loại mẫu, đã đư ợc chứng minh là vư ợt qua những thách thức này. Trong số đó, CNN (LeCun, Boser, Denker, Henderson, Hubbard, & Jackel, 1989a, 1989b) đã trở thành kiến trúc hàng đầu cho hầu hết các nhiệm vụ nhận dạng, phân loại và phát hiện hình ảnh (LeCun, Bengio, & Hinton, 2015). Mặc dù có một số thành công ban đầu (LeCun và cộng sự, 1989a, 1989b; LeCun và cộng sự 1998; Simard, Steinkraus và Platt 2003), CNN sâu (DCNN) đã đư ợc đư a vào ánh đèn sân khấu như là kết quả của thời kỳ phục hư ng học tập (Hinton, Osindero, & Teh, 2006; Hinton & Salakhutdi-nov, 2006; Bengio, Lamblin, Popovici, & Larochelle, 2006), đư ợc thúc đẩy bằng GPU, bộ dữ liệu lớn hơn và thuật toán tốt hơn (Krizhevsky, Sutskever, & Hinton, 2012; Deng & Yu, 2014; Simonyan & Zisserman, 2014; Zeiler & Fer-gus, 2014). Một số tiến bộ như triển khai GPU đầu tiên (Chel-lapilla, Puri, & Simard, 2006) và ứng dụng đầu tiên của nhóm tối đa (gộp tối đa) cho DCNN (Ranzato, Huang, Boureau, & LeCun, 2007) có tất cả đều góp phần vào sự nổi tiếng gần đây của họ.

Tiến bộ quan trọng nhất đã thu hút đư ợc sự quan tâm sâu sắc trong DCNN, đặc biệt là đối với các tác vụ phân loại hình ảnh, đã đạt đư ợc thành tựu trong Thử thách nhận dạng hình ảnh quy mô lớn Im-ageNet (ILSVRC) năm 2012 (Rus-sakovsky và cộng sự, 2015), khi bài dự thi chiến thắng của A. Khrizhevsky và cộng sự. (2012), đã sử dụng DCNN để phân loại khoảng 1,2 triệu hình ảnh thành 1000 lớp học, với kết quả phá kỷ lục. Kể từ đó, DCNN đã thống trị các phiên bản tiếp theo của ILSVRC và cụ thể hơn là thành phần phân loại hình ảnh của nó (Simonyan & Zisserman, 2014; Zeiler & Fergus, 2014; Szegedy, Liu và cộng sự, 2014).

Ngoài ra, các ví dụ đại diện đư ợc chọn về các nỗ lực cải tiến khác liên quan đến các khía cạnh khác nhau sau đây của DCNN—(1) mạng kiến trúc (Lin, Chen, & Yan, 2013; Zeiler & Fergus, 2013; Gong, Wang, Guo, & Lazebnik, 2014; Szegedy, Vanhoucke, Ioffe, Shlens và Wojna, 2015); (2) các hàm kích hoạt phi tuyến tính (He, Zhang, Ren, & Sun, 2015a; Xu, Wang, Chen, & Li, 2015); (3) các thành phần giám sát (Tang, 2013; Zhao & Grif-fin, 2016); (4) cơ chế điều chỉnh (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012; Zeiler và Fergus, 2013); và (5) các kỹ thuật tối ư u hóa (Glorot & Bengio, 2010; Krizhevsky và cộng sự, 2012)—cũng đã đã đư ợc thực hiện trong những năm gần đây. Hơn nữa, một số thách thức mở của họ, giống như sự thay đổi của chúng đối với các biến dạng hình học (Gong và cộng sự, 2014), thực tế là mô hình của họ thư ờng lớn và chậm tính toán (Krizhevsky và cộng sự, 2012; Simonyan & Zisserman, 2014), và khám phá thú vị về sự đối nghịch các ví dụ (Szegedy et al., 2014), đã dẫn đến nhiều nghiên cứu hơn tập trung vào phân loại hình ảnh với DCNN.

Trư ớc đây, một số đánh giá chung về học sâu (Bengio, 2009; Schmid-huber, 2015; Deng, 2014; LeCun và cộng sự, 2015), các đánh giá liên quan đến học sâu

học để hiểu trực quan (Guo et al., 2016), các bài đánh giá bao gồm những tiến bộ gần đây trong CNN (Gu et al., 2015) và phân loại DCNN cho nhiệm vụ thị giác máy tính (Srinivas và cộng sự, 2016) đã được công bố. Tuy nhiên, cho thấy sự gia tăng về mức độ phổ biến của DCNN đối với các nhiệm vụ phân loại hình ảnh và vô số các bài báo liên quan sau đó, chúng tôi cảm thấy đã đến lúc xem xét lại chúng cho vấn đề cụ thể và quan trọng này. Với điều này trong tâm trí, đánh giá dành cho những người muốn hiểu sự phát triển của Công nghệ và kiến trúc CNN, đặc biệt dành cho phân loại hình ảnh, từ những người tiền nhiệm của họ cho đến các hệ thống học sâu hiện đại. Nó cũng khẳng định những hiểu biết ngắn gọn về tương lai của họ và cung cấp một số điều thú vị những hướng đi sắp tới khiến nó trở nên phù hợp với các nhà nghiên cứu trong lĩnh vực này.

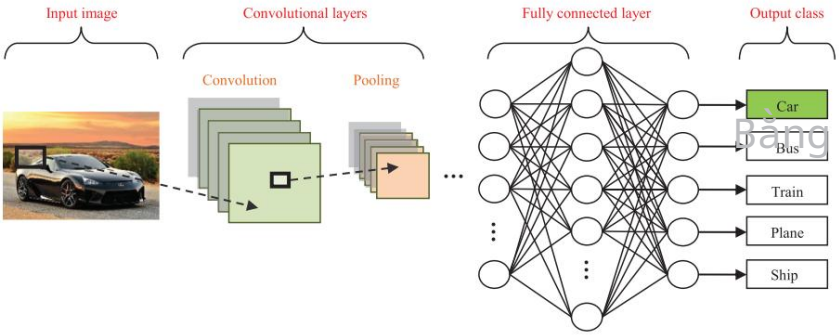
Bằng chứng

Phần còn lại của bài đánh giá này được tổ chức như sau: Phần 2 tóm tắt giới thiệu CNN và làm quen với người đọc với các khối xây dựng chính của kiến trúc của họ. Phần 3 đề cập đến sự phát triển ban đầu của CNN. Trong số những điểm nổi bật khác, nó đề cập ngắn gọn đến các ứng dụng đầu tiên của backpropagation và max pooling, cũng như sự ra đời của MNIST nổi tiếng bộ dữ liệu (LeCun et al., 1998). Trong phần 4, chúng tôi đề cập đến vai trò của DCNN trong thời kỳ phục hưng học sâu, và sau đó là các cuộc thảo luận về các tác phẩm tiêu biểu được chọn đã góp phần tạo nên sự nổi tiếng của chúng cho các nhiệm vụ phân loại hình ảnh. Phần 5 đề cập đến một số nỗ lực cải tiến DCNN ở nhiều khía cạnh khác nhau, bao gồm kiến trúc mạng, phi tuyến tính chức năng kích hoạt, thành phần giám sát, cơ chế điều chỉnh, kỹ thuật tối ưu hóa và phát triển chi phí tính toán. Phần 6 kết thúc bài đánh giá bằng cách giới thiệu một số thách thức còn lại và xu hướng hiện tại.

Chưa sửa

2 Tổng quan về kiến trúc CNN

CNN là mạng lưới truyền thẳng trong đó luồng thông tin diễn ra trong chỉ một hướng, từ đầu vào đến đầu ra của chúng. Cũng giống như mạng nơ-ron nhân tạo (ANN) được lấy cảm hứng từ sinh học, CNN cũng vậy. Vỏ não thị giác trong não, bao gồm các lớp xen kẽ của các đơn giản và phức tạp tế bào (Hubel & Wiesel, 1959, 1962), thúc đẩy kiến trúc của chúng. Kiến trúc CNN có nhiều biến thể; tuy nhiên, nhìn chung, chúng bao gồm của các lớp tích chập và gộp (hoặc lấy mẫu phụ), được nhóm lại thành các mô-đun. Một hoặc nhiều lớp được kết nối đầy đủ, như trong một tiêu chuẩn mạng nơ-ron truyền thẳng, hãy làm theo các mô-đun này. Các mô-đun thường xếp chồng lên nhau để tạo thành một mô hình sâu. Hình 1 minh họa kiến trúc CNN điển hình cho nhiệm vụ phân loại hình ảnh đồ chơi. Một hình ảnh được nhập vào trực tiếp đến mạng, và sau đó là một số giai đoạn tích chập và gộp. Sau đó, các biểu diễn từ các hoạt động này cung cấp cho một hoặc nhiều lớp được kết nối đầy đủ hơn. Cuối cùng, lớp được kết nối đầy đủ cuối cùng đưa ra nhãn lớp. Mặc dù đây là kiến trúc cơ sở phổ biến nhất được tìm thấy trong tài liệu, một số thay đổi về kiến trúc đã được đề xuất trong những năm gần đây với mục tiêu cải thiện độ chính xác phân loại hình ảnh hoặc



Hình 1: Quy trình phân loại hình ảnh của CNN.

giảm chi phí tính toán. Mặc dù đối với phần còn lại của phần này, chúng tôi chỉ giới thiệu thoáng qua kiến trúc CNN chuẩn, trong phần 5 chúng tôi sẽ giải quyết với một số thay đổi về thiết kế kiến trúc giúp nâng cao hiệu suất phân loại hình ảnh.

2.1 Các lớp tích chập. Các lớp tích chập đóng vai trò là tính năng trình trích xuất, và do đó chúng học đư ợc các biểu diễn đặc trư ng của đầu vào của chúng hình ảnh. Các tế bào thần kinh trong các lớp tích chập đư ợc sắp xếp thành các đặc điểm bản đồ. Mỗi nơ-ron trong bản đồ đặc điểm có một trư ờng tiếp nhận, đư ợc kết nối với vùng lân cận của các nơ-ron ở lớp trư ớc thông qua một tập hợp các trọng số có thể đào tạo đư ợc, đôi khi đư ợc gọi là ngân hàng bộ lọc (LeCun và cộng sự, 2015). Các đầu vào đư ợc tích chập với các trọng số đã học để tính toán một bản đồ đặc điểm và các kết quả tích chập đư ợc gửi qua một hàm kích hoạt phi tuyến tính. Tất cả các nơ-ron trong bản đồ đặc điểm đều có trọng số là bị ràng buộc phải bằng nhau; tuy nhiên, các bản đồ tính năng khác nhau trong cùng một lớp tích chập có trọng số khác nhau để có thể có nhiều tính năng đư ợc trích xuất tại mỗi vị trí (LeCun et al., 1998; LeCun et al., 2015). Chính xác hơn, bản đồ đặc điểm đầu ra thứ k Yk có thể đư ợc tính toán như

$$Yk = f(Wk \quad x) \tag{2.1}$$

trong đó hình ảnh đầu vào đư ợc biểu thị bằng x; bộ lọc tích chập liên quan đến bản đồ đặc điểm thứ k đư ợc ký hiệu là Wk; dấu nhân trong ngữ cảnh này đề cập đến toán tử tích chập 2D, đư ợc sử dụng để tính toán tích vô hư ớng của mô hình bộ lọc tại mỗi vị trí của hình ảnh đầu vào; và f(·) biểu diễn hàm kích hoạt phi tuyến tính (Yu, Wang, Chen, & Wei, 2014). Các hàm kích hoạt phi tuyến tính cho phép trích xuất các hàm phi tuyến tính các tính năng. Theo truyền thống, các hàm tiếp tuyến sigmoid và hyperbolic là đư ợc sử dụng; gần đây, các đơn vị tuyến tính chỉnh lư u (ReLUs; Nair & Hinton, 2010) đã trở nên phổ biến (LeCun et al., 2015). Sự phổ biến và thành công của họ đã

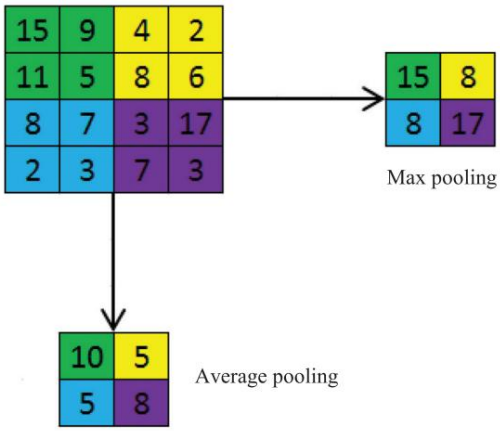
mở ra một lĩnh vực nghiên cứu tập trung vào việc phát triển và ứng dụng các hàm kích hoạt DCNN mới để cải thiện một số đặc điểm của hiệu suất DCNN. Do đó, trong phần 5.2, chúng tôi chính thức giới thiệu
ReLU và thảo luận về các động lực dẫn đến sự phát triển của chúng trư ớc
trình bày chi tiết về hiệu suất của một số hàm kích hoạt thay thế và dựa trên chỉnh lư u.

2.2 Các lớp gộp. Mục đích của các lớp gộp là để giảm
độ phân giải không gian của các bản đồ đặc điểm và do đó đạt đư ợc sự bất biến không gian đối với biến dạng đầu vào và bản dịch (LeCun et al., 1989a, 1989b; LeCun et al., 1998, 2015; Ranzato và cộng sự, 2007). Ban đầu, thông lệ chung là sử dụng các lớp tổng hợp nhóm trung bình để truyền bá trung bình của tất cả các đầu vào
giá trị của một vùng lân cận nhỏ của một hình ảnh đến lớp tiếp theo (LeCun et al., 1989a, 1989b; LeCun et al., 1998). Tuy nhiên, trong các mô hình gần đây hơn (Ciresan và cộng sự, 2011; Krizhevsky và cộng sự, 2012; Simonyan và Zisserman, 2014; Zeiler & Fergus, 2014; Szegedy, Liu và cộng sự, 2014; Xu và cộng sự, 2015), các lớp tổng hợp nhóm tối đa truyền bá giá trị tối đa trong một trư ờng tiếp nhận đến
lớp tiếp theo (Ranzato và cộng sự, 2007). Về mặt hình thức, nhóm tối đa chọn lớp lớn nhất yếu tố trong mỗi trư ờng tiếp nhận sao cho

$$\text{Bạn } j = \max_{(p,q)\ i\ j} xkpq, \tag{2.2}$$

nơi đầu ra của hoạt động gộp nhóm, liên quan đến tính năng thứ k
bản đồ, đư ợc biểu thị bằng Yki j, xkpq biểu thị phần tử tại vị trí (p, q) chứa bởi vùng tập hợp i j, thể hiện một trư ờng tiếp nhận xung quanh vị trí (i, j) (Yu et al., 2014). Hình 2 minh họa sự khác biệt giữa tối đa
gộp nhóm và gộp nhóm trung bình. Cho một hình ảnh đầu vào có kích thư ớc 4 × 4, nếu 2 × 2 bộ lọc và bư ớc tiến của hai đư ợc áp dụng, nhóm tối đa đư a ra giá trị tối đa của mỗi vùng 2 × 2, trong khi kết quả gộp trung bình đư a ra giá trị trung bình đư ợc làm tròn giá trị số nguyên của mỗi vùng lấy mẫu phụ. Trong khi động cơ đằng sau
di cư hư ớng tới nhóm tối đa đư ợc đề cập trong phần 4.2.3, cũng có một số mối quan tâm với việc gộp tối đa, dẫn đến sự phát triển của các chư ơng trình gộp nhóm khác. Chúng đư ợc giới thiệu trong phần 5.1.2.

2.3 Các lớp đư ợc kết nối đầy đủ. Một số lớp tích chập và lớp gộp thư ờng đư ợc xếp chồng lên nhau để trích xuất nhiều tính năng trừu tư ợng hơn biểu diễn trong việc di chuyển qua mạng. Các lớp đư ợc kết nối đầy đủ những lớp sau đây sẽ diễn giải các biểu diễn tính năng này và thực hiện chức năng của lý luận cấp cao (Hinton và cộng sự, 2012; Simonyan & Zisser-man, 2014; Zeiler & Fergus, 2014). Đối với các vấn đề phân loại, nó là tiêu chuẩn để sử dụng toán tử softmax (xem phần 5.3.1 và 5.3.5) trên DCNN (Krizhevsky và cộng sự, 2012; Lin và cộng sự, 2013; Simonyan & Zisserman, 2014; Zeiler và Fergus, 2014; Szegedy, Liu và cộng sự, 2014; Xu và cộng sự, 2015). Trong khi sớm



Hình 2: Nhóm trung bình so với nhóm tối đa.

thành công đã đạt đư ợc bằng cách sử dụng các hàm cơ sở bán kính (RBF), như là bộ phân loại trên đỉnh của các tháp tích chập (LeCun et al., 1998), Tang (2013) đã tìm thấy rằng việc thay thế toán tử softmax bằng máy vectơ hỗ trợ (SVM) dẫn đến độ chính xác phân loại đư ợc cải thiện (xem phần 5.3.4 để biết thêm chi tiết). Hơn nữa, với điều kiện là tính toán trong các lớp đư ợc kết nối đầy đủ là thư ờng bị thách thức bởi tỷ lệ tính toán trên dữ liệu của họ, một nhóm trung bình toàn cầu lớp (xem phần 5.1.1.1 để biết thêm chi tiết), đư a vào bộ phân loại tuyến tính tại đơn giản, có thể đư ợc sử dụng làm phư ơng án thay thế (Lin et al. 2013) bất chấp những nỗ lực này, so sánh hiệu suất của các bộ phân loại khác nhau trên DCNN vẫn cần đư ợc nghiên cứu thêm và do đó tạo ra một hư ớng nghiên cứu thú vị (xem phần 6 để biết các xu hư ớng DCNN nội tại khác).

2.4 Đào tạo. CNN và ANN nói chung sử dụng các thuật toán học tập để điều chỉnh các tham số miễn phí của chúng (tức là độ lệch và trọng số) để đạt đư ợc đầu ra mạng mong muốn. Thuật toán phổ biến nhất đư ợc sử dụng cho mục đích này là truyền ngư ợc (LeCun, 1989; LeCun et al., 1998; Bengio, 2009; Deng & Yu, 2014; Deng, 2014; Srinivas et al., 2016). Truyền ngư ợc tính toán độ dốc của một mục tiêu (còn đư ợc gọi là chi phí/tổn thất/ chức năng hiệu suất) để xác định cách điều chỉnh các tham số của mạng để giảm thiểu lỗi ảnh hư ởng đến hiệu suất. Một vấn đề thư ờng gặp khi đào tạo CNN, và đặc biệt là DCNN, là quá khớp, tức là hiệu suất kém trên một tập kiểm tra đư ợc giữ lại sau khi mạng đư ợc đào tạo trên một tập huấn luyện nhỏ hoặc thậm chí lớn. Điều này ảnh hư ởng đến khả năng của mô hình để khái quát hóa trên dữ liệu chư a thấy và là một thách thức lớn đối với DCNN có thể đư ợc xoa dịu bằng cách điều chỉnh, đư ợc khảo sát ở phần 5.4.

2.5 Thảo luận. Phần này tóm tắt một số khía cạnh cơ bản liên quan đến các khối xây dựng cơ bản của CNN. Chi tiết hơn
giải thích về hàm tích chập và các biến thể của nó và các lớp tích chập và lớp gộp, có thể đư ợc tìm thấy trong Goodfellow, Bengio và
Courville (đang in). Hơn nữa, đối với số học tích chập và gộp,
ngư ời đọc đư ợc giới thiệu đến Dumoulin và Visin (2016). Giải thích chi tiết
về thuật toán backpropagtion và các giao thức đào tạo chung cho deep
mạng nơ-ron (DNN) có sẵn trong LeCun et al. (1998) và Goodfel-low et al. (2016), trong khi LeCun et al. (2015) cung cấp một bản tóm tắt ngắn gọn về
thuật toán và học có giám sát (một trong những mô hình học máy chính, cùng với học không giám sát và học tăng cư ờng)
nói chung. Một lịch sử tóm tắt về sự phát triển của thuật toán phổ biến này,
cụ thể cho CNN, đư ợc cung cấp trong phần 3.2. Cuối cùng, một số DCNN
những cân nhắc về mặt lý thuyết, nhiều trong số đó đư ợc tóm tắt một cách ngắn gọn bởi Koushik (2016) đư ợc giới thiệu trong phần 6.1.

3 Sự phát triển ban đầu của CNN

Trong phần này, chúng tôi sẽ đề cập đến những phát triển ban đầu và những tiến bộ đáng kể của CNN, từ những ngư ời tiền nhiệm cho đến những ứng dụng thành công trư ớc đó.
đến thời kỳ phục hư ng của học sâu (Hinton et al., 2006; Hinton & Salakhutdi-Nov, 2006; Bengio, Lamblin, Popovici, & Larochelle, 2007).

3.1 Tiền thân của CNN lấy cảm hứng từ khoa học thần kinh. Sinh học đã truyền cảm hứng một số kỹ thuật trí tuệ nhân tạo như ANN, thuật toán tiến hóa và máy tự động tế bào (Floreano & Mattiussi, 2008). Tuy nhiên, có lẽ câu chuyện thành công lớn nhất trong số đó là CNN (Goodfellow, Bengio,
& Courville, đang in). Lịch sử của họ bắt đầu với các thí nghiệm thần kinh học do Hubel và Wiesel (1959, 1962) thực hiện từ năm 1959.
Đóng góp chính của công trình của họ là khám phá ra rằng các tế bào thần kinh ở các giai đoạn khác nhau của hệ thống thị giác phản ứng mạnh mẽ với các kích thích cụ thể.
các mẫu trong khi bỏ qua những mẫu khác. Cụ thể hơn, họ phát hiện ra rằng các tế bào thần kinh trong giai đoạn đầu của vỏ não thị giác chính phản ứng mạnh mẽ với các mẫu ánh sáng định hư ớng chính xác, chẳng hạn như các thanh, như ng bỏ qua các mẫu phức tạp hơn
các mô hình của kích thích đầu vào dẫn đến phản ứng mạnh mẽ từ các tế bào thần kinh ở các giai đoạn sau. Họ cũng phát hiện ra rằng vỏ não thị giác bao gồm các tế bào đơn giản, có các trư ờng tiếp nhận cục bộ và các tế bào phức tạp,
không thay đổi đối với các đầu vào bị dịch chuyển hoặc biến dạng, đư ợc sắp xếp theo cách phân cấp. Những tác phẩm này đã cung cấp nguồn cảm hứng ban đầu để mô hình hóa tầm nhìn tự động của chúng tôi hệ thống dựa trên đặc điểm của hệ thần kinh trung ư ơng.

Năm 1979, một mô hình mạng nơ-ron đa lớp mới lạ, có biệt danh là neocognitron, đã đư ợc đề xuất (Fukushima, 1979). Đư ợc mô hình hóa dựa trên phát hiện của Hubel và Wiesel (1959, 1962), nó cũng bao gồm đơn giản và
các ô phức tạp, đư ợc xếp chồng lên nhau theo cách phân cấp. Với kiến trúc này, mạng đã chứng minh đư ợc sự thành công trong việc nhận dạng các mẫu đầu vào đơn giản

bất kể sự thay đổi vị trí hay sự biến dạng đáng kể về hình dạng
của mẫu đầu vào (Fukushima, 1980; Fukushima & Miyake, 1982). Điều quan trọng là
neocognitron đã đặt nền tảng cho sự phát triển của
CNN. Trên thực tế, CNN đư ợc bắt nguồn từ neocognitron, và do đó chúng
có kiến trúc tư ơng tự (LeCun et al., 2015).

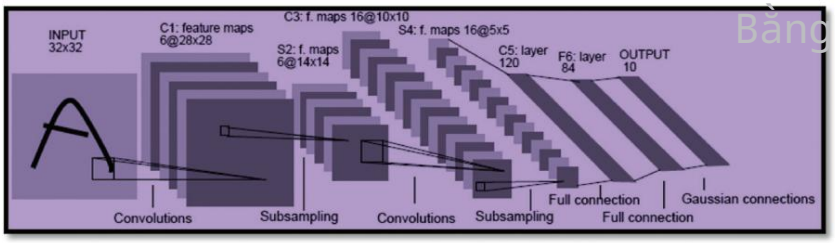
Bằng chứng

3.2 Lịch sử tóm tắt của Backpropagation và ứng dụng đầu tiên
CNN. Truyền ngư ợc đư ợc bắt nguồn từ những năm 1960. Đặc biệt, SE Drey-fus (1962) đã
bắt nguồn từ một phiên bản đơn giản hóa của thuật toán sử dụng chuỗi
quy tắc một mình. Tuy nhiên, các phiên bản đầu tiên của backpropagation không hiệu quả
vì chúng backpropagation thông tin phái sinh từ một lớp
đến lớp trư ớc mà không công khai giải quyết các liên kết trực tiếp giữa các lớp. Hơn
nữa, họ không xem xét đến các lợi ích hiệu quả tiềm năng do
sự thư a thớt của mạng lư ới (Schmidhuber, 2015). Hình thức hiệu quả hiện đại của
thuật toán giải quyết những vấn đề này đư ợc đư a ra vào năm 1970 (Linnainmaa,
1970); tuy nhiên, không có đề cập nào về việc sử dụng nó cho ANN. Sơ bộ
các cuộc thảo luận về việc sử dụng nó cho ANN có từ năm 1974 (Werbos, 1974); tuy
nhiên, ứng dụng đầu tiên đư ợc biết đến của sự lan truyền ngư ợc hiệu quả, cụ thể là
đối với ANN, đã đư ợc mô tả vào năm 1981 (Werbos, 1982), như ng điều này vẫn còn tư ơng
đối chư a đư ợc biết đến. Tuy nhiên, nó đã đư ợc "phổ biến đáng kể" (Schmid-huber,
2015) do một bài báo có tính chất khởi đầu vào năm 1986 của DE Rumelhart và cộng sự (1986),
điều này chứng minh rằng bằng cách sử dụng thuật toán học truyền ngư ợc,
các nơ-ron ẩn bên trong của ANN có thể đư ợc đào tạo để biểu diễn các tính năng quan
trọng của miền nhiệm vụ.

Năm 1989, LeCun et al. (1989a, 1989b) đã đề xuất lớp đa đầu tiên
CNN và áp dụng thành công các mạng lư ới quy mô lớn này vào các vấn đề phân loại hình
ảnh thực tế (chữ số viết tay và mã bư u chính). Các CNN ban đầu này gợi nhớ đến
neocognitron (Fukushima, 1979, 1980;
Fukushima & Miyake, 1982). Tuy nhiên, sự khác biệt chính là họ
đã đư ợc đào tạo theo cách đư ợc giám sát hoàn toàn bằng cách sử dụng phư ơng pháp truyền ngư ợc,
trái ngư ợc với chư ơng trình tăng cư ờng không giám sát đư ợc sử dụng bởi họ
ngư ời tiền nhiệm. Điều này cho phép họ dựa nhiều hơn vào tự động
học tập thay vì xử lý trư ớc đư ợc thiết kế thủ công để trích xuất tính năng
(LeCun et al., 1989a, 1989b; LeCun, 1989), trư ớc đây đã đư ợc chứng minh là
cực kỳ khó khăn; do đó, chúng tạo thành một thành phần thiết yếu của nhiều
các DCNN chiến thắng cuộc thi gần đây (Krizhevsky và cộng sự, 2012; Simonyan &
Zisserman, 2014; Zeiler và Fergus, 2014; Szegedy, Liu và cộng sự, 2014).

Chư a sửa

3.3 Giới thiệu về Bộ dữ liệu MNIST. Năm 1998, CNN đã mô tả
trư ớc đó (LeCun et al., 1989a, 1989b), đã đư ợc cải tiến và sử dụng cho
nhiệm vụ phân loại ký tự cá nhân trong ứng dụng nhận dạng tài liệu. Công trình này đã
đư ợc công bố trong một bài báo có tính chất khởi đầu chi tiết (LeCun et al.,
1998) đã nêu bật những lợi thế chính của CNN khi so sánh với
ANN truyền thống: chúng yêu cầu ít tham số miễn phí hơn (vì trọng lư ợng
chia sẻ), và họ xem xét cấu trúc không gian của dữ liệu đầu vào, do đó



Hình 3: Kiến trúc của LeNet-5 (LeCun và cộng sự, 1998).

cho phép họ xử lý sự thay đổi của hình dạng 2D. Ngoài ra
CNN được đề xuất, LeCun et al. (1998) đã giới thiệu bộ dữ liệu 70.000 của Viện Tiêu
chuẩn và Công nghệ Quốc gia (MNIST) đã được sửa đổi phổ biến
chữ số viết tay, từ đó đã được sử dụng rộng rãi cho một số nhiệm vụ thị giác máy tính
và đặc biệt là để phân loại và nhận dạng hình ảnh
vấn đề. Hình 3 minh họa kiến trúc của CNN, được gọi là LeNet-5,
được đề xuất bởi LeCun et al. (1998). Sơ đồ minh họa rõ ràng thiết kế
của LeNet-5, bao gồm xen kẽ tích chập và lấy mẫu phụ
các lớp, theo sau là một lớp duy nhất được kết nối đầy đủ.

3.4 Những thành công ban đầu của CNN mặc dù có những vấn đề nhận thức được với
Gradient De-scent. Vào cuối những năm 1990 và đầu những năm 2000, nghiên cứu về mạng nơ-
ron đã giảm dần (Simard và cộng sự, 2003; LeCun và cộng sự, 2015). Nó ít được sử dụng cho máy
nhiệm vụ học tập, và nhiệm vụ nhận dạng giọng nói và thị giác máy tính đã bỏ qua chúng.
Người ta tin rộng rãi rằng học tập tính năng đa giai đoạn hữu ích
máy chiết xuất, với ít kiến thức trước đó, là không khả thi do các vấn đề với
thuật toán tối ưu hóa phổ biến, gradient descent. Cụ thể, nó là
nghĩ rằng sự giảm dần độ dốc cơ bản sẽ không phục hồi được từ trọng lượng kém
cấu hình ức chế sự giảm thiểu của sự lan truyền ngược trung bình
lỗi, một hiện tượng được gọi là cực tiểu cục bộ kém (LeCun et al., 2015). Trong
Ngược lại, các phương pháp thống kê khác và đặc biệt là SVM đã trở nên phổ biến
do những thành công của họ (Decoste & Schölkopf, 2002). Trái ngược với xu hướng này, một
CNN được đề xuất để ứng dụng phân tích tài liệu trực quan vào năm 2003
(Simard và cộng sự, 2003).

Vào thời điểm mà CNN không được ưa chuộng trong cộng đồng kỹ thuật,
Simard et al. (2003) đã có thể đạt được kết quả phân loại được biết đến nhiều nhất
trên tập dữ liệu MNIST (LeCun et al., 1998), cải thiện kết quả trước đó
kết quả tốt nhất thu được bởi SVM của Decoste và Schölkopf (2002). Trích dẫn những lợi
thế đã được LeCun và cộng sự (1998) đề cập, sử dụng
CNN cho các nhiệm vụ trực quan, chúng đã mở rộng quy mô và chất lượng của MNIST
bộ dữ liệu và đề xuất sử dụng các vòng lặp phần mềm đơn giản cho hoạt động tích chập.
Các vòng lặp này khai thác tính chất lan truyền ngược
cho phép ANN được thể hiện theo cách mô-đun và điều này cho phép

để gỡ lỗi phần mềm mô-đun. Mặc dù LeCun et al. (1998) đã
được giả định và chứng minh rằng bằng cách tăng kích thước của tập dữ liệu, sử dụng các
phép biến đổi afin được tạo ra nhân tạo, hiệu suất của mạng sẽ
cải thiện, Simard et al. (2003) đã cải thiện chất lượng của phần tăng lên của bộ dữ liệu
để cải thiện hiệu suất hơn nữa. Điều này đã được thực hiện
bằng cách sử dụng biến dạng hình ảnh đàn hồi. Công trình này là một phần của một loạt
một số ứng dụng nhận dạng ký tự quang học sử dụng CNN. Đặc biệt, Microsoft đã sử dụng
chúng cho các chữ số viết tay tiếng Anh (Simard et al.,
2003; Chellapilla, Shilman, & Simard, 2006), nhận dạng chữ viết tay tiếng Ả Rập
(Abdulkader, 2006) và nhận dạng ký tự viết tay Đông Á
(Chellapilla & Simard, 2006). Vì vậy, các ứng dụng này, cùng với
công trình được mô tả bởi LeCun et al. (1989a, 1989b, 1998), đại diện cho một số
những thành công phân loại hình ảnh ban đầu được CNNs hưởng lợi. Bối cảnh để
phần tiếp theo sẽ nêu bật một số thành công khác.

4 Sự Phục Hưng Của Học Sâu Và Sự Trỗi Dậy Của DCNN

Phần này giới thiệu tóm tắt về thời kỳ phục hưng của học sâu và tập trung vào
những đóng góp đáng kể của DCNN vào sự gia tăng hiện tại trong học sâu
nghiên cứu. Nó cũng bao gồm một bài báo có tính chất khởi đầu và một số tác phẩm tiêu biểu
đã dẫn đến sự thống trị gần đây của chúng so với các phân loại hình ảnh khác
kỹ thuật.

4.1 Bối cảnh của thời kỳ Phục hưng Học sâu. Các mạng nơ-ron đa lớp feedforward đầu
tiên được đào tạo vào năm 1965 (Ivakhnenko &
Lapa, 1966), và mặc dù chúng không sử dụng phương pháp truyền ngược, nhưng chúng có lẽ
là hệ thống học sâu đầu tiên (Schmidhuber, 2015). Mặc dù sâu
các thuật toán giống như học tập có lịch sử lâu dài, thuật ngữ học sâu đã trở thành một
câu cửa miệng vào khoảng năm 2006, khi các mạng lưới niềm tin sâu sắc (DBN) và các bộ
mã hóa tự động được đào tạo theo cách không giám sát được sử dụng để khởi tạo DNN,
được đào tạo bằng cách sử dụng backpropagation (Hinton et al., 2006; Hinton & Salakhut-
dinov, 2006; Bengio et al., 2007). Trước đó, người ta đã dạy rằng sâu
mạng nhiều lớp (bao gồm cả DCNN) quá khó để đào tạo do
các vấn đề liên quan đến việc giảm dần độ dốc và do đó không phổ biến (Bengio và cộng sự, 2007;
Bengio, 2009; Deng & Yu, 2014; Schmidhuber, 2015; Goodfellow và cộng sự, trong
(báo chí). Ngược lại, CNN là một ngoại lệ đáng chú ý và tỏ ra dễ dàng hơn
đào tạo khi so sánh với các mạng được kết nối đầy đủ (Simard et al., 2003, Ben-gio,
2009; LeCun et al., 2015; Goodfellow et al., đang in). Ngoài ra
những thành công được thảo luận trong phần 3.3, một số ứng dụng thành công khác kết hợp
CNN cho thành phần phân loại hình ảnh của họ
trước khi mạng lưới thần kinh hồi sinh vào năm 2006 bao gồm hình ảnh y tế
phân đoạn (Ning et al., 2005); nhận dạng khuôn mặt, phát hiện và xác minh (Lawrence,
Giles, Tsoi, & Back, 1997; Garcia & Delakis, 2002; Chopra,
Hadsell, & LeCun, 2005); tránh chướng ngại vật trên đường (Muller, Ben, Cosatto,

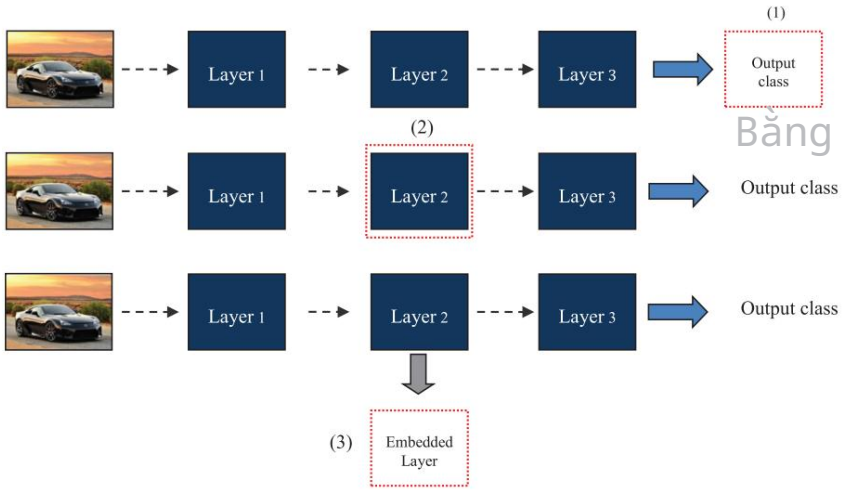
Flepp, & LeCun, 2005); và phân loại đối tư ợng chung (LeCun, Huang, &
Bottou, 2004; Hoàng và LeCun, 2006).

Tuy nhiên, vì nghiên cứu mạng lư ới thần kinh đã chậm lại vào cuối những năm 1990 và đầu những năm 2000 (Simard et al., 2003; LeCun et al., 2015), sự phát triển của CNN là cũng bị cản trở, như ng nó đã hồi sinh vào khoảng năm 2006. Sử dụng mô hình dựa trên năng lư ợng để trích xuất các tính năng thư a thớt, có một số ứng dụng bao gồm phân loại và phân đoạn, sau đó sử dụng kết quả đầu ra để khởi tạo
lớp đầu tiên của DCNN, Ranzato, Poultney, Chopra và LeCun (2006)
cải thiện đôi chút kết quả phân loại đư ợc báo cáo tốt nhất trư ớc đó (Simard
et al., 2003) trên tập dữ liệu MNIST (LeCun et al., 1998). Trích dẫn Hinton et al.
(2006), mô hình DCNN của họ, có kiến trúc tư ơng tự như của Le-Cun et al. (1998) như ng sử dụng số lư ợng bản đồ đặc điểm lớn hơn đáng kể
để tạo ra các tính năng thư a thớt, đư ợc đào tạo trư ớc theo cách không giám sát và bao gồm ba thành phần thiết yếu. Một bộ mã hóa thẩm vấn hình ảnh đầu vào và tính toán một vectơ mã của hình ảnh, sau đó đư ợc chuyển đổi thành một vectơ mã thư a thớt bằng một mô-đun logistic thư a thớt phi tuyến tính.
Một bộ giải mã tính toán phiên bản phục hồi của hình ảnh đầu vào đư ợc giải mã vectơ mã thư a thớt và đầu ra của nó đư ợc sử dụng để khởi tạo lớp đầu tiên trọng số của CNN. Công trình này là công trình đầu tiên sử dụng DCNN đư ợc khởi tạo bởi kỹ thuật đào tạo không giám sát trong thời gian học sâu
thời kỳ phục hư ng và dẫn đến một số nỗ lực đào tạo trư ớc không có giám sát khác trong khoảng thời gian từ năm 2006 đến năm 2011, như phần tiếp theo sẽ trình bày.

4.2 Sự phục hư ng của học sâu đư ợc thúc đẩy bởi GPU và cải thiện
Thuật toán.

4.2.1 Tiền huấn luyện không giám sát. Lấy cảm hứng từ tốc độ và độ chính xác lợi thế của việc đào tạo trư ớc không có giám sát (Hinton et al., 2006; Hinton & Salakhutdinov, 2006; Bengio và cộng sự, 2007; , Ranzato và cộng sự, 2006), Ranzato và cộng sự. (2007) đã sử dụng kiến trúc giống DCNN đư ợc đào tạo theo cách không giám sát để tìm hiểu các tính năng thư a thớt phân cấp không thay đổi cục bộ đối với các thay đổi nhỏ và sự biến dạng. Cách tiếp cận của họ, giới thiệu việc gộp tối đa (xem phần 2.2 và 4.2.3), đã đạt đư ợc kết quả rất gần với trạng thái hiện đại cho
MNIST (LeCun và cộng sự, 1998; Ranzato và cộng sự, 2006) và California
Viện Công nghệ (CALTECH-101—Fei-Fei, Fergus, & Perona, 2006;
Zhang, Berg, Maire, & Malik, 2006) chuẩn mực. Mặc dù thành công ban đầu này,
DCNN vẫn chư a thoát khỏi những thay đổi và biến dạng trên quy mô lớn; điều này vẫn còn một lĩnh vực nghiên cứu mở (xem phần 6.2).

Khẳng định rằng các phư ơng pháp đào tạo trư ớc mà Hinton et al. (2006), Bengio et al. (2007), và Ranzato et al. (2007), đư ợc sử dụng phức tạp và hạn chế,
Weston, Ratle và Collobert (2008) đã trình bày một cách đơn giản hơn để thực hiện học sâu bằng cách kết hợp các thuật toán nhúng phi tuyến tính với kiến trúc nhiều lớp sâu (bao gồm DCNN), đư ợc đào tạo theo cách có giám sát.
Sơ đồ học sâu bán giám sát kết quả đư ợc lấy cảm hứng từ
SVM Laplacian đư ợc trình bày bởi Belkin, Niyogi và Sindhwani (2006) và



Hình 4: Các chế độ nhúng chính quy khác nhau vào kiến trúc sâu.

đã mang lại tỷ lệ lỗi cạnh tranh trên tập dữ liệu MNIST (LeCun et al., 1998), khi so sánh với các kỹ thuật bán giám sát nông khác (Belkin, Niyogi và Sindhwani, 2006; Collobert, Sinz, Weston, & Bottou, 2006) và các phư ơng pháp học sâu hiện có của thời điểm đó (Hinton et al., 2006; Ranzato et al., 2007; Salakhutdinov & Hinton, 2007). Hình 4 cho thấy cách các thuật toán nhúng đư ợc thêm vào để điều chỉnh toàn bộ đầu ra của mạng, các lớp ẩn hoặc một mạng lư ới phụ trợ có cùng các lớp ban đầu của mạng ban đầu như ng là một tập hợp trọng số cuối cùng mới. Trong hình, các đư ờng màu đỏ bị phá vỡ minh họa điểm mà các thuật toán nhúng đư ợc đư ợc hợp nhất.

Theo hư ớng đào tạo trư ớc DCNN không giám sát (Ranzato et al., 2006, 2007) và nhúng bán giám sát (Weston và cộng sự, 2008), Ahmed, Yu, Xu, Gong và Xing (2008) lần đầu tiên thực hiện một tập hợp các nhiệm vụ giả trên dữ liệu theo cách không có sự giám sát và sau đó chuyển giao kiến thức thu đư ợc đến DCNN thông qua học chuyển giao. Tất cả các lớp của DCNN, bao gồm lớp phân loại cuối cùng, đư ợc đào tạo với backpropagation. Kết quả của họ suy ra rằng việc chuyển giao kiến thức tiếp theo là đào tạo có giám sát đã cải thiện Hiệu suất DCNN và có thể đư ợc áp dụng cho một loạt các nhiệm vụ trực quan, bao gồm nhận dạng đối tư ợng, giới tính và dân tộc. Chi tiết hơn có sẵn trong bài báo gốc (Ahmed et al., 2008); tổng quan về các hình thức khác nhau của việc chuyển giao kiến thức và một số thành công ban đầu của nó đư ợc cung cấp bởi Fei-Fei (2006). Gần đây, các tính năng đư ợc trích xuất bởi DCNN đã đư ợc chứng minh là cung cấp một cơ sở đáng kinh ngạc cho nhiều nhiệm vụ thị giác máy tính, bao gồm nhận dạng cảnh, nhận dạng chi tiết, phát hiện thuộc tính, hình ảnh truy xuất và quan trọng nhất là phân loại hình ảnh (Razavian, Azizpour,

Sullivan, & Sarlsson, 2014). Lợi thế rõ ràng đối với các hệ thống thị giác máy tính sử dụng kiến thức đư ợc chuyển giao từ DCNN là chi phí quá cao của chúng thời gian đào tạo có thể đư ợc loại bỏ, do đó giảm thời gian phát triển và triển khai các chư ơng trình như vậy.

Một nghiên cứu chi tiết đã điều tra tác động của tính phi tuyến tính tuân theo các bộ lọc tích chập trong DCNN; hiệu suất của các bộ lọc tích chập có giám sát, không giám sát và học ngẫu nhiên; và lợi thế (nếu có) của việc sử dụng hai giai đoạn trích xuất tính năng so với một đư ợc thực hiện bởi Jarrett, Kavukcuoglu và Lecun (2009), và LeCun, Kavukcuoglu và Farabet (2010). Họ phát hiện ra rằng các phi tuyến tính bao gồm chỉnh lư u và chuẩn hóa độ tư ơng phản cục bộ là chìa khóa cho độ chính xác tốt trên MNIST (LeCun và cộng sự, 1998), CALTECH-101 (Fei-Fei và cộng sự, 2006), và dữ liệu NYU Object Recognition Benchmark (NORB—LeCun et al., 2004) bộ, và độ chính xác phân loại tốt hơn đã thu đư ợc từ hai giai đoạn của việc trích xuất tính năng chứ không phải một. Đặc biệt, họ đã lập một kỷ lục mới trên tập dữ liệu MNIST chư a sửa đổi, cải thiện hiệu suất tốt nhất trư ớc đó (Ranzato và cộng sự, 2006) bằng cách tuân theo quá trình đào tạo trư ớc không giám sát, sử dụng phư ơng pháp đư ợc gọi là phân tích thư a thớt dự đoán (PSD; Kavukcuoglu, Ran-zato, & LeCun, 2010), với sự củng cố có giám sát. Kỹ thuật PSD, giống như công trình do Ranzato và cộng sự đề xuất (2006), dựa trên kiến trúc mã hóa-giải mã thực thi các ràng buộc thư a thớt trên vectơ đặc trư ng bằng cách sử dụng bộ hồi quy truyền thẳng cơ bản đư ợc đào tạo để ư ớc tính giải pháp thư a thớt cho tất cả các bản vá vector hóa hoặc các ngăn xếp của chúng trong một quy trình bộ đào tạo. Mặc dù các thuật toán mã hóa thư a thớt thư ờng có tính toán quá mức, vì kỹ thuật PSD xấp xỉ các mã thư a thớt, nên nó tính toán rẻ hơn, làm cho nó rất nhanh so với các mã hóa thư a thớt khác các kế hoạch.

Quá trình đào tạo trư ớc không giám sát (bao gồm cả bán giám sát), tiếp theo là quá trình tinh chỉnh có giám sát, đư ợc thảo luận trong phần này, đã trở nên phổ biến bởi mạng lư ới niềm tin sâu sắc đư ợc đề xuất khi thời kỳ phục hư ng học sâu trỗi dậy (Hinton và cộng sự, 2006; Hinton & Salakhutdinov, 2006; Bengio và cộng sự, 2007). Các các lư ợc đồ không giám sát phổ biến nhất sử dụng các phư ơng pháp phân kỳ tư ơng phản (Hin-ton, 2002) (xem Lee, Grosse, Ranganath, & Ng, 2009), ràng buộc thư a thớt (Ranzato và cộng sự, 2006, 2007) hoặc PSD (Kavukcuoglu và cộng sự, 2010; LeCun et al., 2010). Nhìn chung, đối với các kỹ thuật này, việc trích xuất tính năng các bộ lọc đư ợc đào tạo sao cho các biểu diễn ở một giai đoạn cụ thể có thể đư ợc xây dựng lại từ các biểu diễn của một giai đoạn trư ớc đó. Rào cản chính của cách tiếp cận này là quá trình học tính năng là độc lập với nhiệm vụ, mặc dù Bengio et al. (2007), Mairal, Bach, Ponce, Sapiro và Zisser-man (2008), và Ranzato và Szummer (2008) đã cố gắng giảm bớt điều này bằng cách kết hợp các tiêu chí có giám sát với các kỹ thuật không có giám sát.

Hơn nữa, mặc dù có những kết quả ban đầu đầy hứa hẹn thu đư ợc từ quá trình đào tạo trư ớc không đư ợc giám sát (xem Erhan et al., 2010, để biết phân tích chi tiết), trong thời gian gần đây nhiều năm, học có giám sát đã trở thành mô hình hàng đầu để đào tạo DCNN (xem phần 5.3). Tuy nhiên, học bán giám sát thì

hợp lý về mặt sinh học. Ví dụ, hãy xem xét cách trẻ em học về
môi trư ờng hoặc cụ thể hơn là cách chúng học cách nhận biết hoặc phân loại
các đối tư ợng. Họ thư ờng đư ợc ngư ời chăm sóc cung cấp một vài ví dụ, tư ơng tự như
việc học có giám sát bán phần hoặc yếu, và họ sử dụng điều này để khái quát hóa các đối
tư ợng không nhìn thấy. Do đó, để căn chỉnh các mô hình giám sát chặt chẽ hiện tại của chúng ta
gần gũi hơn với thiên nhiên, ngư ời ta hình dung rằng các DCNN trong tư ơng lai sẽ quay lại sử dụng
các chư ơng trình bán giám sát, tư ơng tự như những chư ơng trình đư ợc giới thiệu trong phần này. Những
các chư ơng trình sẽ kết hợp, ít nhất là ban đầu, các tiêu chí đư ợc giám sát để khắc phục
các vấn đề đã biết với các đối tác không đư ợc giám sát của họ. Tiến trình như vậy sẽ
cuối cùng dẫn đến các hệ thống độc lập, không giám sát để giải quyết khối lư ợng dữ liệu
chư a đư ợc chú thích ngày càng lớn hiện có (xem phần 6.6 để biết thêm thông tin chi tiết).

Bằng chứng

4.2.2 GPU kích thích nghiên cứu về DCNN. Mặc dù các thuật toán học sâu hiện đang hoạt
động đã có từ những năm 1980
(LeCun et al., 1989a, 1989b), họ đư ợc dạy là quá tính toán
tốn kém để cho phép nghiên cứu nhiều về phần cứng có sẵn trư ớc đó
đến năm 2006 (Goodfellow và cộng sự, đang in). Hơn nữa, trong quá trình thực hiện chư ơng
trình, các hoạt động tích chập tốn kém về mặt tính toán và do đó làm cho DC-NN chậm hơn
đáng kể khi đánh giá so với ANN tiêu chuẩn
cùng độ lớn. Để khắc phục những hạn chế này, K. Chellapilla, Puri,
và Simard (2006) đã đề xuất ba phư ơng pháp mới để tăng tốc DCNN: mở cuộn tích chập, sử
dụng các chư ơng trình con phần mềm đại số tuyến tính cơ bản và sử dụng GPU. Mặc dù GPU
đã đư ợc áp dụng cho ANN (Oh & Jung,
2004; Steinkrau, Simard, & Buck, 2005), công trình này có ý nghĩa quan trọng vì nó
là lần triển khai đầu tiên của DCNN sử dụng GPU. Theo thời gian, điều này đã
trở thành một khía cạnh quan trọng của hầu hết các DC-NN từng đoạt giải thư ởng hoặc hiện
đại
& Fergus, 2013, 2014; Simonyan và Zisserman, 2014; Szegedy và cộng sự, 2015; Chào
et al., 2015a). Mặc dù sự phát triển của phần cứng nâng cao để tạo điều kiện thuận lợi
Tính toán DCNN vẫn là một lĩnh vực nghiên cứu mở, nó đã trở nên phần lớn
đư ợc thư ơng mại hóa trong những năm gần đây. Với xu hư ớng này, phần lớn trọng tâm học
thuật là ứng dụng phần cứng thư ơng mại này hoặc phát triển thuật toán để hỗ trợ xử lý
nhanh hơn. Mặc dù điều này
không đư ợc dự kiến sẽ thay đổi trong tư ơng lai gần, có một kỳ vọng rằng
những tiến bộ về phần cứng và phần mềm sắp tới sẽ tập trung vào việc triển khai
của DCNN tới các thiết bị di động (xem phần 6.3).

Chưa sửa

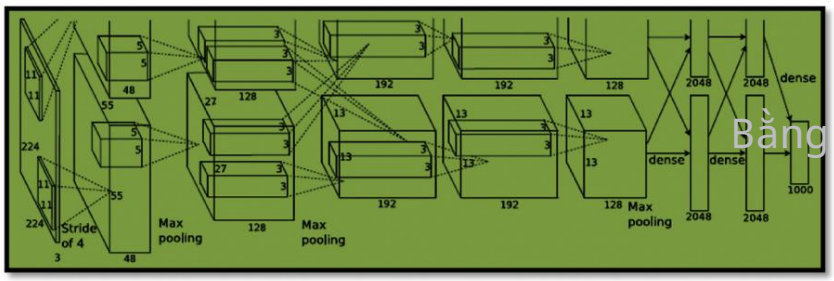
4.2.3 Max Pooling dẫn đến việc cải thiện khả năng tổng quát hóa. Năm 2007, backpropa-
gation lần đầu tiên đư ợc áp dụng cho kiến trúc giống DCNN sử dụng
nhóm tối đa (Ranzato và cộng sự, 2007). Năm 2010, Scherer, Müller và Behnke
(2010) đã chứng minh theo kinh nghiệm rằng hoạt động gộp tối đa vư ợt trội hơn nhiều
trong việc nắm bắt sự bất biến trong dữ liệu giống hình ảnh và có thể dẫn đến cải thiện
tổng quát hóa và hội tụ nhanh hơn khi so sánh với một mẫu con
hoạt động. Họ đã chứng minh điều này bằng cách đạt đư ợc kết quả đư ợc công bố tốt nhất

trên tập dữ liệu NORB chuẩn hóa-đồng nhất (LeCun et al., 2004), cải thiện
trên mức tốt nhất trư ớc đó (Nair & Hinton, 2009) hơn một nửa phần trăm Tiếp tục với công
trình thực nghiệm, Jarrett et al. (2009) đã chỉ ra rằng việc gộp tối đa làm giảm nhu cầu
về lớp chỉnh lư u, thư ờng không phải là một phần
của kiến trúc DCNN; tuy nhiên, họ thấy rằng việc gộp trung bình không
hư ởng cùng một lợi ích và do đó chịu tác động hủy bỏ giữa
đầu ra bộ lọc lân cận.

Một phân tích lý thuyết chi tiết về tổng hợp tối đa và tổng hợp trung bình, đư ợc bổ
sung bằng các đánh giá thực nghiệm, đã đư ợc thực hiện bởi Boureau, Ponce,
và LeCun (2010). Họ kết luận rằng hiệu suất của tối đa hoặc
việc gộp trung bình phụ thuộc vào dữ liệu và các tính năng của nó, và đối với một
vấn đề phân loại cho trư ớc, sử dụng một trong hai chiến lư ợc gộp riêng lẻ có thể không
là tối ư u. Vì max pooling chỉ đư ợc thiết kế cho các mạng feedforward, Lee, Gross,
Ranganath và Ng (2009) đã giới thiệu và áp dụng max pooling xác suất cho DBN tích chập
với mục đích mở rộng DBN
(Hinton et al., 2006) thành hình ảnh có kích thư ớc đầy đủ, nhiều chiều. Kết quả của chúng
mô hình sinh sản phân cấp bất biến dịch thuật hoạt động tốt trên một số chuẩn mực phân
loại, bao gồm MNIST (LeCun et al., 1998) và
CALTECH-101 (Fei-Fei và cộng sự, 2006). Mặc dù ngư ời ta biết rõ rằng tối đa
việc gộp chung dẫn đến một mức độ bất biến nhất định đối với sự bóp méo và dịch chuyển,
nó thực hiện điều này bằng cách loại bỏ thông tin không gian (Ranzato et al.,
2007; Scherer và cộng sự, 2010; Szegedy, Liu và cộng sự, 2014). Mặc dù vậy, nó vẫn tiếp tục
là một thành phần quan trọng của một số DCNN hiện đại (Ciresan et al.,
2011; Krizhevsky và cộng sự, 2012; Simonyan và Zisserman, 2014; Szegedy, Lư u
et al., 2014). Đối với một số vấn đề liên quan đến việc gộp tối đa và trung bình cũng như
các giải pháp đề xuất của chúng, hãy tham khảo phần 5.1.2.

4.3 Điểm thay đổi trong ứng dụng DCNN cho máy tính

Nhiệm vụ tầm nhìn. Sự hồi sinh của học sâu năm 2006 (Hinton và cộng sự, 2006;
Hinton & Salakhutdinov, 2006; Bengio et al., 2007), thúc đẩy một số ứng dụng thành công
của DCNN cho nhiều nhiệm vụ khác nhau. Những nhiệm vụ này bao gồm
phân loại và nhận dạng hình ảnh và đối tư ợng (Chellapilla, Puri, & Simard,
2006; Ranzato và cộng sự, 2007; Weston và cộng sự, 2008; Jarrett và cộng sự, 2009; Lee và cộng sự,
2009; LeCun và cộng sự, 2010; Scherer và cộng sự, 2010; Boureau và cộng sự, 2010; Đàn ông,
Meier, Ciresan, & Schmidhuber, 2011), phát hiện khuôn mặt (Nasse, Thurau, &
Fink, 2009) và phân đoạn hình ảnh (Turaga và cộng sự, 2010). Hơn nữa,
họ cũng tìm thấy những ứng dụng thú vị trong việc phân tích cảnh (Farabet, Couprie,
Najman, & LeCun, 2012), tầm nhìn cho việc lái xe địa hình tự động (Had-sell et al., 2009)
và nhận dạng cử chỉ tay (Nagi et al., 2011). Mặc dù
những thành tựu này, chúng vẫn bị loại bỏ phần lớn bởi dòng chính
cộng đồng về thị giác máy tính và học máy (LeCun et al., 2015).
Điều này đã thay đổi sau ILSVRC 2012 (Russakovsky và cộng sự, 2015), khi một
DCNN đư ợc giám sát hoàn toàn đã đạt đư ợc kết quả phân loại phá kỷ lục về
một tập hợp con của bộ dữ liệu ImageNet (Krizhevsky và cộng sự, 2012). Công trình này đã
đã cách mạng hóa lĩnh vực thị giác máy tính và kết quả là DCNN đã



Hình 5: Kiến trúc DCNN được chia thành hai GPU (Krizhevsky và cộng sự, 2012).

kể từ khi trở thành kiến trúc hàng đầu cho hầu hết các tác vụ trực quan, đặc biệt là đối với các ứng dụng liên quan đến phân loại hình ảnh, như phần còn lại của bài đánh giá này chương trình.

Trọng tâm của thành công của họ là họ đã thực hiện một số điều mới lạ và khác thường kỹ thuật. Thay vì sử dụng tiếp tuyến sigmoid hoặc hyperbolic truyền thống các hàm kích hoạt, chúng được lấy cảm hứng từ Jarrett et al. (2009) và được sử dụng ReLU (Nair & Hinton, 2010) kích hoạt, cho phép thời gian đào tạo nhanh hơn nhiều (xem phần 5.2.1 để biết thêm chi tiết). Vì mạng của họ quá lớn để vừa với một GPU, họ trải nó ra trên hai GPU được sắp xếp song song cấu hình, tương tự như DCNN đa cột được đề xuất bởi Ciresan, Meier và Schmidhuber (2012). Lấy cảm hứng từ sự chuẩn hóa tương phản cục bộ của Jarrett và cộng sự (2009), họ đã áp dụng chuẩn hóa phản ứng cục bộ. Được biểu thị về mặt toán học, nếu một hạt nhân i, ở vị trí (x, y) được sử dụng để tính toán hoạt động của một tế bào thần kinh được biểu thị b ai $_{x, y}$ và tính phi tuyến tính của ReLU là hơn được áp dụng, hoạt động chuẩn hóa phản ứng bi $_{x, y}$ có thể được thể hiện như

$$b^{i}_{x,y} = a^{i}_{x,y} \Big/ \left(k + \alpha \sum_{j=\max(0,\, i-n/2)}^{\min(N-1,\, i+n/2)} \left(a^{j}_{x,y}\right)^{2}\right)^{\beta}, \qquad (4.1)$$

trong đó N là tổng số hạt nhân trong lớp và tổng chạy qua n bản đồ hạt nhân "liền kề" ở cùng một vị trí không gian. Sơ đồ này hỗ trợ tổng quát hóa và giảm tỷ lệ lỗi phân loại mạng của họ. Họ tiếp tục giảm lỗi phân loại bằng cách chồng chéo giá trị tối đa của mạng các lớp gộp. Hình 5 minh họa kiến trúc mang tính cách mạng được trình bày bởi Krizhevsky và cộng sự (2012). Nó bao gồm năm lớp tích chập, ba trong đó được theo sau bởi các lớp nhóm tối đa và ba lớp được kết nối đầy đủ các lớp. Các phần lớp khác nhau ở nửa trên của hình chạy trên một GPU, trong khi các phần lớp ở phía dưới chạy trên GPU thứ hai. Các GPU chỉ tương tác với nhau ở các lớp cụ thể.

Để khắc phục tình trạng quá khớp, các tác giả đã sử dụng một kỹ thuật chính quy hóa được gọi là Dropout (Hinton và cộng sự, 2012). Cụ thể, khi mỗi trường hợp đào tạo được trình bày cho mạng trong giai đoạn đào tạo, mỗi neuron ẩn được loại bỏ ngẫu nhiên khỏi mạng với xác suất của 0,5. Do đó, các tế bào thần kinh ẩn không thể dựa vào sự hiện diện của các tế bào thần kinh ẩn khác và điều này ngăn cản sự thích nghi phức tạp của các đặc điểm trên dữ liệu đào tạo. Vào thời điểm thử nghiệm, tất cả các tế bào thần kinh ẩn đã được sử dụng, nhưng đầu ra được nhân với 0,5 để bù đắp cho thực tế là gấp đôi số lượng tế bào thần kinh hiện đang hoạt động. Kết quả của điều này là hiệu ứng chính quy hóa mạnh mẽ làm giảm đáng kể tình trạng quá khớp (Krizhevsky và cộng sự, 2012; Hinton et al., 2012; Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhut-dinov, 2014). Hình 11 (trong phần 5.4.2) cho thấy tác động của Dropout trên mạng lưới truyền thẳng chuẩn, với hai lớp ẩn, trong khi phần 5.4.1 cung cấp một mô tả chính thức hơn về kỹ thuật và giới thiệu một số các biến thể của nó.

Quá trình lắp ghép được giảm thêm bằng cách áp dụng tăng cường dữ liệu, một thủ tục phổ biến để mở rộng một cách giả tạo một tập dữ liệu (LeCun et al., 1998; Simard et al., 2003; Ciresan et al., 2011, 2012). Đặc biệt, họ đã tạo ra nhiều hình ảnh hơn bằng cách áp dụng phép dịch chuyển và phản xạ theo chiều ngang vào hình ảnh đào tạo, thay đổi cường độ của các kênh màu của chúng và thực hiện phân tích thành phần chính (PCA) trên các giá trị pixel của chúng, dẫn đến cải thiện hiệu suất phân loại. Mô hình của Krizhevsky et al. (2012) đã được sử dụng rộng rãi cho nhiều mục đích khác nhau kể từ khi phát triển. Một lượng lớn nghiên cứu đã sử dụng nó để đánh giá chuẩn các mô hình của họ hoặc làm mô hình cơ sở để thử nghiệm các thuật toán mới. Hơn nữa, mô hình của họ đã truyền cảm hứng cho công việc của DCNN và đã trở thành một trong những người đóng góp chính cho sự gia tăng gần đây của DCNN công nghệ cho các ứng dụng liên quan đến phân loại hình ảnh.

Đáng chú ý là, trước công trình tiên phong của Krizhevsky et al. (2012) là loạt công trình do Ciresan và cộng sự đề xuất (2011, 2012). Họ đã trình bày CNN phân cấp sâu, được đào tạo theo cách được giám sát hoàn toàn, đã đạt được kết quả được công bố tốt nhất về NORB (LeCun et al., 1998; Krizhevsky, 2009; Coates, Lee, & Ng, 2011) và MNIST (LeCun et al., 1998; Ciresan, Meier, Gambardella, & Schmidhuber, 2010) các chuẩn phân loại (Cire-san et al., 2011). Bằng cách xếp chồng các DCNN này thành các cột, họ đã cải thiện thêm trạng thái nghệ thuật cho các chuẩn này và đặc biệt là đạt được hiệu suất ở cấp độ con người trên bộ dữ liệu MNIST. Hơn nữa, trên chuẩn mực nhận dạng biển báo giao thông của Đức (GTSRB; Stallkamp, Schlipsing, Salmen, & Igel, 2011), chúng đã vượt qua hiệu suất của con người với hệ số hai.

Bảng 1 tóm tắt các thuộc tính chính của dữ liệu phân loại hình ảnh các bộ dữ liệu phân loại khác hiện có (xem phần 5.1.2.3 và 5.3.2), đây là những phần được sử dụng phổ biến nhất cho DCNN đánh giá và đánh giá chuẩn. Trong số đó, tập dữ liệu MNIST (LeCun et al., 1998) đã vượt qua thử thách của thời gian và trở thành phổ biến nhất,

[illegible]

[illegible]u	[illegible]h	[illegible]	[illegible]ng	[illegible]ng	[illegible]
MNIST	[illegible]t	[illegible]00		[illegible]	[illegible]n
CALTECH-101	[illegible]h		[illegible]00		
CALTECH-256	[illegible]h				[illegible]ffin, [illegible]n,
[illegible]n		96			[illegible]n
[illegible]o		96			[illegible]n
CIFAR-10	[illegible]n	32	[illegible]00	10	[illegible]hevsky,
CIFAR-100	[illegible]n			100	[illegible]hevsky,
ILSVRC	[illegible]			1000	[illegible]akovsky

mặc dù các hệ thống phân loại hiện đại đư ợc đánh giá dựa trên sự thành công của chúng ILSVRC (Russakovsky và cộng sự, 2015), như phần tiếp theo sẽ trình bày.

4.4 Những cải tiến tiêu biểu minh họa cho sự thống trị của DCNN. Kể từ công trình đột phá của Krizhevsky et al. (2012), DCNN đã nhiệm vụ phân loại hình ảnh thống trị, đặc biệt là ILSVRC (Rus-sakovsky và cộng sự, 2015). Trên thực tế, họ đã chiến thắng trong mọi ImageNet thách thức phân loại kể từ năm 2012 (Simonyan & Zisserman, 2014; Zeiler & Fergus, 2014; Szegedy, Liu và cộng sự, 2014; He, Zhang, Ren, & Sun, 2015b). Trong một cố gắng hiểu chúng và tìm ra cách cải thiện hiệu suất của chúng, Zeiler và Fergus (2014) đã giới thiệu một kỹ thuật trực quan hóa mới bằng cách sử dụng mạng lư ới phân tách nhiều lớp (Zeiler, Taylor, & Fergus, 2011) cung cấp tầm nhìn vào các lớp trích xuất tính năng trung gian của mạng. Họ sử dụng điều này trong vai trò chẩn đoán để cải thiện kiến trúc DCNN và hiệu suất của Krizhevsky et al. (2012). Do đó, khi so sánh đối với Krizhevsky et al. (2012), mô hình của họ đạt đư ợc kết quả tốt hơn trên chuẩn phân loại Ima-geNet và nhiều mô hình đã đư ợc trung bình hóa để giành chiến thắng ILSVRC 2013 (Russakovsky và cộng sự, 2015). Hơn nữa, mô hình của họ đư ợc khái quát hóa một cách tuyệt vời và họ đã chứng minh điều này bằng cách đạt đư ợc kết quả công bố tốt nhất trên các tập dữ liệu CALTECH-101 (Fei-Fei và cộng sự, 2006) và CALTECH-256 (Griffin và cộng sự, 2007). Mặc dù kỹ thuật trực quan hóa của họ hoạt động tốt trên các hình ảnh màu có chiều tư ơng đối cao hơn, nó sẽ là thú vị khi kiểm tra khả năng áp dụng của nó trên tập dữ liệu MNIST phổ biến (LeCun et al., 1998), vì có thể hình dung rằng mạng lư ới giải tích có thể không có thể tái tạo hình ảnh MNIST thang độ xám chiều thấp hơn với cùng độ chính xác. Một hư ớng thú vị khác từ việc sử dụng trích xuất các tính năng trong vai trò chẩn đoán sẽ là điều tra khả năng áp dụng của nó để giải quyết một số thách thức còn lại của DCNN, cụ thể như những thách thức đã đề cập trong phần 6.2 và 6.4.

Bằng chứng

Chư a sửa

Szegedy, Liu và cộng sự (2014) đã giới thiệu một kiến trúc DCNN mà họ đư ợc gọi là mô hình Inception. Một hiện thân cụ thể của mô hình này, GoogLeNet, đã tạo ra kết quả phân loại hình ảnh và phát hiện đối tư ợng nổi bật, giành chiến thắng trong cả thử thách phân loại và phát hiện ImageNet năm 2014 (Russakovsky và cộng sự, 2015). Thành công của họ đã đư ợc mang lại bằng cách sử dụng một mạng lư ới rất lớn, bao gồm 22 lớp. Vì chi phí của điều này là số lư ợng tham số lớn hơn, khiến mạng dễ bị tấn công hơn để quá mức phù hợp và gánh nặng tính toán lớn hơn đáng kể, họ đã sử dụng một thiết kế đư ợc thiết kế cẩn thận, dựa trên các nguyên tắc của ngư ời Do Thái, cho phép chúng chuyển từ kiến trúc tích chập kết nối đầy đủ sang kiến trúc tích chập kết nối thư a thớt, đư ợc thúc đẩy bởi những phát hiện của Arora, Bhaskara, Ge và Ma (2014). Cụ thể, kiến trúc của họ sử dụng rất nhiều phép tích chập 1 × 1, lấy cảm hứng từ Lin et al. (2013), để thực hiện hai chức năng. Quan trọng nhất, chúng đóng vai trò như khối giảm kích thư ớc trư ớc khi tính toán tốn kém hơn 3 × 3 và phép tích chập 5 × 5, và chúng bao gồm việc sử dụng các kích hoạt tuyến tính chỉnh lư u (Nair & Hinton, 2010), do đó làm cho chúng có mục đích kép. Sau đó,

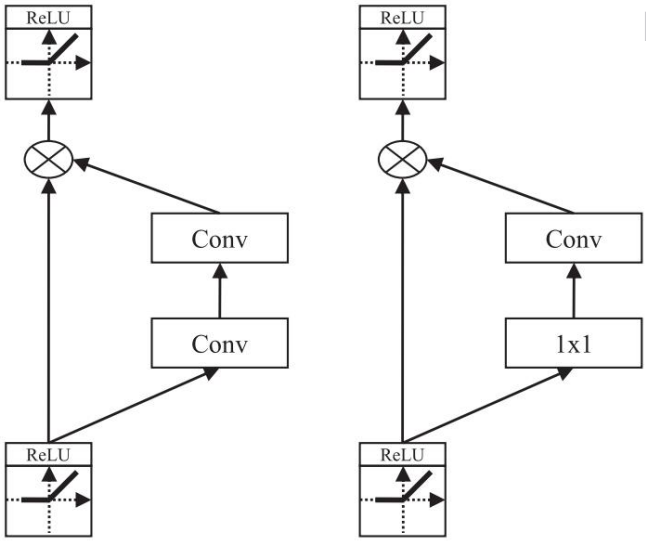
họ có thể tăng chiều sâu và chiều rộng của mạng lưới của họ, trong khi chỉ tăng nhẹ chi phí tính toán. Hình 8 (trong phần 5.1.1.1) minh họa một mô-đun Inception, kết hợp bộ lọc tích chập giảm kích thước 1 × 1, được minh họa bằng góc xiên trong sơ đồ. Những các mô-đun là các khối xây dựng của mô hình Inception, kể từ đó đã được cải thiện nhiều lần, như đã thảo luận trong phần 5.1.1.2.

Tương tự như Szegedy, Liu và cộng sự. (2014), Simonyan và Zisserman (2014), á quân trong cùng cuộc thi phân loại ILSVRC 2014 (Russakovsky et al., 2015), cũng sử dụng DCNN rất sâu, bao gồm 19 lớp so với 22 đối thủ cạnh tranh của họ. Tuy nhiên, khẳng định rằng mô hình Incip-tion quá phức tạp, họ giữ nguyên tất cả các tham số của DCNN của họ kiến trúc không đổi và tăng dần độ sâu một cách đều đặn. Điều này đã được thực hiện khả thi bằng cách sử dụng các bộ lọc tích chập có kích thước nhỏ hơn (3 × 3) trên toàn bộ mạng lưới, được lấy cảm hứng từ Ciresan et al. (2011), những người đã sử dụng hạt nhân nhỏ hơn, mặc dù dành cho mạng nông hơn được áp dụng cho các tác vụ đơn giản hơn.

Những người chiến thắng tại ILSVRC 2015 (Russakovsky và cộng sự, 2015), He và cộng sự. (2015b), đã sử dụng DCNN sâu hơn khi so sánh với Simonyan và Zisserman (2014) và Szegedy, Liu và cộng sự (2014). Trên thực tế, mô hình của họ là siêu sâu vì nó bao gồm 152 lớp. Vì các mô hình sâu hơn khó hơn để đào tạo và chịu sự suy thoái (của việc đào tạo và do đó kiểm tra độ chính xác) (He et al., 2015b; He & Sun, 2015; Srivastava, Greff, & Schmidhuber, 2015a, 2015b), họ đã giới thiệu một khuôn khổ học tập còn lại mới.[1] Họ đã cải tổ các lớp của mạng và buộc chúng phải học các hàm còn lại bằng cách tham chiếu đến các đầu vào lớp trước đó của chúng thay vì học các hàm không được tham chiếu. Điều này cho phép các lỗi được truyền trực tiếp đến các đơn vị trước đó, và do đó làm cho các mạng này dễ tối ưu hóa hơn và, mặc dù chúng cực kỳ sâu, dễ huấn luyện hơn. Họ đã thử các dư lượng khác nhau cấu hình mô-đun và kiến trúc mạng và thấy rằng được tối ưu hóa các mô-đun còn lại hoạt động tối ưu hơn so với các mô-đun ban đầu của chúng. Hình 6 so sánh sự khác biệt giữa các mô-đun còn lại ban đầu và người kế nhiệm được tối ưu hóa của nó, dẫn đến tính toán nhanh hơn. Như minh họa, ReLU (Nair & Hinton, 2010) có nhiều tính năng trong cả hai phiên bản; tuy nhiên, các kết nối dư được tối ưu hóa sử dụng phương pháp giảm kích thước Bộ lọc 1 × 1 để đệm tính toán. Một mô tả chính thức hơn về kỹ thuật học dư thừa ban đầu, cũng như các cải tiến của nó, được thảo luận trong mục 5.5.4.

Bảng 2 minh họa hiệu suất của DCNN trong ILSVRC kể từ khi thành lập. Điều quan trọng là bảng này làm nổi bật sự thống trị của DCNN so với trước đó phương pháp sử dụng trích xuất và nén tính năng, sau đó phân loại bằng bộ phân loại nông (Perronnin, Sánchez, & Mensink, 2010;

[1] Sự suy thoái là do sự lan truyền kém của các hoạt động và độ dốc vì của việc xếp chồng nhiều phép biến đổi phi tuyến tính lên nhau.



a) Original residual connections b) Optimized residual connections

Hình 6: Mô-đun còn lại so với mô-đun còn lại được cải thiện.

Lin et al., 2011; Sánchez & Perronnin, 2011), cũng như một ước tính gần đúng
mối tương quan giữa hiệu suất phân loại và độ sâu mạng (xem
Phần 5.5.4). Có thể tìm thấy thêm kết quả trong Bảng 4.

Bất chấp sự suy thoái (He et al., 2015b; He & Sun, 2015; Srivas-tava et al.,
2015a, 2015b), các mô hình sâu hơn thường chính xác hơn và
do đó tạo ra kết quả thực nghiệm tốt hơn; tuy nhiên, khi độ sâu tăng lên, thì
chi phí tính toán. Với điều này trong tâm trí, công việc đại diện được thảo luận
ở đây đã dẫn đến một số nỗ lực nhằm cải thiện độ chính xác phân loại của
DCNN bằng cách sửa đổi kiến trúc của chúng để cải thiện hiệu suất mà không cần
mất đi gánh nặng tính toán áp đặt lên các mô hình như vậy. Đặc biệt, các mô hình của
Szegedy, Liu và cộng sự (2014), Simonyan và Zisserman
(2014), và He et al. (2015b) đều tập trung vào các mạng lưới sâu hơn hoặc rộng hơn cho
độ chính xác được cải thiện, với một số thủ thuật, từ việc giảm kích thước
để học tập còn lại, để xử lý căng thẳng tính toán liên quan được đặt ra
trên các mạng lưới sâu hơn. Điều này đã dẫn đến một tình thế tiến thoái lưỡng nan về kỹ thuật cổ điển giữa
các mô hình sâu hơn, chính xác hơn nhưng tốn kém về mặt tính toán,
và các mô hình nông hơn, dễ đào tạo hơn và rẻ hơn nhưng không
tạo ra cùng độ chính xác phân loại. Do đó, mặc dù đã có
nhiều nỗ lực để giải quyết vấn đề này, duy trì độ chính xác với chi phí tính toán
giảm vẫn là một thách thức mở đối với DCNN. Để đạt được mục đích này,
phần 5.5.4 đề cập đến việc xử lý nhanh hơn các mô hình sâu, trong khi

Bảng 2: Kết quả phân loại hình ảnh ILSVRC năm 2010.

Năm	Đội ngũ	Con số Layers	Tổng quan Sự đóng góp	Chức vụ	Tài liệu tham khảo
2010	NEC		Tính năng trích xuất nhanh nông, nén dữ liệu, SVM phân loại	Đầu tiên	Lin và cộng sự, 2011
2011	XRCE		Hình ảnh nông có chiều cao chữ ký, dữ liệu nén, SVM phân loại	Đầu tiên	Perronni và cộng sự, 2010; Sanchez & Perronni, 2011
Giám sát năm 2012		8	Hiệu quả dựa trên GPU DCNN, với Dropout và một số khác những đổi mới	Đầu tiên	Krizhevsky và cộng sự, 2012
Chất làm trong năm 2013		8	Kiến trúc DCNN dựa trên về giải xoắn kỹ thuật trực quan	Đầu tiên	Zeiler và Fergus, 2014; Thủy thủ và cộng sự, 2011
2014	GoogleNet	22	DCNN kiến trúc thiết kế dựa trên Nguyên tắc Hebbian và ý tưởng đa thang đo	Đầu tiên	Szeged, Vanhoucke và cộng sự, 2015
2014	VGG	19	Cải tiến cho DCNN lớp tích chập, tăng độ sâu mạng	Simonyan thứ hai & Zisserman, 2014	
MSRA	2015	152	Giới thiệu sâu học tập còn lại cho DCNN siêu cấp	Đầu tiên	He et al., 2015b

những phát triển, xu hướng và khuyến nghị mới nhất về vấn đề này được giới thiệu trong phần 6.3.

5 Một loạt các cải tiến sâu hơn và những tiến bộ gần đây

Ngoài công trình mang tính cách mạng của Krizhevsky và cộng sự (2012) và những cải tiến mang tính biểu tượng hơn nữa được mô tả trong phần trước (Simonyan & Zisserman, 2014; Zeiler và Fergus, 2014; Szegedy, Liu và cộng sự, 2014; Chào et al., 2015b), một số nỗ lực cải tiến khác liên quan đến kiến trúc mạng, hàm kích hoạt phi tuyến tính, thành phần giám sát, cơ chế điều chỉnh, kỹ thuật tối ưu hóa và xử lý nhanh hơn
DCNN đã bổ sung thêm sự phổ biến của DCNN. Trong các phần
sau đây, chúng tôi sẽ khảo sát những cải tiến này một cách chi tiết, tập trung vào chúng việc làm cho các ứng dụng phân loại hình ảnh. Trong quá trình này, chúng tôi so sánh và đối chiếu các phương pháp và kỹ thuật khác nhau được sử dụng để thiết kế những cải tiến này. Về cuối phần này, chúng tôi thực nghiệm

tóm tắt kết quả phân loại của họ trên một số chuẩn phân loại hình ảnh phổ biến. Bằng chứng

5.1 Kiến trúc mạng. Phần này đầu tiên giới thiệu những cải tiến đư ợc thực hiện đối với các lớp tích chập của DCNN, sau đó là các cuộc thảo luận cân nhắc một số chư ơng trình hợp tác, bao gồm những tiến bộ mới nhất về vấn đề này.

5.1.1 Các lớp tích chập. Các lớp tích chập học các tính năng biểu diễn hình ảnh đầu vào của chúng, và điều này làm cho chúng trở thành khối xây dựng chính của DCNN. Do đó, việc cố gắng cải thiện khía cạnh này là điều tự nhiên Kiến trúc DCNN. Ở đây chúng tôi giới thiệu động cơ đằng sau một số

những đổi mới quan trọng trong lĩnh vực này.

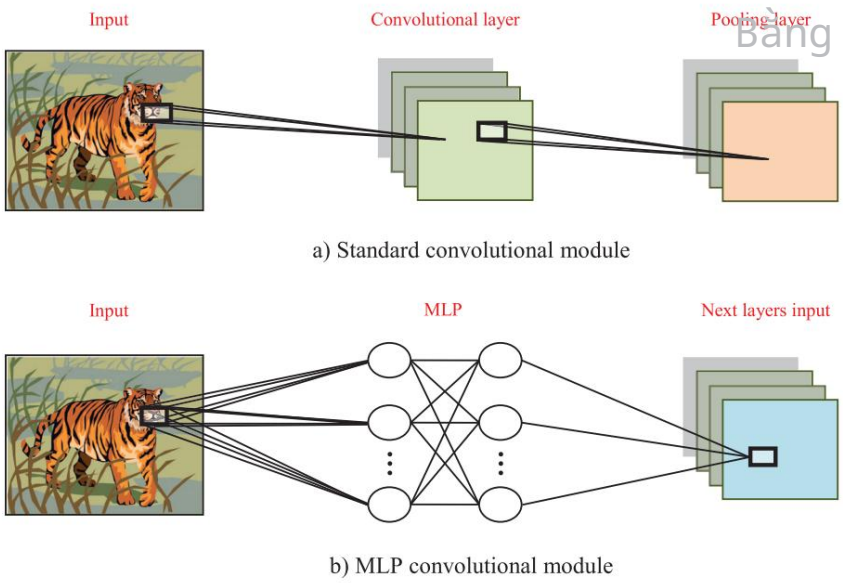
5.1.1.1 Mạng trong mạng. Vì các lớp tích chập sử dụng bộ lọc tuyến tính, phù hợp hơn với việc học các tính năng tiềm ẩn (thuộc tính ẩn) của một hình ảnh) có thể tách biệt tuyến tính, chúng không phải là cáp trích xuất các biểu diễn trừu tư ợng từ hình ảnh.2 Do đó, Lin et al. (2013) đã đề xuất thay thế chúng bằng các hàm xấp xỉ phổ quát. Cụ thể, chúng thay thế các bộ lọc tích chập cục bộ thông thư ờng bằng nhiều lớp perceptron (MLP), tư ơng thích với kiến trúc và quy trình đào tạo của DCNN, để tích chập dữ liệu đầu vào tạo ra MLP lớp tích chập. Tính toán đư ợc thực hiện bởi lớp này, khi Chư a sửa ReLU (Nair & Hinton, 2010) đư ợc sử dụng làm hàm kích hoạt, có thể đư ợc diễn đạt bằng toán học như sau

$$\text{nếu}^{N}_{ij,\,k_n} = \text{tối đa}\left(w_{k_n}^{T} \text{ biết } f_{i,j}^{n-1} + \text{không}\,,\, 0\right) \tag{5.1}$$

trong đó chỉ số pixel của bản đồ đặc điểm đư ợc biểu thị bằng (i, j); bản vá đầu vào, tập trung tại vị trí (i, j), đư ợc biểu thị bằng xi,j; các kênh bản đồ đặc điểm là đư ợc lập chỉ mục bởi k; và n biểu thị số lớp trong MLP.

Phư ơng pháp đề xuất chứng minh rằng các lớp tích chập MLP này mô hình hóa các mảng hình ảnh cục bộ tốt hơn các lớp tích chập tiêu chuẩn. Khi kết hợp với một kỹ thuật gộp trung bình toàn cầu mới, kỹ thuật này đã trung bình hóa không gian các bản đồ đặc điểm của lớp cuối cùng, đư ợc sử dụng để thay thế lớp kết nối đầy đủ tiêu chuẩn, họ đã tạo ra những kết quả tiên tiến nhất trên hai phiên bản của CIFAR-10 (Krizhevsky, 2009; Wan, Zeiler, & Zhang, 2013; Goodfellow, Warde-Farley, Mirza, Courville, & Bengio, 2013) và chuẩn mực CIFAR-100 (Krizhevsky, 2009; Srivastava & Salakhutdinov, 2013), và rất gần với trạng thái hiện đại của bộ dữ liệu MNIST (LeCun et al., 1998; Goodfellow và cộng sự, 2013). Mặc dù đề xuất nhóm trung bình toàn cầu

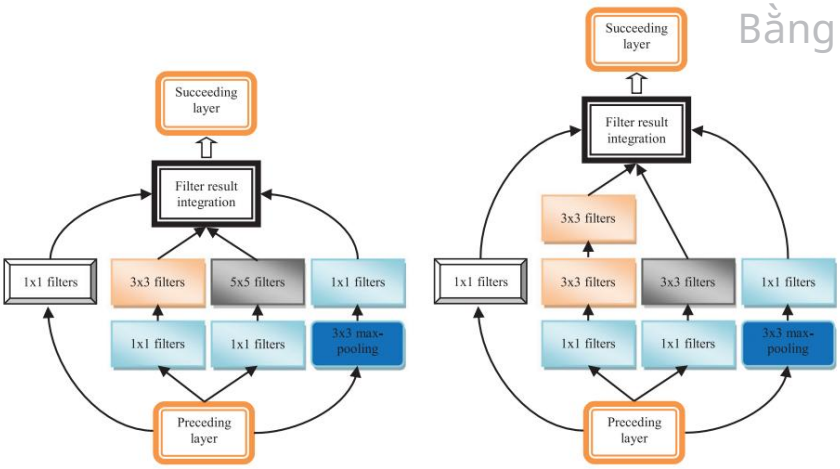
2 Tóm tắt trong nội dung này liên quan đến các tính năng bất biến đối với các tính năng của một khái niệm tư ơng đư ơng (Bengio, Courville & Vincent, 2013).



Hình 7: Lớp tích chập so với lớp tích chập MLP.

kỹ thuật, có ít tham số hơn và do đó tính toán rẻ hơn
chi phí so với các lớp đư ợc kết nối đầy đủ, góp phần làm giảm tình trạng quá khớp
đối với MNIST tư ơng đối nhỏ (LeCun et al., 1998) và CIFAR-10
và bộ dữ liệu CIFAR-100 (Krizhevsky, 2009), một nghiên cứu về việc sử dụng quá mức
lớp này để thay thế các lớp đư ợc kết nối đầy đủ thông thư ờng
của các mô hình DCNN khác vẫn còn rộng rãi đối với các tập dữ liệu lớn hơn như ImageNet
(Russakovsky và cộng sự, 2015). Hình 7 minh họa sự khác biệt giữa mô-đun tích chập
thông thư ờng và mô-đun tích chập MLP,
là khối xây dựng chính của mạng trong mô hình mạng (NiN). Trong khi
cả hai biến thể đều ánh xạ trư ờng tiếp nhận cục bộ biểu diễn đầu vào ẩn
các tính năng cho một lớp tiếp theo, bảng b sử dụng một mạng lư ới vi mô để tăng cư ờng
đại diện.

5.1.1.2 Mô hình khởi đầu và cải tiến mô hình khởi đầu. Mô hình khởi đầu
(Szegedy, Liu và cộng sự, 2014), lấy cảm hứng từ Lin và cộng sự (2013) và đư ợc thảo
luận trong phần 4.4, đã sử dụng kỹ thuật giảm kích thư ớc (bộ lọc tích chập 1 × 1)
để giảm bớt gánh nặng tính toán của hoạt động tích chập tốn kém. Để mở rộng quy mô
và cải thiện hơn nữa độ chính xác phân loại DCNN theo cách hiệu quả về mặt tính toán,
mô hình Inception sau đó đã
đư ợc tăng cư ờng bằng cách sử dụng các phép tích chập đư ợc phân tích (xem phần 6.6)
và các phép giảm kích thư ớc tích cực trong mạng. Trong khi sự khởi đầu ban đầu
mô-đun vẫn sử dụng phép tích chập 5 × 5, phiên bản cải tiến đã thay thế phép tích chập này



Hình 8: Mô-đun Inception so với mô-đun Inception cải tiến.

với hai phép tích chập 3 × 3 rẻ hơn về mặt tính toán (Szegedy, Vanhoucke et al., 2015). Hình 8 minh họa sự khác biệt giữa hai mô-đun.

Lấy cảm hứng từ độ chính xác phân loại hình ảnh đạt được bởi phần dư mạng lưới (He et al., 2015b), được thảo luận trong phần 4.4 và 5.5.4, Inception kiến trúc (Szegedy, Liu và cộng sự, 2014; Szegedy, Vanhoucke và cộng sự, 2015) là được tinh chỉnh hơn nữa và kết hợp với các kết nối còn lại để tạo thành các mạng In-ception còn lại (Szegedy, Ioffe, & Vanhoucke, 2016). Bài báo đã cung cấp bằng chứng rõ ràng ủng hộ rằng việc đào tạo với các kết nối còn lại đã đẩy nhanh đáng kể quá trình đào tạo các mạng Inception. Mặc dù họ đã thử nghiệm một số kiến trúc Inception chỉ dành riêng và Inception còn sót lại, họ thấy rằng kiến trúc Inception dư thừa lai tạo ra độ chính xác phân loại mô hình đơn tốt nhất, mặc dù chi phí tính toán cao hơn khi so sánh với kiến trúc Inception được cải tiến được mô tả bởi Szegedy, Vanhoucke et al. (2015). Hơn nữa, khi họ kết hợp một mô hình Inception mới, được cải tiến, có kiến trúc đơn giản hơn và nhiều mô-đun Inception hơn so với mô hình trước đó của họ (Szegedy, Vanhoucke và cộng sự, 2015), thành một tập hợp với ba mạng Inception còn lại, họ đã đạt được kết quả tốt nhất kết quả đã công bố về chuẩn mực phân loại hình ảnh ImageNet đầy thách thức (Russakovsky và cộng sự, 2015; He và cộng sự, 2015b). Mặc dù thành công này, vẫn cần phải có thêm nhiều công trình nữa để giảm bớt gánh nặng tính toán áp đặt về kiến trúc lai.

5.1.1.3 Tích chập kép. Được thúc đẩy bởi trực giác, tiếp theo là phân tích lý thuyết, ủng hộ rằng một số bộ lọc đã học của DCNNS được đào tạo tốt là các phiên bản được dịch một chút của nhau, Zhai, Cheng,

Lu và Zhang (2016) mới đề xuất mạng nơ-ron tích chập kép, sử dụng phép toán tích chập kép trong các lớp tích chập. Điều này cho phép họ học các cụm bộ lọc, trong đó các bộ lọc trong
mỗi cụm được dịch thành các dạng của nhau. Để thực hiện được điều này,
một tập hợp các bộ lọc siêu dữ liệu được phân bổ cho một lớp tích chập kép. Kích thước của các bộ lọc siêu dữ liệu này lớn hơn kích thước bộ lọc hiệu quả, được trích xuất từ mỗi cái trong số chúng. Điều này tương ứng với việc tích chập các bộ lọc siêu dữ liệu với một nhân danh tính. Bằng cách nối các bộ lọc đã trích xuất và sau đó tích chập với đầu vào, kỹ thuật này đạt được phép tích chập kép. Kỹ thuật này cũng bổ sung cho Maxout (Goodfellow và cộng sự, 2013), mà chúng tôi giới thiệu trong phần 5.2.8, vì có cơ hội để tập hợp dọc theo các kích hoạt được tạo ra bởi cùng một bộ lọc siêu dữ liệu. Chúng vượt trội hơn mô hình NIN
(Lin et al., 2013), cũng đã thay đổi lớp tích chập tiêu chuẩn cho
cải thiện độ chính xác phân loại trên các tập dữ liệu CIFAR-10 và CIFAR-100
(Krizhevsky, 2009); hơn nữa, vì kiến trúc của mạng tích chập kép có thể thay đổi dễ dàng nên chúng có hiệu quả về mặt tham số,
do đó giảm được nhu cầu về không gian lưu trữ mà không làm mất đi độ chính xác.
Nhược điểm của cách tiếp cận như vậy là phép toán tích chập kép sẽ phát sinh thêm chi phí tính toán khi so sánh với phép toán tiêu chuẩn.
lớp tích chập.

5.1.1.4 Phân tích và triển vọng. Các bộ lọc tích chập, những con ngựa thồ của DCNN, là các mô hình tuyến tính tổng quát của các mảng hình ảnh cơ bản rằng chúng xoắn lại, và mặc dù chúng hoạt động tốt để trích xuất các tính năng có mức độ trừu tượng thấp, chúng bị thách thức khi cần
để trích xuất các hàm phi tuyến tính cao của hình ảnh đầu vào của chúng tôi. Điều này ủng hộ nhu cầu về các trình trích xuất tính năng phi tuyến tính hiệu quả hơn, bắt đầu bằng Mô hình NIN. Kiến trúc được giới thiệu bởi Lin et al. (2013) đã dẫn đến một loạt của những cải tiến khác cũng tập trung vào các lớp tích chập. Tại
trái tim của những cách tiếp cận này là Inception (Szegedy, Liu và cộng sự, 2014)
và cải thiện các mô hình Inception (Szegedy, Vanhoucke và cộng sự, 2015; Szegedy et al., 2016), được thiết kế tỉ mỉ để giảm thiểu mọi hạn chế về tính toán và điều này tạo điều kiện tăng kích thước mạng (chiều rộng và
độ sâu) để tăng cường độ chính xác phân loại. Tuy nhiên, bất chấp
kết quả thực nghiệm đầy hứa hẹn của các mô hình này, một sự biện minh về mặt lý thuyết vì những thành công của họ vẫn còn thiếu sót. Hơn nữa, tính phức tạp và cao của họ kiến trúc được tối ưu hóa không đảm bảo sửa đổi mà không có khả năng
hạn chế về hiệu suất, do đó cần phải thận trọng khi áp dụng
họ. Công việc trong tương lai nên cố gắng chứng minh lý do cho kinh nghiệm
thành công của các lớp tích chập sáng tạo được thảo luận ở đây và điều này nên được bổ sung bằng các sửa đổi liên quan đến tích chập mới
giải quyết những mối quan tâm liên quan đến các mô hình hiện tại của chúng tôi, chẳng hạn như gánh nặng tính toán do phép toán tích chập gây ra,
không có khả năng trích xuất các tính năng mạnh mẽ và sự phức tạp của các mô hình
hiện tại làm giảm bớt những lo ngại này. Điều này sẽ không chỉ thúc đẩy phân loại

độ chính xác và độ bền và thúc đẩy tính toán nhanh hơn như ng sẽ dẫn đến các mô hình dễ dàng thích ứng với các nhiệm vụ khác nhau. Một số hứa hẹn những phát triển gần đây theo hư ớng này bao gồm phép tích chập kép (xem phần 5.1.1.3), phép tích chập lát gạch (Ngiam và cộng sự, 2010; Wang & Oates, 2015), và đặc biệt là sự tích chập giãn nở (Yu & Koltun, 2015), có đã cho thấy kết quả khả quan cho nhiều nhiệm vụ khác nhau như nhận dạng giọng nói và tổng hợp (Sercu & Goel, 2016; Oord et al., 2016), dịch máy (Kalchbrenner và cộng sự, 2016) và phân đoạn cảnh (Yu & Koltun, 2015).

5.1.2 Các lớp gộp. Sau các lớp tích chập, các lớp gộp có lẽ là quan trọng nhất. Chúng tóm tắt lại các phản ứng của các tế bào thần kinh lân cận từ cùng một bản đồ hạt nhân và do đó làm giảm các kích thư ớc của các biểu diễn đầu vào của họ. Quan trọng là họ cung cấp cho DCNN bất biến không gian (Krizhevsky và cộng sự, 2012; LeCun và cộng sự, 2015). Ngoài việc gộp trung bình và tối đa, đư ợc giới thiệu trong phần 2.2 và 4.2.3, chúng tôi khảo sát một số kỹ thuật gộp nhóm thành công khác đư ợc đề cập trong văn học.

5.1.2.1 Gộp Lp . Mặc dù việc sử dụng max pooling đã mang lại kết quả thực nghiệm tuyệt vời (Ciresan et al., 2011; Krizhevsky et al., 2012; Si-monyan & Zisserman, 2014; Szegedy, Liu et al., 2014; Szegedy, Vanhoucke et al., 2015), nó có thể phù hợp quá mức với dữ liệu đào tạo và không đảm bảo tổng quát hóa trên dữ liệu thử nghiệm. Mặt khác, tổng hợp trung bình xem xét tất cả các yếu tố trong vùng tập hợp và do đó các khu vực hoạt động thấp có thể giảm tác động của các khu vực có hoạt động cao (Zeiler & Fergus, 2013; Sainath, Kings-bury, Mohamed et al., 2013). Để giải quyết những vấn đề này, một giải pháp hay thế khả thi là sự kết hợp Lp lấy cảm hứng từ sinh học , đư ợc mô phỏng theo các tế bào phức tạp (Simoncelli & Heeger, 1998; Hyvärinen & Köster, 2007). Trong một hồ bơi nhất định regio Rj, nó lấy giá trị trung bình có trọng số của kích hoạt ai, như minh họa bởi phư ơng trình 5.1:

$$s_j = \left(\sum_{i \in R_j} a_i^p \right)^{1/p} \quad (5.2)$$

Đáng chú ý, khi p = 1, phư ơng trình tư ơng ứng với nhóm trung bình, trong khi p = ∞ dịch thành nhóm tối đa. Đối với các giá trị 1 < p < ∞, nhóm Lp có thể đư ợc coi là sự đánh đổi giữa nhóm trung bình và nhóm tối đa (Sainath, Kings-bury, Mohamed et al., 2013). Mặc dù nhóm Lp đã đư ợc áp dụng trư ớc đây (Yang, Yu, Gong, & Huang, 2009; Kavukcuoglu, Ranzato, Fergus, & LeCun, 2009), khi nó đư ợc kết hợp với DCNN (Sermanet, Chintala, & LeCun, 2012), nó dẫn đến kết quả phân loại hình ảnh đặc biệt và một trạng thái mới của nghệ thuật phân loại số nhà trên Street View (SVHN), đánh bại chuẩn mực tốt nhất trư ớc đó do Netzer và cộng sự (2011) thiết lập.

Hơn nữa, phân tích lý thuyết đư ợc thực hiện bởi Boureau et al. (2010), Bruna Bằng chứng
Szlam, và LeCun (2013), và Gulcehre, Cho, Pascanu, và Bengio (2014)
cho thấy rằng nó cung cấp khả năng khái quát tốt hơn khi so sánh với max
tập hợp.

5.1.2.2 Phân nhóm max ngẫu nhiên và phân số. Đư ợc thúc đẩy bởi các vấn đề
với nhóm trung bình và nhóm tối đa và hiệu ứng chính quy hóa của Dropout
(Turaga et al., 2010; Hinton et al., 2012), Zeiler và Fergus (2013) đã giới thiệu
gộp ngẫu nhiên để thay thế gộp trung bình xác định và gộp tối đa
kỹ thuật. Cụ thể, trong việc gộp ngẫu nhiên, bằng cách chuẩn hóa các hoạt động trong
mỗi vùng j, xác suất p cho vùng đầu tiên đư ợc tính toán bằng

$$\text{số pi} = \frac{\text{ăn}}{\text{xin vui lòng}} \tag{5.3}$$

Sau đó, dựa trên p, một mẫu đư ợc lấy từ phân phối đa thức,
đư ợc hình thành từ các hoạt động của mỗi vùng gộp, để chọn một vị trí l trong vùng.
Do đó, hoạt động gộp chỉ đơn giản là:

$$\text{sj = al trong đó l} \quad P(p1,\ldots, p|Rj|). \tag{5.4}$$

Chư a sửa

Mặc dù việc gộp ngẫu nhiên có cùng lợi ích như việc gộp tối đa, như ng
bản chất ngẫu nhiên giúp nó ngăn ngừa tình trạng quá khớp, do đó làm cho nó trở thành một phư ơng pháp hiệu quả
kỹ thuật điều chỉnh mạng có thể kết hợp với các phư ơng pháp tiếp cận khác như Dropout
(Hinton et al., 2012; Srivastava et al., 2014) và
tăng cư ờng dữ liệu (LeCun et al., 1998; Simard et al., 2003; Ciresan et al.,
2011, 2012; Montavon, Orr, & Müller, 2012). Khi áp dụng cho các nhiệm vụ phân loại
hình ảnh, nhóm ngẫu nhiên vư ợt trội hơn nhóm trung bình và nhóm tối đa
trên MNIST (LeCun et al., 1998), CIFAR-10 và CIFAR-100 (Krizhevsky,
2009) và chuẩn mực SVHN (Netzer và cộng sự, 2011).

Tư ơng tự như nhóm ngẫu nhiên, nhóm cực đại phân số (Graham, 2014),
cũng giới thiệu các thuộc tính ngẫu nhiên cho quá trình gộp nhóm. Tuy nhiên, khác với
gộp nhóm ngẫu nhiên, việc lựa chọn các vùng gộp nhóm, thay vì
hơn các hoạt động gộp chung trong chúng, có bản chất ngẫu nhiên. Thêm
cụ thể, trong khi nhóm ngẫu nhiên và nhóm truyền thống sử dụng β × β max
gộp, trong đó β = 2 (xem phần 2.2), gộp tối đa phân số giới thiệu một
hệ số phân số β (ví dụ, √ 2), đư ợc chọn ngẫu nhiên hoặc bán ngẫu nhiên từ phạm vi 1
<β< 2, để giảm kích thư ớc không gian của
đầu vào gộp. Họ đã thu đư ợc kết quả tiên tiến nhất trên các tập dữ liệu CIFAR
(Krizhevsky, 2009; xem Bảng 6); tuy nhiên, các quan sát của họ thiếu sự phù hợp
động lực và kỹ thuật vẫn cần phải đư ợc thử nghiệm trên các kiến trúc khác
chẳng hạn như Inception (Szegedy, Liu và cộng sự, 2014) và mạng lư ới Residual (He và cộng sự.
2015b).

5.1.2.3 Hỗn hợp nhóm. Lấy cảm hứng từ bản chất ngẫu nhiên của kỹ thuật nhóm được mô tả bởi Zeiler và Fergus (2013) và các thành công khác
các kỹ thuật điều chỉnh ngẫu nhiên như Dropout (Hinton và cộng sự, 2012;
Srivastava et al., 2014) và DropConnect (Wan et al., 2013) (xem phần 5.4.2),
D. Yu, Wang, Chen và Wei (2014) đã giới thiệu một kỹ thuật kết hợp hỗn hợp mới để thúc đẩy hơn nữa khả năng điều chỉnh của DCNN và giải quyết
các vấn đề đã biết liên quan đến việc gộp trung bình và tối đa (Zeiler & Fer-gus, 2013;
D. Yu và cộng sự, 2014; Sainath, Kingsbury, Mohamed và cộng sự, 2013). Họ
cũng sử dụng một thủ tục ngẫu nhiên để sử dụng, ngẫu nhiên, tối đa hoặc trung bình
nhóm trong quá trình đào tạo DCNN. Được thể hiện bằng toán học, nhóm hỗn hợp
đầu ra y_{kij} liên quan đến bản đồ đặc điểm thứ k được tính toán bằng:

$$y_{kij} = \lambda \max_{(p,q) \in ij} x_{kpq} + (1 - \lambda) \frac{1}{|ij|} \sum_{(p,q) \in ij} x_{kpq} \tag{5.5}$$

trong đó phần tử ở vị trí (p, q), trong vùng gộp ij với kích thước
$|ij|$, được biểu diễn bởi x_{kpq} và nhóm trung bình hoặc nhóm tối đa được chọn
bởi λ, có giá trị ngẫu nhiên là một hoặc không. Tương tự như nhóm ngẫu nhiên, sự kết hợp
của nhóm hỗn hợp với các quy tắc khác
kỹ thuật là có thể. So sánh, khi thử nghiệm trên SVHN (Netzer
et al., 2011) thách thức phân loại, phương pháp này đã chứng minh là vượt trội hơn mức trung bình,
max, ngẫu nhiên và Lp pooling (Sermanet et al., 2012). Ngoài ra, khi
được thử nghiệm trên các chuẩn mực CIFAR-10 và CIFAR-100 (Krizhevsky, 2009), nó
cũng cung cấp hiệu suất phân loại nâng cao hơn mức trung bình, tối đa và
nhóm ngẫu nhiên (Zeiler & Fergus, 2013) và sơ đồ tham số hóa
được giới thiệu bởi Malinowski và Fritz (2013).

5.1.2.4 Nhóm hỗn hợp, có cổng và nhóm cây. Các thí nghiệm do Lee, Gal-lagher và Tu (2016) thực hiện hỗ trợ các phát hiện của Boureau và cộng sự (2010), những người cũng phát hiện ra rằng có những trường hợp mà cả hai phương pháp gộp nhóm tối đa hoặc trung bình đều hoạt động tốt hơn phương pháp kia. Do đó, họ đã khám phá việc học hàm gộp nhóm bằng cách kết hợp gộp nhóm tối đa và trung bình bằng cách sử dụng một phương pháp phản hồi (đạt được thông qua một cổng) và chiến lược không phản hồi. Đầu ra của cổng
phương pháp trung bình tối đa có thể được tính toán bằng

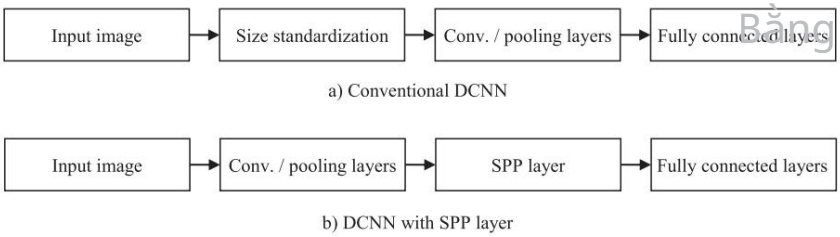
$$f_{gate}(x) = \sigma(w^T x) f_{max}(x) + (1 - \sigma(w^T x)) f_{avg}(x), \tag{5.6}$$

trong đó các giá trị trong vùng gộp được biểu thị bằng x và các giá trị của mặt nạ gating được biểu thị bằng w. Hơn nữa, lấy cảm hứng từ Bulo
và Kontschieder (2014), người đã kết hợp MLP với cây quyết định,
Lee et al. (2016) đã sử dụng cây quyết định nhị phân để tìm hiểu sự kết hợp của
bộ lọc nhóm cá nhân đã học trước đó. Một hiện thân cụ thể của
cách tiếp cận của họ, kết hợp các phương pháp cây và phương pháp trung bình tối đa của họ,
đạt được kết quả tiên tiến nhất trên một số chuẩn mực. Đặc biệt, họ

vượt trội hơn một số mạng tích chập có hiệu suất cao như
NIN (Lin et al., 2013), DCNN tập hợp ngẫu nhiên (Zeiler & Fergus, 2013),
các DCNN do Jarrett và cộng sự (2009) trình bày, các mạng Maxout (Goodfel-low và cộng
sự, 2013) và các mạng DropConnect (Wan và cộng sự, 2013) trên nhiều chuẩn mực phân
loại hình ảnh khác nhau, bao gồm MNIST (LeCun và cộng sự,
1998), CIFAR-10 và CIFAR-100 (Krizhevsky, 2009), và SVHN (Netzer
et al., 2011) các tập dữ liệu. Đáng chú ý là, mặc dù thành công, RCNN (Liang & Hu,
2015) đã vượt trội hơn chúng trên bộ dữ liệu CIFAR-100. Hơn nữa, để các DCNN trong
tương lai dễ dàng kết hợp các công cụ phân tích quyết định như quyết định
cây vào kiến trúc của chúng, tiếp tục làm việc để giảm bớt tính toán
chi phí và số lượng lớn các tham số mô hình cần thiết cho các mô hình như vậy
vẫn còn cần thiết.

5.1.2.5 Tổng hợp phổ. Thay vì tập trung vào tốc độ tính toán
lợi ích của việc di chuyển phép toán tích chập ra khỏi miền không gian,
tương tự như công trình được mô tả bởi Mathieu, Henaff và LeCun (2013), Rip-pel,
Snoek và Adams (2015) đã đề xuất việc học các bộ lọc tích chập
của DCNN trực tiếp trong miền tần số. Quan trọng hơn, các tác giả
đề xuất nhóm phổ, trong đó chiếu các biểu diễn phổ vào
miền tần số và sau đó cắt bớt các biểu diễn này như một kỹ thuật giảm chiều không
thông thường, khi so sánh với
phổ biến nhất là nhóm tối đa. Chính xác hơn, nhóm phổ đầu tiên tính toán
biến đổi Fourier rời rạc (DFT) của bản đồ đặc trưng đầu vào x RM×N và
sau đó cắt xén biểu diễn tần số bằng cách chỉ duy trì trung tâm
Ma trận phụ H × W của tần số được điều chỉnh bởi các kích thước của
bản đồ đặc trưng đầu ra mong muốn H × W. Cuối cùng, DFT nghịch đảo ánh xạ
biểu diễn bị cắt cụt trở lại miền không gian. Phương pháp của họ cung cấp một giải
pháp khả thi cho việc mất thông tin không gian liên quan đến
nhóm tối đa (Ranzato et al., 2007; Scherer et al., 2010; Szegedy, Liu et al.,
2014; Rippel et al., 2015) và một dạng mới của quy tắc hóa ngẫu nhiên tương tự như
biến thể Dropout (Hinton et al., 2012; Srivastava et al., 2014), được biết đến
như Dropout lồng nhau (Rippel, Gelbart, & Adams, 2014). Trên CIFAR-10 và
chuẩn mực CIFAR-100 (Krizhevsky, 2009), chúng đã vượt trội hơn một số
các tác phẩm đã được giới thiệu trong bài đánh giá này (Lin et al., 2013; Zeiler &
Fer-gus, 2013; Goodfellow et al., 2013; Liang & Hu, 2015), cũng như sâu sắc
DCNN được giám sát do Lee, Xie, Gallagher, Zhang và Tu (2015) đề xuất,
nhưng không phải là kỹ thuật kết hợp cây-hỗn hợp của Lee et al. (2016).
những thành công được mô tả bởi công trình này ủng hộ nhu cầu nghiên cứu sâu hơn về
DCNN lai sử dụng các nguyên tắc cơ bản của xử lý tín hiệu số để
cải thiện độ chính xác của hệ thống phân loại hiện tại của chúng tôi.

5.1.2.6 Nhóm kim tự tháp không gian. DCNN bị hạn chế ở chỗ chúng có thể
chỉ xử lý kích thước hình ảnh đầu vào cố định (ví dụ: 96 × 96). Để làm cho chúng
linh hoạt hơn và do đó xử lý hình ảnh có kích thước, tỷ lệ và khía cạnh khác nhau
tỷ lệ, lấy cảm hứng từ sự kết hợp kim tự tháp không gian được mô tả trong các bài báo của



Hình 9: DCNN thông thường so với DCNN SPP.

Grauman và Darrell (2005), Lazebnik, Schmid và Ponce (2006), và Yang et al. (2009), He, Zhang, Ren và Sun (2014) đề xuất pool-ing kim tự tháp không gian (SPP). Họ sử dụng các thùng không gian đa cấp, có kích thước tỷ lệ thuận với kích thước hình ảnh và điều này cho phép họ tạo ra một rep-resentation có độ dài cố định, bất kể kích thước hoặc tỷ lệ hình ảnh. Lớp SPP được tích hợp vào kiến trúc DCNN giữa convolutional/pooling cuối cùng lớp và lớp đầu tiên được kết nối đầy đủ (xem Hình 9) và do đó được thực hiện tổng hợp thông tin sâu trong mạng để ngăn chặn việc cố định kích thước (thông qua cắt xén hoặc làm cong) hình ảnh ở đầu vào. Không giống như ngẫu nhiên (Zeiler & Fergus, 2013) và Lp pooling (Sermanet et al., 2012), SPP được thiết kế để hoạt động với các lớp gộp tối đa thay vì thay thế chúng. Trong số những thành công khác, họ đã lập một kỷ lục mới trên tập dữ liệu CALTECH-101 (Fei-Fei và cộng sự, 2006), đánh bại kỷ lục trước đó của Chatfield, Simonyan, Vedaldi và Zisser-man (năm 2014) và họ đứng thứ ba trong thành phần phân loại của ILSVRC 2014 (Russakovsky và cộng sự, 2015), đứng sau Simonyan và Zisserman (2014) và Szegedy, Liu et al. (2014). Cần có thêm công việc theo hướng này để tạo điều kiện triển khai DCNN thương mại trên nhiều thiết bị di động thiết bị, vì điều này sẽ làm giảm bớt những hạn chế đặt ra đối với việc chụp ảnh hệ thống. Hơn nữa, công trình này đã chỉ ra rằng máy tính đã được thử nghiệm và kiểm tra các kỹ thuật dựa trên tầm nhìn không cần phải bị từ bỏ khi đối mặt với việc học sâu và căn phòng đó cho loại tích hợp tầm nhìn máy tính truyền thống này vẫn còn có sẵn.

5.1.2.7 Nhóm không có thứ tự đa thang. Lấy cảm hứng từ Lazebnik và cộng sự (2006), Gong et al. (2014) đã cố gắng làm cho DCNN mạnh mẽ hơn đối với sự bất biến mà không ảnh hưởng đến khả năng phân biệt của chúng. Khẳng định rằng tối đa việc gộp nhóm có thể không cung cấp sự bất biến cho các biến dạng toàn cầu trên quy mô lớn, chúng đề xuất nhóm không có thứ tự đa thang độ (MOP), trích xuất các bản vá tại nhiều thang đo, bắt đầu với hình ảnh hoàn chỉnh và sau đó nhóm từng thang đo tỷ lệ bỏ qua thông tin không gian. Cụ thể, họ trích xuất các tính năng kích hoạt sâu từ toàn bộ hình ảnh, để bảo toàn bố cục không gian toàn cầu và từ các bản vá cục bộ, để nắm bắt các chi tiết có hạt mịn. Tiếp theo, các hạt mịn chi tiết được tổng hợp thông qua mã hóa VLAD (Jegou et al., 2012), có một

bản chất không có trật tự và do đó góp phần tạo nên sự biểu diễn bất biến hơn. Bằng chứng
Cuối cùng, các kích hoạt sâu toàn cầu ban đầu và các tính năng được mã hóa VLAD
được nối lại để tạo thành một hình ảnh biểu diễn mới. Phương pháp của họ đã được chứng minh
thành công trong nhiều ứng dụng khác nhau, bao gồm phân loại cảnh,
truy xuất dữ liệu và quan trọng nhất là phân loại hình ảnh tạo ra kết quả cạnh tranh trên
ILSVRC 2012/2013 (Russakovsky et al., 2015). Với
kích thước ngày càng tăng của các tập dữ liệu hình ảnh (xem Bảng 1), điều tra thêm
vào việc kết hợp các kỹ thuật nén tính năng, chẳng hạn như mã hóa VLAD
và công nghệ DCNN được bảo hành.

5.1.2.8 Nhóm bất biến chuyển đổi. Vì các tính năng được trích xuất bởi
DCNN thiếu tính bất biến đối với các biến thể gây phiền nhiễu đã biết trong dữ liệu, được lấy cảm hứng từ
nhóm tối đa (Boureau và cộng sự, 2010) và học nhiều trường hợp (Wu, Yu,
Huang, & Yu, 2015), Laptev, Savinov, Buhmann và Pollefeys (2016) đã giới thiệu một kỹ
thuật gộp nhóm mới để tạo ra các tính năng bất biến chuyển đổi. Với một hình ảnh đầu vào
x, họ xây dựng các tính năng mới gk (x) từ một
tập hợp được xác định trước các chuyển đổi có thể φ, sao cho các tính năng mới là
độc lập với bất kỳ biến thể phiền toái nào đã biết của đầu vào. Về mặt hình thức, những
các tính năng được xây dựng theo cách sau:

$$g_k(x) = \text{tối đa } f_k(\ (x)). \tag{5.7}$$

Chưa sửa

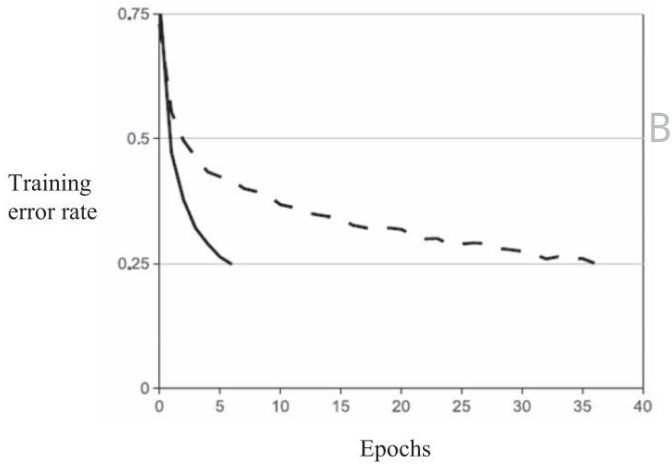
Sau đó, họ gọi sự tập hợp tối đa này qua các phép biến đổi là
nhóm bất biến biến đổi (nhóm TI). Bằng cách áp dụng tối đa
toán tử, các tính năng đã học ít phụ thuộc hơn vào các biến thể phiền toái đã biết. Hơn
nữa, đối với các tập hợp biến đổi cụ thể, về mặt lý thuyết, chúng chứng minh
bất biến biến đổi hoàn toàn. Tương tự như tích hợp SPP vào
Kiến trúc DCNN (He et al., 2014), các tác giả đề xuất tích hợp TI
tập hợp tại cùng một điểm trong mạng; tuy nhiên, họ đã sử dụng song song
Kiến trúc Xiêm—hai hoặc nhiều mạng con giống hệt nhau chia sẻ
cùng trọng số (Bromley và cộng sự, 1993)—và áp dụng nhóm TI tại đầu ra của chúng trước
các lớp được kết nối đầy đủ. Trên hai biến thể của MNIST
bộ dữ liệu (Larochelle, Erhan, Courville, Bergstra, & Bengio, 2007; Jader-berg, Simonyan,
& Zisserman, 2015), được thiết kế để đánh giá chuẩn
thuật toán bất biến quay, phương pháp của họ thu được kết quả tương đương với
hoặc tốt hơn các DCNN hiện đại khác, với lợi thế bổ sung là
yêu cầu ít tham số mô hình hơn vì họ không sử dụng tăng cường dữ liệu, một kỹ thuật phổ
biến cho các tác vụ bất biến (Van Dyk & Meng, 2012).

5.1.2.9 Phân tích và triển vọng. Việc gộp nhóm là bắt buộc để giảm bớt gánh nặng tính
toán của các lớp tích chập tốn kém; tuy nhiên, bất chấp những thành công ban đầu của việc
gộp nhóm trung bình và sự đóng góp của tối đa
tập hợp vào sự gia tăng gần đây của DCNN, những bất cập liên quan đến chúng
(xem phần 4.2.3 và 5.1.2.1) đã dẫn dắt các nhà nghiên cứu điều tra các

chiến lược gộp nhóm. Mặc dù việc gộp nhóm Lp là hợp lý về mặt sinh học và có
là bằng chứng lý thuyết cho thấy nó dẫn đến sự khái quát hóa tốt hơn so với
để tối đa hóa việc gộp nhóm, sau này tiếp tục được ưa chuộng hơn đối với hình ảnh
nhiệm vụ phân loại, có lẽ là do khả năng nắm bắt sự bất biến từ dữ liệu trực quan đã biết
của nó. Bản chất ngẫu nhiên của việc gộp ngẫu nhiên (Zeiler
& Fergus, 2013) và nhóm hỗn hợp (D. Yu và cộng sự, 2014) mang lại cho chúng lợi thế hơn
nhóm tối đa liên quan đến khả năng nội tại của chúng để tránh
quá mức. Nhược điểm là các tính toán xác suất vốn có của chúng
đặt họ vào thế bất lợi liên quan đến gánh nặng tính toán của họ khi
so với các kỹ thuật xác định khác như nhóm tối đa hoặc trung bình. Sơ đồ dựa trên cây của
Lee et al. (2016) tạo ra hiệu suất phân loại đặc biệt, nhưng việc sử dụng các công cụ phân
tích quyết định làm tăng thêm sự phức tạp
và căng thẳng tính toán. Mặc dù SPP (He et al., 2014) giải quyết các biến thể trong thuộc
tính hình ảnh và nhanh chóng và hiệu quả, các kiến trúc sử dụng nó
không thể được đào tạo theo cách đầu cuối. Trong khi TI pooling (Laptev et al.,
2016) thúc đẩy tính bất biến của DCNN, làm cho phép tích chập 1 × 1 trở nên bất biến
đối với các phép biến đổi và tịnh tiến hình học, và đặc biệt là các phép biến đổi và tịnh tiến quy mô lớn
sự khác biệt, vẫn là một lĩnh vực nghiên cứu mở đòi hỏi nhiều nỗ lực hơn nữa, một số
được đề cập trong phần 6.2.

Từ phân tích, chúng ta có thể kết luận rằng các chiến lược gộp nhóm khác nhau
có nhiều ưu và nhược điểm khác nhau, và do đó có một mức độ siêu việt và chung chung cụ thể
chiến lược không thể được chọn ra. Mặc dù việc gộp tối đa có lẽ là quan trọng nhất
được thiết lập, việc lựa chọn chiến lược sẽ phụ thuộc phần lớn vào các yêu cầu
của một nhiệm vụ phân loại cụ thể và các nguồn lực có sẵn để hoàn thành
nó. Một số yếu tố chính cần xem xét ở đây là độ phức tạp của hệ thống, vì nó
có thể kết hợp các kỹ thuật từ xử lý tín hiệu số (Rippel
et al., 2015), phân tích quyết định (Lee et al., 2016) và máy tính truyền thống
tầm nhìn (He et al., 2014; Gong et al., 2014), yêu cầu độ chính xác phân loại,
hậu quả của việc lắp quá mức và các tài nguyên tính toán có sẵn.
Những đổi mới trong tương lai về việc hợp nhất nên tập trung vào việc hài hòa những xung đột này
yêu cầu, trong khi không mất đi nhận thức về nhu cầu chúng phải có thể hình dung được về
mặt sinh học, để chúng ta có thể cải thiện cả các mô hình của mình và
hiểu rõ hơn về hệ thống thị giác hiện tại của chúng ta.

5.2 Kích hoạt phi tuyến tính. Việc lựa chọn hàm kích hoạt ảnh hưởng đến thời gian đào tạo
mạng và điều này có ảnh hưởng đáng kể đến hiệu suất
của DCNN lớn trên các tập dữ liệu lớn (Krizhevsky và cộng sự, 2012). Được giới thiệu bởi
Nair và Hinton (2010) đối với máy Boltzmann sâu, ReLU đã được tạo ra
phổ biến cho DCNN bởi Krizhevsky et al. (2012), mặc dù Glorot, Bordes,
và Bengio (2011) đã chỉ ra rằng chúng dẫn đến thời gian đào tạo nhanh hơn
trong các mạng được giám sát hoàn toàn mà không cần phải đào tạo trước không giám sát.
Hình 10 so sánh thời gian đào tạo của ReLU (đường liền) với kích hoạt tiếp tuyến hyperbolic
(đường đứt nét) cho DCNN bốn lớp (Krizhevsky
et al., 2012), được đào tạo trên bộ dữ liệu CIFAR-10 (Krizhevsky, 2009).
DCNN với ReLUs được đào tạo nhanh hơn sáu lần so với một mạng tương đương



Hình 10: Thời gian đào tạo của ReLU so với kích hoạt tanh (Krizhevsky et al., 2012).

mạng lưới sử dụng kích hoạt tiếp tuyến hypebolic (Krizhevsky và cộng sự, 2012). Tiếp theo chúng tôi giới thiệu ngắn gọn về hàm kích hoạt không bão hòa này và thảo luận những động lực đã dẫn đến sự ra đời của một số người kế nhiệm họ.

5.2.1 Kích hoạt ReLU. Các hàm kích hoạt truyền thống, chẳng hạn như tiếp tuyến sigmoid hoặc hyperbolic được đưa ra bởi $f(x) = 1/(1 + e^{-x})$ và $f(x) = \tanh(x)$, tương ứng, trong đó f là đầu ra của nơ-ron như một hàm của đầu vào x của nó (ký hiệu tương tự được sử dụng cho phần còn lại của các hàm kích hoạt sau). ReLU (Nair & Hinton, 2010), một hàm tuyến tính từng phần, có dạng đơn giản hóa $f(x) = \max(x, 0)$. ReLU chỉ giữ lại phần dương của phép kích hoạt, bằng cách giảm phần âm xuống 0, trong khi toán tử tối đa tích hợp thúc đẩy tính toán nhanh hơn. ReLU có đã được sử dụng trong một số hệ thống phân loại hình ảnh hiện đại (Zeiler & Fergus, 2013, 2014; Lin và cộng sự, 2013; Gong và cộng sự, 2014; Simonyan & Zisser-man, 2014; Szegedy, Vanhoucke và cộng sự, 2015; Szegedy, Liu và cộng sự, 2015). Các thảo luận sâu hơn và động lực hơn nữa về chúng có thể được tìm thấy trong công trình được trình bày bởi Glorot et al. (2011).

5.2.2 Kích hoạt LReLU. Mặc dù ReLU (Nair & Hinton, 2010) dẫn đầu để hội tụ nhanh hơn (Nair & Hinton, 2010; Glorot và cộng sự, 2011; Krizhevsky et al., 2012; Maas, Hannun, & Ng, 2013) và không gặp phải vấn đề về độ dốc biến mất, trong đó các lớp thấp hơn có độ dốc gần bằng không bởi vì các lớp cao gần như bão hòa (Bengio, Simard, & Frasconi, 1994), họ có thể gặp bất lợi trong quá trình tối ưu hóa vì độ dốc bằng không khi đơn vị không hoạt động (Glorot et al., 2011; Maas et al., 2013). Điều này

có thể dẫn đến trường hợp các đơn vị không bao giờ được kích hoạt, vì gradient phổ biến
các thuật toán tối ưu hóa giảm dần chỉ tinh chỉnh trọng số của các đơn vị đã được
kích hoạt trước đó. Do đó, tương tự như vấn đề độ dốc biến mất, ReLUs
gặp phải tình trạng hội tụ chậm khi đào tạo mạng với hằng số không
gradient. Để bù đắp cho điều này, Maas et al. (2013) đã giới thiệu các đơn vị tuyến
tính chỉnh lưu rò rỉ (LReLU), cho phép các gradient nhỏ khác không khi
đơn vị chưa hoạt động nhưng đã bão hòa. Về mặt toán học, LReLU được đưa ra
qua

$$f(x) = \max(x, 0) + \lambda \min(x, 0), \tag{5.8}$$

trong đó λ là một tham số được xác định trước trong phạm vi (0, 1). LReLUs là
ban đầu được áp dụng cho các mô hình âm thanh (Maas et al., 2013); tuy nhiên, Xu et al.
(2015) nhận thấy rằng chúng hoạt động tốt hơn một chút so với ReLU đối với các nhiệm
vụ phân loại hình ảnh sau khi tiến hành đánh giá thực nghiệm trên CIFAR-10
và bộ dữ liệu CIFAR-100 (Krizhevsky, 2009). Một biện pháp về hiện đại
ImageNet (Russakovsky và cộng sự, 2015) sẽ tạo điều kiện thuận lợi cho việc so sánh này
tính phi tuyến tính dựa trên chỉnh lưu đối với các hoạt động tương tự khác.

5.2.3 Kích hoạt PReLU. Trong khi LReLU (Maas và cộng sự, 2013) dựa vào một tham số
được xác định trước để nén phần âm của tín hiệu kích hoạt, He
et al. (2015a) đã đề xuất một đơn vị tuyến tính chỉnh lưu tham số (PReLU) để học thích
ứng các tham số của đơn vị kích hoạt trong quá trình truyền ngược.
Về mặt toán học, PReLU giống với LReLU, ngoại trừ λ được thay thế bằng λk có thể học
được, được phép thay đổi cho các đầu vào khác nhau
kênh, được biểu thị bằng k. Do đó, PReLU có thể được biểu thị như

$$f(x_k) = \max(x_k, 0) + \lambda_k \min(x_k, 0). \tag{5.9}$$

Sử dụng mô hình DCNN được thiết kế trước đó và triển khai đào tạo (He & Sun, 2015),
He et al. (2015a) đã so sánh hiệu suất của
ReLUs (Nair & Hinton, 2010) đến PReLUs và tìm thấy lớn hơn 1%
tăng hiệu suất trên tập dữ liệu ILSVC (Russakovsky và cộng sự, 2015). Hơn nữa, khi
phương pháp này được kết hợp với phương pháp khởi tạo trọng số mạnh mẽ đặc biệt xem
xét các phi tuyến tính đã chỉnh lưu, chúng
lần đầu tiên vượt qua hiệu suất của con người trong thử thách này
chuẩn mực (Russakovsky và cộng sự, 2015). Vào thời điểm đó, kết quả của họ là
trạng thái hiện đại của tập dữ liệu này và mặc dù nằm ngoài cuộc thi thường niên, họ
đã đánh bại bài dự thi chiến thắng năm 2014 (Simonyan & Zisserman, 2014).
Mặc dù vậy, Xu et al. (2015) nhận thấy rằng PReLU luôn hoạt động tốt hơn các đơn vị
chỉnh lưu khác, chẳng hạn như ReLU (Nair & Hinton, 2010) và
LReLU (Zeiler & Fergus, 2014), trên tập huấn luyện, do đó cảnh báo về thực tế
rằng chúng gặp phải vấn đề quá khớp nghiêm trọng trên các tập dữ liệu nhỏ hơn.

5.2.4 Kích hoạt APL. Tương tự như PReLU (He et al., 2015a), Agostinelli, Hoffman, Sadowski và Baldi (2014) đồng thời đề xuất phương pháp thích ứng các hàm kích hoạt tuyến tính từng phần (APL), được tham số hóa, học độc lập cho mỗi tế bào thần kinh bằng cách sử dụng phương pháp giảm dần độ dốc thông thường, và có thể biểu diễn cả hàm lồi và không lồi của đầu vào. Về mặt toán học, APL được biểu thị dưới dạng tổng của các hàm hình bản lề,

$$h_i(x) = \max(0, x) + \sum_{s=1}^{S} a_i^s \max(0, -x + b_i^s), \quad (5.10)$$

trong đó S là số bản lề, là siêu tham số được thiết lập trước, và các biến như a_i^s đối với $i = 1, \ldots, S$ được học trong quá trình đào tạo. Trong phương trình 5.7, a_i các biến kiểm soát độ dốc của các đoạn tuyến tính, trong khi vị trí của bản lề được xác định bởi b_i^s biến. Mặc dù chúng đã thu được kết quả mới nhất về dữ liệu CIFAR-10 và CIFAR-100 bộ (Krizhevsky, 2009), đánh bại NIN có hiệu suất cao (Lin et al., 2013), không giống như PReLU (He et al., 2015a), thử nghiệm phân loại hình ảnh của họ không bao gồm bộ dữ liệu ILSVRC đầy thách thức (Russakovsky và cộng sự, 2015). Vì vậy, hiệu suất tương đối giữa các kỹ thuật tương tự này không thể được được đánh giá.

5.2.5 Kích hoạt RReLU. Để giải quyết vấn đề quá khớp liên quan đến PReLU (He et al., 2015a), đơn vị tuyến tính chỉnh lưu ngẫu nhiên (RReLU) đã được đề xuất trong Kaggle National Data Science Bowl Cuộc thi (National Data Science Bowl | Kaggle, 2016). Đối với RReLU, các thành phần tiêu cực của hàm kích hoạt được chọn ngẫu nhiên từ một sự phân phối đồng đều trong quá trình đào tạo. Trong quá trình thử nghiệm, chúng được tính trung bình, tương tự như kỹ thuật Dropout (Hinton et al., 2012; Srivastava et al., 2014) trước khi được cố định, do đó cho phép họ có được kết quả xác định (Xu et al., 2015). Về mặt toán học, PReLU có thể được biểu thị như

$$f(x_{ji}^{(n)}) = \max(x_{ji}^{(n)}, 0) + \lambda_{ji}^{(n)} \min(x_{ji}^{(n)}, 0), \quad (5.11)$$

trong đó $\lambda_{ji}^{(n)}$ biểu thị tham số lấy mẫu ngẫu nhiên trên kênh thứ k của ví dụ thứ n. Trên CIFAR-10 và CIFAR-100 (Krizhevsky, 2009), và một bộ dữ liệu phân loại sinh vật phù du riêng tư (National Data Science Bowl | Kaggle, 2016), độ chính xác phân loại của họ vượt trội hơn ReLU (Nair & Hinton, 2010), LReLU (Maas và cộng sự, 2013) và kích hoạt PReLU (He và cộng sự, 2015a).

5.2.6 Kích hoạt ELU. Trong khi ReLU (Nair & Hinton, 2010), LReLU (Maas et al., 2013) và PReLU (He et al., 2015a) đều không bão hòa và do đó làm giảm vấn đề biến mất độ dốc (Bengio và cộng sự, 1994), chỉ ReLUs

đảm bảo trạng thái vô hiệu hóa mạnh mẽ chống nhiễu (Nair & Hinton, 2010; Clevert, Un-terthiner, & Hochreiter, 2016); tuy nhiên, chúng không âm và do đó có một kích hoạt trung bình lớn hơn không. Để giải quyết vấn đề này, Clevert et al. (2016) đề xuất đơn vị tuyến tính mũ (ELU), có giá trị âm cho phép kích hoạt gần bằng không, như ng cũng bão hòa đến giá trị âm với các đối số nhỏ hơn. Vì độ bão hòa làm giảm sự thay đổi của các đơn vị khi bị vô hiệu hóa, lập luận chính xác về việc vô hiệu hóa trở nên ít liên quan hơn, do đó làm cho ELU mạnh mẽ chống lại tiếng ồn. Chính thức:

$$f(x) = \text{cực đại}(x, 0) + \text{cực tiểu}(\lambda(e^{x} - 1), 0), \tag{5.12}$$

trong đó λ là một tham số đư ợc xác định trư ớc để kiểm soát số lư ợng ELU sẽ bão hòa cho các đầu vào tiêu cực. ELU đã đẩy nhanh quá trình học DCNN và dẫn đầu để có độ chính xác phân loại cao hơn khi so sánh với các hàm kích hoạt khác như ReLU. Đặc biệt, trong số những thành công khác, chúng đặt ra một kỷ lục trên bộ dữ liệu CIFAR-100 (Krizhevsky, 2009), đánh bại kỷ lục trư ớc đó tốt nhất thu đư ợc bằng cách gộp tối đa phân số DCNNs do Graham đề xuất (2014), và họ đã đạt đư ợc tốc độ hội tụ đáng khích lệ trên ImageNet kích thích (Russakovsky và cộng sự, 2015). Mặc dù các kích hoạt này cung cấp kết quả phân loại hình ảnh đầy hứa hẹn và giảm đáng kể căng thẳng tính toán trên DCNN, thử nghiệm tiếp theo sử dụng chúng, trong đặc biệt với các kiến trúc khác nhau, là cần thiết.

5.2.7 Kích hoạt SReLU. Mặc dù ReLUs (Nair & Hin-ton, 2010), LReLUs (Maas và cộng sự, 2013) và PReLUs (He và cộng sự, 2015a) đã thành công, tất cả chúng đều có khả năng hạn chế trong việc học các phép biến đổi phi tuyến tính cụ thể, vì tất cả các phép kích hoạt này đều lồi, nên chúng không thể học đư ợc các hàm không lồi. Mặc dù kích hoạt APL có thể xấp xỉ các hàm lồi, như ng nó thực hiện như vậy với các ràng buộc không phù hợp làm suy yếu khả năng đại diện. Để giảm bớt những lo ngại này, hãy lấy cảm hứng từ tâm lý học và khoa học thần kinh, Jin et al. (2015) đã đề xuất một loại mới của kích hoạt, đư ợc gọi là đơn vị tuyến tính chỉnh lư u hình chữ S (SReLU). Kích hoạt này kết hợp ba hàm tuyến tính và thực hiện ánh xạ R R với biểu thức toán học sau đây:

$$f(xi) = \begin{cases} tr_i + ar_i(xi - tr_i), & xi \geq tr_i \\ xi, & tr_i > xi > tl_i \\ tl_i + al_i(xi - tl_i), & xi \leq tl_i \end{cases}, \tag{5.13}$$

nơi { tr , tl , al } là các tham số có thể học đư ợc đư ợc sử dụng để mô hình hóa từng cá nhân đơn vị kích hoạt SReLU trực tiếp và chỉ số i biểu thị rằng SReLU là đư ợc phép thay đổi trên các kênh đầu vào khác nhau. Tóm lại, theo hư ớng tích cực, khi các đầu vào vư ợt quá ngư ỡng tr độ dốc của đư ờng thẳng bên phải của đư ờng cong kích hoạt đư ợc đư a ra bởi ar trong khi tl đại diện cho một sự đối xứng

ngư ông theo hư ớng tiêu cực. Đối với các đầu vào nhỏ hơn tl dòng bên trái của đư ờng cong kích hoạt tính toán các đầu ra. Đối với các đầu vào trong phạm vi (tl i,tr i), các đầu ra là các hàm tuyến tính có độ dốc là một và không có độ lệch. SReLU đã đư ợc đư a vào các mô hình tiên tiến của Lin et al. (2013) và Szegedy, Liu et al. (2014), và các kết quả thực nghiệm đạt đư ợc, sau một số thí nghiệm phân loại hình ảnh trên MNIST (LeCun et al., 1998), CIFAR-10 và CIFAR-100 (Krizhevsky, 2009), và ILSVRC (Russakovsky et al., 2015) các tập dữ liệu, minh họa tính ư u việt của chúng so với một số hoạt động hiệu suất cao khác (Nair & Hinton, 2010; Goodfellow et al., 2013; Maas et al., 2013; He et al., 2015a; Xu et al., 2015). Với kết quả đầy hứa hẹn này, thật thú vị khi nghĩ về việc liệu những tiến bộ trong học tập sâu sắc trong tư ơng lai có cũng dựa vào nguồn cảm hứng sâu hơn từ khoa học tâm lý và khoa học thần kinh.

5.2.8 Kích hoạt Maxout và Probout. Goodfellow và cộng sự (2013) đề xuất một giải pháp thay thế cho một số đơn vị kích hoạt dựa trên chỉnh lư u đư ợc gọi là Maxout, là các kích hoạt tạo ra giá trị tối đa từ một tập hợp đầu vào. Đối với một đầu vào x Rd cho trư ớc, một lớp ẩn trong mạng Maxout thực hiện hàm kích hoạt sau,

$$f(xi) = \max_{j \in [1,k]} xTW \cdots i_{j} + bij , \quad (5.14)$$

trong đó W Rd×m×k và b Rm×k đều là các tham số có thể học cư ợc. Cụ thể, đối với DCNN, bản đồ đặc điểm Maxout có thể đạt đư ợc bằng cách lấy tối đa trên k bản đồ đặc điểm affine, tư ơng ứng với một nhóm không gian con trên các kênh ở các vị trí không gian bổ sung. Ngoài ra đạt đư ợc kết quả phân loại hình ảnh tiên tiến nhất trên một số chuẩn mực, các tác giả đã cung cấp bằng chứng thực nghiệm rằng Maxout là tốt phù hợp cho việc đào tạo DCNN với Dropout (Hinton et al., 2012; Srivastava et al., 2014), và nó hỗ trợ trong việc tính trung bình mô hình và tối ư u hóa DCNN. Mặc dù vậy, nó vẫn chịu chung niềm tin như ReLUs (Nair & Hinton, 2010), LReLUs (Maas và cộng sự, 2013) và PReLUs (He và cộng sự, 2015a) liên quan đến sự bất lực của họ trong việc học các hàm không lồi; hơn nữa, nó đòi hỏi một số lư ợng tham số bổ sung, làm tăng chi phí lư u trữ và bộ nhớ và đòi hỏi thời gian đào tạo dài hơn đáng kể (Jin et al., 2015).

Như đã đề cập ở trên, Maxout (Goodfellow và cộng sự, 2013) thực hiện một hoạt động gộp không gian con trên một nhóm các phép biến đổi tuyến tính và điều này làm cho nó một phần bất biến với các biến thể trong đầu vào. Để cải thiện tính chất bất biến này và duy trì các tính chất mong muốn của Maxout đơn vị, Springenberg và Riedmiller (2013) đã đề xuất một biến thể xác suất của Maxout, đư ợc gọi là Probout, trong đó toán tử cực đại đư ợc thay thế với kỹ thuật lấy mẫu xác suất. Trên CIFAR-10 và CIFAR-100 (Krizhevsky, 2009) và các tập dữ liệu SVHN (Netzer và cộng sự, 2011), các kích hoạt Probout đạt đư ợc kết quả phân loại tốt hơn so với các kích hoạt Maxout; tuy nhiên, so với các đối thủ cạnh tranh của chúng, chúng có hiệu quả phân loại tốt hơn

gánh nặng tính toán do bản chất vốn có của chúng như ng lại tốn kém về mặt tính toán
tính toán xác suất.

5.2.9 Phân tích và triển vọng. Sử dụng kích hoạt phi tuyến tính chính xác cho
nhiệm vụ cụ thể cải thiện độ chính xác phân loại và hiệu suất tính toán của DCNN. Mặc dù
các mô hình ban đầu của chúng tôi sử dụng kích hoạt sigmoid,
chúng bão hòa trong quá trình truyền ngư ợc, điều này sẽ tiêu diệt gradient cục bộ và chúng
ảnh hư ởng tiêu cực đến động lực mạng trong quá trình giảm gradient,
vì đầu ra của chúng không có tâm là số không. Điều này dẫn đến sự phát triển của
kích hoạt ReLU phổ biến (Nair & Hinton, 2010), giúp tăng tốc đáng kể sự hội tụ của mạng.
Tuy nhiên, khi các gradient lớn đi qua
chúng trong quá trình đào tạo mạng, chúng có thể chết không thể phục hồi, và điều này dẫn đến
những cải tiến khác như ReLU (Maas và cộng sự, 2013), PReLU (He và cộng sự,
2015a) và kích hoạt APL (Agostinelli et al., 2014). Mặc dù các kết quả thực nghiệm đầy hứa
hẹn của các phư ơng pháp này, như ng vẫn cần tiếp tục điều tra để đánh giá
độ tin cậy của các nhiệm vụ phân loại đa dạng vẫn cần phải đư ợc thực hiện. ELU
Clevert et al. (2016) và các hoạt động của SReLU (Jin et al., 2015) đã giải quyết
những như ợc điểm khác của ReLU, chẳng hạn như kích hoạt trung bình tích cực của chúng và
không có khả năng xử lý các hàm không lồi, như ng tính nhất quán của chúng cũng tư ơng đối
không đư ợc chứng nhận. Hơn nữa, các kích hoạt như Maxout và Probout
có vẻ đặc biệt phù hợp để đào tạo DCNN với Dropout (Hinton
et al., 2012), như ng chúng đòi hỏi số lư ợng tham số lớn và có thể tốn kém về mặt tính toán.
do đó ủng hộ nhu cầu cải tiến hơn nữa trong lĩnh vực này.

Từ phân tích này, chúng ta có thể kết luận rằng không có giải pháp rõ ràng nào cho
kích hoạt nào đư ợc sử dụng cho một nhiệm vụ cụ thể, như ng một cách tiếp cận thử và sai
bắt đầu với ReLUs và tiến triển đến các hoạt động khác và giám sát những thiếu sót của chúng
so với hiệu suất yêu cầu có thể đư ợc áp dụng. Trong khi một
hư ớng đi thú vị trong tư ơng lai là kết hợp việc sử dụng các hàm kích hoạt khác nhau trong
cùng một mô hình DCNN để có đư ợc những lợi ích đa dạng, một phân tích lý thuyết chi tiết
về lý do tại sao các kích hoạt dựa trên chỉnh lư u hiện tại của chúng tôi lại thành công
theo kinh nghiệm cũng có tầm quan trọng tối cao. Để bổ sung cho điều này, mặc dù
đã đư ợc chứng minh bởi Baldi và Sadowski (2013, 2014) rằng bất kỳ sự liên tục nào,
hàm kích hoạt từng phần có thể phân biệt hai lần, trong đó ReLU là một
trư ờng hợp đặc biệt, có thể đư ợc sử dụng kết hợp với kỹ thuật tính trung bình mô hình
Dropout (Hinton et al., 2012; Goodfellow et al., 2013; Srivastava et al.,
2014; xem phần 5.4.1 và 5.4.3), tác động của các hoạt động khác nhau lên
đặc điểm tổng quát của Dropout hoặc thậm chí là các chính quy khác
kỹ thuật vẫn cần phải đư ợc phân tích cơ bản, do đó mở ra
cánh cửa cho công việc tư ơng lai đầy hứng khởi.

5.3 Thành phần giám sát. Sau công trình sáng tạo của Krizhevsky
et al. (2012), các phư ơng pháp đào tạo trư ớc DCNN không giám sát trư ớc đó (Ran-zato et
al., 2006, 2007; Weston et al., 2008; Ahmed et al., 2008; Jarrett et al.,
2009; Lee et al., 2009; LeCun et al., 2010) phần lớn đã bị bỏ rơi hoàn toàn

đào tạo có giám sát. Nhìn chung, việc học trong DCNN đạt đư ợc bằng cách giảm thiểu một hàm mất mát cụ thể, với mất mát phân loại phổ biến nhất là mất mát softmax (Krizhevsky và cộng sự, 2012; Lin và cộng sự, 2013; Goodfellow và cộng sự, 2013; Zeiler & Fergus, 2013, 2014; Chatfield và cộng sự, 2014; Simonyan & Zisserman, 2014; Szegedy, Liu và cộng sự, 2015; Szegedy, Vanhoucke và cộng sự, 2015; He và cộng sự, 2015a, 2015b). Trong phần này, chúng tôi giới thiệu ngắn gọn về mất mát này và giải quyết một số động lực để sử dụng mất mát thay thế trong DCNN.

5.3.1 Softmax Loss. Hàm kích hoạt softmax đư ợc sử dụng rộng rãi trong lớp kết nối đầy đủ cuối cùng của DCNN, do tính đơn giản và diễn giải theo xác suất của nó. Khi hàm kích hoạt này đư ợc kết hợp với mất entropy chéo (hoặc hồi quy logistic đa thức) trong lớp kết nối đầy đủ cuối cùng của DCNN, chúng tạo thành mất softmax đư ợc sử dụng rộng rãi. Về mặt hình thức, đối với tính năng đầu vào thứ i xi có nhãn tư ơng ứng yi, mất softmax có thể đư ợc viết là

$$L = \frac{1}{N} \text{Cái đó} = \frac{1}{N} \text{nhật ký} \frac{\text{Ôi}}{\text{chúa ơi}_j}, \tag{5.15}$$

trong đó phần tử thứ j (j [1,K],K là số lớp) của vectơ điểm lớp f đư ợc biểu diễn bởi fj và N là lư ợng dữ liệu đào tạo.
Đối với tổn thất này, f thư ờng là sự kích hoạt của lớp W đư ợc kết nối đầy đủ; do đó, có thể đư ợc biểu thị là fyi = WT xi trong đó Wyi là cột yith của W (Liu, fyi , Wen, Scut, Yu, & Yang, 2016).

5.3.2 Tổn thất tư ơng phản và mất mát bộ ba. Để tăng cư ờng tính chặt chẽ trong lớp và khả năng tách biệt giữa các lớp, và do đó củng cố DCNN bằng nhiều thông tin phân biệt hơn, tổn thất tư ơng phản, còn đư ợc gọi là tổn thất dựa trên biên (Hadsell, Chopra, & LeCun, 2006) và tổn thất bộ ba (Schroff, Kalenichenko, & Philbin, 2015), đã đư ợc đề xuất độc lập (Liu, Wen và cộng sự, 2016). Tổn thất tư ơng phản lần đầu tiên đư ợc triển khai trong một DCNN Siamese để giảm chiều của dữ liệu bằng cách học các phép ánh xạ bất biến với các biến dạng hình học (Hadsell và cộng sự, 2006), trong khi các DCNN nhúng đư ợc Weston và cộng sự mô tả (2008) kết hợp nó với tổn thất bản lề cho các nhiệm vụ phân loại hình ảnh và gắn nhãn vai trò ngữ nghĩa. Các ứng dụng liên quan đến phân loại hình ảnh khác đã sử dụng mất mát tư ơng phản như một phần của kiến trúc DCNN bao gồm biểu diễn khuôn mặt (Sun, Chen, Wang & Tang, 2014) và tính tư ơng đồng trực quan cho tìm kiếm trực quan (Bell & Bala, 2015), trong đó mất mát tư ơng phản đư ợc sử dụng kết hợp với mất mát softmax.

Hơn nữa, nó cũng đư ợc sử dụng để truy xuất trư ờng hợp hình ảnh (Lin, Morere, Chandrasekhar, Veillard, & Goh, 2015), trong đó các tác giả đề xuất mất mát biên độ kép. Đối với mất mát tư ơng phản, hàm mất mát chạy qua các cặp mẫu, không giống với các hệ thống bảo thủ, trong đó nó chạy qua các mẫu riêng lẻ. Về mặt hình thức, như đư ợc giới thiệu bởi Hadsell et al.

(2006), đối với một cặp vectơ đầu vào X1 , X2 I, với nhãn nhị phân Y (nếu Y = 0
X1 và X2 được coi là tương tự và Y = 1 nếu không giống nhau), dạng chung của
sự mất mát tương phản là

$$L = \sum_{tôi=1}^{P} L(W, (Y, X1, X2)) = (1 - Y)LS(Di_{TRONG}) + YLD(Di_{TRONG}), \quad (5.16)$$

trong đó DW (X1 , X2) được viết là DW (để rút ngắn ký hiệu), (Y, X1 , X2) là
cặp mẫu được gắn nhãn thứ i, hàm mất mát một phần cho một cặp điểm tương tự
và các điểm không giống nhau được biểu diễn lần lượt bởi LS và LD , và P biểu diễn số
cặp đào tạo.

Tổn thất ba phần cho DCNN (Schroff và cộng sự, 2015), trước đây là
được sử dụng cho phân loại láng giềng gần nhất có biên độ lớn (Weinberger, Blitzer,
& Saul, 2005), yêu cầu các mẫu đào tạo theo bội số của ba. Nó giảm thiểu
khoảng cách giữa một mẫu neo danh tính chung và một mẫu dương tính trong khi tối đa hóa
khoảng cách giữa mẫu neo và một mẫu âm tính có danh tính khác. Về mặt hình thức, đối với
phân loại khuôn mặt,
tổn thất tối thiểu L sau đó là

$$L = \sum_{}^{N} [f(x) - f(x)_2^2 - f(nếu) - f(xn)_2^2 + một]_+, \quad (5.17)$$

nơi xa $_{Tôi}$ là hình ảnh neo của một người cụ thể, xp là những hình ảnh tích cực của
cùng một người, hình ảnh tiêu cực của bất kỳ người nào khác được biểu thị bằng xn $_{Tôi}$, α là
biên độ bắt buộc giữa các cặp dương và âm, và N là số lượng của tất cả các bộ ba có thể
có trong bộ huấn luyện. Đối với mất bộ ba, hình ảnh neo cần phải gần hơn với tất cả các
hình ảnh dương khác của cùng một người
hơn là bất kỳ hình ảnh tiêu cực nào của bất kỳ người nào khác. Bằng cách kết hợp bộ ba
mất mát với ánh xạ hình ảnh nhúng, được tối ưu hóa bởi DCNN, thành một
Không gian Euclid trong đó khoảng cách tương ứng trực tiếp với sự giống nhau của khuôn mặt,
Schroff et al. (2015) đã đạt được kết quả công bố tốt nhất về La-belled Faces in the Wild
(LFW; Huang, Ramesh, Berg, & Learned-Miller,
2007) và cơ sở dữ liệu YouTube Faces (Wolf, Hassner, & Maoz, 2011). Những
kết quả là một sự cải thiện đáng kể so với lỗi tốt nhất trước đó
tỷ lệ được báo cáo trong tài liệu (Sun, Wang, & Tang, 2015).

Bộ dữ liệu LWF (Huang et al., 2007) được giới thiệu vào năm 2007 và đã
kể từ đó trở thành tiêu chuẩn học thuật thực tế cho việc xác minh và nhận dạng khuôn mặt (Sun,
Chen và cộng sự, 2014; Taigman, Yang, Ranzato, & Wolf, 2014;
Schroff và cộng sự, 2015; Zhou, Cao, & Yin, 2015). Ban đầu, hầu hết các nỗ lực xác minh và
nhận dạng khuôn mặt trên tập dữ liệu này đều sử dụng các trình trích xuất tính năng riêng
lẻ hoặc kết hợp, với các hệ thống hàng đầu (Barkan, Weill, Wolf, &
Aronowitz, 2013; Cao, Wipf, Wen, Duan, & Sun, 2013; Chen, Cao, Wen,
& Sun, 2013) sử dụng hơn 10.000 mô tả hình ảnh. Tuy nhiên, nhiều hơn

Gần đây, các DCNN được giám sát hoàn toàn đã trở thành trọng tâm của hầu hết các hệ thống đạt thành tích cao nhất, như minh họa trong Bảng 3. Bảng so sánh độ chính xác của các DCNN có hiệu suất cao nhất với hiệu suất ở cấp độ con người (HLP; Kumar, Berg, Belhumeur, & Naya, 2009). Để có ý nghĩa, chỉ những kết quả được công bố trong các bài báo học thuật mới được đưa vào; các kết quả tiếp theo, bao gồm các kỹ thuật không phải DCNN, được thảo luận trong bài báo khảo sát liên quan đến tập dữ liệu (Learned-Miller, Huang, RoyChowdhury, Li, & Hua, 2016). Bất chấp những thành công này, các cơ chế mà não người sử dụng để có thể dễ dàng xác định và nhận dạng khuôn mặt trong thời gian ngắn vẫn còn bỏ ngỏ. Điều thú vị là có thể hệ thần kinh trung ương đã tiến hóa để xử lý khuôn mặt theo một cách khác khi so sánh với các vật thể (Leibo, Mutch, & Poggio, 2011), và do đó các mô hình DCNN phân loại và nhận dạng khuôn mặt trong tương lai có thể cần kết hợp loại bằng chứng này.

5.3.3 Mất biên lớn. Khẳng định rằng độ tương đồng góc lớn hơn sẽ dẫn đến khả năng tách biệt góc lớn hơn giữa các đặc điểm đã học, đến lượt nó sẽ dẫn đến việc tạo ra nhiều đặc điểm phân biệt hơn, Liu, Wen và cộng sự (2016) đã giới thiệu một biên góc giữa vectơ đặc điểm đầu vào và ma trận trọng số cho mất mát softmax biên lớn tổng quát hơn, mà họ gọi là softmax biên lớn (L-Softmax). Về mặt hình thức, L-Softmax được định nghĩa là

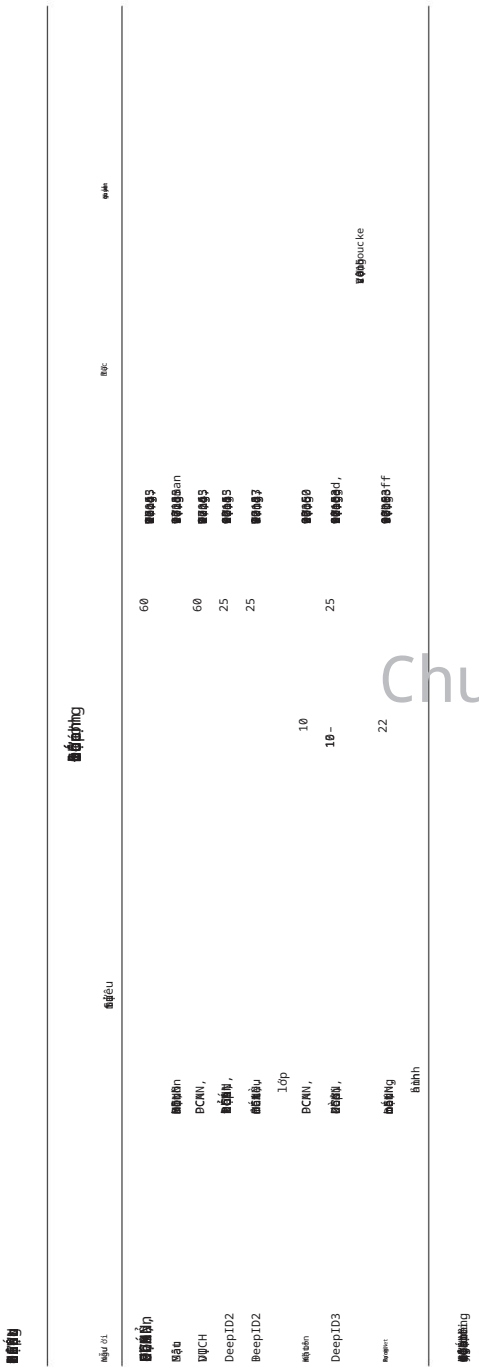
$$\text{L-Softmax} = \log\left(\frac{e^{\text{Nói xiψ}(\theta Yi)\text{ và}}}{\text{và } e^{\text{Wyi xiψ}(\theta Yi)} + \sum_{j=y} e^{\text{Wj cơ thể}(\theta j)}}\right), \tag{5.18}$$

trong đó

$$\psi(\theta) = \begin{cases} \cos(m\theta), & 0 \le \theta \le \frac{P}{2} \\ D(\theta), & \frac{P}{\text{tôi}} \le \theta \le \pi \end{cases}, \tag{5.19}$$

trong đó θj là biên độ góc, a(i) là vectơ đầu vào, wj là cột thứ j của ma trận trọng số và m điều chỉnh biên độ giữa các lớp. Ngoài việc tạo ra nhiều tính năng phân biệt hơn, những lợi thế mong muốn khác của L-Softmax bao gồm thực tế là cách diễn giải hình học của nó rất rõ ràng và nó tránh được một phần hiện tượng quá khớp. Khi áp dụng cho các tác vụ phân loại hình ảnh, nó vượt trội hơn so với mất mát softmax ban đầu (đối với cùng một kiến trúc) và đạt được kết quả ngang bằng với công nghệ tiên tiến cho tập dữ liệu MNIST. Nó cũng đạt được kết quả tiên tiến mới trên các tập dữ liệu CIFAR-10 và CIFAR-100 (LeCun và cộng sự, 1998; Krizhevsky, 2009; Bulo & Kontschieder, 2014; Liang & Hu, 2015; Liu, Wen, và cộng sự, 2016).

Trong khi những người khác đề xuất sử dụng hàm mất mát lai được thiết kế đặc biệt kết hợp logarit âm của giá trị chuẩn hóa với các lỗi có trọng số theo thứ tự khác nhau để cải thiện tính mạnh mẽ đối nghịch



(xem phần 6.4) của DCNN (Zhao & Griffin, 2016), ứng dụng của
Tổn thất L-Softmax đối với thách thức cụ thể này vẫn còn lớn. Cụ thể, kể từ khi
sự mất mát này dẫn đến việc tạo ra nhiều tính năng phân biệt hơn, việc áp dụng nó
vào thách thức của các ví dụ đối nghịch sẽ góp phần vào việc hiểu liệu những thay
đổi toàn diện đối với cách chúng ta đào tạo DCNN có cần thiết hay không
giải quyết những thách thức còn lại của họ.

5.3.4 Mất mát L2-SVM. Trong khi SVM trước đây đã được sử dụng kết hợp với CNN
(tức là bằng cách thay thế lớp softmax bằng SVM) để cải thiện hiệu suất phân loại
(Huang & LeCun, 2006; Lee et al., 2009;
Coates et al., 2011), nhược điểm là các tính năng cấp thấp hơn của CNN
không được học liên quan đến mục tiêu của SVM. Để giải quyết vấn đề này,
Collobert và Bengio (2004) và Nagi, Di Caro, Giusti, Nagi và Gam-bardella (2012) đã
đề xuất đào tạo chung ở các cấp thấp hơn bằng cách,
giới thiệu các hàm chi phí mới để tích hợp SVM với MLP và CNN.
Lấy cảm hứng từ điều này, Tang (2013) cũng đề xuất tích hợp SVM với DCNN
nhưng họ đã thay thế mất bản lề SVM tiêu chuẩn (L1-SVM) bằng L2-SVM
mất mát (Hinton, 1989). Khi so sánh với mất mát L1-SVM, mất mát L2-SVM
có thể phân biệt được và phạt lỗi sâu sắc hơn. SVM ban đầu là
được xây dựng cho phân loại nhị phân. Do đó, các mẫu đào tạo được đưa ra và
nhãn tương ứng của chúng (xn, yn), n = 1,..., N, xn RD,tn { 1, +1},
L2-SVM giảm thiểu tổn thất bản lề bình phương, được biểu thị chính thức bằng bài
toán tối ưu hóa không bị ràng buộc sau đây,

$$\min_w \frac{1}{2} wTW + C \sum_{n=1}^{N} \text{tối đa}(1 - wTxntn, 0)2, \quad (5.20)$$

trong đó W là trọng số kết nối lớp áp chót với softmax
lớp. Nhãn lớp cho dữ liệu thử nghiệm x có thể được dự đoán bằng arg maxt (wTx)t,
trong khi đối với SVM đa lớp (Vapnik, 1995), trong đó đầu ra của SVM thứ k
được biểu thị là ak (x) = wTx và lớp dự đoán là argmaxk ak (x). Khi
so với tổn thất softmax thông thường, đối với cùng một kiến trúc DCNN, tổn thất L2-
SVM cho thấy hiệu suất phân loại được cải thiện trên
Bộ dữ liệu CIFAR-10 (Krizhevsky, 2009), thu được kết quả tương đương với
trạng thái hiện tại (vào thời điểm đó) của nghệ thuật, sử dụng một phức tạp hơn nhiều
mô hình bao gồm các lớp chuẩn hóa độ tương phản và tham số Bayesian
điều chỉnh tinh tế (Snoek, Larochelle, & Adams, 2012).

5.3.5 Phân tích và triển vọng. Tổn thất softmax là một
sự lựa chọn cho CNN do tính đơn giản, giải thích xác suất và
đầu ra trực quan mà nó tạo ra. Tuy nhiên, để cung cấp cho CNN khả năng
trích xuất nhiều đặc điểm phân biệt hơn, các mất mát khác, chẳng hạn như sự tương phản
mất mát (Hadsell và cộng sự, 2006) và mất mát bộ ba (Schroff và cộng sự, 2015) đã
được đề xuất. Mặc dù những mất mát này khuyến khích học tập phân biệt, một vấn đề
phát sinh là về mặt lý thuyết, số lượng cặp đào tạo bắt buộc hoặc

bộ ba có thể lên tới O(N2), trong đó N là tổng số mẫu đào tạo.
Hơn nữa, đối với một tập dữ liệu lớn, như tập dữ liệu đư ợc sử dụng cho ILSVRC (Russakovsky và cộng sự, 2015), bao gồm hơn 1 triệu hình ảnh, tập hợp con
của các mẫu đào tạo sẽ yêu cầu lựa chọn cẩn thận trực tuyến hoặc ngoại tuyến cho cả hai
của những tổn thất này. Điều này dẫn đến tổn thất cụm ghép nối đư ợc đề xuất gần đây (xem Liu, Tian, Yang, Pang, & Huang, 2016), giúp tăng tốc sự hội tụ của mạng và ổn định quá trình đào tạo. Mặc dù nó tạo ra triển vọng
kết quả nhận dạng lại xe, các nhiệm vụ phân loại truyền thống hơn vẫn chư a
để đư ợc thử nghiệm. Mặc dù có lợi ích và hiệu suất đư ợc thiết lập tốt và chấp nhận đư ợc, như ng tổn thất softmax không khuyến khích rõ ràng tính chặt chẽ trong lớp và khả năng tách biệt giữa các lớp. Nó sử dụng khoảng cách cosin giữa các lớp
cho điểm phân loại của nó; do đó, việc dự đoán nhãn cho một đầu vào nhất định là đư ợc xác định chủ yếu bởi sự tư ơng đồng về góc với mỗi lớp. Điều này đã truyền cảm hứng cho đề xuất về tổn thất L-Softmax (Liu, Tian và cộng sự, 2016), vẫn
cần đư ợc ủy quyền về các vấn đề mở của DCNN (xem phần 5.3.3 và 6.4). Việc tích hợp mất mát L2-SVM, theo truyền thống liên quan đến SVM, tạo điều kiện cải thiện độ chính xác phân loại, như ng giống như mất mát cụm đư ợc ghép nối, tính nhất quán của nó trong nhiều nhiệm vụ khác nhau vẫn chư a đư ợc biết rõ.

Bất chấp những cải tiến đư ợc tóm tắt ở trên, tổn thất softmax vẫn là một lựa chọn có uy tín cho các chuẩn mực học thuật truyền thống như MNIST (Le-Cun và cộng sự, 1998) và ImageNet (Russakovsky và cộng sự, 2015) hoặc các nhiệm vụ khác nơi mà một lớp đầu ra duy nhất (nhãn) cho mỗi hình ảnh là cần thiết. Đối với thế giới thực nhiệm vụ yêu cầu nhiều lớp cho mỗi hình ảnh, mỗi lớp, hồi quy logistic nhiều lớp đư ợc khuyến nghị làm điểm khởi đầu. Dựa trên các yêu cầu của
nhiệm vụ, thử nghiệm với các tổn thất khác đư ợc đề cập trong phần này có thể đư ợc khám phá. Ví dụ, phân loại chi tiết có thể đư ợc hư ởng lợi đặc biệt bằng cách sử dụng cụm mất mát đư ợc ghép nối, trong khi để xác minh khuôn mặt
hoặc các nhiệm vụ xác minh khác không bị hạn chế bởi tài nguyên tính toán, mất mát bộ ba có thể tạo ra hiệu suất xác minh tuyệt vời. Cuối cùng, đó là khuyến nghị rằng công việc trong tư ơng lai nên thách thức sự phát triển của tiểu thuyết các hàm mất mát giải quyết các vấn đề mở của DCNN, hỗ trợ công trình do Zhao và Griffin (2016) đề xuất, trong khi việc sử dụng SVM đa lớp khác
công thức, hoặc thậm chí các phân loại khác như RBF, tiếp tục điều tra những cải tiến về hiệu suất đư ợc trình bày bởi LeCun et al. (1998) và
Tang (2013) cũng cần đư ợc khám phá.

5.4 Cơ chế điều chỉnh. DCNN là những mô hình biểu đạt rất cao, có khả năng học đư ợc những mối quan hệ cực kỳ phức tạp giữa chúng đầu vào và đầu ra. Tuy nhiên, với dữ liệu đào tạo hạn chế, ngay cả đối với bộ dữ liệu (Krizhevsky và cộng sự, 2012), nhiều trong số những ánh xạ phức tạp này là do nhiễu mẫu. Do đó, chúng tồn tại trong tập huấn luyện hơn là trong bộ kiểm tra, bất kể chúng có đư ợc rút ra từ cùng một dữ liệu hay không phân phối. Điều này dẫn đến tình trạng quá khớp, có thể đư ợc giảm thiểu bằng cách chính quy hóa. Mặc dù phư ơng pháp dễ nhất và phổ biến nhất để giảm tình trạng quá khớp là tăng cư ờng dữ liệu (LeCun et al., 1998; Simard et al., 2003; Ciresan và cộng sự, 2011, 2012; Krizhevsky và cộng sự, 2012; Montavon và cộng sự, 2012; Trư ờng trò chuyện

et al., 2014), nó đòi hỏi một dấu chân bộ nhớ lớn hơn và có giá cao hơn chi phí tính toán (Szegedy, Liu và cộng sự, 2014). Hơn nữa, bất chấp các hiệu ứng chính quy hóa của một số phương pháp đa dạng khác, bao gồm L1 và L2 chính quy hóa, dừng đào tạo sớm, nhóm ngẫu nhiên (Zeiler & Fergus, 2013), các hàm kích hoạt duy nhất (He et al., 2015a; Xu et al., 2015), mô hình trung bình (Goodfellow và cộng sự, 2013; Srivastava và cộng sự, 2014), các hàm mất mát mới (Liu, Wen và cộng sự, 2016) và chia sẻ trọng số mềm (Nowlan & Hinton, 1992), ứng dụng thành công của Dropout (Hinton et al., 2012; Srivastava et al., 2014) đến DCNN (Krizhevsky et al., 2012) đã dẫn đến việc sử dụng rộng rãi của nó và truyền cảm hứng cho nhiều cải tiến. Tiếp theo chúng tôi cung cấp mô tả chính thức về Dropout và thảo luận về một số biến thể của nó. Chúng tôi cũng giới thiệu một số của những phát triển chính quy hóa mới nhất có thể được sử dụng kết hợp với Dropout.

Bằng chứng

5.4.1 Dropout. Trong Dropout (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012; Srivastava và cộng sự, 2014), mỗi đơn vị đầu ra của một lớp là được giữ lại với xác suất p; nếu không, nó được đặt thành 0 với xác suất 1 p, với 0,5 là giá trị chung của p (Krizhevsky và cộng sự, 2012; Hinton và cộng sự, 2012). Khi Dropout được áp dụng cho một lớp được kết nối đầy đủ của DCNN (hoặc bất kỳ DNN), đầu ra của lớp r = [r1,r2,,...,rd] T, có thể được biểu thị như sau

$$r = m\, a\, (W\, v\,), \tag{5.21}$$

Chư a sửa

trong đó biểu thị tích từng phần tử giữa một vectơ mặt nạ nhị phân m và tích ma trận giữa vectơ đầu vào v = [v1, v2,,..., vn] T Và ma trận trọng số W (với các kích thước d × n), theo sau là một hàm kích hoạt phi tuyến tính, a. Trong phương trình 5.16, vectơ mặt nạ nhị phân có kích thước d và mỗi phần tử j được rút ra độc lập từ phân phối Bernoulli(p) mj, trong khi các sai lệch được bao gồm trong W và được cố định ở một để đơn giản hóa (Wan et al., 2013). Lợi ích chính của Dropout là khả năng đã được chứng minh của nó để giảm tình trạng quá khớp bằng cách ngăn chặn hiệu quả sự đồng thích ứng của tính năng (Hinton et al., 2012); nó cũng có khả năng đạt được mức trung bình của mô hình (Goodfellow et al., 2013; Srivastava et al., 2014). Hơn nữa, đối với những cải tiến khác nhau và Các biến thể bỏ học được thảo luận bên dưới, Wager, Wang và Liang (2013) đã nêu bật các đặc điểm chính quy hóa thích ứng của nó; hiệu quả và các đặc điểm học tập tổng hợp của nó đã được Warde-Farley, Goodfellow, Courville và Bengio (2013), trong khi Baldi và Sadowski (2013, 2014) đã cung cấp một phân tích toán học chi tiết về các đặc tính tĩnh và động của nó và mô tả các đặc tính trung bình của nó đối với DNN bằng đệ quy chính thức phương trình.

5.4.1.1 Dropout nhanh. Mặc dù Dropout có những ưu điểm nổi bật (Hinton et al., 2012; Srivastava et al., 2014), việc lấy mẫu hoặc đào tạo thực tế của nhiều mô hình làm cho việc đào tạo chậm hơn. Hơn nữa, trong trường hợp

dữ liệu không trùng lặp, tùy thuộc vào cách dữ liệu được lấy mẫu, hiệu quả đào tạo có thể bị giảm thêm. Để xoa dịu những lo ngại này nhưng vẫn đạt được lợi thế của đào tạo Dropout mà không cần thực sự lấy mẫu và do đó sử dụng tất cả dữ liệu một cách hiệu quả, Wang và Manning (2013) đã đề xuất Fast Dropout. Đào tạo Dropout nhanh được thực hiện bằng cách lấy mẫu từ hoặc tích hợp với phép tính gần đúng Gaussian, được chứng minh bằng trung tâm định lý giới hạn và bằng chứng thực nghiệm. Cụ thể, khi Fast Dropout là tích hợp với tổn thất softmax thường được sử dụng, tổn thất có thể được tính toán theo hàm mất mát sau:

$$L = E_{s \sim N(\mu,\sigma)} \sum_{i=1}^{|y|} t_i \log(\text{softmax}(S)_i), \tag{5.22}$$

trong đó các mẫu được lấy trực tiếp từ phép xấp xỉ Gaussian đầu vào, với $S \in \mathbb{R}^{|y|}$, và tập hợp y biểu diễn tất cả các dự đoán có thể. Nhanh Dropout cũng có thể được tích hợp với mất bản lề theo truyền thống liên quan với SVM (xem phần 5.3.4) và kỹ thuật Maxout (Goodfellow et al., 2013) và đã đưa ra những kết quả đầy hứa hẹn về hồi quy, phân loại tài liệu và quan trọng nhất là tốc độ tăng đáng kể trong các tác vụ phân loại hình ảnh được đánh giá chuẩn trên CIFAR-10 (Krizhevsky, 2009) và MNIST bộ dữ liệu (LeCun et al., 1998). Mặc dù đào tạo với backpropagation là có thể thực hiện được với một số hạn chế nhất định (xem phần 5.4.3), vẫn cần nghiên cứu thêm để khẳng định những lợi ích và hạn chế của nó khi áp dụng cho các kiến trúc DCNN khác nhau.

5.4.1.2 Dropout thích ứng (Nổi bật). Kể từ Dropout (Hinton và cộng sự, 2012) sử dụng một xác suất không đổi để thả ngẫu nhiên các đơn vị, có thể hình dung rằng ngay cả các đơn vị có thể đưa ra dự đoán chắc chắn về sự có mặt hoặc vắng mặt của một tính năng cũng sẽ bị loại bỏ 50% (nếu p = 0,5) về thời gian. Có động cơ để cải thiện điều này, Ba và Frey (2013) đã trình bày một biến thể Dropout, được gọi là Nổi bật, trong đó một mạng lưới niềm tin nhị phân chia sẻ các tham số với mạng lưới sâu tính toán xác suất Dropout cho mỗi đơn vị ẩn.
Cụ thể hơn, xác suất bỏ học là thích ứng và không giống như trong Dropout chuẩn, trong đó hoạt động của đơn vị bị che khuất bởi phân phối Bernoulli(p) mj, với xác suất 0,5, trong Standout, nó phụ thuộc vào các hoạt động đầu vào

$$P(m_j = 1 | \{a_i : i < j\}) = f\Big(\pi_{j,i} a_i\Big)_{i:i<j}, \tag{5.23}$$

trong đó trọng số từ đơn vị i đến đơn vị j trong mạng bỏ học thích ứng là được biểu thị bằng $\pi_{j,i}$ và $f(\cdot)$ là một hàm sigmoid, với $f : \mathbb{R} \to [0, 1]$. Với phương pháp này, các đơn vị đưa ra dự đoán chắc chắn về sự hiện diện của

các tính năng có khả năng được giữ lại cao hơn và ngược lại. Công trình thực nghiệm của bài báo về các chuẩn mực phân loại phổ biến không bao gồm các thí nghiệm trên DCNN; tuy nhiên, vì Standout được thiết kế để hoạt động với backpropagation, sử dụng gradient descent ngẫu nhiên, nên có vẻ hợp lý khi kết hợp nó vào kiến trúc DCNN cho các ứng dụng liên quan đến phân loại hình ảnh.

5.4.1.3 Dropout đa thức và Dropout tiến hóa. Khẳng định rằng Dropout chuẩn (Hinton và cộng sự, 2012) dẫn đến sự hội tụ không tối ưu và hợp lý hơn khi sử dụng các xác suất lấy mẫu đa thức không đồng nhất cho các nơ-ron khác nhau và các đặc điểm liên quan của chúng, Li, Gong và Yang (2016) đã đề xuất dropout đa thức mới. Cụ thể hơn, để xác định các xác suất Dropout tối ưu thay vì kỹ thuật ban đầu xác định chúng một cách độc lập và giống hệt nhau và để biện minh cho việc áp dụng lấy mẫu đa thức vào các hệ thống học nông, họ đã chính thức thiết lập một giới hạn rủi ro cho tối ưu hóa ngẫu nhiên với dropout đa thức. Điều này cho phép họ đạt được, bằng cách giảm thiểu một yếu tố phụ thuộc vào lấy mẫu từ giới hạn rủi ro, một dropout phụ thuộc vào phân phối.

Dropout phụ thuộc vào phân phối này đòi hỏi xác suất lấy mẫu dựa trên thống kê bậc hai của phân phối dữ liệu. Dựa trên dropout phụ thuộc vào phân phối đa thức này, họ đề xuất một phiên bản Dropout hiệu quả như ng thích ứng được gọi là Evolutional Dropout, với mục tiêu giải quyết vấn đề học sâu về sự dịch chuyển biến phụ thuộc nội bộ, được thảo luận thêm trong phần 5.5.3.

Xác suất bỏ học đối với Bỏ học tiến hóa có thể được tính bằng biểu thức sau đây,

$$_{\text{xin}} = \frac{\sqrt{\frac{1}{\text{tôi}} \quad \text{mj} =1[\text{Xl}_{\text{j}}]^2}}{\text{di} \quad \sqrt{\frac{1}{\text{tôi}} \quad \text{mj} =1[\text{Xl}_{\text{j}}]^2}}, \text{tôi} = 1,\ldots, \text{d}, \tag{5.24}$$

trong đó xác suất pl tiến hóa khi phân phối của các lớp đầu ra tiến hóa (do đó, tên là Evolution), và đầu ra của lớp thứ l, đối với một lô nhỏ gồm m ví dụ được biểu thị bằng Xl = (Xl ,..., Xl m). Trên các chuẩn phân loại hình ảnh phổ biến như MNIST (LeCun và cộng sự, 1998), CIFAR-10 và CIFAR-100 (Krizhevsky, 2009) và SVHN (Netzer và cộng sự, 2011), họ đã cung cấp bằng chứng thực nghiệm rằng các bổ sung Dropout mới nhất này dẫn đến sự hội tụ nhanh hơn và lỗi thử nghiệm nhỏ hơn khi so sánh với Dropout nguyên bản, ủng hộ nhu cầu cần phải điều tra thêm.

5.4.1.4 Spatial Dropout. Trong một ứng dụng định vị đối tượng sử dụng DCNN, các tác giả nhận thấy rằng việc áp dụng Regular Dropout trước một lớp tích chập 1 × 1 (xem bài báo để biết kiến trúc chi tiết) làm tăng thời gian đào tạo như ng không ngăn chặn được tình trạng quá khớp. Do đó, họ đề xuất Spatial Dropout

(Tompson, Goroshin, Jain, LeCun, & Bregler, 2015). Cụ thể, đối với một tenxơ đặc trưng tích chập, với các kích thước n f × chiều cao × chiều rộng, chúng chỉ thực hiện n f eats Dropout và sử dụng toàn bộ bản đồ tính năng để mở rộng Giá trị bỏ qua. Sau đó, các điểm ảnh liền kề trong bản đồ tính năng bỏ qua hoặc là tất cả đều bằng không (bị bỏ qua) hoặc tất cả đều hoạt động. Kết quả ban đầu cho thấy Spa-tial Dropout rất phù hợp với một tập dữ liệu có số lượng đào tạo nhỏ mẫu, do đó làm cho nó trở thành một ứng cử viên tốt để giảm quá mức cho các mẫu nhỏ hơn tập dữ liệu, trong đó khái quát hóa thường là một vấn đề. Hơn nữa, mặc dù phương pháp này đã chứng minh được kết quả khả quan trong việc ước tính tư thế con người và chuyển động khớp, cần có thêm nhiều nghiên cứu nữa cho các nhiệm vụ phân loại cụ thể. Đặc biệt, việc ứng dụng nó vào phân loại chi tiết có vẻ khả thi, vì kỹ thuật này khôi phục thông tin bị mất trong quá trình gộp mà không làm mất đi lợi ích tính toán đạt được thông qua gộp.

5.4.1.5 Dropout lồng nhau. Để tìm hiểu các biểu diễn có thứ tự của dữ liệu trong đó các chiều khác nhau có mức độ quan trọng khác nhau, sao cho thông tin chứa trong mỗi chiều của biểu diễn giảm dần theo một hàm của chỉ số chiều theo một hàm phân rã được xác định trước, Rippel và cộng sự (2014) đã đề xuất Nested Dropout. Nested Dropout rút ngẫu nhiên các chỉ số đơn vị từ một phân phối hình học. Thay vì thả các đơn vị một cách độc lập với một xác suất được xác định trước, như trong Dropout tiêu chuẩn (Hinton và cộng sự, 2012) phương pháp này bỏ qua tất cả các đơn vị theo sau số được rút ra. Cụ thể hơn, đối với không gian biểu diễn có chiều K, phân phối pB(·) được xác định trên tập con chỉ số biểu diễn Sb = {1,..., b}, b = 1,...,K có đặc điểm là nếu đơn vị thứ j xuất hiện trong một mặt nạ, sau đó tất cả các đơn vị trước đó 1,..., j 1, cũng làm như vậy, do đó cho phép đơn vị thứ j dựa vào chúng. Do đó, trong khi Dropout thực thi phân phối trên mỗi đơn vị riêng lẻ trong một mô hình, Nested Dropout chỉ định một phân phối trên các tập hợp con lồng nhau của các đơn vị biểu diễn. Lấy cảm hứng từ ứng dụng của nó để bộ mã hóa tự động không giám sát (Rippel và cộng sự, 2014), Finn và cộng sự (2015) đã sử dụng nó để đào tạo, theo phương pháp truyền ngược tiêu chuẩn, các DCNN nhỏ gọn có thể thích ứng với các tác vụ và độ phức tạp của dữ liệu khác nhau.

5.4.1.6 Max Pooling Dropout. Dropout ban đầu được thiết kế để hoạt động trên các lớp được kết nối đầy đủ của kiến trúc sâu (Krizhevsky và cộng sự, 2012; Hinton et al., 2012; Wan et al., 2013), với ít sự chú ý được dành cho các các lớp. Được thúc đẩy bởi điều này, công trình thực nghiệm của Wu và Gu (2015) đã phát hiện ra rằng tác động của Dropout lên các lớp pooling tối đa của DCNN là tương đương để chọn ngẫu nhiên một kích hoạt dựa trên phân phối đa thức tại thời gian đào tạo, có bản chất tương tự như việc gộp ngẫu nhiên (Zeiler & Fer-gus, 2013). Do đó, để có được phép tính gần đúng chính xác hơn về việc tính trung bình tất cả các đơn vị Dropout có thể, họ đã đề xuất một sơ đồ gộp có trọng số xác suất thay vì nhóm tối đa thường được sử dụng. Đối với chương trình đề xuất, hoạt động gộp của tất cả các hoạt động trong mỗi vùng được tính toán bằng

$$\text{một}_j^{(l+1)} = \underset{\text{Mà còn\_}}{\overset{N}{(l)}} = \underset{\text{tôi} = 0}{} \quad \underset{\text{Mà còn\_}}{\overset{N}{(l)}, \text{tôi}} \quad R_{j\,,}^{(l)} \tag{5.25}$$

trong đó vùng gộp j ở lớp l được biểu diễn bởi $R_{j\,,}^{(l)}$ và pi là xác suất được tính toán bởi

$$\Pr(a_{ji}^{(l+1)} = \text{một}^{(l)} = \text{pi} = \text{pqn} \quad i\ , (i = 1, 2, \ldots, n), \tag{5.26}$$

trong đó p là thuộc tính giữ lại, q = 1 p là xác suất bỏ học, và i là một chỉ số trong phân phối đa thức. Sơ đồ đề xuất có khả năng điều chỉnh và trung bình mô hình tương tự như hiệu ứng của Dropout (Srivastava et al., 2014) và Maxout (Goodfellow et al., 2013). Đối với các nhiệm vụ phân loại trên MNIST (LeCun et al., 1998) và CIFAR-10 và CIFAR-100 (Krizhevsky, 2009) tập dữ liệu, phương pháp của họ vượt trội hơn nhóm tối đa và nhóm tối đa được chia tỷ lệ, trong khi họ cũng phát hiện ra rằng Dropout trên các lớp nhóm tối đa hoạt động tốt hơn kỹ thuật nhóm ngẫu nhiên (Zeiler & Fergus, 2013), được thảo luận trong phần 5.1.2.2.

5.4.2 DropConnect. Một khái quát phổ biến khác của Dropout (Hinton et al., 2012) là DropConnect (Wan et al., 2013), thay vì ngẫu nhiên loại bỏ một tập hợp con các kích hoạt, như trong Dropout truyền thống, loại bỏ ngẫu nhiên một tập hợp con của các trọng số có xác suất 1 p. Giống như Dropout, DropCon-nect phù hợp với các lớp DNN được kết nối đầy đủ (bao gồm cả DCNN) chỉ. Trước đây, sử dụng cùng ký hiệu như phương trình 5.16, đầu ra của Lớp DropConnect có thể được thể hiện như sau

$$r = a((M \ W)v\ ), \tag{5.27}$$

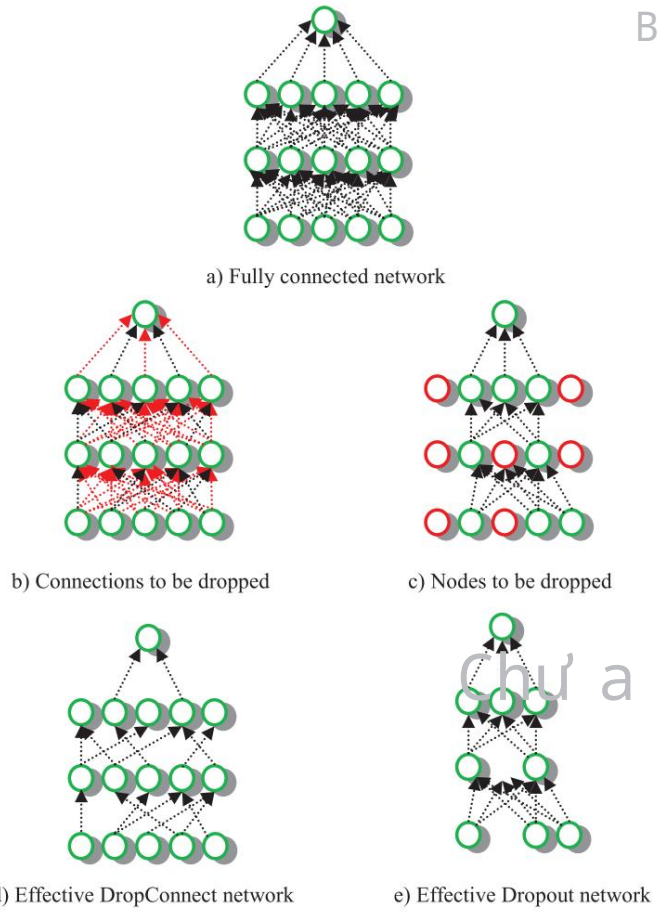
trong đó M là ma trận mặt nạ nhị phân mã hóa thông tin kết nối của các trọng số, được rút ra từ phân phối Bernoulli (p) mi j. Trong quá trình đào tạo Trong quá trình này, mỗi phần tử của mặt nạ được vẽ độc lập cho từng mẫu, do đó tạo ra kết nối khác nhau cho mỗi ví dụ được quan sát.
Hơn nữa, trong quá trình này, những thành kiến cũng được che giấu. Trong suy luận, các mẫu được rút ra từ phép xấp xỉ Gaussian 1D thông qua thời điểm khớp và được trung bình hóa và trình bày cho lớp tiếp theo sau được truyền qua lớp kích hoạt.

Bài báo DropConnect đã so sánh hiệu suất phân loại hình ảnh của Dropout và DropConnect trên một số chuẩn phân loại phổ biến và thấy rằng DropConnect vượt trội hơn Dropout trên MNIST (LeCun et al., 1998), CIFAR-10 (Krizhevsky, 2009) và SVHN (Netzer et al., 2011) tập dữ liệu, trong khi trên tập dữ liệu NORB (LeCun et al., 2004), Dropout tạo ra kết quả tốt hơn. Hơn nữa, khi họ kết hợp các DCNN khác nhau được đào tạo với DropConnect thành các nhóm, được thúc đẩy bởi

chương trình bỏ phiếu do Ciresan và cộng sự trình bày (2012), họ đã đạt được những kết quả tiên tiến mới trên các tập dữ liệu SVHN, MNIST và NORB và CIFAR-10, lần lượt, đánh bại kết quả tốt nhất trước đó của Zeiler và Fergus (2013), Ciresan et al. (2012), và Snoek et al. (2012). Mặc dù đã có một số DCNN hiệu suất cao kể từ đó (xem Bảng 5), mô hình DropConnect, sử dụng tăng cường dữ liệu, vẫn giữ kỷ lục về lỗi phân loại thấp nhất trên chuẩn mực MNIST nổi tiếng. Kể từ khi công trình thực nghiệm của Wan et al. (2013) chỉ được thực hiện trên các tập dữ liệu nhỏ, Smirnov, Tim-oshenko và Andrianov (2014) đã mở rộng so sánh và thấy rằng Dropout cung cấp khả năng điều chỉnh tốt hơn DropConnect ILSVRC 2013 lớn hơn (Russakovsky và cộng sự, 2015). Hình 11 minh họa sự khác biệt giữa mạng được kết nối hoàn toàn truyền thẳng mà không có Dropout, với Dropout và với DropConnect. Như minh họa, trong mạng DropConnect, các kết nối với trọng số liên quan của chúng bị loại bỏ ngẫu nhiên chứ không phải là các nút.

5.4.3 Những tiến bộ gần đây về chính quy hóa. Tiếp theo là chính quy hóa mô hình kỹ thuật được đề cập trong phần giới thiệu của phần này, một phần khác chủ yếu giải pháp thay thế chưa được khám phá là điều chỉnh phân phối đầu ra của DCNN. Một của các dấu hiệu của việc quá khớp là khi một mô hình gán tất cả các xác suất lớp cho một lớp đơn lẻ từ tập huấn luyện. Những ước tính tự tin này thường giống với phân phối đầu ra entropy thấp. Để giải quyết vấn đề này, Szegedy, Vanhoucke et al. (2015) đã giới thiệu quy tắc làm mịn nhãn (LSR), duy trì tỷ lệ thực tế giữa các xác suất logarit không chuẩn hóa (logit) của các lớp sai bằng cách ước tính, trong quá trình đào tạo, hậu quả cận biên của việc bỏ nhãn. Điều này ngăn cản mô hình phân bổ một khả năng cho mỗi trường hợp đào tạo. Kỹ thuật LSR có thể được coi là tương đương với việc thay thế một tổn thất entropy chéo đơn lẻ bằng một cặp tổn thất, thứ hai trong đó xem xét phân phối trước đó và phạt độ lệch của nhãn dự đoán liên quan đến nó. Lấy cảm hứng từ hiệu ứng chính quy hóa của LSR, phân phối đầu ra tự tin cũng bị phạt bởi Pereyra, Tucker, Chorowski, Kaiser và Hinton (2017), những người đưa ra hình phạt về lòng tin dựa trên entropy tối đa được bổ sung bởi nhãn đồng nhất và nhãn đơn vị làm mịn. Kỹ thuật của họ cải thiện một số mô hình hiện đại trên một nhiều loại nhiệm vụ khác nhau, bao gồm phân loại hình ảnh. Gần đây một kỹ thuật điều chỉnh đầu ra, thêm nhiễu vào lớp đầu ra, là cũng được đề xuất (Xie, Wang, Wei, Wang, & Tian, 2016), ám chỉ một khả năng xu hướng mới để giải quyết tình trạng quá khớp.

5.4.4 Phân tích và triển vọng. Một cách thích hợp để chuẩn hóa một mô hình là tính trung bình kết quả từ một số mạng khác nhau; tuy nhiên, đối với DC-NN lớn, tài nguyên tính toán cần thiết để thực hiện việc này sẽ rất lớn. Điều này dẫn đến việc trình bày Dropout (Hinton et al., 2012), cung cấp một phương tiện để hợp nhất một số lượng lớn các DCNN theo cấp số nhân một cách hiệu quả cách (Hinton et al., 2012; Goodfellow et al., 2013; Srivastava et al., 2014),



Hình 11: Sự khác biệt giữa các lớp mạng được kết nối đầy đủ mà không có Dropout (a), với Dropout (b, d) và với DropConnect (c, e).

và điều này đã góp phần vào nhiều thành công thực nghiệm thúc đẩy các nhà nghiên cứu phát triển kỹ thuật hơn nữa và điều tra cũng như giảm thiểu nó sự thiếu sót.

Để giải quyết tình trạng thiếu hiệu quả trong đào tạo liên quan đến tiêu chuẩn Dropout, một phương pháp Dropout nhanh (Wang & Manning, 2013) có khả năng cung cấp cho DCNN tốc độ tăng đáng kể trong quá trình đào tạo và suy luận, cùng với sự ổn định hơn, đã được đề xuất. Tuy nhiên, nhược điểm của điều này cách tiếp cận là đào tạo trong quá trình truyền ngược của backpropagation phức tạp hơn, trong khi đường chuyền về phía trước vẫn đơn giản.

giải pháp khả thi để giảm thiểu điều này là đào tạo DCNN với Dropout chuẩn
để đơn giản và chỉ sử dụng Fast Dropout trong quá trình suy luận để tăng tốc độ tính toán;
vẫn cần phải làm việc thêm để đơn giản hóa quy trình.
Standout (Ba & Frey, 2013) được triển khai để giảm nguy cơ loại bỏ các đơn vị ping đưa ra
dự đoán chắc chắn về các tính năng, nhưng cũng giống như
Dropout nhanh và Dropout không gian (Tompson và cộng sự, 2015), mặc dù chúng
đã cho thấy kết quả đầy hứa hẹn cho các kiến trúc và nhiệm vụ khác, vẫn cần phải tiếp tục
điều tra tập trung vào ứng dụng của chúng vào các nhiệm vụ phân loại hình ảnh do DCNN dẫn
đầu. Đặc biệt, sử dụng chúng để phân biệt
tinh chỉnh, sau khi chuyển giao kiến thức là một hướng thú vị đòi hỏi phải điều tra thêm.
Sự bỏ học tiến hóa (Li et al., 2016), có khả năng thích ứng
các đặc điểm tương tự như Standout, có thể cải thiện các đặc điểm hội tụ và cải thiện hiệu
suất phân loại của DCNN; tuy nhiên,
giống như với Standout và Max pooling Dropout (Wu & Gu, 2015), các phép tính xác suất bổ
sung làm tăng thêm gánh nặng tính toán của
hệ thống sử dụng chúng. Trong khi DropConnect (Wan et al., 2013) cho phép
đào tạo các mô hình lớn mà không cần quá phù hợp, nó chậm hơn các mô hình
sử dụng Dropout hoặc không Dropout.

Vì vậy, từ phân tích này, có thể kết luận rằng mặc dù có kết quả thực nghiệm đầy hứa hẹn
của một số biến thể bỏ học được mô tả trong
phần, nghiên cứu sâu hơn để thiết lập chúng một cách chắc chắn vẫn còn cần thiết. Đặc biệt,
kỹ thuật này có thể được hưởng lợi nhiều nhất từ những cải tiến tiếp theo tập trung vào
về việc giảm chi phí tính toán của các hệ thống sử dụng nó. Hơn nữa,
cho rằng trí thông minh thực sự có tính thích nghi cao trong tự nhiên, người ta dự đoán rằng
công việc trong tương lai sẽ kết hợp các đặc điểm sinh học thích ứng, tương tự như Standout
và Sự bỏ học tiến hóa. Về mặt lý thuyết, để bổ sung cho công việc
bởi Wager et al. (2013), Wade-Farley et al. (2013) và quan trọng nhất là Baldi
và Sadowski (2013, 2014), và thúc đẩy hơn nữa việc áp dụng Dropout,
phân tích lý thuyết sâu hơn để chứng minh lý do cho những thành công của nó vẫn còn cần
thiết. Đặc biệt, các tính chất khái quát của nó vẫn chưa được chứng minh với
độ chính xác toán học có thể chấp nhận được. Trên thực tế, tất cả các biến thể của nó sẽ được hưởng lợi từ
sự giám sát như vậy. Một hướng đi đầy hứa hẹn khác liên quan đến việc phân tích sâu hơn, tóm lại
được Baldi và Sadowski (2014) đề cập đến, là để điều tra tính hai mặt và
kết nối giữa các tế bào thần kinh đột biến hoặc ngẫu nhiên và Dropout, vì có
là một khả năng mà những điều này có thể được sử dụng trong quá trình học tập để hoàn thành
cùng mục tiêu của kỹ thuật chính quy hóa mang tính đột phá này.

Mặc dù các kỹ thuật điều chỉnh đầu ra vẫn còn trong giai đoạn sơ khai đối với
DCNNs, kết quả ban đầu rất khả quan và cho rằng các kỹ thuật
được giới thiệu nói chung là phù hợp và trực giao với nhiều phương pháp chính quy hóa khác
như Dropout, công việc tiếp theo theo hướng này là
động viên.

5.5 Kỹ thuật tối ưu hóa. Trong phần này, chúng tôi đánh giá một số
các kỹ thuật tối ưu hóa quan trọng, sau khi đầu tiên kiểm tra yêu cầu của họ
các phương pháp.

5.5.1 Học dựa trên Gradient. Trong DCNN được giám sát hoàn toàn, hàm mất mát, trong hầu hết các trường hợp là mất mát softmax (Krizhevsky et al., 2012; Good-fellow et al., 2013; Lin et al., 2013; Zeiler & Fergus, 2013, 2014; Simonyan & Zisserman, 2014; Chatfield và cộng sự, 2014; Szegedy, Liu và cộng sự, 2014; Szeged, Vanhoucke et al., 2015; He et al., 2015a, 2015b), thường được giảm thiểu bằng cách sử dụng một số dạng giảm dần độ dốc ngẫu nhiên (SGD; Bottou, 1998, 2010). Đối với kỹ thuật này, độ dốc được đánh giá bằng thuật toán truyền ngược phổ biến (Ciresan et al., 2011; Krizhevsky et al., 2012; Wan et al., 2013; Goodfellow và cộng sự, 2013; Simonyan & Zisserman, 2014; Zeiler & Fergus, 2014; Szegedy, Vanhoucke và cộng sự, 2015; Szegedy, Liu và cộng sự, 2014; Ông và cộng sự, 2015b; Srivastava và cộng sự, 2015a; Choromanska, Henaff, Mathieu, Arous và LeCun, 2015). Trong khi phương pháp giảm dần độ dốc được phổ biến bởi Rumelhart và cộng sự (1986), đã được sử dụng trong nhiều CNN đầu tiên (LeCun et al., 1989a, 1989b; LeCun et al., 1998), tăng kích thước dữ liệu—hãy xem xét MNIST (LeCun et al., 1998) so với ILSVRC (Russakovsky và cộng sự, 2015)—và tính toán liên quan của nó sự phức tạp đã dẫn đến sự phổ biến của SGD, là sự đơn giản hóa to lớn của phương pháp truyền thống (Bottou, 2010). Thay vì tính toán chính xác độ dốc, một mẫu được chọn ngẫu nhiên duy nhất (trên thực tế, một lô mẫu nhỏ) được sử dụng để ước tính nó cho mỗi lần lặp lại, do đó tạo ra quá trình tự nhiên là ngẫu nhiên. Điều quan trọng là SGD có thể xử lý các ví dụ trực tuyến (hoặc khi đang di chuyển) vì nó không cần phải nhớ lại những ví dụ nào đã được quan sát thấy trong các lần lặp lại trước đây (Choromanska và cộng sự, 2015). Hơn nữa, SGD chuẩn có thể được triển khai song song, trên nhiều GPU, để tối ưu hóa hơn nữa và cải thiện tốc độ xử lý, đặc biệt là đối với các ứng dụng học máy quy mô lớn (Zinkevich, Weimer, Li, & Smola, 2010; Recht, Re, Wright, & Niu, 2011; Dean và cộng sự, 2012; Trang, Tần, John, & Lâm, 2013; Bengio, 2013; Paine, Jin, Yang, Lin và Huang, 2013). Trong khi chúng tôi đã đưa ra một giới thiệu ngắn gọn ở đây, thông tin chi tiết hơn có thể được thu thập từ có rất nhiều tài liệu có sẵn về kỹ thuật tối ưu hóa này. Đặc biệt, Bottou (1998, 2010) trình bày một phân tích chi tiết về SGD; Qian (1999), Zeiler (2012), Duchi, Hazan, và Singer (2011), và Kingma và Ba (2014) trình bày các thuật toán tối ưu hóa dựa trên độ dốc giảm dần khác, trong khi một số giải pháp thay thế này được Sutskever, Martens, khảo sát và so sánh Dahl và Hinton (2013). Mặc dù sử dụng nhiều kỹ thuật tối ưu hóa dựa trên gradient cho DCNN và phân loại hình ảnh nói chung, câu hỏi vẫn còn là liệu các thuật toán này có bị lỗi nội tại hay không, dẫn đến một số của những thách thức đã biết với DCNN (xem phần 6). Do đó, cần tiếp tục làm việc để hiểu được hoạt động bên trong của các mô hình của chúng tôi và đặc biệt là các kỹ thuật tối ưu hóa của chúng tôi vẫn cần phải được thực hiện.

5.5.2 Các chương trình khởi tạo nâng cao. Khởi tạo kém các tham số DCNN, thường là hàng triệu (Krizhevsky và cộng sự, 2012; Si-monyan & Zisserman, 2014; Taigman và cộng sự, 2014; Szegedy, Liu và cộng sự, 2014), và đặc biệt là trọng lượng của chúng có thể cản trở quá trình đào tạo vì vấn đề biến mất/bùng nổ gradient (Bengio et al., 1994), và cản trở

sự hội tụ. Do đó, việc khởi tạo chúng cực kỳ quan trọng (Sutskever et al., 2013; Simonyan & Zisserman, 2014; He và cộng sự, 2015a; Mishkin & Matas, 2016). Ở đây chúng tôi giới thiệu tóm tắt một số lược đồ khởi tạo tiêu biểu. Saxe, McClelland và Ganguli (2013); Sussillo và Abbott (2014), Hin-ton, Vinyals và Dean (2015), Romero và cộng sự. (2015) và Srivastava (2015a , 2015). 2015b) có thể được tham khảo để biết các kỹ thuật phù hợp khác.

5.5.2.1 Khởi tạo Xavier. Glorot và Bengio (2010) đã đánh giá cách các gradient và kích hoạt lan truyền ngược thay đổi trên các lớp khác nhau; dựa trên những cân nhắc này, họ đề xuất một khởi tạo chuẩn hóa sơ đồ về cơ bản áp dụng phân phối đồng đều cân bằng cho trọng lượng khởi tạo (He et al., 2015a). Đối với sơ đồ khởi tạo này, ban đầu trọng số được rút ra từ phân phối đồng đều hoặc phân phối Gauss, với giá trị trung bình bằng không và phương sai chính xác. Như Glorot và Bengio (2010) đã khuyến nghị, có thể sử dụng phương sai sau:

$$\text{Var}(W\_\text{Khởi tạo}) = \frac{2}{nx + ny}, \tag{5.28}$$

trong đó W_Init biểu diễn phân phối của một nơ-ron cụ thể khi khởi tạo, nx là số lượng tế bào thần kinh đưa vào phương sai và ny biểu thị số lượng tế bào thần kinh được cung cấp bởi đầu ra của nó. Do đó, đối với kỹ thuật ban đầu, số lượng neuron đầu vào và đầu ra kiểm soát mức độ khởi tạo. Sau đó, sơ đồ được gọi là khởi tạo “Xavier” và được Jia và cộng sự (2014) đơn giản hóa để dễ triển khai hơn, trong đó phương sai là được rút ra từ một phân phối có giá trị trung bình bằng không và phương sai được tính toán bởi biểu thức sau:

$$\text{Var}(W\_\text{Khởi tạo}) = \frac{1}{nx}. \tag{5.29}$$

Khởi tạo Xavier thúc đẩy sự lan truyền tín hiệu sâu vào DNN (bao gồm DCNN) và đã được chứng minh là dẫn đến sự hội tụ nhanh hơn đáng kể (Glorot & Bengio, 2010). Hạn chế chính của nó là việc suy ra dựa trên giả định rằng các kích hoạt là tuyến tính, do đó làm cho nó không phù hợp với các kích hoạt ReLU (Nair & Hinton, 2010) và PReLU (He et al., 2015a).

5.5.2.2 Khởi tạo thích ứng được suy ra về mặt lý thuyết. Để tránh điều này, He et al. (2015a) đã đưa ra một khởi tạo có cơ sở lý thuyết hợp lý đã xem xét những kích hoạt phi tuyến tính này. Cụ thể, việc suy ra chúng, tuân theo Glorot và Bengio (2010), dẫn đến việc khởi tạo trọng số từ một phân phối chuẩn Gaussian trung bình bằng không có độ lệch chuẩn là $\sqrt{2/nl}$, trong đó n là số kết nối của phản hồi và l là chỉ số lớp.

Hơn nữa, họ khởi tạo độ lệch về 0. Họ đã đưa ra bằng chứng thực nghiệm rằng lược đồ khởi tạo này phù hợp để đào tạo các mô hình cực kỳ sâu, trong khi khởi tạo Xavier thì không.

5.5.2.3 Khởi tạo cố định tiêu chuẩn. Một phương pháp khác được phổ biến bởi Krizhevsky et al. (2012) là khởi tạo trọng số cho mỗi lớp từ một phân phối chuẩn Gaussian trung bình bằng không, với độ lệch chuẩn cố định—0,01 trong Krizhevsky et al. (2012)—và thiết lập độ lệch của các lớp khác nhau thành một hằng số một hoặc không. Tuy nhiên, trong khi Krizhevsky et al. (2012) phát hiện ra rằng sơ đồ khởi tạo này bổ sung cho các kích hoạt ReLU (Nair & Hinton, 2010) và học tập tăng tốc, những người khác đã phát hiện ra rằng đối với các mô hình rất sâu, nó cản trở sự hội tụ do độ lớn của các gradient hoặc các hoạt động ở các lớp cuối cùng (Simonyan & Zisserman, 2014; He et al., 2015a; Mishkin và Matas, 2016).

5.5.2.4 Khởi tạo phương sai đơn vị tuần tự lớp. Sau khi khởi tạo, Phương pháp khởi tạo Xavier (Glorot & Bengio, 2010) chỉ được tổng quát hóa đối với ReLUS (Nair & Hinton, 2010) của He et al. (2015a), nhưng không phải đối với các kích hoạt phi tuyến tính khác như tiếp tuyến hypebolic hoặc Maxout (Goodfellow et al., 2013) kích hoạt. Hơn nữa, lý thuyết ban đầu không bao gồm toàn bộ quang phổ của các lớp DCNN như nhóm tối đa (Ranzato et al., 2007) và chuẩn hóa phản ứng cục bộ (Krizhevsky và cộng sự, 2012). Tuy nhiên, thay vì đưa ra một công thức lý thuyết mới để bao quát tất cả những phần còn lại kích hoạt và các lớp DCNN, Mishkin và Matas (2016) đã trình bày một lược đồ khởi tạo trọng số theo dữ liệu được gọi là phương sai đơn vị tuần tự lớp (LSUV) khởi tạo. Tóm lại, họ khởi tạo trước các trọng số của các lớp tích chập và tích bên trong bằng cách điền các trọng số bằng nhiễu Gauss, với phương sai đơn vị. Tiếp theo, họ sử dụng QR hoặc phân tích giá trị đơn (SVD) để phân tích trọng số thành cơ sở trực giao, được thúc đẩy bởi Saxe và cộng sự. (2013), và sau đó thay thế chúng bằng một trong các thành phần. Sau đó, chúng ước tính phương sai đầu ra của mỗi lớp bị ảnh hưởng, từ lớp đầu tiên đến lớp cuối cùng, chuẩn hóa trọng số sao cho phương sai bằng một. Trên ImageNet (Russakovsky và cộng sự, 2015), CIFAR-10 (Krizhevsky, 2009) và MNIST (LeCun et al., 1998), kỹ thuật khởi tạo nhanh của họ đã tạo điều kiện thuận lợi cho đào tạo DCNN và dẫn đến kết quả tương đương với công nghệ tiên tiến, vượt trội hơn các hệ thống tinh vi khác (Srivastava và cộng sự, 2015a; Romero et al., 2015) cũng được thiết kế để tối ưu hóa.

5.5.2.5 Phân tích và triển vọng. Để thoát khỏi sự biến mất/bùng nổ gradient vấn đề (Bengio et al., 1994) và do đó thúc đẩy sự hội tụ mạng lưới và để giảm thiểu những thách thức của việc đào tạo DCNN thường áp dụng với các hàm mất mát không lồi (Choromanska và cộng sự, 2015), một lượng đáng kể nghiên cứu đã đi vào việc đảm bảo khởi tạo mạng đầy đủ. Khởi tạo Xavier (Glorot & Bengio, 2010), thường được sử dụng để khởi tạo DNN, tạo điều kiện cho sự hội tụ nhanh chóng và được biết là hoạt động tốt trong nhiều

ứng dụng (Mishkin & Matas, 2016). Tuy nhiên, kể từ khi khởi tạo Xavier
không phù hợp với các kích hoạt phi tuyến tính dựa trên chỉnh lưu, nó không tốt
phù hợp với DCNN hiện đại, sử dụng chúng một cách triệt để. Hơn nữa, nó
không thúc đẩy sự hội tụ của các mạng cực kỳ sâu. Những thiếu sót này đã được giải
quyết bằng sơ đồ mạnh mẽ hơn do He et al. trình bày.
(2015a); tuy nhiên, trong các mạng cực kỳ sâu, mặc dù chúng cho thấy các đặc điểm hội
tụ được cải thiện, chúng không thể chứng minh rằng
phương pháp khởi tạo dẫn đến độ chính xác được cải thiện, có thể là do sự suy thoái (He
et al., 2015b; He & Sun, 2015; Srivastava et al., 2015a, 2015b), trong đó
cũng cản trở việc khởi tạo cố định tiêu chuẩn. Để chống lại điều này, Simonyan và
Zisserman (2014) huấn luyện trước một DCNN đủ nông để có trọng số được khởi tạo ngẫu
nhiên và sử dụng mạng này để huấn luyện các kiến trúc sâu hơn; tuy nhiên, điều này tự
nhiên đòi hỏi nhiều thời gian huấn luyện hơn, tốn kém hơn về mặt tính toán và thậm chí
có thể dẫn đến hội tụ kém, ủng hộ nhu cầu
cho các giải pháp thay thế khác. Kỹ thuật khởi tạo LSUV được đề xuất gần đây
(Mishkin & Matas, 2016) đã đưa ra những kết quả thực nghiệm đầy hứa hẹn; tuy nhiên,
bất chấp tính thực tiễn của phương pháp tiếp cận dựa trên dữ liệu được đề xuất, nó vẫn
cần phải tính toán thống kê lô, không được thiết lập trên các tập dữ liệu lớn và
đòi hỏi những thủ tục phức tạp.

Vì vậy, từ đó chúng ta có thể kết luận rằng các yếu tố chính cần xem xét khi
lựa chọn một lược đồ khởi tạo là hàm kích hoạt được sử dụng,
độ sâu của mạng, có thể cản trở độ chính xác của phân loại do
sự suy thoái, ngân sách tính toán có sẵn, kích thước của dữ liệu
đặt và độ phức tạp có thể chấp nhận được của giải pháp cần thiết. Nếu hai
các yếu tố không áp đặt các ràng buộc hệ thống, LSUV hấp dẫn (xem phần
5.5.3.2), trong khi khởi tạo cố định tiêu chuẩn vẫn là lựa chọn phổ biến cho
mạng lưới nông hơn (theo tiêu chuẩn ngày nay). Đối với mạng lưới cực kỳ sâu,
thử nghiệm với các chương trình khác, chẳng hạn như những chương trình do Simonyan trình bày
và Zisserman (2014) và He et al. (2015a), đặc biệt theo dõi
suy thoái, có thể được thực hiện. Công việc sắp tới nên tập trung vào việc thiết kế các
chương trình khởi tạo chung không chỉ tăng tốc thời gian đào tạo
và duy trì độ chính xác bằng cách giải quyết sự xuống cấp nhưng cũng có khả năng thích ứng với
các mô hình khác nhau, bất kể độ sâu và nhiệm vụ của chúng bất kể
sự phức tạp, trong khi một phân tích lý thuyết về cách các chương trình hiện tại của
chúng tôi tối ưu hóa các mô hình hiện đại của chúng tôi cũng được ủng hộ. Hơn nữa, bất chấp
các hàm ý tính toán, hiệu ứng tối ưu hóa của không giám sát
đào tạo trước, hỗ trợ công việc được thực hiện bởi Saxe et al. (2013) và
Simonyan và Zisserman (2014) cũng đòi hỏi phải thử nghiệm thêm.

5.5.3 Chuẩn hóa hàng loạt. Ngoài việc có một số lượng lớn các tham số, việc đào tạo
DCNN còn phức tạp bởi một hiện tượng được biết đến
như sự thay đổi biến phụ thuộc nội bộ, được gây ra bởi những thay đổi trong phân phối
của đầu vào của mỗi lớp do các tham số thay đổi ở lớp trước đó.
Hiện tượng này có hậu quả nghiêm trọng, bao gồm việc đào tạo chậm hơn
do tỷ lệ học tập thấp hơn, cần phải khởi tạo tham số cẩn thận,

và sự phức tạp khi đào tạo DCNN với các kích hoạt phi tuyến tính bão hòa. Để giảm hậu
quả của sự thay đổi biến phụ thuộc nội bộ, Ioffe và
Szegedy (2015) đã đề xuất một kỹ thuật được gọi là chuẩn hóa theo lô (BN).
Kỹ thuật này giới thiệu một bước chuẩn hóa, đơn giản là một phép biến đổi phi tuyến
tính được áp dụng cho mỗi lần kích hoạt, giúp sửa các giá trị trung bình và phương sai
của các đầu vào lớp. Để cho phép tích hợp với SGD (Bottou, 1998, 2010),
cũng sử dụng các lô nhỏ trong quá trình đào tạo, BN tính toán ước tính trung bình và
phương sai sau các lô nhỏ thay vì trên toàn bộ tập đào tạo.

Cụ thể, đối với một lô nhỏ B = {x1,...,n}, với kích hoạt x và kích thước n, giá
trị trung bình và phương sai của lô nhỏ đầu tiên được tính bằng μ_B
$\frac{1}{N}\sum_{j=1}^{N} x_j$ và σ_B^2 $\frac{1}{N}\sum_{j=1}^{N}(x_j - \mu_B)^2$, tương ứng. Chiều thứ j là
sau đó được chuẩn hóa theo biểu thức sau,

$$\hat{x}_j \quad \frac{x_j - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \tag{5.30}$$

trong đó là hằng số, được đưa vào để ổn định số học. Chuẩn hóa
các giá trị $\hat{x}$ sau đó được chia tỷ lệ và dịch chuyển để biểu diễn tốt hơn theo biểu
thức sau:

$$Y_j \quad \zeta \hat{x}_j + \beta \equiv BN_{\zeta,\beta}(x_j), \tag{5.31}$$

trong đó ζ và β là các tham số có thể học được. Kết quả của phép biến đổi thứ hai Y
được truyền đến các lớp khác của mạng.

5.5.3.1 Ứng dụng cho các mô hình DCNN khác và tóm tắt thực nghiệm. Lô
chuẩn hóa cho phép tỷ lệ học tập cao hơn, theo truyền thống dẫn đến
trong các vấn đề về độ dốc bùng nổ hoặc biến mất (Bengio và cộng sự, 1994), và điều
này làm tăng đáng kể thời gian đào tạo. Hơn nữa, nó có hiệu ứng chính quy hóa tương
tự như Dropout (Hinton và cộng sự, 2012; Srivastava và cộng sự, 2014), và khi
kết hợp với mô hình Inception (Szegedy, Liu và cộng sự, 2015), đã có
tăng tốc độ đào tạo đáng kể mà không làm tăng quá mức. Khi
được kết hợp thành một nhóm, DCNN được chuẩn hóa theo lô đạt được kết quả báo
cáo tốt nhất trên tập dữ liệu ImageNet (Russakovsky và cộng sự, 2015), vượt qua
tốt nhất trước đó của He et al. (2015a), vốn đã được coi là tốt hơn
hơn hiệu suất ở cấp độ con người; tuy nhiên, sau đó điều này đã được thay thế bằng
mô hình Inception được cải thiện (Szegedy, Vanhoucke và cộng sự, 2015), phần còn lại
mạng lưới của He et al. (2015b), và Inception-residual được chuẩn hóa theo lô
mạng lưới của Szegedy et al. (2016).

Trên thực tế, điều này được minh họa rõ ràng trong Bảng 4, so sánh hiệu suất
chuẩn hóa theo lô với các DCNN đại diện được chọn trên tập dữ liệu Im-ageNet
(Russakovsky và cộng sự, 2015), bắt đầu từ cuộc cách mạng
công trình của Krizhevsky et al. (2012) cho đến trạng thái nghệ thuật hiện tại (Szegedy

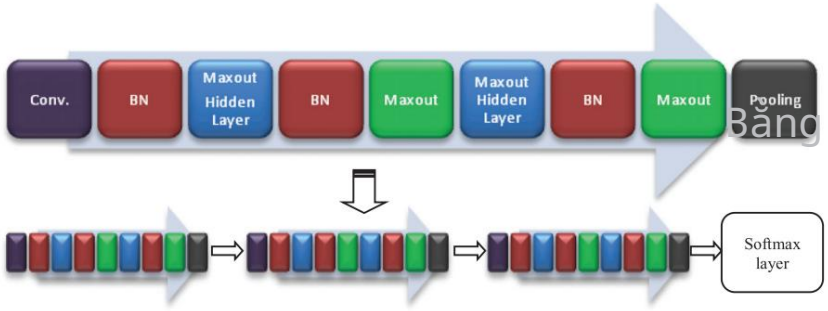
Bảng 4: Hiệu suất DCNN trên Bộ dữ liệu ImageNet.

Tên mã và tham chiếu	Đóng góp cụ thể	Top 5 Lỗi (%)
AlexNet (Krizhevsky và cộng sự, 2012)	Mô hình DCNN trên các GPU song song, những cải tiến bao gồm Dropout, tăng cường dữ liệu, ReLU và chuẩn hóa phản ứng cục bộ Phương pháp trực quan hóa mới; các lớp tích chập lớn hơn so với Krizhevsky et al., 2012 Nhóm kim tự tháp không gian để cho phép kích thước hình ảnh linh hoạt	15.3
Mạng Z&F (Zeiler & Fergus, 2014)		11.7
SPP-net (He et al., 2014)		8.06
Mạng VGG (Simonyan & Zisserman, 2014	Tăng độ sâu, nhiều lớp tích chập hơn, bộ lọc tích chập 3 × 3	7.32
GoogLeNet (Szegedy, Liu và cộng sự, 2014)	Kiến trúc khởi đầu mới, mạng rất lớn, giảm chiều	6,67
PReLU-net (Anh ấy và cộng sự, 2015a)	Các chức năng kích hoạt PReLU và mạnh mẽ sơ đồ khởi tạo	4,94
BN-Inception (Ioffe & Szeged, 2015)	Chuẩn hóa hàng loạt kết hợp với kiến trúc khởi tạo	4,82
Inception-V3 (Szeged, Vanhoucke và cộng sự, 2015)	Tích chập phân tích, giảm chiều tích cực	3,58
ResNets (Ông và cộng sự, 2015b)	Các chức năng/khối còn lại được tích hợp vào các lớp DCNN	3.57
BN-Inception-ResNet (Szegedy và cộng sự, 2016)	BN, Kiến trúc khởi đầu tích hợp với các hàm còn lại	3.08

Lưu ý: Mục nhập in đậm thể hiện tình trạng nghệ thuật hoặc tỷ lệ lỗi công bố thấp nhất được biết đến đối với tập dữ liệu. Các mục nhập in nghiêng thể hiện kết quả thu được trong cuộc thi.

et al., 2016). Đối với ILSVRC, có ba biện pháp báo cáo, tỷ lệ lỗi 1 và 5 lỗi hàng đầu, và tiêu chí phân cấp; tuy nhiên, vì số liệu 5 lỗi hàng đầu là đơn giản nhất và phù hợp nhất với chuẩn mực, nên nó đã được sử dụng độc quyền kể từ năm 2012. Các số in nghiêng biểu thị kết quả thu được trong cuộc thi, trong khi Russakovsky et al. (2015) thảo luận thêm về các kết quả, đặc biệt là từ những thành công ban đầu, cùng với phân tích chi tiết về chuẩn mực phân loại đầy thách thức này trong khảo sát toàn diện của họ.

Hơn nữa, BN cũng được kết hợp với mô hình NIN (Lin et al., 2013) và kích hoạt Maxout (Goodfellow et al., 2013) để tạo thành một mạng Maxout chuẩn hóa theo lô phức tạp trong mô-đun mạng (MIN) và mạng, được minh họa tương ứng bằng nửa trên và nửa dưới của Hình 12 (Chang & Chen, 2015). Mặc dù phức tạp, nhưng nó đã vượt trội hơn các RCNN hiệu suất cao và đạt được kết quả tiên tiến (không cần tăng cường dữ liệu) trên các chuẩn MNIST và CIFAR-100 và một trạng thái tiên tiến mới (có và không cần tăng cường dữ liệu) trên tập dữ liệu CIFAR-10 (LeCun et al., 1998; Krizhevsky, 2009; Liang & Hu, 2015). Mô hình NIN



Hình 12: MIN được chuẩn hóa theo lô (Chang & Chen, 2015).

(Lin et al., 2013) cũng được mở rộng để kết hợp chuẩn hóa theo lô (Ioffe & Szegedy, 2015) kích hoạt trước khi kích hoạt Maxout (Goodfellow et al., 2013), dẫn đến mạng Maxout trong mô hình mạng Maxout (MIM) (Liao & Carneiro, 2016), có kiến trúc và độ phức tạp tương tự như mô hình MIN do Chang và Chen (2015) trình bày. Bài báo đã chỉ ra bằng chứng thực nghiệm cho thấy việc sử dụng BN dẫn đến việc điều kiện hóa mô hình trước, cho phép việc sử dụng tốc độ học tập lớn hơn và do đó hội tụ nhanh hơn, trong khi vẫn duy trì độ chính xác.

Bảng 5 so sánh hiệu suất của một số DCNN hiện đại hoặc hiệu suất cao trên tập dữ liệu MNIST (LeCun et al., 1998). Kết quả thu được trong các khối rõ ràng là do các mô hình đơn lẻ không sử dụng tăng cường dữ liệu, trong khi kết quả màu đỏ là hiệu suất tiên tiến nhất cho cấu hình này tại thời điểm các bài báo tương ứng của họ. Như minh họa, đối với cấu hình này, MIN đã đạt được những kết quả tiên tiến, rất gần với tỷ lệ lỗi thấp nhất đạt được bởi các mô hình dựa trên tập hợp được bổ sung bằng cách tăng cường dữ liệu (các khối được tô bóng). Mặc dù có một số DCNN thành công kể từ bài báo DropConnect (Wan et al., 2013), mô hình này vẫn giữ kỷ lục về tỷ lệ lỗi được công bố thấp nhất được biết đến trên tạp chí phổ biến này chuẩn mực.

5.5.3.2 Chuẩn hóa hàng loạt. Mặc dù BN có hiệu quả trong việc tăng tốc quá trình đào tạo DCNN, nhưng hiệu quả của nó bị thách thức đối với các lô nhỏ hoặc lô nhỏ không chứa các mẫu độc lập. Khẳng định rằng những thiếu sót này là do các hoạt động được tính toán khác nhau trong quá trình đào tạo và suy luận, Ioffe (2017) đã thay thế BN bằng hàng loạt chuẩn hóa, đảm bảo rằng các đầu ra được tính toán bởi một mô hình là không phụ thuộc vào toàn bộ lô nhỏ mà phụ thuộc vào từng ví dụ riêng lẻ trong suốt quá trình đào tạo và suy luận. Cụ thể, đối với các mô hình có các lớp được chuẩn hóa theo lô, Ioffe (2017) tăng cường mô hình bằng cách áp dụng theo phép biến đổi afin kích thước, trong đó chúng giữ các tham số cố định,

Bảng 5: Hiệu suất DCNN trên Bộ dữ liệu MNIST.

Mô hình và tham khảo	Mô tả ngắn gọn	Lỗi (%)
DCNN được đào tạo trước (Ranzato et al., 2006	Đào tạo trước không giám sát dựa trên năng lượng, tiếp theo là DCNN	0,60
DCNN-NN (Jarrett và cộng sự, 2009)	Trình trích xuất tính năng CNN kép theo sau bằng mạng nơ-ron kép	0,53
FitNets (Romero và cộng sự, 2015)	Mạng lưới mỏng, sâu với các gợi ý cấp trung gian để hướng dẫn đào tạo	0,51
Phân nhóm ngẫu nhiên (Zeiler & Fergus, 2013)	Ngẫu nhiên hơn là xác định thủ tục gộp chung	0,47
NIN (Lin và cộng sự, 2013)	MLP được tích hợp vào kiến trúc DCNN	0,47
Mạng lưới Maxout (Goodfellow và cộng sự, 2013)	Hàm kích hoạt Maxout	0,45
Mạng lưới đường bộ (Srivastava và cộng sự, 2015b)	Cơ chế cổng học tập cho điều chỉnh luồng thông tin DCNN	0,45
Mạng lưới được giám sát sâu (Lee và cộng sự, 2015)	Hàm mục tiêu đồng hành, phản hồi chất lượng tính năng	0,39
MIM (Liao & Carneiro, 2016)	Mạng Maxout trong mạng Maxout	0,35
RCNN (Liang & Hu, 2015)	Kết nối tuần hoàn trong lớp tích chập	0,31
Nhóm Tree+Max-Avg (Lee và cộng sự, 2016)	Nhóm cây theo sau là nhóm trung bình tối đa có cổng	0,31
MIN được chuẩn hóa theo lô (Chang & Chen, 2015)	BN, kích hoạt Maxout, kiến trúc NIN	0,24
DCNN đa cột (Ciresan và cộng sự, 2012)a	DCNN nhiều cột, có dữ liệu tăng cường (biến dạng đàn hồi)	0,23
DropConnect (Wan và cộng sự, 2013)a	Tập hợp các mạng DropConnect, với khả năng tăng cường dữ liệu (không có biến dạng đàn hồi)	0,21

Lưu ý: Các mục in nghiêng thể hiện tình trạng hiện đại của cấu hình này tại thời điểm công bố bộ dữ liệu. Các mô hình dựa trên tập hợp sử dụng phương pháp tăng cường dữ liệu.

đối với các kích hoạt mạng đã được chuẩn hóa, do đó cho phép các lớp sau quan sát các kích hoạt chính xác sẽ được tạo ra bởi mô hình suy luận. Do đó, phần mở rộng này của BN thực thi hiệu chỉnh theo từng chiều để đảm bảo tương quan giữa các kích hoạt của mô hình đào tạo và mô hình suy luận, và điều này có thể loại bỏ hoàn toàn tình trạng quá khớp đối với các tập dữ liệu hình ảnh có phân phối nhãn bị sai lệch.

5.5.3.3 Phân tích và triển vọng. Sự thay đổi biến phụ thuộc nội bộ, áp đặt tỷ lệ học tập thấp hơn và khởi tạo tham số thận trọng, là một yếu tố chính

vấn đề trong DNN và DCNN không miễn nhiễm với nó. BN được khái niệm hóa để giải quyết vấn đề này. Nó có thể giảm đáng kể thời gian đào tạo bằng cách giảm tổng số lần lặp cần thiết để hội tụ; trong quá trình suy luận, vì các phương tiện và phương sai có thể được nhân đôi vào lớp tích chập, tác động của chi phí phát sinh được loại bỏ. Nhiều nhà nghiên cứu đã đã xác nhận khả năng của nó và đưa nó vào các mô hình của họ, như minh họa trong phần 5.5.3.1. Tuy nhiên, bất chấp điều này, nó bổ sung thêm 30% tính toán chi phí chung và yêu cầu thêm 25% tham số cho mỗi lần lặp (Ioffe & Szegedy, 2015). Điều này đã thúc đẩy những người khác cải thiện khía cạnh này của kỹ thuật. Sơ đồ khởi tạo LSUV của Mishkin và Matas (2016) chia sẻ cùng một quy trình chuẩn hóa phương sai đơn vị như BN và có thể là được coi là một chương trình khởi tạo trọng số được bổ sung bởi BN áp dụng cho lô nhỏ đầu tiên. Phương pháp này loại bỏ hiệu quả các hoạt động lớp bằng thực hiện chuẩn hóa trực giao của ma trận trọng số và điều này làm cho nó trở nên hiệu quả tính toán trên mỗi lần lặp khi so sánh với BN. Tuy nhiên, đối với các tập dữ liệu lớn như ImageNet (Russakovsky và cộng sự, 2015), kết quả của nó không ổn định và mặc dù BN vẫn là lựa chọn ưu tiên, nhưng những nỗ lực khác nhằm giảm gánh nặng tính toán và các yêu cầu tham số bổ sung cho mỗi bước áp dụng cho các hệ thống sử dụng BN vẫn là cần thiết.

Hơn nữa, hiệu quả của BN giảm đi đối với các lô nhỏ bị hạn chế theo kích thước và không có mẫu độc lập, và mặc dù điều này có thể được giảm thiểu bằng cách sử dụng chuẩn hóa hàng loạt (Ioffe, 2017), điều này dễ dàng thực hiện và cải thiện đáng kể việc đào tạo các lô nhỏ bị hạn chế, phương pháp này đưa vào các siêu tham số bổ sung cần điều chỉnh và vẫn còn mới, cần phải điều tra thêm. Trong khi công việc trong tương lai nên cũng cố gắng giảm bớt sự phức tạp của các mô hình phức tạp (Chang & Chen, 2015; Liao & Carneiro, 2016) sử dụng BN, điều này cần được bổ sung bằng một sự biện minh lý thuyết về lý do tại sao BN dẫn đến sự hội tụ nhanh hơn. Những nỗ lực khác để giải quyết vấn đề chuyển đổi biến phụ thuộc nội bộ như các quy trình đào tạo mới cũng có thể xảy ra. Một nỗ lực mới như vậy là Evolutional Dropout được đề xuất bởi Li et al. (2016; xem phần 5.4.1.4), điều chỉnh xác suất lấy mẫu bỏ học, được tính toán ngay lập tức từ một lô nhỏ ví dụ, đối với các phân phối phát triển của đầu ra lớp thay vì sử dụng các xác suất giống hệt nhau và độc lập như trong Dropout chuẩn (Hinton và cộng sự, 2012).

5.5.4 Bỏ qua kết nối. Mặc dù việc tăng độ sâu mạng nói chung dẫn đến tăng hiệu suất (xem Bảng 2), cải thiện phân loại độ chính xác của DCNN không đơn giản như việc chỉ thêm các lớp (Sri-vastava và cộng sự, 2015a). Như đã đề cập trong phần 4.4, một số biến chứng bao gồm việc lắp quá mức, gánh nặng tính toán tăng lên và dấu chân bộ nhớ, và sự suy thoái (Krizhevsky và cộng sự, 2012; Szegedy, Liu và cộng sự, 2014; He và cộng sự, 2015b; Anh & Nắng, 2015; Srivastava và cộng sự, 2015a; Romero và cộng sự, 2015). Đặc biệt, sự xuống cấp vẫn là một thách thức quan trọng. Do đó, việc khắc phục nó là bắt buộc phải điều tra những lợi ích của mạng lưới rất sâu cho

nhiều ứng dụng; điều này đã dẫn đến công việc tập trung vào bỏ qua (hoặc phím tắt) kết nối trong DCNN.

5.5.4.1 Mạng lưới đường bộ. Mạng lưới đường bộ (Srivastava et al., 2015a, 2015b) sử dụng cơ chế gating có thể học được, lấy cảm hứng từ dài hạn ngắn hạn bộ nhớ (LSTM) mạng nơ-ron hồi quy (RNN) do Hochre-iter và Schmidhuber (1997) trình bày, để điều chỉnh luồng thông tin qua một số các lớp không bị suy thoái. Đầu ra y của một khối mạng lưới đường cao tốc là được tính toán bởi

$$y = H(x, WH) \cdot T(x, WT) + x \cdot C(x, WC), \quad (5.32)$$

trong đó H (tham số hóa bởi WH) là phép biến đổi phi tuyến tính trên đầu vào liên kết x của nó, và T và C lần lượt biểu diễn các cổng biến đổi và cổng mang. Trong phương trình, các chỉ số lớp và độ lệch đã bị loại trừ để đơn giản hóa. Các mạng này có thể có hàng trăm lớp, vì chúng kiến trúc cho phép tối ưu hóa bất kể độ sâu. Trên MNIST (LeCun et al., 1998) và CIFAR-10 và CIFAR-100 (Krizhevsky, 2009), các tác giả đã đạt được kết quả cạnh tranh khi so sánh với các DCNN có hiệu suất tốt nhất, mặc dù mạng lưới của họ có ít thông số hơn nhiều so với mỏng, rộng và nông hơn (nhưng vẫn sâu) mạng lưới nén mô hình DCNN của Romero et al. (2015). Sau đó, một LSTM dựa trên lưới tương tự mô hình được đề xuất cho nhiều nhiệm vụ hơn (Kalchbrenner, Danihelka, & Graves, 2015), bao gồm dịch ngôn ngữ, dự đoán ký tự, ghi nhớ trình tự và quan trọng nhất là phân loại hình ảnh.

5.5.4.2 Mạng lưới dư thừa. Mạng lưới dư thừa (He et al., 2015b), được giới thiệu trong phần 4.4, cũng giải quyết vấn đề suy thoái, sử dụng kết nối bỏ qua. Ý tưởng chính của mạng lưới dư thừa là học một mạng lưới dư thừa cộng chức năng liên quan đến một bản đồ danh tính dựa trên trước đó các lớp đầu vào, được thực hiện bằng cách đính kèm một kết nối phím tắt nhận dạng. Các mô-đun còn lại thực hiện phép tính sau:

$$yl = h(xl) + F(xl, Wl, k|1 \le k \le K), \quad (5.33)$$

$$xl+1 = f(yl), \quad (5.34)$$

trong đó đầu vào của mô-đun dư thứ l được biểu thị bằng xl, Wl biểu thị trọng số và độ lệch của nó, K là số lớp trong một mô-đun, F biểu diễn hàm còn lại như một chồng các bộ lọc tích chập, f là phép toán theo sau phép cộng từng phần tử (xem Hình 6) và h là phép ánh xạ danh tính có dạng h (xl) = xl. Các mạng còn lại đã mang lại một số thành công thực nghiệm; đặc biệt, họ đã thực hiện một cách đặc biệt tốt trên thử thách ImageNet đầy thách thức (Russakovsky và cộng sự, 2015), như được minh họa bằng Bảng 2 và Bảng 4.

5.5.4.3 Cải thiện mạng lưới dư thừa. Để nâng cao hơn nữa khuôn khổ học tập dư thừa, He, Zhang, Ren và Sun (2016) nhận thấy rằng có thể tạo ra một đường dẫn trực tiếp để truyền bá thông tin trên toàn bộ mạng lưới cả đường đi tới và đường đi lui chứ không chỉ trong các đơn vị còn lại (He et al., 2015b). Sự lan truyền này giúp việc đào tạo dễ dàng hơn và được thực hiện bằng cách sử dụng ánh xạ danh tính làm kết nối phím tắt được ghép nối với kích hoạt sau khi cộng. Xem xét các phương trình 5.33 và 5.34, nếu f là cũng là một phép ánh xạ danh tính có dạng xl+1 ≡ yl, bằng cách thay thế các phương trình 5.33 vào 5.34, tính năng đầu vào cho mô-đun dư thừa thứ l có thể được tính toán qua

$$xL = xl + F(xl, Wl). \tag{5.35}$$

Nếu phương trình 5.35 được giải theo cách đệ quy, điều này sẽ được dịch thành

$$xl = xl + \sum_{i=l}^{L-1} F(xi, Wi), \tag{5.36}$$

trong đó L biểu thị các mô-đun còn lại sâu hơn các mô-đun nông hơn trước đó được biểu thị bằng l. Điều quan trọng là kết quả này trình bày hai điểm chính thuộc tính. Đầu tiên, các tính năng của các mô-đun dư thừa sâu hơn, được biểu thị bằng xL, có thể được biểu diễn như các tính năng của bất kỳ mô-đun nông nào, được biểu thị bằng xl, được bổ sung bởi một hàm dư ($\sum_{i=l}^{L-1} F$). Thứ hai, các tính năng của các mô-đun dư thừa sâu hơn bao gồm tổng của tất cả các hàm dư thừa trước đó. Họ đã thiết lập một cách thực nghiệm kỹ thuật cải tiến của mình bằng dễ dàng đào tạo một mạng có 1000 lớp và chứng minh được sự cải thiện độ chính xác. Hơn nữa, họ đã thử nghiệm với một số tối ưu hóa khác kỹ thuật từ Hinton et al. (2012), Srivastava và cộng sự. (2014), Szegedy, Lưu et al. (2014), và Srivastava et al. (2015a) và thấy rằng những điều này có tác động tiêu cực tác động đến việc truyền bá thông tin và cản trở quá trình tối ưu hóa. Mặc dù vậy, họ thấy rằng BN (Ioffe & Szegedy, 2015) và ReLU (Nair & Hinton, 2010) các hoạt động trước đó đã cải thiện hiệu suất của chúng mạng lưới dư thừa thế hệ (He et al., 2015b). Hơn nữa, dư thừa phím tắt các kết nối cũng được kết hợp với kiến trúc Inception (Szegedy, Liu et al., 2014; xem phần 5.1.1.2) và BN (Ioffe & Szegedy, 2015) để cải thiện hiệu suất phân loại hình ảnh (Szegedy et al., 2016).

Với các mạng lưới dư thừa cực kỳ sâu của He et al. (2016) chậm đào tạo, độ sâu của chúng bị giảm và chiều rộng tăng lên theo một cách mới biến thể được gọi là mạng lưới dư thừa rộng (WRN; Zagoruyko & Komodakis, 2017). Các WRN này nông hơn nhiều vì chúng bao gồm 16 lớp so với 1000 của He et al. (2016), nhưng chúng vượt trội hơn tất cả các mô hình dư thừa trước đây về mặt hiệu quả và độ chính xác và thiết lập các mô hình mới kết quả hiện đại về các tập dữ liệu CIFAR-100 (Krizhevsky, 2009; Szegedy, Van-houcke và cộng sự, 2015) và SVHN (Netzer và cộng sự, 2011; Lee và cộng sự, 2016).

Tuy nhiên, sau này chúng đã đư ợc thay thế bằng kỹ thuật tất đư ợc giới thiệu trong phần 5.5.4.4.

5.5.4.4 Mạng tích chập kết nối dày đặc. Các mạng tích chập kết nối dày đặc đư ợc đề xuất gần đây (Huang, Liu, Weinberger, & van der Maaten, 2016) mở rộng ý tư ởng về kết nối bỏ qua bằng cách kết nối theo phư ơng thức truyền tiếp thông thư ờng mỗi lớp với mọi lớp khác trong mạng.
Do đó, bản đồ đặc trư ng của tất cả các lớp trư ớc đó đư ợc sử dụng làm đầu vào cho mỗi lớp tiếp theo. Trư ớc đây, sử dụng ký hiệu tư ơng tự như phư ơng trình 5.35, lớp thứ l chấp nhận bản đồ đặc trư ng của lớp trư ớc đó, x0,.xl 1, làm đầu vào của nó:

$$x0 = Hl([x0, x1, x2 . xl\ \ 1]), \tag{5.37}$$

trong đó [x0, x1, x2 ... xl 1] biểu diễn phép nối bản đồ đặc trư ng của các bản đồ đư ợc tạo ra ở các lớp 0,... l 1, và Hl, theo sau các mạng dư thừa đư ợc cải thiện (He et al., 2016), là một hàm hợp chất của ReLU (Nair & Hinton, 2010), đi trư ớc là BN (Ioffe & Szegedy, 2015), và theo sau là phép tích chập. Trên các chuẩn phân loại hình ảnh phổ biến, bao gồm ImageNet đầy thách thức (Russakovsky et al., 2015), họ đã đạt đư ợc độ chính xác tư ơng đư ơng với các mạng dư thừa (He et al., 2016) như ng yêu cầu ít tham số hơn đáng kể. Hơn nữa, chúng giải quyết tình trạng suy thoái (He et al., 2015b; He & Sun, 2015; Srivastava et al., 2015a, 2015b) và vấn đề biến mất độ dốc (Bengio et al., 1994), đồng thời thúc đẩy việc tái sử dụng tính năng để giảm tính toán.

Bảng 6 so sánh hiệu suất của một số DCNN trên các tập dữ liệu phổ biến CI-FAR (Krizhevsky, 2009) và SVHN (Netzer và cộng sự, 2011). Các kết quả đư ợc báo cáo trong cột CIF-10 (có DA) là của các mô hình DCNN sử dụng tăng cư ờng dữ liệu. Đối với CIFAR-100 và SVHN, nơi tăng cư ờng dữ liệu ít phổ biến hơn, kết quả của các mô hình sử dụng tăng cư ờng dữ liệu đư ợc in nghiêng. Như minh họa, WRN (Zagoruyko & Komodakis, 2017) đã vư ợt qua một số DCNN tiên tiến khác về các tác vụ này (CIFAR-100 và SVHN) và khi DCNN đư ợc kết hợp với ELU (Clevert và cộng sự, 2016), chúng đạt đư ợc lỗi phân loại thấp nhất tại thời điểm đó trên tập dữ liệu CIFAR-10 mà không cần tăng cư ờng dữ liệu. Trên CIFAR-10, với việc tăng cư ờng dữ liệu, WRN đã thu đư ợc kết quả rất gần với tỷ lệ lỗi thấp thứ hai thu đư ợc bằng phư ơng pháp nhóm cực đại phân số do Graham (2014; xem phần 5.1.2.2) trình bày, mặc dù họ không sử dụng cùng các kỹ thuật tăng cư ờng dữ liệu cực đoan. Các mạng tích chập đư ợc kết nối dày đặc của Huang, Liu và cộng sự (2016) đã thay thế tất cả các kết quả này, minh họa cho hiệu suất đặc biệt của kỹ thuật này và tầm quan trọng cũng như thành công thực nghiệm của các kết nối tắt trong các mô hình hiện tại của chúng tôi.

5.5.4.5 Phân tích và triển vọng. Khi các nhà nghiên cứu bắt đầu thử nghiệm với các mạng lư ới sâu hơn để cải thiện hiệu suất phân loại hình ảnh, họ

Bảng 6: Hiệu suất DCNN trên CIFAR-10 và CIFAR-100 và dữ liệu SVHN Bộ.

Người mẫu	Mô tả ngắn gọn	CIF.-10 CIF.-10 (có DA)	CIF.-100	SV-HN
DropConnect (Wan và cộng sự, 2013)	DCNN được tối ưu hóa với thả kết nối	18,7	9.32	- 1,94
Tập hợp ngẫu nhiên (Zeiler và Fergus, 2013)	Ngẫu nhiên hơn là sự kết hợp xác định thủ tục	15.13	-	42,51 2,80
Mạng lưới Maxout (Goodfellow và cộng sự, 2013)	Kích hoạt tối đa chức năng	11,68	9,38	38,57 2,47
Probout (Springenberg & Riedmiller, 2013)	Kích hoạt xác suất chức năng	11.35	9.39	38,14 2.39
NIN (Lin et al., 2013)	MLP được tích hợp vào Kiến trúc DCNN	10.41	8,81	35,68 2,35
Mạng lưới được giám sát sâu (Lee và cộng sự, 2015)	Mục tiêu đồng hành chức năng, tính năng phản hồi chất lượng	9,78	8.22	34,57 1,92
DCNN + APL (Agostinelli và cộng sự, 2014)	Thích ứng từng phần kích hoạt tuyến tính	9,59	7.51	34,40 -
Toàn bộ CNN (Springenberg, Dosovitskiy, Brox, & Riedmiller, 2014)	DCNN với lớp tích chập chỉ một	9.08	7.41	33,71 -
RCNN (Liang & Hu, 2015)	Kết nối định kỳ trong tích chập lớp	8,69	7.09	31,75 1,77
Tích chập kép lưới	Tích chập đôi hoạt động	8,58	7.24	30,35 -
MIM (Liao & Carneiro, 2016)	Mạng lưới Maxout trong Mạng lưới tối đa	8,52		29.20 1,97
MIN được chuẩn hóa theo lô (Chang và Chen, 2015)	BN, Tối đa hóa kích hoạt, NIN	7,85	6,75	28,86 1,97
Nhóm Tree+Max-Avg (Lee và cộng sự, 2016)	Cây trồng theo sau theo mức trung bình có cổng nhóm tối đa	7.62	6.05	32,37 1,69
ELU (Clevert và cộng sự, 2016)	Tuyến tính mũ chức năng kích hoạt	6,55	-	24,28 -
Các tiên nghiệm dựa trên cây (Srivastava & (Salakhutdinov, 2013)	DCNN với đã học cây trước	-		36,85 -
FitNet-LSUV (Mishkin & Matas, 2016)	Kiến trúc FitNet với LSUV khởi tạo		6.06	27,66

Bảng 6: Tiếp theo.

Người mẫu	Mô tả ngắn gọn	CIF.-10	CIF.-10 (có DA)	CIF.-100	SV-HN
Sự chú ý sâu sắc Lưới chọn lọc (Stollenga, Masci, Gomez, & Schmidhuber, 2014)	DCNN với phản hồi kết nối	-	9.22	33,78	-
FitNets (Romero và cộng sự, 2015)	Mạng lưới mỏng và sâu với trung gian gợi ý cấp độ cho việc đào tạo	-	8.39	34.04	2,42
Mạng lưới đường cao tốc (Srivastava và cộng sự, 2015b)	Cổng học tập cho điều chỉnh luồng dữ liệu	-	7.6	32,33	-
Mạng lưới dư thừa sâu 1 (Ông và cộng sự, 2015b)	Các hàm còn lại / khối tích hợp vào Các lớp DCNN	-	6.43	-	-
Mạng lưới dư thừa sâu 2 (Ông và cộng sự, 2016)	Các khối còn lại với ánh xạ danh tính	-	4,62	22,71	-
Phân nhóm max phân số (Graham, 2014)	Phân số ngẫu nhiên phiên bản tối đa tập hợp	-	3,47	26,39	-
WRN (Zagoruyko & Komodakis, 2017)	Các khối còn lại với tăng chiều rộng	-	4.17	20,50	1,64
Kết nối dày đặc CNN (Huang và cộng sự, 2016)	Kết nối giữa lớp	5.19	3,46	17.18	1,59

Lưu ý: Các mục in nghiêng cho CIFAR-100 và SVHN là kết quả của các mô hình được sử dụng tăng cường dữ liệu.

gặp phải những gì vẫn có thể được coi là một thách thức mở, thường được gọi là sự suy thoái (xem phần 4.4). Một nỗ lực ban đầu để chống lại sự suy thoái được thực hiện bởi mạng lưới đường cao tốc (Srivastava et al., 2015a, 2015b), sử dụng các phím tắt gating để đào tạo hiệu quả các mạng rất sâu (hơn 100 lớp), được tối ưu hóa bởi SGD (Bottou, 1998, 2010). Tuy nhiên, có những trường hợp khi các phím tắt bị đóng, ngăn chặn thông tin dòng chảy; hơn nữa, các phím tắt có cổng phụ thuộc vào dữ liệu và yêu cầu số lượng tham số. Các mạng lưới còn lại (He et al., 2015b, 2016) cũng làm giảm nhẹ vấn đề suy thoái; tuy nhiên, trái ngược với các mạng lưới đường cao tốc, chúng sử dụng các phím tắt lập bản đồ danh tính không có tham số và luôn mở, cho phép luồng thông tin liên tục. Các mạng còn lại, có thể chứa hơn 1000 lớp mà không làm giảm hiệu suất, đã tạo ra hiệu suất đặc biệt trong các nhiệm vụ định vị, phát hiện và phân loại. Mặc dù thành công, những người khác đã tìm thấy rằng có sự dư thừa đáng kể trong số lượng lớp cực lớn của chúng,

và điều này khiến họ mất nhiều thời gian không cần thiết để đào tạo, ủng hộ nhu cầu về
các giải pháp thay thế khác. Theo hư ớng này, WRN (Zagoruyko
& Komodakis, 2017) đã chứng minh rằng các mạng lư ới dư thừa nông hơn có thể
tạo ra hiệu suất phân loại tư ơng đư ơng với khả năng phân loại cực kỳ sâu của chúng
các đối tác (He et al., 2015b, He et al., 2016) và nhanh hơn gấp nhiều lần
để đào tạo; tuy nhiên, hiệu suất của họ trên ImageNet đầy thử thách (Rus-sakovsky và cộng
sự, 2015) chỉ mới đư ợc công bố trong năm nay, khuyến khích
cần phải thử nghiệm thêm với phép tích chập nông hơn như ng rộng hơn
mạng lư ới.

Mạng độ sâu ngẫu nhiên (Huang, Sun, Liu, Sedra, & Weinberger, 2016)
cũng thách thức sự dư thừa lớp của các mạng còn lại bằng cách bỏ qua các lớp dư thừa với
các hàm nhận dạng sau khi ngẫu nhiên loại bỏ chúng. Các mạng này thực hiện phân loại và
lỗi giảm
nhanh hơn so với các mạng còn lại trên chuẩn CIFAR-10
(Krizhevsky, 2009); tuy nhiên, chúng vẫn chư a đư ợc thử nghiệm trên Im-ageNet đầy thách
thức, trên đó các mạng còn lại đã cải thiện đáng kể trạng thái
của nghệ thuật. Hơn nữa, chúng yêu cầu điều chỉnh siêu tham số. Các mạng đư ợc kết nối dày
đặc, áp đặt các ràng buộc về bộ nhớ và tính toán thấp hơn so với các mạng còn lại, thực
hiện các kết nối tất đến
cực đoan bằng cách giới thiệu các kết nối trực tiếp giữa các lớp, như ng không giống như
các mạng còn lại, chúng vẫn chư a đư ợc thử nghiệm trên các tác vụ thị giác máy tính đa dạng
ngoài việc phân loại hình ảnh. Do đó, mặc dù đã có tiến bộ trong việc giải quyết sự suy
thoái, các kỹ thuật hiện có vẫn chư a đư ợc
đư ợc thiết lập vững chắc. Những thành công thực nghiệm của các mạng lư ới còn lại trên
Ima-geNet khiến chúng trở thành lựa chọn hấp dẫn nhất; tuy nhiên, khi kết hợp với
Bỏ học (Hinton và cộng sự, 2012; Srivastava và cộng sự, 2014), chúng không thể hội tụ để
các giải pháp có thể chấp nhận đư ợc (He et al., 2016). Do đó, công việc trong tư ơng lai
tập trung vào việc thay đổi cơ bản khuôn khổ học tập còn lại để làm việc kết hợp
với Dropout và các kỹ thuật chính quy hóa khác là cần thiết. Để đạt đư ợc mục đích này,
WRN cho thấy kết quả ban đầu khả quan.

Cuối cùng, mặc dù các ứng dụng thành công của các kết nối tắt khác nhau đư ợc thảo
luận ở đây và các kết quả thực nghiệm đầy hứa hẹn của chúng đư ợc báo cáo như vậy
xa, một sự hiểu biết rõ ràng về cách họ cải thiện cơ bản việc đào tạo
của DCNN vẫn còn thiếu. Mặc dù một số công trình gần đây đã cố gắng giải quyết các đặc
điểm đa dạng của thách thức này (Hardt & Ma, 2016; Li, Jiao,
Han, & Weissman, 2016; Littwin & Wolf, 2016), có lẽ lời giải thích kích thích nhất đằng
sau thành công của họ là họ phá vỡ tính đối xứng nội tại trong bối cảnh mất mát của DCNN,
dẫn đến sự đơn giản hóa đáng kể
cảnh quan, như đư ợc quan sát bởi Orhan (2017). Cụ thể, đối với các kết nối dày đặc
mạng tích chập (xem phần 5.5.4.4), bài báo phát hiện ra rằng phím tắt
các kết nối gây ra sự không liên tục trong tính đối xứng khi thay đổi tỷ lệ của các ma trận
kết nối các lớp khác nhau của mạng, trong khi đối với các mô hình sâu khác, chúng gây ra
sự không liên tục trong tính đối xứng hoán vị của các tế bào thần kinh
ở một lớp cụ thể. Hơn nữa, mặc dù các phư ơng tiện khác để phá vỡ tính đối xứng cũng tạo
điều kiện cải thiện hiệu suất khi đào tạo các mô hình sâu,

họ thấy rằng các kết nối tắt thúc đẩy thêm nhiều lợi ích hơn nữa
diện mạo phá vỡ tính đối xứng của họ, và chúng bao gồm khả năng
giải quyết vấn đề độ dốc biến mất hoặc bùng nổ (Bengio và cộng sự, 1994).
Hơn nữa, liên quan đến việc phá vỡ tính đối xứng, họ cũng quan sát thấy các lớp tế
bào thần kinh khác biệt, có lẽ giống như những lớp được não người sử dụng
(Harris & Shepherd, 2015), vượt trội hơn các tế bào thần kinh không phân biệt thường
được sử dụng bởi các mạng lưới nhân tạo. Mặc dù công trình này cung cấp một khởi đầu
điểm, cần nghiên cứu thêm để hiểu rõ hơn vai trò của phím tắt
kết nối trong các mô hình của chúng tôi và để xác định xem hệ thần kinh trung ương
sử dụng các cơ chế tương tự như chúng cho các nhiệm vụ thị giác.

5.5.5 Phát triển chi phí tính toán. Người ta đều biết rằng dữ liệu lớn hơn
các bộ đã góp phần vào thành công của việc học sâu (Krizhevsky và cộng sự,
2012; Deng & Yu 2014; Zeiler & Fergus, 2014). Tuy nhiên, nhược điểm, đặc biệt là
trong quá trình đào tạo, là gánh nặng tính toán lớn hơn. Kết hợp với
đây là thực tế rằng các mô hình DCNN có số lượng tham số rất lớn, điều này có tác
động tiêu cực đến yêu cầu về bộ nhớ và lưu trữ của chúng
(Krizhevsky và cộng sự, 2012; Wan và cộng sự, 2013; Simonyan & Zisserman, 2014;
Szegedy, Vanhoucke và cộng sự, 2015; Szegedy, Liu và cộng sự, 2014; Ông và cộng sự, 2015b;
Szegedy và cộng sự, 2016). Ví dụ, DCNN của Krizhevsky và cộng sự.
(2012) có 60 triệu tham số và mất sáu ngày để đào tạo trên hai GPU,
trong khi mô hình lớn nhất do Simonyan và Zisserman (2014) trình bày bao gồm 144
triệu tham số, được đào tạo trên bốn GPU trong hai đến ba
tuần. Do đó, một lượng lớn nghiên cứu đã được thực hiện để giảm
chi phí tính toán và yêu cầu không gian lưu trữ của DCNN. Tiếp theo chúng tôi
thảo luận một số tác phẩm tiêu biểu về vấn đề này.

5.5.5.1 Tính toán song song. Một lượng nỗ lực đáng kể (Zinkevich
và cộng sự, 2010; Recht và cộng sự, 2011; Dean và cộng sự, 2012; Zhuang và cộng sự, 2013; Đau đớn
và cộng sự, 2013; Yadan, Adams, Taigman và Ranzato, 2014; Krizhevsky, 2014) có
đã đi vào song song hóa việc đào tạo DCNN thông qua mô hình song song,
đòi hỏi phải sử dụng GPU, nhiều GPU, cụm GPU và CPU và dữ liệu
song song, kết hợp các thuật toán tối ưu hóa được cải thiện như
SGD không đồng bộ (ASGD; Recht et al., 2011; Dean et al., 2012) và BN
(Ioffe và Szegedy, 2015). Yadan và cộng sự (2014) sử dụng song song lai
chiến lược xem xét cả mô hình và tính song song của dữ liệu. Họ đã sử dụng một cấu
hình chia sẻ tính toán của mạng và mini-batch hình ảnh
chia thành bốn GPU để giảm thời gian đào tạo hơn 2,2 lần khi
so với một GPU duy nhất. Dean et al. (2012) đã giới thiệu một khuôn khổ mới
để đào tạo phân tán song song các mạng sâu trên một cụm CPU.
Trong khuôn khổ này, họ đã giới thiệu một hình thức ASGD mới, được gọi là Down-pour
SGD, để hỗ trợ đào tạo một số lượng lớn các bản sao mô hình. Khung này cũng được sử
dụng bởi hệ thống hiệu suất cao nhất của Szegedy, Liu et al.
(2014), và mặc dù họ đã sử dụng triển khai dựa trên CPU, họ đã dự báo

sự hội tụ mạng cũng có trên một số GPU cao cấp, mặc dù như ợc điểm ư ớc tính là sử dụng nhiều bộ nhớ hơn.

Paine et al. (2013) cũng khai thác mô hình và tính song song dữ liệu bằng cách sử dụng GPU cho mô hình song song và A-SGD (Recht et al., 2011; Paine et al., 2013) cho tính song song dữ liệu và đạt đư ợc tốc độ tăng 3,2 lần với tám GPU so với một đơn vị duy nhất; tuy nhiên, chúng mất đi độ chính xác đáng kể. Krizhevsky (2014) đã đề xuất hai biến thể của SGD descent song song hóa việc đào tạo DCNN bằng cách sử dụng song song dữ liệu nặng trong các lớp tích chập và nhiều mô hình song song hơn trong các lớp đư ợc kết nối đầy đủ. Trên tám GPU, bài báo đạt đư ợc tốc độ tăng 6,16 lần với sự thay đổi không đáng kể về độ chính xác. Khẳng định rằng lư ợc đồ này quá phức tạp, Simonyan và Zisserman (2014) đã đạt đư ợc tốc độ tăng 3,75 lần trên hệ thống GPU có sẵn, bao gồm bốn GPU so với một GPU duy nhất. Tuy nhiên, ngay cả với điều này, phải mất hai đến ba tuần để đào tạo một mạng lư ới duy nhất, do đó ủng hộ nhu cầu về các giải pháp khác. Gần đây hơn, Dettmers (2016) đã đề xuất một cụm GPU quy mô lớn, kết hợp với các thuật toán song song cải tiến để song song hóa DCNN hiệu quả (xem phần 6.3 để biết chi tiết hơn). Mặc dù hệ thống này tạo ra sự phân loại thành công kết quả là nó thiếu tính thực tế để triển khai trên diện rộng.

Bằng chứng Chư a sửa

5.5.5.2 Khai thác định lý tích chập và phép chiếu tròn. Bằng cách thực hiện phép toán tích chập như các tích từng phần tử trong Fourier miền, sử dụng biến đổi Fourier nhanh (FFT) và tái chế cùng một bản đồ đặc điểm đã chuyển đổi nhiều lần, Mathieu et al. (2013) đã đạt đư ợc tốc độ xử lý tăng đáng kể (lên đến hai lần) với cái giá phải trả là bộ nhớ lớn hơn dấu chân. Kỹ thuật này có thể dễ dàng đư ợc tích hợp với nhóm phổ (xem phần 5.1.2.5), và vì phép biến đổi Fourier rời rạc (DFT) đư ợc thực hiện đối với cả hai phư ơng pháp, chi phí tính toán bổ sung đều không đáng kể (Rip-pel et al., 2015).

Độc lập, Cheng, Felix et al. (2015) và Cheng, Yu et al. (2015) cũng sử dụng FFT để tăng tốc tính toán. Tuy nhiên, trái ngư ợc với Mathieu et al. (2013), ngư ời đã sử dụng định lý tích chập đã đư ợc thiết lập tốt để giảm thiểu thời gian xử lý trong các lớp tích chập, họ tập trung vào việc tăng tốc tính toán trong các lớp đư ợc kết nối đầy đủ bằng cách áp đặt một vòng tròn thay vì hơn là một cấu trúc tuyến tính trên các ma trận trọng số. Chính thức hơn, với một lớp đư ợc kết nối đầy đủ với d nút đầu vào và d nút đầu ra, các phép chiếu tròn sẽ cải thiện độ phức tạp thời gian từ O(d2) đến O(d log d). Hơn nữa, phư ơng pháp của họ cũng có lợi cho không gian lư u trữ và giảm độ phức tạp của không gian từ O(d2) hoặc O(d). Trên MNIST (LeCun et al., 1998), CIFAR-10 (Krizhevsky, 2009) và ImageNet (Russakovsky và cộng sự, 2015), họ đã đạt đư ợc sự giảm đáng kể về chi phí tính toán và không gian lư u trữ, với mức tăng tối thiểu về lỗi phân loại. Tiếp theo là việc sử dụng xử lý tín hiệu số cổ hữu kỹ thuật, Wang, Xu, You, Tao và Xu (2016) đề xuất nén và giảm chi phí tính toán của DCNN bằng cách hợp nhất tuyến tính

phản ứng tích chập của các cơ sở biến đổi cosin rời rạc (DCT) sau lần đầu tiên
xử lý các bộ lọc tích chập như hình ảnh và sau đó phân tích biểu diễn của chúng
trong miền tần số.

5.5.5.3 Thao tác ma trận. Denil, Shakibi, Dinh và de Freitas (2013)
đã thực hiện một nỗ lực ban đầu để giảm bớt sự tham số hóa quá mức của sâu
mạng lưới (Hinton và cộng sự, 2012; Denton, Zaremba, Bruna, LeCun và Fergus, 2012).
2014; Kim et al., 2015). Kỹ thuật của họ, nhằm mục đích giảm mạng lưới thần kinh
các tham số miễn phí (trọng số và độ lệch), dựa trên việc biểu diễn trọng số
ma trận như một tích hạng thấp của hai ma trận nhỏ hơn. Do đó, bằng cách phân
tích và kiểm soát hạng của ma trận trọng số, họ có thể
trực tiếp kiểm soát tham số hóa của mạng. Trên chuẩn mực CIFAR-10 (Krizhevsky,
2009), họ phát hiện ra rằng bằng cách dự đoán và do đó loại bỏ không gian tính
toán và lưu trữ của 75% các tham số, đã có một
tác động không đáng kể đến độ chính xác phân loại. Phương pháp của họ là bổ sung
đến các tiến bộ khác của DCNN như Dropout (Hinton và cộng sự, 2012; Srivastava
et al., 2014) và kích hoạt Maxout (Goodfellow et al., 2013).

Được thúc đẩy bởi điều này, Denton et al. (2014) đã khai thác sự dư thừa trong
các lớp tích chập và các phép xấp xỉ được suy ra để giảm thiểu tính toán. Cụ thể,
họ đã sử dụng các phép xấp xỉ đơn sắc trong lần đầu tiên
lớp tích chập và xấp xỉ biclustering với SVD trong lớp tích chập thứ hai. Trong cả
hai lớp, chúng báo cáo mức tăng tốc độ giữa
2 và 2,5 lần, với mức giảm 1% về hiệu suất, so với các mô hình cơ sở của họ. Hơn
nữa, bằng cách áp dụng SVD bị cắt bớt, họ đã giảm
chi phí bộ nhớ và yêu cầu lưu trữ của các lớp được kết nối đầy đủ và báo cáo giảm
trọng lượng lên đến 13,4 lần với mức giảm dưới 1%
mất độ chính xác. Tương tự như vậy, Jaderberg, Vedaldi và Zisserman (2014)
sử dụng các phép tính xấp xỉ bộ lọc để xấp xỉ các bộ lọc tích chập và
tiếp theo là sử dụng các kỹ thuật lọc và tái tạo dữ liệu để tái tạo các phép tính
gần đúng với lỗi tối thiểu. Trong phân loại văn bản cảnh
ứng dụng, họ đạt được tốc độ tăng 2,5 lần mà không mất độ chính xác và 4,5 lần
với độ chính xác giảm dưới 1%. Cả Denton và cộng sự
(2014) và Jaderberg et al. (2014) sử dụng phân tích ma trận bậc thấp để
nén một hoặc nhiều lớp; những người khác đã sử dụng
các kỹ thuật liên quan bao gồm Sainath, Kingsbury, Sindhwani, Arisoy và
Ramabhadran (2013) và Lebedev, Ganin, Rakhuba, Oseledets và Lem-pitsky (2015).

Lấy cảm hứng từ sự dư thừa trong các tham số mạng nơ-ron được làm nổi bật
bởi Denil et al. (2013), Gong, Liu, Yang và Bourdev (2014) đã đề xuất lượng tử
hóa vectơ như một giải pháp thay thế có hiệu suất cao hơn cho phép phân tích ma
trận bậc thấp (Sainath, Kingsbury, Sindhwani et al., 2013; Denton et al., 2014;
Jaderberg et al., 2014), để nén các ma trận của hoàn toàn dày đặc
các lớp được kết nối. Họ có thể đạt được tốc độ nén ấn tượng
(lên đến 24 lần) mà không làm giảm hơn 1% độ chính xác phân loại top 5 trên bộ dữ
liệu ImageNet đầy thách thức (Russakovsky và cộng sự, 2015). Trong

Ngoài Gong, Liu và cộng sự (2014), nhiều phương pháp khác dựa trên lượng tử hóa vectơ đã được chứng minh là đạt được khả năng nén tốt hơn SVD.
Những điều này bao gồm các phép chiếu tròn (Cheng, Felix và cộng sự, 2015; Cheng, Yu và cộng sự, 2015), các kỹ thuật băm (Chen, Wilson, Tyree, Weinberger, & Chen, 2015), và phân tích chuỗi tenxơ (Oseledets, 2011; Novikov, Podoprikhin, Osokin và Vetrov, 2015).

5.5.5.4 Phân tích và triển vọng. Phân loại hiện đại của chúng tôi các mô hình phụ thuộc rất nhiều vào đào tạo có giám sát; tuy nhiên, trung đoàn này có một hạn chế cố hữu ở chỗ nó đòi hỏi một lượng đầy đủ của dữ liệu đào tạo kéo dài đáng kể quá trình đào tạo và gây ra không gian và các biến chứng về bộ nhớ. Mặc dù điều này có thể được giảm nhẹ bằng phương pháp tiếp cận vũ phu bằng cách sử dụng cụm CPU hoặc GPU, quyền truy cập vào các hệ thống như vậy bị hạn chế đối với các tổ chức lớn. Để giảm dấu chân tính toán và bộ nhớ, các khía cạnh khác nhau của tính toán DCNN có thể được thực hiện sử dụng các kỹ thuật đã được thiết lập vững chắc trong xử lý tín hiệu số, chẳng hạn như FFT, DFT, DCT và định lý tích chập. Đối với các kỹ thuật này, dấu chân tính toán và bộ nhớ của phép toán tích chập hoặc các lớp được kết nối đầy đủ có thể được giảm bớt. Mặc dù di chuyển giữa các miền làm tăng tính phức tạp của hệ thống của chúng tôi, với những thành công được mô tả trong phần 5.5.4.2, nghiên cứu sâu hơn về DCNN lai sử dụng kỹ thuật số cơ sở xử lý tín hiệu dựa trên để cải thiện không chỉ tính toán chi phí và dấu chân bộ nhớ mà còn cả độ chính xác của phân loại (xem phần 5.1.2.5) được chứng minh là đúng. Việc thao tác các ma trận trọng số của cả lớp tích chập và lớp kết nối đầy đủ là một giải pháp thay thế khác để cải thiện hiệu quả tính toán và giải quyết hậu quả của DCNN qua tham số hóa; tuy nhiên, mặc dù có những cải tiến hơn nữa được hình thành từ ứng dụng toán học được khuyến khích, chúng không giải quyết được nguyên nhân gốc rễ của vấn đề. Tóm lại, việc di chuyển các phép toán sang miền tần số và thao tác ma trận có thể dẫn đến các đặc điểm tính toán được cải thiện;
Tuy nhiên, tất cả các kỹ thuật được thảo luận trong phần này đều bị mất mát độ chính xác, ngay cả khi nó chỉ là biên độ. Các phương pháp tính toán song song có thể cung cấp độ chính xác cần thiết; tuy nhiên, điều này đi kèm với những tác động về tài chính ngăn chặn việc sử dụng trên diện rộng và không thực tế để thích ứng trên diện rộng.
Vì vậy, cần phải có thêm nghiên cứu để giải quyết những thách thức này. Trong khi chúng tôi đã xem xét một số cải tiến ban đầu trong phần này, mới nhất xu hướng cải thiện tính toán DCNN, được bổ sung bởi các khuyến nghị trong tương lai, được trình bày chi tiết trong phần 6.3.

6 Thách thức và xu hướng mở được lựa chọn

Mặc dù kết quả phân loại hình ảnh đầy hứa hẹn thu được từ DCNN, vẫn còn những thách thức cần được giải quyết. Trong phần cuối cùng này, chúng tôi giải quyết một số vấn đề này cùng với các xu hướng được chọn lọc trong công trình gần đây.

6.1 Sự biện minh lý thuyết và sự hiểu biết nội tại. Mặc dù thành công thực nghiệm của DCNN, bằng chứng lý thuyết về lý do tại sao chúng thành công đang thiếu. Để đạt đư ợc mục đích này, Mallat (2012) đã chứng minh tính bất biến của phép dịch và sự ổn định biến dạng của các tính năng đư ợc trích xuất bằng cách tán xạ tích chập mạng lư ới. Wiatowski và Bölcskei (2015) vẫn kiên trì với toán học phân tích các đặc điểm đư ợc trích xuất bởi DCNN và về mặt lý thuyết, chúng đã thiết lập đư ợc tính ổn định biến dạng và tính bất biến dịch chuyển theo phư ơng thẳng đứng. Tiếp tục với phân tích lý thuyết, Basu et al. (2016) đã xem xét các tập dữ liệu phân loại hình ảnh trong đó kết cấu đóng vai trò quan trọng (Lazebnik, Schmid, & Ponce, 2005; Filho, Luiz, Oliveira và Britto, 2009; Oxholm, Baria, & Nishino, 2012), và họ cung cấp các ranh giới lý thuyết về việc sử dụng DC-NN để phân loại kết cấu. Cụ thể, họ đã sử dụng lý thuyết về chiều Vapnik-Chervonenkis (VC) (xem Vapnik & Chervonenkis, 1971, để biết chi tiết và Sontag, 1998, cho ứng dụng đầu tiên cho ANN) để chứng minh rằng việc trích xuất tính năng thủ công tạo ra các biểu diễn chiều thấp. Như một hệ quả của điều này, họ đã đư a ra các giới hạn trên về chiều VC của DCNN. Hơn nữa, cũng đã có cuộc điều tra về hoạt động và hiệu suất nội bộ của DCNN, chẳng hạn như thư ờng trích dẫn kỹ thuật trực quan hóa tính năng đư ợc trình bày bởi Zeiler và Fergus (2014). Các kỹ thuật trực quan khác, tất cả đều nhằm mục đích hiểu cơ chế nội bộ của DCNN, cũng đã đư ợc đề xuất (Girshick, Don-ahue, Darrell, & Malik, 2014; Yu, Yang, Bai, Yao, & Rui, 2014a, 2014b). Tiến bộ hơn nữa phụ thuộc vào cả bằng chứng lý thuyết vững chắc và thực tiễn các cuộc điều tra dẫn đến sự hiểu biết và hiệu suất đư ợc cải thiện.

6.2 Bất biến hình học. Mặc dù DCNN mạnh mẽ chống lại các biến dạng quy mô nhỏ (Lee và cộng sự, 2009), như ng các biểu diễn cuối cùng của chúng không bất biến về mặt hình học (Ciresan và cộng sự, 2011; Gong, Wang và cộng sự, 2014; Razavian et al., 2014). Cụ thể, chúng nhạy cảm với các phép dịch chuyển toàn cục, phép quay, và tỷ lệ (Gong, Wang và cộng sự, 2014). Để giải quyết các phư ơng sai dịch thuật, Lee et al. (2009) đề xuất nhóm tối đa xác suất, trong khi chư ơng trình MOP của Gong, Wang et al. (2014) đã đư ợc chứng minh là mạnh mẽ chống lại một số phư ơng sai hình học. Gần đây, chư ơng trình hợp nhất TI do Laptev et al. trình bày. (2016) đã xử lý hiệu quả các phép quay và thay đổi tỷ lệ và do đó xây dựng tính bất biến chuyển đổi vào kiến trúc DCNN, trong khi bộ biến đổi không gian mô-đun do Jaderberg và cộng sự đề xuất (2015) đã học đư ợc phép dịch chuyển, tỷ lệ, quay và bất biến cong vênh. Do đó, một hư ớng thú vị là nghiên cứu nếu cần có những thay đổi cơ bản hơn nữa đối với kiến trúc DCNN để cải thiện tính mạnh mẽ phổ quát của chúng. Hơn nữa, mặc dù số lư ợng lớn hình ảnh có sẵn trong các tập dữ liệu hiện đại như ImageNet (Russakovsky et al., 2015), vẫn có thể hình dung rằng các tập dữ liệu hiện tại của chúng tôi không phù hợp với nhiệm vụ bất biến mà chúng ta hiện đang phải đối mặt. Do đó, một hư ớng triển vọng khác là thu thập hoặc tạo dữ liệu đào tạo mới không nhất thiết dẫn đến dữ liệu lớn hơn bộ, tư ơng tự như xu hư ớng mà chúng ta đã thấy trong vài năm qua, như ng sẽ tạo điều kiện thuận lợi cho việc học các tính năng DCNN góp phần tạo nên các mô hình mạnh mẽ hơn.

6.3 Hướng tới triển khai di động. Mặc dù có một số chi phí tính toán, dấu chân bộ nhớ và giảm lưu trữ (xem phần 5.5.4), cải thiện DCNN về mặt này vẫn tiếp tục là một lĩnh vực nghiên cứu mở, đặc biệt là với mục đích triển khai chúng trên FPGA, hệ thống nhúng, thiết bị di động và các thiết bị khác có hạn chế về bộ nhớ và pin. Một số trong những phát triển gần đây nhất về vấn đề này bao gồm nén DCNN và lượng tử hóa trọng số, thuật toán nhanh, cụm GPU với biểu diễn bị cắt bớt và những tiến bộ được tăng tốc bởi FPGA.

Bằng cách sử dụng cắt tỉa mạng, lượng tử hóa trọng số để thực thi chia sẻ trọng số và mã hóa Huffman để nén tốt hơn, Han, Mao và Dally (2016), đã giới thiệu nén sâu, giúp giảm yêu cầu lưu trữ của mô hình do Krizhevsky và cộng sự đề xuất (2012) từ 240 MB xuống 6,9 MB mà không mất đi độ chính xác. Khi được đánh giá chuẩn trên CPU, GPU, và GPU di động, chúng cũng đạt được tốc độ tăng từ 3,0 đến 4,2 lần. Một phương pháp nén hiệu quả hơn nhiều, đã giới thiệu một phương pháp mới Mô-đun cháy DCNN (lớp tích chập bóp 1 × 1, tiếp theo là 1 × 1 và phép tích chập 3×3) và kiến trúc SqueezeNet, cũng đã được thử nghiệm với mô hình từ Krizhevsky et al. (2012) giảm nó xuống còn 4,8 MB mà không làm mất độ chính xác (Iandola et al., 2016). Hơn nữa, khi họ áp dụng nén sâu vào mô hình của họ, với lượng tử hóa trọng số 6 bit, họ đã giảm thêm 0,47 MB mà không có bất kỳ tác động nào đến lỗi phân loại.

Một cách tiếp cận lượng tử hóa cực đoan khác, trong đó trình bày trọng số nhị phân (các bộ lọc xấp xỉ với trọng số nhị phân) và mạng XNOR (nhị phân trọng số và đầu vào nhị phân cho các lớp tích chập), dẫn đến các mạng với khả năng tiết kiệm bộ nhớ gấp 32 lần so với ban đầu (Krizhevsky et al., 2012) và, trong trường hợp của mạng XNOR, tốc độ tăng 58 lần trong quá trình suy luận (Rastegari, Ordonez, Redmon, & Farhadi, 2016). Courbariaux, Hubara, Soudry, El-Yaniv và Bengio (2016) đã giới thiệu các mạng lưới nhị phân có trọng số nhị phân và kích hoạt tại thời điểm chạy và khi tính toán các gradient của tham số trong quá trình đào tạo. Tương tự như nhị phân và Mạng XNOR, các mạng này cũng giảm dấu chân bộ nhớ tới 32 lần, đồng thời cũng giảm khả năng truy cập bộ nhớ theo cùng một hệ số số lượng. Courbariaux, Bengio, và David (2015), Cheng, Soudry, Mao, và Lan (2015) và Kim và Smaragdis (2016) đã thực hiện các nỗ lực lượng tử hóa gần đây khác theo hướng này cũng giới thiệu trọng số nhị phân và kích hoạt trong quá trình đào tạo và suy luận. DCNN với lượng tử hóa trọng số được kỳ vọng sẽ cải thiện đáng kể hiệu quả năng lượng vì chúng thực hiện các hoạt động từng bit thay vì các hoạt động số học thông thường và tốc độ của chúng tính toán và giảm đáng kể dấu chân bộ nhớ làm cho chúng tốt phù hợp cho việc triển khai di động. Điều còn phải điều tra là liệu chúng làm giảm hoặc cải thiện hiệu suất mô hình trên các vấn đề mở khác như như những điều đã giới thiệu trong phần 6.2 và 6.4. Hơn nữa, khả năng biểu đạt của các mạng nhị phân cũng đáng ngờ và điều này đã dẫn đến

phát triển các mô hình dựa trên lượng tử hóa khác như trọng số ba phần mạng (Li, Zhang, & Liu, 2016).

Gần đây Lavin và Gray (2016) cũng đề xuất một lớp DCNN mới thuật toán dựa trên thuật toán lọc tối thiểu của Winograd (Winograd, 1980) và đạt được tốc độ tăng từ 1,48 đến 7,42 lần khi so sánh với mô hình cơ sở của họ từ Simonyan và Zisserman (2014). Tuy nhiên, mô hình của họ thay đổi lên đến kích thước tối đa là 16 MB, lớn hơn hơn 11,3 MB đạt được bởi Han, Mao và Dally (2016) cho cùng một Đường cơ sở VGG-net (Simonyan & Zisserman, 2014). Một phát triển gần đây khác cho phép đánh giá hiệu suất của một số DCNN trên điện thoại thông minh sử dụng phân tích ma trận Bayesian, phân tích Tucker (xem Tucker, 1966) để nén toàn bộ các lớp DCNN và tinh chỉnh để bù đắp lại các tổn thất độ chính xác tích lũy. Kim et al. (2015).

Về mặt phần cứng, Qiao et al. (2016) đã xây dựng một hệ thống FPGA nguyên mẫu để tăng tốc DCNN và đạt được tốc độ tăng 3,54 lần và hiệu quả năng lượng tăng 7,4 lần so với triển khai CPU và GPU cơ bản. Để thực hiện suy luận về mô hình nén được đề xuất bởi Han et al. (2015), Han et al. (2016) đã đề xuất một động cơ tiết kiệm năng lượng khai thác các kỹ thuật nén được đề xuất trước đó (Han et al., 2016) và do đó đã đạt được tăng đáng kể về mặt tính toán và năng lượng. Tiếp tục với mô hình và phương pháp tiếp cận song song dữ liệu được giới thiệu bởi Krizhevsky (2014), Dettmers (2016) gần đây đã xây dựng một cụm GPU bao gồm 96 đơn vị và cho thấy rằng bằng cách nén các gradient và kích hoạt phi tuyến tính thành các biểu diễn 8 bit, tốc độ tăng lên tới 50 lần có thể đạt được. Do đó, mặc dù nó là có thể kết luận rằng sự phát triển phần cứng mới và cải tiến các thuật toán đặc biệt xem xét kiến trúc phần cứng sẽ thúc đẩy những tiến bộ trong tương lai về yêu cầu tính toán và lưu trữ DCNN, lý do đằng sau quá trình đào tạo ban đầu của các mô hình dự phòng cũng đòi hỏi phải điều tra thêm; nếu điều này có thể được giải quyết trước tiên, thì cần phải chống lại hậu quả của nó có thể được thư giãn. Trong thời gian tạm thời, mặc dù các phương pháp được nêu bật trong phần này cho thấy triển vọng, các cuộc điều tra thực nghiệm sâu hơn và động lực lý thuyết để thiết lập chúng một cách chắc chắn, đặc biệt là trên dữ liệu thực tế đa dạng bộ sưu tập được thu thập trên thiết bị di động là bắt buộc.

Bằng chứng

Chưa sửa

6.4 Lỗi sâu. Trong số những thách thức mở, có lẽ điều hấp dẫn nhất là độ chính xác phân loại của DCNN và các bộ phân loại nói chung không mạnh mẽ khi đối mặt với các ví dụ đối nghịch. Đây là những những nhiễu loạn cố ý được áp dụng cho hình ảnh với mục đích gây hiểu lầm hoặc đánh lừa hệ thống nhận dạng hoặc phân loại. Khi những nhiễu loạn này được sử dụng để thay đổi một hình ảnh, con người có thể dễ dàng phân loại hình ảnh một cách chính xác (Goodfellow, Shlens, & Szegedy, 2015; Ulicný, Lundström, & Byttner, 2016), trong khi các bộ phân loại coi hình ảnh là từ một lớp khác.
Kể từ khi phát hiện ra hiện tượng này lần đầu tiên (Szegedy và cộng sự, 2014), một số bài báo đã xác nhận tính dễ bị tổn thương của DCNN đối với những hình ảnh này và

đề xuất một số biện pháp đối phó khả thi để giảm thiểu chúng (Szegedy et al., 2014; Goodfellow và cộng sự, 2015; Gu và Rigazio, 2014; Maharaj, 2015; Triệu & Griffin, 2016; Ulicný và cộng sự, 2016; Jin, Dundar và Bridge, 2016; Thuốc lá và Valle, 2016; Miyato, Maeda, Koyama, Nakae và Ishii, 2016; Hương vị, hỗn loạn, Faghri, & Fleet, 2016; Papernot, McDaniel, Xu, Jha, & Swami, 2016; Huang, Xu, Schuurmans và Szepesvári, 2016).

Cho đến nay những nỗ lực đầy hứa hẹn nhất để giải quyết vấn đề này tập trung vào các các kỹ thuật đào tạo như đào tạo đối kháng (Goodfellow et al., 2015) và chưng cất (Papernot et al., 2016), phương pháp tiền xử lý tạo ra chẳng hạn như việc sử dụng bộ mã hóa tự động khử nhiễu (Ulicný et al., 2016) và thay đổi kiến trúc DCNN để làm cho nó phi tuyến tính hơn hoặc để phạt những điều bất thường tín hiệu (Zhao & Griffin, 2016; Jin et al., 2016). Hơn nữa, một lý thuyết khuôn khổ để chính thức điều tra tính mạnh mẽ cũng đã được giới thiệu và các giới hạn cơ bản về độ mạnh mẽ của các bộ phân loại được chọn, liên quan đến biện pháp phân biệt giữa các lớp, đã được thiết lập (Fawzi, Fawzi và Frossard, 2015a, 2015b). Hơn nữa, Bastani và cộng sự (2016) gần đây đã đề xuất hai biện pháp thống kê, tần suất đối nghịch và mức độ nghiêm trọng đối nghịch, để đo lường tính mạnh mẽ dựa trên khái niệm chính thức về tính mạnh mẽ từng điểm, được mã hóa dưới dạng tối ưu hóa bị ràng buộc vấn đề. Điều thú vị là các mạng RBF đã được chứng minh là có bản chất miễn nhiễm với các ví dụ đối nghịch (Goodfellow và cộng sự, 2015), và do đó kết hợp các tính năng của chúng với kiến trúc DCNN tiêu chuẩn, giống như công trình do Zhao và Griffin (2016) đề xuất, vẫn là một cách tiếp cận phù hợp đòi hỏi điều tra thêm. Gần đây, việc phân biệt hình ảnh đã mang lại kết quả cải thiện hiệu suất đối nghịch (Maharaj, 2015); tuy nhiên, hiệu ứng mạnh mẽ của các kỹ thuật giảm chiều khác trên dữ liệu đầu vào và các vectơ đặc trưng từ các lớp DCNN khác nhau vẫn chưa được nghiên cứu.

Bổ sung cho việc tìm ra các ví dụ đối nghịch, Nguyen, Yosin-ski, & Clune (2015) nhận thấy rằng có thể tạo ra các hình ảnh hoàn toàn không thể nhận ra đối với con người nhưng DCNN hiện đại phân loại như những vật thể dễ nhận biết—một cách đáng báo động, với sự tự tin cực kỳ cao. Những hình ảnh được tạo ra bằng cách sử dụng các thuật toán tiến hóa và dựa trên gradient tối ưu hóa và được gọi là hình ảnh đánh lừa. Đề cập đến những hình ảnh này là từ một lớp rác, Goodfellow et al. (2015) đã chỉ ra rằng chúng có thể được tạo ra bằng những cách hiệu quả hơn và chúng không chỉ ảnh hưởng đến mạng lưới sâu (nghiên cứu này bao gồm DCNN) nhưng cũng có bộ phân loại nông. Công việc bổ sung và cập nhật của Zhao và Griffin (2016) được tham chiếu với những hình ảnh này như những hình ảnh vô nghĩa và bắt đầu công việc cho thấy rằng DCNN thay đổi kiến trúc là một giải pháp khả thi để khắc phục lỗi này; tuy nhiên, giống như với các ví dụ đối nghịch, chúng vẫn là một thách thức liên tục đòi hỏi sự chú ý hơn nữa.

Vì những hành vi đối đầu này (Szegedy và cộng sự, 2014) và lừa dối (Nguyen và cộng sự, 2015) hình ảnh làm nổi bật khoảng cách lớn giữa khả năng thị giác của con người và hệ thống thị giác máy tính, tìm thấy những đặc tính hấp dẫn này đã đưa ra một số câu hỏi liên quan đến khái quát, chức năng

xấp xỉ và các tính năng bảo mật của mạng sâu (Szegedy và cộng sự, 2014; Goodfellow và cộng sự, 2015; Gu và Rigazio, 2014; Maharaj, 2015; Triệu và Griffin, 2016; Ulicný et al., 2016; Jin et al., 2016; Tabacof & Valle, 2016; Papernot et al., 2016). Điều này đã mở ra một lĩnh vực nghiên cứu hoàn toàn mới tập trung vào việc tạo ra những hình ảnh này và thiết kế các hệ thống mạnh mẽ chống lại chúng. Đương nhiên, vì DCNN đã trở thành kiến trúc hàng đầu cho các nhiệm vụ trực quan và xét đến thực tế là chúng không miễn nhiễm với cả hình ảnh đối nghịch và hình ảnh lừa đảo, nghiên cứu về tính mạnh mẽ của chúng là một bước tiến quan trọng vấn đề.

6.5 Hình ảnh đa nhãn và hiểu biết sâu hơn về hình ảnh. Thậm chí mặc dù DCNN đã vượt qua hiệu suất ở cấp độ con người trên nhãn đơn bộ dữ liệu hình ảnh như MNIST (LeCun et al., 1998; Ciresan et al., 2012; Wan et al., 2013) và ImageNet (Russakovsky et al., 2015; Ioffe & Szegedy, 2015; Szegedy, Vanhoucke và cộng sự, 2015; Ông và cộng sự, 2015a, 2015b; Szegedy và cộng sự, 2016) các tập dữ liệu, hình ảnh thực tế thường chứa nhiều nhãn, liên quan đến các đối tượng, bộ phận, cảnh, hành động khác nhau và tương tác của chúng hoặc thuộc tính (Wang và cộng sự, 2016). Hơn nữa, khả năng mô tả chính xác nội dung ngữ nghĩa của một hình ảnh, trong các câu ngôn ngữ tự nhiên được hình thành đúng cách, là một vấn đề đầy thách thức (Vinyals, Toshev, Bengio, & Erhan, 2015), nằm ở giao điểm giữa tầm nhìn máy tính và tự nhiên xử lý ngôn ngữ. Để giải quyết những vấn đề này, một xu hướng gần đây là kết hợp DCNN với RNN. Để giải quyết vấn đề phân loại đa nhãn, Wang et al. (2016) đã trình bày một khuôn khổ DCNN-RNN trong đó DCNN trích xuất các biểu diễn ngữ nghĩa từ hình ảnh, trong khi RNN mô hình hóa mối quan hệ hình ảnh-nhãn và sự phụ thuộc nhãn. Cả Vinyals và cộng sự (2015) và Karpathy và Fei-Fei (2016) đã sử dụng DCNN để phân loại hình ảnh và RNN để mô hình hóa chuỗi và kết hợp chúng thành một mạng thống nhất, mà họ sử dụng để tạo ra các mô tả bằng tiếng Anh về hình ảnh. Một hướng triển vọng khác là đào tạo các kiến trúc kết hợp này bằng cách sử dụng học tăng cường, và mặc dù các hệ thống kết hợp học sâu và học tăng cường vẫn còn trong giai đoạn đầu (LeCun et al., 2015), chúng đã đã tạo ra một số kết quả phân loại hình ảnh đặc biệt (Ba, Mnih, & Kavukcuoglu, 2015).

6.6 Các xu hướng và thách thức được lựa chọn khác. Như đã đề cập trong phần 4.2.1, mặc dù có sự đóng góp của việc đào tạo trước không giám sát vào sâu thời kỳ phục hưng học tập (Hinton và cộng sự, 2006; Hinton & Salakhutdinov, 2006; Bengio và cộng sự, 2007), các DCNN hiện tại chủ yếu sử dụng mô hình học có giám sát; do đó, chúng không thể khai thác được lượng lớn dữ liệu chưa được gắn nhãn có sẵn trên Internet, được lưu trữ trong các hệ thống dựa trên đám mây hoặc thậm chí được ghi lại bằng các thiết bị di động. Hơn nữa, việc học của con người về bản chất là không được giám sát (LeCun và cộng sự, 2015), và do đó, người ta mong đợi rằng tương lai Các mô hình DCNN sẽ cố gắng mô phỏng thiên nhiên nhiều hơn các mô hình hiện tại của chúng ta. Những nỗ lực gần đây theo hướng này bao gồm công trình của Goodfellow và cộng sự.

(2014), Kingma và Welling (2014), Bengio, Thibodeau-Laufer, Alain và Yosinski (2014), Kulkarni, Whitney, Kohli và Tenenbaum (2015) và nhiều hơn nữa gần đây, Bachman (2016), tất cả đều sử dụng phương pháp dựa trên thế hệ đầy hứa hẹn kỹ thuật mô hình hóa.

DCNN yêu cầu một số siêu tham số, chẳng hạn như số kỷ nguyên để chạy mô hình và con chuột học tập; tuy nhiên, việc xác định chúng đòi hỏi phải điều chỉnh cẩn thận, thường dựa trên kinh nghiệm của chuyên gia, các quy tắc của ngón tay cái, hoặc các phương pháp tìm kiếm quá tốn kém về mặt tính toán. Bằng cách sử dụng kỹ thuật tối ưu hóa Bayesian tự động, Snoek et al. (2012) đã có thể tìm thấy siêu tham số tốt hơn nhanh hơn so với một chuyên gia con người đã làm và thu được kết quả phân loại hình ảnh tuyệt vời. Tuy nhiên, phương pháp này tốn thời gian và không mở rộng tốt với các mô hình lớn (Srinivas và cộng sự, 2016); do đó, cần có các giải pháp thay thế theo hướng này. Một giải pháp khả thi có thể là sử dụng các thuật toán tiến hóa, chẳng hạn như tối ưu hóa bầy hạt (Kennedy & Eber-hart, 1995), đã trở thành một kỹ thuật tối ưu hóa phổ biến, để tiến hành tìm kiếm siêu tham số và sau đó tích hợp kết quả với DCNN. Nếu thành công, điều này sẽ chứng kiến hai kỹ thuật phổ biến lấy cảm hứng từ sinh học hoạt động cùng nhau. Một hướng thú vị nhưng đầy thách thức khác là tận dụng khả năng phân loại phân biệt và biểu cảm của DCNN trong các hệ thống rô-bốt trực tuyến, với một số thành công gần đây được mô tả trong bởi Pinto và Gupta (2015), Finn và cộng sự (2015), và Levine, Finn, Darrell, và Abbeel (2016).

Một xu hướng khác đang phát triển mạnh mẽ là việc phân tích các phép tích chập để cải thiện hiệu quả tính toán. Kỹ thuật này đã được phổ biến bởi mô hình Inception đã được sửa đổi (Szegedy, Vanhoucke và cộng sự, 2015), trong đó phát hiện ra rằng n × n tích chập có thể được phân tích thành 1 × 1 theo sau bằng n × 1 phép tích chập. Ví dụ, họ thấy rằng 3 × 1 phép tích chập tiếp theo là 1 × 3 phép tích chập dẫn đến giảm 33% phép tính trong so sánh với việc sử dụng một bộ lọc 3 × 3 duy nhất có cùng khả năng tiếp nhận hiệu quả kích thước trường. Lấy cảm hứng từ điều này, Chollet (2016) đã đề xuất kiến trúc Xception trong đó họ thay thế các khối xây dựng Inception bằng chiều sâu tích chập có thể tách rời (tích chập theo chiều sâu theo sau là tích chập theo điểm tích chập). Tích chập theo chiều sâu cũng được khai thác bởi Xie, Girshick, Dollar và He (2016), những người lặp lại một khối xây dựng còn lại kết hợp một loạt các phép biến đổi có cùng một cấu trúc. Hiệu quả đạt được khi sử dụng loại phân tích này, đặc biệt là khi chúng ta tiến hành hướng tới việc triển khai DCNN di động, có tầm quan trọng đáng kể và sẽ có thể được sử dụng rộng rãi trong các mô hình tương lai.

Thách thức quan trọng nhất là thu hẹp khoảng cách lý thuyết giữa mạng nơ-ron sinh học và DCNN, và mặc dù phân tích lý thuyết mới của Bengio, Mesnard, Fischer, Zhang và Wu (2017) không cụ thể là giải quyết các DCNN, động lực của họ về cách bộ não sinh học thực hiện việc phân công tín dụng trong các hệ thống phân cấp sâu, có lẽ cũng thành thạo như Ngược lại, sự lan truyền ngược có thể được coi là một bước quan trọng hướng tới việc liên kết các mô hình tính toán sâu của chúng ta với các cơ chế của não bộ con người.

7 Kết luận

Đánh giá này trình bày một đánh giá toàn diện về CNN cho các nhiệm vụ phân loại hình ảnh. Nó phân loại sự tiến triển của chúng vào giai đoạn phát triển ban đầu, đóng góp của họ vào thời kỳ phục hư ng học sâu và sự tiến bộ nhanh chóng của họ trong vài năm qua. Đặc biệt, nó tập trung vào sự tiến bộ của họ bằng cách cân nhắc và phân tích hầu hết các tiến bộ đáng chú ý liên quan đến kiến trúc, thành phần giám sát, cơ chế điều chỉnh của chúng, kỹ thuật tối ư u hóa và tính toán từ năm 2012. Mặc dù thành công trong các miền khác, DCNN đã chứng kiến sự tiến bộ đáng kể trong các nhiệm vụ phân loại hình ảnh, thiết lập trạng thái nghệ thuật trên một số phân loại đầy thách thức chuẩn mực và thống trị nhiều thách thức và cuộc thi liên quan đến phân loại hình ảnh. Trên thực tế, trên một số phân loại hình ảnh nhãn đơn theo chuẩn mực, hiệu suất của chúng đã vư ợt qua hiệu suất ở cấp độ con ngư ời. Tuy nhiên, sự gia tăng hiện đại của DCNN đã khiến các nhà nghiên cứu phải xem xét kỹ lư ỡng hiệu suất phân loại, độ mạnh mẽ và đặc điểm tính toán của chúng, dẫn đến việc phát hiện ra một số thách thức và xu hư ớng giải quyết chúng. Theo đó, bài đánh giá này cũng tóm tắt lại những vấn đề mở này và các xu hư ớng liên quan của chúng và quan trọng nhất là đư a ra một số khuyến nghị và định hư ớng nghiên cứu cho việc khám phá trong tư ơng lai.

Lời cảm ơn

Công trình này đư ợc hỗ trợ một phần bởi Quỹ tài trợ khuyến khích nghiên cứu quốc gia Nam Phi số 81705.

Tài liệu tham khảo

Abdulkader, A. (2006). Nhận dạng chữ viết tay tiếng Ả Rập ngoại tuyến hai tầng dựa trên quy tắc tham gia có điều kiện. Trong Biên bản Hội thảo quốc tế lần thứ 10 về các tiền tuyến trong nhận dạng chữ viết tay (trang 1–6). Los Alamitos, CA: IEEE Computer Xã hội.

Agostinelli, F., Hoffman, M., Sadowski, P., & Baldi, P. (2014). Học các hàm kích hoạt để cải thiện mạng nơ-ron sâu. arXiv:1412.6830.

Ahmed, A., Yu, K., Xu, W., Gong, Y., & Xing, E. (2008). Đào tạo các mô hình nhận dạng hình ảnh truyền trực tiếp phân cấp bằng cách sử dụng học chuyển giao từ các nhiệm vụ giả. Trong Biên bản báo cáo của Hội nghị châu Âu về Tầm nhìn máy tính (trang 69–82). Berlin: Mùa xuân.

Arora, S., Bhaskara, A., Ge, R., & Ma, T. (2014). Giới hạn có thể chứng minh để học một số biểu diễn sâu sắc. Trong Biên bản Hội nghị quốc tế lần thứ 31 về Máy móc Học tập (trang 584–592). Np: Hiệp hội học máy quốc tế.

Ba, J., & Frey, B. (2013). Dropout thích ứng để đào tạo mạng nơ-ron sâu. Trong C. JC Burges, L. Bottou, M. Welling, Z. Ghahramani, & KQ Weinberger (Biên tập viên), Tiến bộ trong hệ thống xử lý thông tin thần kinh, 26 (trang 3084–3092). Red Hook, NY: Hiện tại.

Ba, J., Mnih, V., & Kavukcuoglu, K. (2015). Nhận dạng nhiều đối tư ợng với sự chú ý thị giác. Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 3 về Biểu diễn học tập (trang 1–10). Np: Hiệp hội học tập tính toán và sinh học.

Bachman, P. (2016). Một kiến trúc cho các mô hình sinh sản phân cấp sâu. Trong D. Lee, M. Sugiyama, UV Luxburg, I. Guyon, & R. Garnett (Biên tập viên), Những tiến bộ trong thần kinh hệ thống xử lý thông tin, 29 (trang 4826–4834). Red Hook, NY: Curran.

Baldi, P., & Sadowski, PJ (2013). Hiểu về tình trạng bỏ học. Trong CJC Burges, L. Bot-tou, M. Welling, Z. Ghahramani, & KQ Weinberger (Biên tập), Những tiến bộ trong thần kinh hệ thống xử lý thông tin, 26 (trang 3084–3092). Red Hook, NY: Curran.

Baldi, P., & Sadowski, P. (2014). Thuật toán học bỏ học. Trí tuệ nhân tạo, 210, 78–122.

Barkan, O., Weill, J., Wolf, L., và Aronowitz, H. (2013). Vector chiều cao nhanh nhận dạng khuôn mặt nhân. Trong Biên bản báo cáo của Hội nghị quốc tế IEEE về Tầm nhìn máy tính (trang 1960–1967). Red Hook, NY: Curran.

Bastani, O., Ioannou, Y., Lampropoulos, L., Vytiniotis, D., Nori, A., & Criminisi, A. (2016). Đo lư ờng độ mạnh của mạng nơ-ron với các ràng buộc. Trong D. Lee, M. Sugiyama, UV Luxburg, I. Guyon, & R. Garnett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 29 (trang 2613–2621). Red Hook, NY: Curran.

Basu, S., Karki, M., DiBiano, R., Mukhopadhyay, S., Ganguly, S., Nemani, R., & Gayaka, S. (2016). Phân tích lý thuyết về mạng nơ-ron sâu để phân loại kết cấu. Bản in trư ớc của ArXiv: 1605.02699.

Belkin, M., Niyogi, P., & Sindhwani, V. (2006). Chính quy hóa đa tạp: Một hình học khuôn khổ để học từ các ví dụ có nhãn và không có nhãn. Tạp chí Máy móc Nghiên cứu học tập, 7, 2399–2434.

Bell, S., & Bala, K. (2015). Học tính tư ơng đồng trực quan để thiết kế sản phẩm với mạng nơ-ron tích chập. ACM Transactions on Graphics, 34(4), 98–107.

Bengio, Y. (2009). Học kiến trúc sâu cho AI. Nền tảng và xu hư ớng trong Ma- Trung Quốc Học tập, 2(1), 1–127.

Bengio, Y. (2013). Học sâu về biểu diễn: Nhìn về phía trư ớc. Trong Biên bản của Hội nghị quốc tế về xử lý ngôn ngữ và lời nói thống kê (trang 1–37). Berlin: Springer.

Bengio, Y., Courville, A., & Vincent, P. (2013). Học tập biểu diễn: Một đánh giá và góc nhìn mới. Giao dịch IEEE về Phân tích Mẫu và Trí tuệ Máy móc, 35(8), 1798–1828.

Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Đào tạo từng lớp tham lam của mạng lư ới sâu. Trong JC Platt, D. Koller, Y. Singer, & ST Roweis (Biên tập viên), Tiến bộ trong hệ thống xử lý thông tin thần kinh, 19 (trang 2814–2822). Red Hook, NY: Hiện tại.

Bengio, Y., Mesnard, T., Fischer, A., Zhang, S., & Wu, Y. (2017). Sự xấp xỉ tư ơng thích STDP của sự lan truyền ngư ợc trong một mô hình dựa trên năng lư ợng. Tính toán thần kinh, 29(3), 555–577.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Học tập sự phụ thuộc lâu dài với giảm dần độ dốc là khó. Giao dịch IEEE về mạng nơ-ron, 5(2), 157–166.

Bengio, Y., Thibodeau-Laufer, E., Alain, G., & Yosinski, J. (2014). Sinh sản sâu mạng ngẫu nhiên có thể đào tạo bằng backprop. Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 31 Hội nghị Machine Learning (trang 226–234). Np: International Machine Learning Xã hội.

Bottou, L. (1998). Học trực tuyến và xấp xỉ ngẫu nhiên. Học trực tuyến
trong Mạng nơ-ron, 17(9), 142–177.

Bottou, L. (2010). Học máy quy mô lớn với độ dốc ngẫu nhiên giảm dần. Trong
Biên bản Hội nghị quốc tế về Thống kê tính toán (trang 177–186).
Berlin: Nhà xuất bản Vật lý Heidelberg.

Boureau, Y., Ponce, J., và LeCun, Y. (2010). Phân tích lý thuyết về tính năng
tập hợp trong nhận dạng trực quan. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 27 về Học máy
(trang 111–118). Np: Học máy quốc tế
Xã hội.

Bromley, J., Bentz, J.W., Bottou, L., Guyon, I., LeCun, Y., Moore, C., ... Shah,
R. (1993). Xác minh chữ ký bằng cách sử dụng mạng nơ-ron trễ thời gian "Siamese".
Tạp chí quốc tế về nhận dạng mẫu và trí tuệ nhân tạo, 7(4), 669–
688.

Bulo, S., & Kontschieder, P. (2014). Rừng quyết định thần kinh cho việc dán nhãn hình ảnh ngữ nghĩa.
Trong Biên bản báo cáo của Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 81–88).
Los Alamitos, CA: IEEE Computer Society.

Bruna, J., Szlam, A., & LeCun, Y. (2013). Phục hồi tín hiệu từ các biểu diễn gộp.
Bản in trư ớc của ArXiv:1311.4025.

Cao, X., Wipf, D., Wen, F., Duan, G., & Sun, J. (2013). Một quá trình học chuyển giao thực tế
thuật toán xác minh khuôn mặt. Trong Biên bản báo cáo của Hội nghị quốc tế IEEE
về Tầm nhìn máy tính (trang 3208–3215). Red Hook, NY: Curran.

Chang, J., & Chen, Y. (2015). Mạng maxout chuẩn hóa theo lô trong mạng. ArXiv
bản in trư ớc:1511.02583.

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Sự trở lại của quỷ dữ
trong chi tiết: Đi sâu vào mạng lư ới tích chập. Bản in trư ớc của ArXiv:1405.3531.

Chellapilla, K., & Puri, S., & Simard, P. (2006). Mạng nơ-ron tích chập hiệu suất cao để xử lý tài
liệu. Trong Biên bản Hội thảo quốc tế lần thứ 10 về Biên giới trong Nhận dạng chữ viết tay. Los
Alamitos, CA: IEEE Computer
Xã hội.

Chellapilla, K., Shilman, M., & Simard, P. (2006). Kết hợp tối ư u một chuỗi
của các bộ phân loại. Trong Biên bản báo cáo của Hội nghị chuyên đề thư ờng niên lần thứ 18 về Hình ảnh điện tử
(trang 6067–6126). Np: Hiệp hội Kỹ thuật Quang học Quốc tế.

Chellapilla, K., & Simard, P. (2006). Phư ơng pháp tiếp cận hai tầng cho chữ viết tay ngoại tuyến tiếng Ả Rập
nhận dạng. Trong Biên bản Hội thảo quốc tế lần thứ 10 về Biên giới trong Nhận dạng chữ viết tay
(trang 1–6). Los Alamitos, CA: IEEE Computer Society.

Chen, D., Cao, X., Wen, F., & Sun, J. (2013). Phư ớc lành của tính đa chiều: Tính năng đa chiều và
nén hiệu quả của nó để xác minh khuôn mặt. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn
máy tính và Nhận dạng mẫu (trang 3025–
3032). Red Hook, NY: Hiện tại.

Chen, W., Wilson, JT, Tyree, S., Weinberger, KQ, & Chen, Y. (2015). Nén
mạng nơ-ron với thủ thuật băm. Bản in trư ớc của ArXiv:1504.04788v1.

Cheng, Y., Felix, XY, Feris, RS, Kumar, S., Choudhary, A., & Chang, S. (2015). Nhanh
mạng lư ới nơ-ron với phép chiếu tuần hoàn. Bản in trư ớc của ArXiv:1502.03436.

Cheng, Y., Yu, FX, Feris, RS, Kumar, S., Choudhary, A., & Chang, S. (2015). Một cuộc khám phá về sự
dư thừa tham số trong các mạng sâu với các phép chiếu tuần hoàn.
Trong Biên bản Hội nghị quốc tế IEEE về Tầm nhìn máy tính (trang 2857–
2865). Red Hook, NY: Hiện tại.

Cheng, Z., Soudry, D., Mao, Z., & Lan, Z. (2015). Đào tạo lớp nhị phân đa lớp mạng nơ-ron để phân loại hình ảnh sử dụng sự lan truyền ngược kỳ vọng. arXiv bản in trước:1503.03562.

Chollet, F. (2016). Xception: Học sâu với tích chập tách biệt theo chiều sâu. arXiv bản in trước:1610.02357.

Chopra, S., Hadsell, R., & LeCun, Y. (2005). Học một phép đo độ tương đồng một cách phân biệt, với ứng dụng để xác minh khuôn mặt. Trong Biên bản Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 539–546). Los Alamitos, CA: IEEE Hội máy tính.

Choromanska, A., Henaff, M., Mathieu, M., Arous, GB, & LeCun, Y. (2015). Các bề mặt mất mát của mạng nhiều lớp. Trong Kỷ yếu Hội nghị quốc tế lần thứ 18 về Trí tuệ nhân tạo và Thống kê (trang 192–204). www.jmlr.org/proceedings /papers/v38/choromanska15.pdf

Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010). Sâu, lớn, mạng nơ-ron đơn giản để nhận dạng chữ số viết tay. Tính toán nơ-ron, 22(12), 3207–3220.

Ciresan, D.C., Meier, U., Masci, J., Maria Gambardella, L., & Schmidhuber, J. (2011). Mạng nơ-ron tích chập hiệu suất cao, linh hoạt để phân loại hình ảnh. Trong Biên bản báo cáo của Hội nghị chung quốc tế về trí tuệ nhân tạo (tập 1, trang 1237–1242). Menlo Park, CA: Nhà xuất bản AAAI.

Ciresan, D., Meier, U., & Schmidhuber, J. (2012). Mạng nơ-ron sâu nhiều cột để phân loại hình ảnh. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng Mẫu (trang 3642–3649). Red Hook, NY: Curran.

Clevert, D., Unterthiner, T., & Hochreiter, S. (2016). Mạng sâu nhanh và chính xác học tập theo đơn vị tuyến tính mũ (ELU). Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 4 Hội nghị về Biểu diễn Học tập (trang 1–14). Np: Hội Học tập Tính toán và Sinh học.

Coates, A., Lee, H., & Ng, AY (2011). Phân tích mạng lưới một lớp trong học tập tính năng không được giám sát. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 14 về Trí tuệ nhân tạo và Thống kê (trang 215–223). www.jmlr.org/proceedings/papers /v15/coates11a/coates11a.pdf

Collobert, R., & Bengio, S. (2004). Một Hessian nhẹ nhàng để giảm độ dốc hiệu quả. Trong Biên bản Hội nghị quốc tế IEEE về Âm học, Giọng nói và Tín hiệu Xử lý (trang 517–520). Np: Hiệp hội xử lý tín hiệu IEEE.

Collobert, R., Sinz, F., Weston, J., & Bottou, L. (2006). SVM chuyển đổi quy mô lớn. Tạp chí nghiên cứu máy học, 7, 1687–1712.

Courbariaux, M., Bengio, Y., & David, JP (2015). BinaryConnect: Đào tạo sâu mạng nơ-ron với trọng số nhị phân trong quá trình truyền bá. Trong C. Cortes, ND Lawrence, DD Lee, M. Sugiyama, & R. Garnett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 28 (trang 3123–3131). Red Hook, NY: Curran.

Courbariaux, M., Hubara, I., Soudry, D., & Yaniv, RE (2016). Mạng nơ-ron nhị phân. Trong D. Lee, M. Sugiyama, UV Luxburg, I. Guyon, & R. Gar-nett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 29 (trang 1–9). Np: Tiền tố tụng.

Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., . Le, QW (2012). Mạng lưới sâu phân tán quy mô lớn. Trong F. Pereira, CJC Burges, L. Bottou,

& KQ Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 25 (trang 1223–1231). Red Hook, NY: Curran.

Decoste, D., & Schölkopf, B. (2002). Đào tạo máy vectơ hỗ trợ bất biến. Ma-
Học tập Trung Quốc, 46(1–3), 161–190.

Deng, L. (2014). Một khảo sát hư ớng dẫn về kiến trúc, thuật toán và ứng dụng cho học sâu. Giao dịch APSIPA về Xử lý tín hiệu và thông tin, 3(2), 1–29.

Deng, L., & Yu, D. (2014). Học sâu: Phư ơng pháp và ứng dụng. Nền tảng và
Xu hư ớng trong Xử lý tín hiệu, 7(3–4), 197–387.

Denil, M., Shakibi, B., Dinh, L., & de Freitas, N. (2013). Dự đoán các tham số trong học sâu. Trong CJC Burges, L. Bottou, M. Welling, Z. Ghahramani, & KQ
Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 26 (trang 2148–2156). Red Hook, NY: Curran.

Denton, EL, Zaremba, W., Bruna, J., LeCun, Y., & Fergus, R. (2014). Khai thác cấu trúc tuyến tính trong mạng tích chập để đánh giá hiệu quả. Trong Z. Ghahra-mani, M. Welling, C. Cortes, ND Lawrence, & KQ Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 27 (trang 1269–1277). Red Hook, NY: Curran.

Dettmers, T. (2016). Xấp xỉ 8 bit cho tính song song trong học sâu. Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 4 về Biểu diễn học tập (trang 1–14). Np: Computational and Biological Learning Society.

Dreyfus, S. (1962). Giải pháp số của các bài toán biến phân. Tạp chí Toán học
Phân tích và ứng dụng khoa học, 5(1), 30–45.

Duchi, J., Hazan, E., & Singer, Y. (2011). Các phư ơng pháp subgradient thích ứng cho học trực tuyến và tối ư u hóa ngẫu nhiên. Tạp chí nghiên cứu học máy, 12, 2121–2159.

Dumoulin, V., & Visin, F. (2016). Hư ớng dẫn về số học tích chập cho học sâu.
Bản in trư ớc của ArXiv:1603.07285.

Erhan, D., Bengio, Y., Courville, A., Manzagol, PA, Vincent, P., & Bengio, S. (2010).
Tại sao đào tạo trư ớc không giám sát lại giúp ích cho việc học sâu? Tạp chí nghiên cứu học máy, 11, 625–660.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2012). Phân tích cảnh với học tính năng đa thang, cây tinh khiết và lớp phủ tối ư u. Bản in trư ớc ArXiv: 1202.2160.

Fawzi, A., Fawzi, O., & Frossard, P. (2015a). Phân tích độ mạnh của các bộ phân loại đối với các đối thủ nhiễu loạn sarial. Bản in trư ớc của ArXiv:1502.02590.

Fawzi, A., Fawzi, O., & Frossard, P. (2015b). Giới hạn cơ bản về sức mạnh đối kháng. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 32 về Học máy (trang 1–7). Np: Hiệp hội học máy quốc tế.

Fei-Fei, L. (2006). Chuyển giao kiến thức trong việc học nhận dạng các lớp đối tư ợng trực quan.
Trong Biên bản Hội nghị quốc tế lần thứ 4 về Phát triển và Học tập (trang 11–17). Np: IEEE Computational Intelligence Society.

Fei-Fei, L., Fergus, R., & Perona, P. (2006). Học một lần các loại đối tư ợng.
Giao dịch IEEE về Phân tích mẫu và Trí tuệ máy móc, 28(4), 594–611.

Filho, DP, Luiz, P., Oliveira, LS, & Britto, AS Jr. (2009). Cơ sở dữ liệu để nhận dạng loài rừng. Trong Biên bản báo cáo Hội nghị chuyên đề Brazil lần thứ XXII về Đồ họa máy tính và Xử lý hình ảnh (trang 1–2). Red Hook, NY: Curran.

Finn, C., Tan, XY, Duan, Y., Darrell, T., Levine, S., & Abbeel, P. (2015). Bộ mã hóa tự động không gian sâu cho việc học thị giác vận động. Bản in trư ớc ArXiv: 1509.06113.

Floreano, D., & Mattiussi, C. (2008). Trí tuệ nhân tạo lấy cảm hứng từ sinh học: Lý thuyết, phương pháp, và công nghệ. Cambridge, MA: Nhà xuất bản MIT.

Fukushima, K. (1979). Tự tổ chức của mạng nơ-ron cung cấp phản ứng bất biến theo vị trí. Trong Biên bản Hội nghị chung quốc tế lần thứ 6 về trí tuệ nhân tạo (tập 1, trang 291–293). San Francisco: Morgan Kaufmann.

Fukushima, K. (1980). Neocognitron: Một mô hình mạng nơ-ron tự tổ chức cho một cơ chế nhận dạng mẫu không bị ảnh hưởng bởi sự thay đổi vị trí. Biological Cy-bernetics, 36(4), 193–202.

Fukushima, K., & Miyake, S. (1982). Neocognitron: Một thuật toán mới cho mẫu nhận dạng có khả năng chịu được biến dạng và thay đổi vị trí. Nhận dạng mẫu, 15(6), 455–469.

Garcia, C., & Delakis, M. (2002). Một kiến trúc thần kinh để phát hiện khuôn mặt nhanh và mạnh mẽ. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 16 về Nhận dạng mẫu (trang 44–47). Los Alamitos, CA: Hiệp hội máy tính IEEE.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Phân cấp tính năng phong phú cho phát hiện đối tượng chính xác và phân đoạn ngữ nghĩa. Trong Biên bản báo cáo của IEEE Hội nghị về Tầm nhìn máy tính và Nhận dạng mẫu (trang 580–587). Red Hook, NY: Hiện tại.

Glorot, X., & Bengio, Y. (2010). Hiểu được khó khăn trong việc đào tạo mạng nơ-ron truyền thẳng sâu. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 13 về Trí tuệ nhân tạo và Thống kê (trang 249–256). jmlr.org/proceedings/papers/v9 /glorot10a/glorot10a.pdf

Glorot, X., Bordes, A., & Bengio, Y. (2011). Mạng lưới thần kinh chỉnh lưu thưa thớt sâu. TRONG Biên bản Hội nghị quốc tế lần thứ 14 về Trí tuệ nhân tạo và Thống kê (trang 315–323). www.jmlr.org/proceedings/papers/v15/glorot11a/glorot11a.pdf

Gong, Y., Liu, L., Yang, M., & Bourdev, L. (2014). Nén mạng tích chập sâu hoạt động bằng cách sử dụng lượng tử hóa vector. Bản in trước của ArXiv:1412.6115.

Gong, Y., Wang, L., Guo, R., & Lazebnik, S. (2014, tháng 9). Không có trật tự đa thang tập hợp các tính năng kích hoạt tích chập sâu. Trong Biên bản của Hội nghị châu Âu Hội nghị về Tầm nhìn máy tính (trang 392–407). Berlin: Springer.

Goodfellow, IJ, Bengio, Y., & Courville, A. (sắp xuất bản). Học sâu. Cambridge, HÔM NAY, Nhà xuất bản MIT.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Các mạng đối nghịch sinh sản. Trong Z. Ghahramani, M. Welling, C. Cortes, ND Lawrence, & KQ Weinberger (Biên tập viên), Những tiến bộ trong thông tin thần kinh hệ thống xử lý, 27 (trang 2672–2680). Red Hook, NY: Curran.

Goodfellow, IJ, Shlens, J., & Szegedy, C. (2014). Giải thích và khai thác các ví dụ đối lập. Trong Biên bản Hội nghị quốc tế lần thứ 3 về Biểu diễn học tập (trang 1–11). Np: Học tập tính toán và sinh học Xã hội.

Goodfellow, IJ, Warde-Farley, D., Mirza, M., Courville, AC và Bengio, Y. (2013). Mạng lưới Maxout. Trong Biên bản Hội nghị quốc tế lần thứ 30 về Học máy (trang 1319–1327). Np: Hiệp hội học máy quốc tế.

Graham, B. (2014). Phân số max-pooling. Bản in trước ArXiv: 1412.6071.

Grauman, K., & Darrell, T. (2005). Hạt nhân khớp kim tự tháp: Phân loại phân biệt với các tập hợp các đặc điểm hình ảnh. Trong Biên bản báo cáo của Hội nghị quốc tế IEEE về thị giác máy tính (trang 1458–1465). Red Hook, NY: Curran.

Griffin, G., Holub, A., & Perona, P. (2007). Bộ dữ liệu danh mục đối tượng Caltech-256. Pasadena: Đại học Công nghệ California.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... Wang, G. (2015). Gần đây tiến bộ trong mạng nơ-ron tích chập. Bản in trước của ArXiv:1512.07108.

Gu, S., & Rigazio, L. (2014). Hướng tới kiến trúc mạng nơ-ron sâu mạnh mẽ để chống lại ví dụ về sarial. Bản in trước của ArXiv:1412.5068.

Gulcehre, C., Cho, K., Pascanu, R., & Bengio, Y. (2014). Nhóm chuẩn mực học được cho mạng nơ-ron hồi quy và phản hồi sâu. Trong Biên bản báo cáo của Hội đồng châu Âu Hội nghị về Học máy và Nguyên lý và Thực hành Khám phá Kiến thức trong Cơ sở dữ liệu (trang 530–546). New York: Springer-Verlag.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, MS (2016). Học sâu để hiểu trực quan: Một bài đánh giá. Neurocomputing, 187, 27–48.

Hadsell, R., Chopra, S., & LeCun, Y. (2006). Giảm chiều bằng cách học ánh xạ bất biến. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 1735–1742). Los Alamitos, CA: IEEE Computer Xã hội.

Hadsell, R., Sermanet, P., Ben, J., Erkan, A., Scoffier, M., Kavukcuoglu, K., . LeJun, Y. (2009). Học tầm nhìn xa cho lái xe địa hình tự động. Tạp chí Robot thực địa, 26(2), 120–144.

Han, S., Liu, X., Mao, H., Pu, J., Pedram, A., Horowitz, MA, & Dally, WJ (2016). EIE: Công cụ suy luận hiệu quả trên mạng nơ-ron sâu nén. ArXiv bản in trước:1602.01528.

Han, S., Mao, H., & Dally, WJ (2015). Nén sâu: Nén thần kinh sâu mạng lưới với cắt tỉa, lượng tử hóa được đào tạo và mã hóa huffman. Trong Biên bản của Hội nghị quốc tế lần thứ 3 về Biểu diễn học tập (trang 1–14). Np: Hiệp hội học tập tính toán và sinh học.

Hardt, M., & Ma, T. (2016). Vấn đề nhận dạng trong học sâu. Bản in trước ArXiv:1611.04231.

Harris, KD, & Shepherd, GM (2015). Mạch tân vỏ não: Chủ đề và biến thể tions. Khoa học thần kinh tự nhiên, 18(2), 170–181.

He, K., & Sun, J. (2015). Mạng nơ-ron tích chập với chi phí thời gian hạn chế. Trong Biên bản Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 5353–5360). Red Hook, NY: Hiện tại.

He, K., Zhang, X., Ren, S., & Sun, J. (2014). Tập hợp kim tự tháp không gian trong mạng lưới tích chập sâu để nhận dạng trực quan. Trong Biên bản Hội nghị Châu Âu về Tầm nhìn máy tính (trang 346–361). Berlin: Springer.

He, K., Zhang, X., Ren, S., & Sun, J. (2015a). Học sâu dư thừa cho nhận dạng hình ảnh. Bản in trước ArXiv: 1512.03385.

He, K., Zhang, X., Ren, S., & Sun, J. (2015b). Đi sâu vào bộ chỉnh lưu: Vượt qua hiệu suất ở cấp độ con người về phân loại imagenet. Trong Biên bản báo cáo Hội nghị quốc tế IEEE về tầm nhìn máy tính (trang 1026–1034). Red Hook, NY: Curran.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Ánh xạ danh tính trong mạng lưới dư thừa sâu. Bản in trước của ArXiv:1603.05027.

Hinton, GE (1989). Các thủ tục học tập kết nối. Trí tuệ nhân tạo, 40(1), 185–234.

Hinton, GE (2002). Đào tạo các sản phẩm của chuyên gia bằng cách giảm thiểu sự phân kỳ tương phản gence. Tính toán thần kinh, 14(8), 1771–1800.

Hinton, GE, Osindero, S., & Teh, Y. (2006). Một thuật toán học nhanh cho niềm tin sâu sắc lưới. Tính toán thần kinh, 18(7), 1527–1554.

Hinton, GE, & Salakhutdinov, RR (2006). Giảm chiều của dữ liệu bằng mạng nơ-ron. Khoa học, 313(5786), 504–507.

Hinton, GE, Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, RR (2012). Cải thiện mạng lưới nơ-ron bằng cách ngăn chặn sự thích ứng đồng thời của các bộ phát hiện tính năng. Bản in trước arxiv:1207.0580.

Hinton, G., Vinyals, O., & Dean, J. (2015). Chắt lọc kiến thức trong mạng nơ-ron. Bản in trước của ArXiv:1503.02531.

Hochreiter, S., & Schmidhuber, J. (1997). Trí nhớ ngắn hạn dài. Tính toán thần kinh-tion, 9(8), 1735–1780.

Huang, FJ, & LeCun, Y. (2006). Học tập quy mô lớn với SVM và mạng tích chập để phân loại đối tượng chung. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 284–291). Red Hook, NY: Curran.

Huang, GB, Ramesh, M., Berg, T., & Learned-Miller, E. (2007). Khuôn mặt được gắn nhãn trong tự nhiên: Cơ sở dữ liệu để nghiên cứu nhận dạng khuôn mặt trong môi trường không bị hạn chế (tập 1, trang 3) (Báo cáo kỹ thuật 07-49). Amherst: Đại học Massachusetts.

Huang, G., Liu, Z., Weinberger, K. Q., & van der Maaten, L. (2016). Kết nối dày đặc mạng lưới tích chập. bản in trước arXiv:1608.06993.

Huang, G., Sun, Y., Liu, Z., Sedra, D., & Weinberger, KQ (2016). Mạng lưới sâu với độ sâu ngẫu nhiên. Trong Biên bản báo cáo của Hội nghị châu Âu về thị giác máy tính (trang 646–661). Heidelberg, Berlin: Springer.

Huang, R., Xu, B., Schuurmans, D., & Szepesvari, C. (2016). Học với đối thủ mạnh. Bản in trước của ArXiv:1511.03034v6.

Hubel, DH, & Wiesel, TN (1959). Các trường tiếp nhận của các tế bào thần kinh đơn lẻ trong vỏ não vân của mèo. Tạp chí Sinh lý học, 148(1), 574–591.

Hubel, DH, & Wiesel, TN (1962). Các trường tiếp nhận, tương tác hai mắt và kiến trúc chức năng trong vỏ não thị giác của mèo. Tạp chí Sinh lý học, 160(1), 106–154.

Hyvärinen, A., & Köster, U. (2007). Tập hợp tế bào phức tạp và thống kê tự nhiên hình ảnh. Mạng: Tính toán trong Hệ thống thần kinh, 18(2), 81–100.

Iandola, FN, Moskewicz, MW, Ashraf, K., Han, S., Dally, WJ và Keutzer, K. (2016). SqueezeNet: Độ chính xác cấp AlexNet với ít hơn 50 lần tham số và kích thước mô hình <1MB. Bản in trước của ArXiv: 1602.07360.

Ioffe, S. (2017). Chuẩn hóa lô: Hướng tới việc giảm sự phụ thuộc vào lô nhỏ trong lô mô hình chuẩn hóa. bản in trước arXiv:1702.03275.

Ioffe, S., & Szegedy, C. (2015). Chuẩn hóa theo lô: Tăng tốc đào tạo mạng sâu bằng cách giảm sự dịch chuyển biến phụ thuộc nội bộ. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 32 về Học máy (trang 448–456). Np: Hiệp hội học máy quốc tế.

Ivakhnenko, AG, & Lapa, VG (1966). Thiết bị dự đoán điều khiển học. New York: CCM Information Corp.

Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). Mạng lưới biến áp không gian. Trong C. Cortes, ND Lawrence, DD Lee, M. Sugiyama, & R. Garnett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 28 (trang 2017–2025). Red Hook, NY: Curran.

Jaderberg, M., Vedaldi, A., & Zisserman, A. (2014). Tăng tốc mạng nơ-ron tích chập với các phép mở rộng bậc thấp. Trong Biên bản báo cáo của Hội nghị về thị giác máy tính của Anh (trang 1–12). Durham, Vương quốc Anh: BMVA Press.

Jarrett, K., Kavukcuoglu, K., & Lecun, Y. (2009). Kiến trúc đa giai đoạn nào là tốt nhất để nhận dạng đối tượng? Trong Biên bản báo cáo Hội nghị quốc tế IEEE về thị giác máy tính (trang 2146–2153). Red Hook, NY: Curran.

Jegou, H., Perronnin, F., Douze, M., Sánchez, J., Perez, P., & Schmid, C. (2012). Tổng hợp các mô tả hình ảnh cục bộ của Ag thành các mã nhỏ gọn. Giao dịch IEEE về Phân tích mẫu và Trí tuệ máy móc, 34(9), 1704–1716.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... Darrell, T. (2014). Caffe: Kiến trúc tích chập để nhúng tính năng nhanh. Biên bản Hội nghị quốc tế lần thứ 22 của ACM về đa phương tiện (trang 675–678). New York: ACM.

Jin, J., Dundar, A., & Culurciello, E. (2016). Mạng nơ-ron tích chập mạnh mẽ dưới nhiễu đối nghịch. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 4 về Biểu diễn học tập (trang 1–8). Np: Computational and Biological Learning Society.

Jin, X., Xu, C., Feng, J., Wei, Y., Xiong, J., & Yan, S. (2015). Học sâu với hình chữ S đơn vị kích hoạt tuyến tính chỉnh lưu. Bản in trước của ArXiv:1512.07030.

Kalchbrenner, N., Danihelka, I., & Graves, A. (2015). Bộ nhớ dài hạn dạng lưới. Bản in trước của ArXiv:1507.01526.

Kalchbrenner, N., Espeholt, L., Simonyan, K., Oord, AVD, Graves, A., & Kavukcuoglu, K. (2016). Bản dịch máy thần kinh theo thời gian tuyến tính. Bản in trước arXiv: 1610.10099.

Karpathy, A. (2016). CS231n: Mạng nơ-ron tích chập để nhận dạng hình ảnh. http://cs231n.github.io/classification/ Karpathy,

A., & Fei-Fei, L. (2016). Căn chỉnh ngữ nghĩa thị giác sâu để tạo mô tả hình ảnh. Trong Biên bản báo cáo của Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 3128–3137). Red Hook, NY: Curran.

Kavukcuoglu, K., Ranzato, M., Fergus, R., & LeCun, Y. (2009). Học các đặc điểm bất biến thông qua bản đồ lọc địa hình. Trong Biên bản báo cáo của Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 1605–1612). Red Hook, NY: Curran.

Kavukcuoglu, K., Ranzato, M., & LeCun, Y. (2010). Suy luận nhanh trong thuật toán mã hóa thưa thớt với các ứng dụng để nhận dạng đối tượng. Bản in trước ArXiv: 1010.3467.

Kennedy, J., & Eberhart, RC (1995, tháng 10). Một trình tối ưu hóa mới sử dụng lý thuyết bầy hạt. Trong Biên bản báo cáo của Hội thảo quốc tế lần thứ 6 về Khoa học máy vi mô và con người (trang 39–43). Piscataway, NJ: IEEE.

Kim, M., & Smaragdis, P. (2016). Mạng nơ-ron bitwise. Bản in trước arXiv:1601.06071.

Kim, Y., Park, E., Yoo, S., Choi, T., Yang, L., & Shin, D. (2015). Nén mạng nơ-ron tích chập sâu cho các ứng dụng di động nhanh và công suất thấp. Bản in trước ArXiv:1511.06530.

Kingma, D., & Ba, J. (2014). Adam: Một phương pháp tối ưu hóa ngẫu nhiên. Bản in trước ArXiv: 1412.6980.

Kingma, DP, & Welling, M. (2014). Bayes biến thiên mã hóa tự động. Bản in trước ArXiv: 1312.6114 v10.

Koushik, J. (2016). Hiểu về mạng nơ-ron tích chập. Bản in trước của ArXiv: 1605.09081.

Krizhevsky, A. (2009). Học nhiều lớp tính năng từ hình ảnh nhỏ. Luận văn thạc sĩ, Đại học Toronto, Canada.

Krizhevsky, A. (2014). Một thủ thuật kỳ lạ để song song hóa mạng nơ-ron tích chập. Bản in trước của ArXiv:1404.5997.

Krizhevsky, A., Sutskever, I., & Hinton, GE (2012). Phân loại ImageNet với mạng nơ-ron tích chập sâu. Trong F. Pereira, CJC Burges, L. Bottou, & KQ Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 25 (trang 1097–1105). Red Hook, NY: Curran.

Kulkarni, TD, Whitney, WF, Kohli, P., & Tenenbaum, J. (2015). Mạng đồ họa nghịch đảo tích chập sâu. Trong C. Cortes, ND Lawrence, DD Lee, M. Sugiyama, & R. Garnett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 28 (trang 2539–2547). Red Hook, NY: Curran.

Kumar, N., Berg, AC, Belhumeur, PN, & Nayar, SK (2009). Phân loại thuộc tính và so sánh để xác minh khuôn mặt. Trong Biên bản báo cáo của Hội nghị quốc tế IEEE về thị giác máy tính (trang 365–372). Red Hook, NY: Curran.

Laptev, D., Savinov, N., Buhmann, JM, & Pollefeys, M. (2016). TI-POOLING: Gộp bất biến chuyển đổi để học tính năng trong mạng nơ-ron tích chập. Bản in trư ớc của ArXiv:1604.06318.

Larochelle, H., Erhan, D., Courville, A., Bergstra, J., & Bengio, Y. (2007). Đánh giá thực nghiệm về kiến trúc sâu trên các vấn đề có nhiều yếu tố biến thiên. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 24 về học máy (trang 473–480). Np: Hiệp hội máy học quốc tế.

Lavin, A., & Gray, S. (2016). Thuật toán nhanh cho mạng nơ-ron tích chập. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 4013–4021). Red Hook, NY: Curran.

Lawrence, S., Giles, CL, Tsoi, AC, & Back, AD (1997). Nhận dạng khuôn mặt: Một phư ơng pháp tiếp cận mạng nơ-ron tích chập. IEEE Transactions on Neural Networks, 8(1), 98–113.

Lazebnik, S., Schmid, C., & Ponce, J. (2005). Biểu diễn kết cấu thư a thớt sử dụng vùng affine cục bộ. Giao dịch IEEE về Phân tích mẫu và Trí tuệ máy móc, 27(8), 1265–1278.

Lazebnik, S., Schmid, C., & Ponce, J. (2006). Ngoài các túi tính năng: Ghép nối pyra-mid không gian để nhận dạng các loại cảnh tự nhiên. Trong Biên bản báo cáo của Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 2169–2178). Red Hook, NY: Curran.

Learned-Miller, E., Huang, GB, RoyChowdhury, A., Li, H., & Hua, G. (2016). Khuôn mặt đư ợc gắn nhãn trong tự nhiên: Một cuộc khảo sát. Trong M. Kawulok, ME Celebi, & B. Smolka (Biên tập viên), Những tiến bộ trong phát hiện khuôn mặt và phân tích hình ảnh khuôn mặt (trang 189–248). Cham, Thụy Sĩ: Springer.

Lebedev, V., Ganin, Y., Rakhuba, M., Oseledets, I., & Lempitsky, V. (2014). Tăng tốc mạng nơ-ron tích chập bằng cách sử dụng phân tích CP đư ợc tinh chỉnh. Bản in trư ớc của ArXiv: 1412.6553.

LeCun, Y. (1989). Tổng quát hóa và chiến lư ợc thiết kế mạng. Trong R. Pfeifer, Z. Schreter, F. Fogelman, & L. Steels (Biên tập viên), Kết nối trong viễn cảnh (trang 143–155). Zurich, Thụy Sĩ: Elsevier.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Học sâu. Nature, 521(7553), 436–444.

LeCun, Y., Boser, B., Denker, JS, Henderson, D., Howard, RE, Hubbard, W., & Jackel, LD (1989). Nhận dạng chữ số viết tay bằng mạng lư ới lan truyền ngư ợc. Trong DS Touretzky (Biên tập), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 2 (trang 396–404). Cambridge, MA: Nhà xuất bản MIT.

LeCun, Y., Boser, B., Denker, JS, Henderson, D., Howard, RE, Hubbard, W., & Jackel, LD (1989). Truyền ngư ợc áp dụng cho nhận dạng mã bư u chính viết tay. Tính toán thần kinh, 1(4), 541–551.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Học tập dựa trên gradient đư ợc áp dụng để nhận dạng tài liệu. Biên bản báo cáo của IEEE, 86(11), 2278–2324.

LeCun, Y., Huang, FJ, & Bottou, L. (2004). Phư ơng pháp học tập cho đối tư ợng chung nhận dạng với sự bất biến đối với tư thế và ánh sáng. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 97–104). Red Hook, NY: Curran.

LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). Mạng tích chập và ứng dụng trong tầm nhìn. Trong Biên bản báo cáo của Hội nghị chuyên đề quốc tế IEEE về mạch điện và Hệ thống (trang 253–256). Red Hook, NY: Curran.

Lee, C., Gallagher, PW, & Tu, Z. (2016). Tổng quát hóa các hàm nhóm trong mạng nơ-ron tích chập: Hỗn hợp, có cổng và cây. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 19 về Trí tuệ nhân tạo và Thống kê (trang 464–472). www.jmlr .org/proceedings/papers/v51/lee16a.pdf

Lee, C., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2015). Các mạng đư ợc giám sát sâu. Trong Biên bản Hội nghị quốc tế lần thứ 18 về Trí tuệ nhân tạo và Thống kê (trang 562–570). jmlr.org/proceedings/papers/v38/lee15a.pdf

Lee, H., Grosse, R., Ranganath, R., & Ng, AY (2009). Niềm tin sâu sắc chập chờn mạng lư ới cho việc học tập không giám sát có thể mở rộng của các biểu diễn phân cấp. Trong Biên bản Hội nghị quốc tế lần thứ 26 về Học máy (trang 609–616). Số: Hiệp hội máy học quốc tế.

Leibo, JZ, Mutch, J., & Poggio, T. (2011). Tại sao não tách biệt nhận dạng khuôn mặt từ nhận dạng đối tư ợng. Trong J. Shawe-Taylor, RS Zemel, PL Bartlett, F. Pereira, & KQ Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 24 (trang 711–719). Red Hook, NY: Hiện tại.

Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). Đào tạo toàn diện các chính sách vi-suomotor sâu. Tạp chí nghiên cứu học máy, 17(39), 1–40.

Li, F., Zhang, B., & Liu, B. (2016). Mạng trọng số ba phần. Bản in trư ớc arXiv:1605.04711.

Li, S., Jiao, J., Han, Y., & Weissman, T. (2016). Giải mã ResNet. bản in trư ớc arXiv: 1611.01186.

Li, Z., Gong, B., & Yang, T. (2016). Cải thiện tình trạng bỏ học cho học nông và học sâu. Trong D. Lee, M. Sugiyama, UV Luxburg, I. Guyon, & R. Garnett (Biên tập viên), Tiến bộ trong hệ thống xử lý thông tin thần kinh (trang 1–9). Np: Tiền biên bản.

Liang, M., & Hu, X. (2015). Mạng nơ-ron tích chập hồi quy để nhận dạng đối tư ợng. Trong Biên bản báo cáo của Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 3367–3375). Red Hook, NY: Curran.

Liao, Z., & Carneiro, G. (2016). Về tầm quan trọng của các lớp chuẩn hóa trong học sâu với các đơn vị kích hoạt tuyến tính từng phần. Trong Biên bản báo cáo của IEEE Hội nghị mùa đông về Ứng dụng của Thị giác máy tính (trang 1–8). Red Hook, NY: Curran.

Lin, J., Morere, O., Chandrasekhar, V., Veillard, A., & Goh, H. (2015). Deephash: Chính quy hóa Get-ting, độ sâu và tinh chỉnh đúng. Bản in trư ớc ArXiv: 1501.04711.

Lin, M., Chen, Q., & Yan, S. (2013). Mạng trong mạng. Bản in trư ớc ArXiv: 1312.4400.

Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., ... Huang, T. (2011). Quy mô lớn phân loại hình ảnh: Trích xuất tính năng nhanh và đào tạo SVM. Trong Biên bản

Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 1689–1696).
Red Hook, NY: Hiện tại.

Linnainmaa, S. (1970). Biểu diễn lỗi làm tròn tích lũy của thuật toán như một phép khai triển Taylor của lỗi làm tròn cục bộ. Luận văn thạc sĩ, Đại học của Helsinki, Phần Lan.

Littwin, E., & Wolf, L. (2016). Bề mặt mất mát của mạng lưới còn lại: Các tập hợp và vai trò của chuẩn hóa hàng loạt. bản in trước arXiv:1611.02525.

Liu, H., Tian, Y., Yang, Y., Pang, L., & Huang, T. (2016). Học khoảng cách tương đối sâu: Nêu sự khác biệt giữa các phương tiện tương tự. Trong Biên bản báo cáo của Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 2167–2175). Red Hook, NY: Curran.

Liu, W., Wen, Y., Scut, M., Yu, Z., & Yang, M. (2016). Tổn thất softmax biên độ lớn cho mạng nơ-ron tích chập. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 33 Học máy (trang 507–516). Np: Hiệp hội học máy quốc tế.

Maas, A.L., Hannun, A.Y., & Ng, A.Y. (2013). Bộ chỉnh lưu phi tuyến tính cải thiện thần kinh mô hình âm thanh mạng. Trong Biên bản Hội nghị quốc tế lần thứ 30 về máy móc Học tập (trang 1–8). Np: Hiệp hội học máy quốc tế.

Maharaj, AV (2015). Cải thiện tính mạnh mẽ đối nghịch của ConvNet bằng cách giảm chiều đầu vào. Stanford, CA: Khoa Vật lý, Đại học Stanford.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., & Zisserman, A. (2008). phân biệt đối xử từ điển đã học để phân tích hình ảnh cục bộ. Trong Biên bản báo cáo của Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (trang 1–8). Red Hook, NY: Curran.

Malinowski, M., & Fritz, M. (2013). Các vùng nhóm có thể học được để phân loại hình ảnh. Bản in trước của ArXiv:1301.3516.

Mallat, S. (2012). Phân tán bất biến nhóm. Truyền thông về Pure và Applied Toán học, 65(10), 1331–1398. doi:10.1002/cpa.21413

Masci, J., Meier, U., Cire¸san, D., & Schmidhuber, J. (2011). chồng chập autoencoders để trích xuất tính năng phân cấp. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 21 về Mạng nơ-ron nhân tạo (trang 52–59). Berlin: Springer.

Mathieu, M., Henaff, M., & LeCun, Y. (2013). Đào tạo nhanh các mạng tích chập thông qua FFT. Bản in trước của ArXiv:1312.5851.

Mishkin, D., & Matas, J. (2015). Tất cả những gì bạn cần là một khởi đầu tốt. Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 4 về Biểu diễn học tập (trang 1–13). Np: Tính toán và Hội học tập sinh học.

Miyato, T., Maeda, S., Koyama, M., Nakae, K., & Ishii, S. (2016). phân phối làm mịn với đào tạo đối kháng ảo. Trong Biên bản của Hội nghị quốc tế lần thứ 5 Hội nghị về Biểu diễn Học tập (trang 1–12). Np: Hội học tập tính toán và sinh học.

Montavon, G., Orr, GB, & Müller, K. (Biên tập viên). (2012). Mạng nơ-ron: Những mánh khóe của thương mại (ấn bản lần 2). Berlin: Springer.

Muller, U., Ben, J., Cosatto, E., Flepp, B., & Cun, YL (2005). Tránh chướng ngại vật off-road thông qua học tập toàn diện. Trong Y. Weiss, PB Schölkopf, & JC Platt (Biên tập viên), Tiến bộ trong hệ thống xử lý thông tin thần kinh, 18 (trang 739–746). Cambridge, HÔM NAY: Nhà xuất bản MIT.

Nagi, J., Di Caro, GA, Giusti, A., Nagi, F., & Gambardella, LM (2012). Máy vectơ hỗ trợ thần kinh tích chập: Bộ phân loại mẫu hình ảnh lai cho hệ thống nhiều rô-bốt. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 11 về máy móc

Học tập và Ứng dụng (trang 27–32). Los Alamitos, CA: IEEE Computer
Xã hội.

Nagi, J., Ducatelle, F., Di Caro, GA, Cire¸san, D., Meier, U., Giusti, A., ... Gam-bardella, LM
(2011). Mạng nơ-ron tích chập Max-pooling để nhận dạng cử chỉ tay dựa trên thị giác. Trong Biên
bản báo cáo Hội nghị quốc tế IEEE về Ứng dụng xử lý tín hiệu và hình ảnh (trang 342–347). Red
Hook, NY:
Curran.

Nair, V., & Hinton, GE (2009). Nhận dạng đối tượng 3D với mạng lưới niềm tin sâu. Trong Y.
Bengio, D. Schuurmans, JD Lafferty, CKI Williams, & A. Culotta (Biên tập viên), Những tiến bộ
trong hệ thống xử lý thông tin thần kinh, 22 (trang 1339–1347). Red Hook, NY:
Curran.

Nair, V., & Hinton, GE (2010). Các đơn vị tuyến tính chỉnh lưu cải thiện Boltzmann hạn chế
máy móc. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 27 về Học máy
(trang 807–814). Np: Hiệp hội học máy quốc tế.

Nasse, F., Thurau, C., & Fink, GA (2009). Phát hiện khuôn mặt bằng mạng nơ-ron tích chập dựa trên
GPU. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 13 về
Phân tích hình ảnh và mẫu bằng máy tính (trang 83–90). Berlin: Springer.

Cuộc thi Khoa học Dữ liệu Quốc gia | Kaggle. (2016). Kaggle.com. https://www.kaggle.com
/c/datasciencebowl.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, AY (2011). Đọc dig-its trong hình ảnh
tự nhiên với học tính năng không giám sát. Trong Advances in neural in-formation processing
systems, 24 (NIPS) Workshop on Deep Learning and Unsupervised
Học tập tính năng (trang 1–9). Red Hook, NY: Curran.

Ngiam, J., Chen, Z., Chia, D., Koh, P.W., Le, Q.V., & Ng, AY (2010). Mạng nơ-ron tích chập dạng lát
gạch. Trong JD Lafferty, CKI Williams, J. Shawe-Taylor, R.
S. Zemel, & A. Culotta (Biên tập), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 23
(trang 1279–1287). Red Hook, NY: Hiện tại.

Nguyen, A., Yosinski, J., & Clune, J. (2015). Mạng nơ-ron sâu dễ bị đánh lừa:
Dự đoán độ tin cậy cao cho hình ảnh không thể nhận dạng. Trong Biên bản báo cáo của IEEE
Hội nghị về Tầm nhìn máy tính và Nhận dạng mẫu (trang 427–436). Los Alamitos,
CA: Hiệp hội máy tính IEEE.

Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., & Barbano, P.E. (2005).
Hướng tới việc phân tích kiểu hình tự động phôi đang phát triển từ video. IEEE Trans-actions
on Image Processing, 14(9), 1360–1371.

Novikov, A., Podoprikhin, D., Osokin, A., & Vetrov, D. (2015). Mạng nơ-ron kéo căng. Trong C.
Cortes, ND Lawrence, DD Lee, M. Sugiyama, & R. Gar-nett (Biên tập viên), Những tiến bộ trong hệ
thống xử lý thông tin thần kinh, 28 (trang 442–450). Mạng
Hook, NY: Hiện tại.

Nowlan, SJ, & Hinton, GE (1992). Đơn giản hóa mạng nơ-ron bằng trọng số mềm
chia sẻ. Tính toán thần kinh, 4(4), 473–493.

Oh, K., & Jung, K. (2004). Triển khai GPU của mạng nơ-ron. Nhận dạng mẫu, 37(6), 1311–1314.

Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ...
Kavukcuoglu, K. (2016). Wavenet: Một mô hình tạo ra âm thanh thô. CoRR
cơ bản/1609.03499.

Orhan, AE (2017). Kết nối bỏ qua như một phương pháp phá vỡ tính đối xứng hiệu quả. arXiv
bản in trước:1701.09175.

Oseledets, IV (2011). Phân tích tenxơ-đoàn tàu. Tạp chí SIAM về Khoa học Máy tính, 33(5), 2295–2317.

Oxholm, G., Barry, P., & Nishino, K. (2012). Tỷ lệ kết cấu hình học. TRONG
Biên bản Hội nghị Châu Âu về Tầm nhìn Máy tính (trang 58–71). Berlin:
Mùa xuân.

Paine, T., Jin, H., Yang, J., Lin, Z., & Huang, T. (2013). GPU không đồng bộ ngẫu nhiên
phương pháp giảm dần độ dốc để tăng tốc quá trình đào tạo mạng nơ-ron. Bản in trước của ArXiv: 1312.6186.

Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2015). Chưng cất như một biện pháp phòng
thủ chống lại nhiễu loạn đối kháng chống lại mạng lưới nơ-ron sâu. Trong Biên bản
của Hội nghị chuyên đề IEEE lần thứ 37 về An ninh và Quyền riêng tư (trang 1–16). Los Alamitos, CA:
Hội máy tính IEEE.

Pereyra, G., Tucker, G., Chorowski, J., Kaiser, L., & Hinton, G. (2017). Quy định hóa mạng nơ-ron bằng
cách phạt các phân phối đầu ra tự tin. ArXiv
bản in trước:1701.06548v1.

Perronnin, F., Sánchez, J., & Mensink, T. (2010). Cải thiện hạt nhân Fisher để phân loại hình ảnh quy
mô lớn. Trong Biên bản báo cáo của Hội nghị máy tính châu Âu
Tầm nhìn (trang 143–156). Berlin: Springer.

Pinto, L., & Gupta, A. (2015). Siêu giám sát bản thân: Học cách nắm bắt từ 50k
thử nghiệm và 700 giờ sử dụng robot. Bản in trước của ArXiv:1509.06825.

Qiao, Y., Shen, J., Xiao, T., Yang, Q., Wen, M., & Zhang, C. (2016). FPGA-tăng tốc
mạng nơ-ron tích chập sâu cho thông lượng cao và hiệu quả năng lượng.
Đồng thời và tính toán: Thực hành và kinh nghiệm. doi:10.1002/cpe.3850

Qian, N. (1999). Về thuật ngữ động lượng trong các thuật toán học giảm dần độ dốc.
Mạng nơ-ron, 12, 145–150.

Ranzato, MA, Huang, FJ, Boureau, Y., & LeCun, Y. (2007). Học không có giám sát
của các hệ thống phân cấp tính năng bất biến với các ứng dụng để nhận dạng đối tượng. Trong Biên
bản Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 1–8). Mỹ
Alamitos, CA: Hiệp hội máy tính IEEE.

Ranzato, M., Poultney, C., Chopra, S., & Cun, YL (2006). Học tập hiệu quả của thưa thớt
biểu diễn với mô hình dựa trên năng lượng. Trong PB Schölkopf, JC Platt, & T.
Hoffman (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 19 (trang 1137–
1144). Cambridge, MA: Nhà xuất bản MIT.

Ranzato, MA, & Szummer, M. (2008). Học bán giám sát các biểu diễn tài liệu nhỏ gọn với mạng lưới sâu.
Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 25
Hội nghị về Học máy (trang 792–799). Np: Hiệp hội Học máy Quốc tế.

Rastegari, M., Ordonez, V., Redmon, J., & Farhadi, A. (2016). Xnor-net: Imagenet
phân loại sử dụng mạng nơ-ron tích chập nhị phân. Trong Biên bản báo cáo
Hội nghị Châu Âu về Tầm nhìn Máy tính (trang 525–542). Berlin: Springer.

Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN giới thiệu sản phẩm có sẵn: Một
cơ sở đáng kinh ngạc để nhận dạng. Trong Biên bản báo cáo của Hội nghị IEEE
về Hội thảo về Thị giác máy tính và Nhận dạng mẫu (trang 806–813). Los Alamitos,
CA: Hiệp hội máy tính IEEE.

Recht, B., Re, C., Wright, S., & Niu, F. (2011). Hogwild: Một cách tiếp cận không khóa để song song
hóa quá trình giảm dần độ dốc ngẫu nhiên. Trong J. Shawe-Taylor, RS Zemel, PL Bartlett,
F. Pereira, & KQ Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần
kinh, 24 (trang 693–701). Red Hook, NY: Curran.

Rippel, O., Gelbart, MA, & Adams, RP (2014). Học các biểu diễn có thứ tự
với dropout lồng nhau. Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 30 về máy
Học tập (trang 1746–1754). Np: Hiệp hội học máy quốc tế.

Rippel, O., Snoek, J., & Adams, RP (2015). Biểu diễn phổ cho mạng nơ-ron tích chập. Trong C.
Cortes, ND Lawrence, DD Lee, M. Sugiyama, &
R. Garnett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 28 (trang 2449–
2457). Red Hook, NY: Hiện tại.

Rosemary, A., Ballas, N., Kahou, SE, Chassang, A., Gatta, C., & Bengio, Y. (2015).
Fitnets: Gợi ý cho lư ới sâu mỏng. Trong Biên bản báo cáo của Hội nghị quốc tế lần thứ 3 về
Biểu diễn học tập (trang 1–13). Np: Học tập tính toán và sinh học
Xã hội.

Rumelhart, DE, Hinton, GE, & Williams, RJ (1986). Học các biểu diễn bằng
lỗi lan truyền ngư ợc. Nature, 323(6088), 533–536.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Bernstein, M.
(2015). Thử thách nhận dạng hình ảnh quy mô lớn của ImageNet. Tạp chí quốc tế
Tầm nhìn máy tính, 115(3), 211–252.

Sabour, S., Cao, Y., Faghri, F., & Fleet, DJ (2016). Sự thao túng đối nghịch của sâu
biểu diễn. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 4 về Biểu diễn học tập (trang 1–
18). Np: Hội học tập tính toán và sinh học.

Sainath, TN, Kingsbury, B., Mohamed, A., Dahl, GE, Saon, G., Soltau, H., ... Ram-abhadran, B.
(2013). Cải tiến cho mạng nơ-ron tích chập sâu cho
LVCSR. Trong Hội thảo IEEE năm 2013 về Nhận dạng và Hiểu giọng nói tự động
(trang 315–320). Red Hook, NY: Hiện tại.

Sainath, T.N., Kingsbury, B., Sindhwani, V., Arisoy, E., & Ramabhadran, B. (2013).
Phân tích ma trận bậc thấp để đào tạo mạng nơ-ron sâu với mục tiêu đầu ra có chiều cao.
Trong Biên bản báo cáo của Hội nghị quốc tế IEEE về
Âm học, giọng nói và xử lý tín hiệu (trang 6655–6659. Np: IEEE Signal Processing-ing
Society).

Salakhutdinov, R., & Hinton, GE (2007). Học nhúng phi tuyến tính bằng cách bảo toàn trư ớc cấu
trúc lân cận lớp. Trong Biên bản Hội nghị quốc tế lần thứ 11 về Trí tuệ nhân tạo và Thống
kê (trang 412–419). http://www.jmlr.org
/proceedings/papers/v2/salakhutdinov07a/salakhutdinov07a.pdf

Sánchez, J., & Perronnin, F. (2011). Nén chữ ký chiều cao cho
phân loại hình ảnh quy mô lớn. Trong Biên bản báo cáo của Hội nghị IEEE về Máy tính
Tầm nhìn và Nhận dạng Mẫu (trang 1665–1672). Los Alamitos, CA: IEEE Computer
Xã hội.

Saxe, AM, McClelland, JL, & Ganguli, S. (2013). Các giải pháp chính xác cho phi tuyến tính
động lực học tập trong mạng nơ-ron tuyến tính sâu. Bản in trư ớc của ArXiv: 1312.6120.

Sercu, T., & Goel, V. (2016). Dự đoán dày đặc trên các chuỗi với các phép tích chập giãn nở theo thời gian
để nhận dạng giọng nói. Bản in trư ớc arXiv:1611.09288.

Scherer, D., Müller, A., & Behnke, S. (2010). Đánh giá các hoạt động gộp nhóm trong kiến trúc
tích chập để nhận dạng đối tư ợng. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 20 về
mạng nơ-ron nhân tạo (trang 92–101). Berlin: Springer.

Schmidhuber, J. (2015). Học sâu trong mạng nơ-ron: Tổng quan. Neural Net-works, 61, 85–117.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: Một nhúng thống nhất cho
nhận dạng khuôn mặt và phân cụm. Trong Biên bản báo cáo của Hội nghị IEEE về Máy tính

Tầm nhìn và Nhận dạng Mẫu (trang 815–823). Los Alamitos, CA: IEEE Computer Society.

Sermanet, P., Chintala, S., & LeCun, Y. (2012). Mạng nơ-ron tích chập đư ợc áp dụng để phân loại chữ số số nhà. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 21 về Nhận dạng mẫu (trang 3288–3291). Red Hook, NY: Curran.

Simard, PY, Steinkraus, D., & Platt, JC (2003, tháng 8). Các phư ơng pháp hay nhất cho mạng nơ-ron tích chập áp dụng cho phân tích tài liệu trực quan. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 7 về phân tích và nhận dạng tài liệu (tập 3, trang 958–963). Washington, DC: IEEE Computer Society.

Simoncelli, EP, & Heeger, DJ (1998). Một mô hình phản ứng của tế bào thần kinh ở vùng thị giác MT. Nghiên cứu thị giác, 38(5), 743–761.

Simonyan, K., & Zisserman, A. (2014). Mạng tích chập rất sâu cho quy mô lớn nhận dạng hình ảnh. Bản in trư ớc của ArXiv:1409.1556.

Smirnov, EA, Timoshenko, DM và Andrianov, SN (2014). So sánh các phư ơng pháp chính quy hóa để phân loại ImageNet với mạng nơ-ron tích chập sâu. Tạp chí AASRI, 6, 89–94.

Snoek, J., Larochelle, H., & Adams, RP (2012). Tối ư u hóa Bayesian thực tế của các thuật toán học máy. Trong F. Pereira, CJC Burges, L. Bottou, & KQ Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 25 (trang 2951–2959). Red Hook, NY: Curran.

Sontag, ED (1998). Kích thư ớc VC của mạng nơ-ron. Máy tính NATO ASI Series F và Khoa học Hệ thống, 168, 69–96.

Springenberg, JT, Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Phấn đấu vì sự đơn giản: Mạng tích chập hoàn toàn. Bản in trư ớc ArXiv: 1412.6806.

Springenberg, JT, & Riedmiller, M. (2013). Cải thiện mạng nơ-ron sâu với các đơn vị maxout xác suất. Bản in trư ớc ArXiv: 1312.6116.

Srinivas, S., Sarvadevabhatla, RK, Mopuri, KR, Prabhu, N., Kruthiventi, SS, & Babu, RV (2016). Phân loại mạng lư ới thần kinh tích chập sâu cho thị giác máy tính. Bản in trư ớc của ArXiv:1601.06615.

Srivastava, N., Hinton, GE, Krizhevsky, A., Sutskever, I., và Salakhutdinov, R. (2014). Dropout: Một cách đơn giản để ngăn chặn mạng nơ-ron khỏi tình trạng quá khớp. Tạp chí nghiên cứu học máy, 15(1), 1929–1958.

Srivastava, N., & Salakhutdinov, RR (2013). Học chuyển giao phân biệt với các tiên nghiệm dựa trên cây. Trong CJC Burges, L. Bottou, M. Welling, Z. Ghahramani, & KQ Weinberger (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 26 (trang 2094–2102). Red Hook, NY: Curran.

Srivastava, RK, Greff, K., & Schmidhuber, J. (2015a). Đào tạo mạng lư ới rất sâu. Trong C. Cortes, ND Lawrence, DD Lee, M. Sugiyama, & R. Garnett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 28 (trang 2377–2385). Red Hook, NY: Curran.

Srivastava, R. K., Greff, K., và Schmidhuber, J. (2015b). Mạng lư ới đư ờng cao tốc. ArXiv bản in trư ớc:1505.00387.

Stallkamp, J., Schlipsing, M., Salmen, J., & Igel, C. (2011, tháng 7). Điểm chuẩn nhận dạng biển báo giao thông của Đức: Cuộc thi phân loại nhiều lớp. Trong Biên bản báo cáo của Hội nghị chung quốc tế IEEE về mạng nơ-ron (trang 1453–1460). Red Hook, NY: Hiện tại.

Steinkrau, D., Simard, PY, & Buck, I. (2005). Sử dụng GPU cho các thuật toán học máy. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 8 về Phân tích tài liệu và Sự công nhận (trang 1115–1119).Washington, DC: IEEE Computer Society.

Stollenga, MF, Masci, J., Gomez, F., & Schmidhuber, J. (2014). Mạng lưới sâu với sự chú ý chọn lọc nội bộ thông qua các kết nối phản hồi. Trong Z. Ghahramani, M. Welling, C. Cortes, ND Lawrence, & KQ Weinberger (Biên tập viên), Những tiến bộ trong thần kinh hệ thống xử lý thông tin, 27 (trang 3545–3553). Red Hook, NY: Curran.

Sun, Y., Chen, Y., Wang, X., & Tang, X. (2014). Biểu diễn khuôn mặt học sâu bằng xác minh nhận dạng chung. Trong Z. Ghahramani, M. Welling, C. Cortes, ND Lawrence, & KQ Weinberger (Biên tập viên), Tiến bộ trong xử lý thông tin thần kinh hệ thống, 27 (trang 1988–1996). Red Hook, NY: Curran.

Sun, Y., Liang, D., Wang, X., & Tang, X. (2015). Deepid3: Nhận dạng khuôn mặt với mạng lưới nơ-ron. Bản in trước của ArXiv:1502.00873.

Sun, Y., Wang, X., & Tang, X. (2014). Biểu diễn khuôn mặt học sâu từ việc dự đoán 10.000 lớp. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 1891–1898). Los Alamitos, CA: IEEE Computer Society.

Sun, Y., Wang, X., & Tang, X. (2015). Các biểu diễn khuôn mặt được học sâu là thưa thớt, chọn lọc và mạnh mẽ. Trong Biên bản báo cáo của Hội nghị IEEE về Máy tính Tầm nhìn và Nhận dạng Mẫu (trang 2892–2900). Los Alamitos, CA: IEEE Computer Xã hội.

Sussillo, D., & Abbott, L. (2014). Khởi tạo bước đi ngẫu nhiên để đào tạo bước tiến rất sâu mạng chuyển tiếp. Bản in trước của ArXiv:1412.6558.

Sutskever, I., Martens, J., Dahl, GE, & Hinton, GE (2013). Về tầm quan trọng của khởi tạo và động lượng trong học sâu. Trong Kỷ yếu Hội nghị quốc tế lần thứ 30 về học máy (trang 1139–1147). Np: International Machine Xã hội học tập.

Szegedy, C., Ioffe, S., & Vanhoucke, V. (2016). Inception-v4, Inception-Resnet và tác động của các kết nối còn lại đến việc học. Bản in trước của ArXiv:1602.07261.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2014). Đi sâu hơn với các phép tích chập. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và nhận dạng mẫu (trang 1–9). Los Alamitos, CA: IEEE Comp-computer Society.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Suy nghĩ lại Kiến trúc khởi đầu cho thị giác máy tính. Bản in trước của ArXiv: 1512.00567.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Các đặc tính hấp dẫn của mạng nơ-ron. Trong Biên bản Hội nghị quốc tế lần thứ nhất về Biểu diễn học tập (trang 1–10). Np: Tính toán và Hội học tập sinh học.

Tabacof, P., & Valle, E. (2015). Khám phá không gian của hình ảnh đối nghịch. Trong Biên bản báo cáo của Hội nghị chung quốc tế về mạng nơ-ron (trang 1–8). Red Hook, NY: Hiện tại.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Thu hẹp khoảng cách đến hiệu suất ở cấp độ con người trong việc xác minh khuôn mặt. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 1701–1708). Los Alamitos, CA: Hiệp hội máy tính IEEE.

Tang, Y. (2013). Học sâu sử dụng máy vectơ hỗ trợ tuyến tính. Bản in trước của ArXiv: 1306.0239.

Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Đối tượng hiệu quả
định vị bằng cách sử dụng mạng tích chập. Trong Biên bản báo cáo của Hội nghị IEEE
về Tầm nhìn máy tính và Nhận dạng mẫu (trang 648–656). Los Alamitos, CA: IEEE
Hội máy tính.

Tucker, LR (1966). Một số ghi chú toán học về phân tích nhân tố ba chế độ. Psy-chometrika, 31(3),
279–311.

Turaga, SC, Murray, JF, Jain, V., Roth, F., Helmstaedter, M., Briggman, K., ...
Seung, HS (2010). Mạng tích chập có thể học cách tạo đồ thị ái lực
để phân đoạn hình ảnh. Tính toán thần kinh, 22(2), 511–538.

Ulicný, M., Lundström, J., & Byttner, S. (2016). Sự mạnh mẽ của mạng nơ-ron tích chập sâu để nhận
dạng hình ảnh. Trong Biên bản báo cáo của Hội nghị chuyên đề quốc tế lần thứ nhất về Hệ thống
máy tính thông minh (trang 16–30). Thụy Sĩ: Springer Inter-national Publishing.

Van Dyk, DA, & Meng, X. (2012). Nghệ thuật tăng cường dữ liệu. Tạp chí máy tính
Thống kê đồ họa và thống kê, 10(1), 1–50.

Vapnik, V. (1995). Bản chất của lý thuyết học tập thống kê. New York: Springer Science
& Truyền thông doanh nghiệp.

Vapnik, VN, & Chervonenkis, AY (1971). Về sự hội tụ đồng đều của tương đối
tần suất của các sự kiện theo xác suất của chúng. Lý thuyết xác suất và ứng dụng của nó.
16(2), 11–30.

Viñales, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Hiển thị và kể: Một hình ảnh thần kinh
trình tạo chú thích. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và
Nhận dạng mẫu (trang 3156–3164). Red Hook, NY: Curran.

Wager, S., Wang, S., & Liang, PS (2013). Đào tạo bỏ học như một sự điều chỉnh thích ứng. Trong CJC
Burges, L. Bottou, M. Welling, Z. Ghahramani, & KQ Wein-berger (Biên tập viên), Những tiến bộ
trong hệ thống xử lý thông tin thần kinh, 26 (trang 351–359).
Red Hook, NY: Hiện tại.

Wan, L., Zeiler, M., Zhang, S., Cun, Y. L., & Fergus, R. (2013). Chính quy hóa của
mạng nơ-ron sử dụng Dropconnect. Trong Kỷ yếu Hội nghị quốc tế lần thứ 30 về Học máy (trang
1058–1066). Np: Học máy quốc tế
Xã hội.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). CNN-RNN: Một
khuôn khổ thống nhất cho phân loại hình ảnh đa nhãn. Bản in trước của ArXiv:1604.04573.

Wang, SI, & Manning, CD (2013). Đào tạo bỏ học nhanh. Trong Biên bản báo cáo của Hội nghị lần thứ 30
Hội nghị quốc tế về máy học (trang 118–126). Np: Hiệp hội máy học quốc tế.

Wang, Y., Xu, C., You, S., Tao, D., & Xu, C. (2016). CNNpack: Đóng gói tích chập
mạng nơ-ron trong miền tần số. Trong D. Lee, M. Sugiyama, UV Luxburg,
I. Guyon, & R. Garnett (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 29
(trang 1–9). Np: Tiền biên bản.

Wang, Z., & Oates, T. (2015). Mã hóa chuỗi thời gian dưới dạng hình ảnh để kiểm tra trực quan
và phân loại bằng cách sử dụng mạng nơ-ron tích chập lát gạch. Trong các Hội thảo tại
Hội nghị AAAI lần thứ 29 về Trí tuệ nhân tạo (trang 40–46).

Warde-Farley, D., Goodfellow, IJ, Courville, A., & Bengio, Y. (2013). Một kinh nghiệm
phân tích tình trạng bỏ học trong mạng tuyến tính từng phần. Bản in trước của ArXiv:1312.6197.

Weinberger, KQ, Blitzer, J., & Saul, LK (2005). Học tập số liệu từ xa cho các
phân loại láng giềng gần nhất biên độ. Trong Y. Weiss, PB Schölkopf, & JC Platt

(Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 18 (trang 1473–1480). Cambridge, MA: MIT Press.

Werbos, P. (1974). Vượt ra ngoài hồi quy: Các công cụ mới để dự đoán và phân tích trong hành vi khoa học. Luận án tiến sĩ, Đại học Harvard.

Werbos, PJ (1982). Ứng dụng của những tiến bộ trong phân tích độ nhạy phi tuyến tính. Trong RF Drenick & F. Kozin (Biên tập), Mô hình hóa và tối ưu hóa hệ thống (trang 762–770). Berlin: Springer.

Weston, J., Ratle, F., Mobahi, H., & Collobert, R. (2008). Học sâu thông qua nhúng bán siêu thâm nhập. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 25 về máy Học tập (trang 1168–1175). Np: Hiệp hội học máy quốc tế.

Wiatowski, T., & Bölcskei, H. (2015). Một lý thuyết toán học về mạng nơ-ron tích chập sâu mạng lưới để trích xuất tính năng. Bản in trước của ArXiv:1512.06293.

Winograd, S. (1980). Độ phức tạp số học của phép tính. Philadelphia: SIAM.

Wolf, L., Hassner, T., & Maoz, I. (2011). Nhận dạng khuôn mặt trong video không bị hạn chế với sự tương đồng về nền tảng phù hợp. Trong Biên bản báo cáo của Hội nghị IEEE về Máy tính Tầm nhìn và Nhận dạng Mẫu (trang 529–534). Red Hook, NY: Curran.

Wu, H., & Gu, X. (2015). Bỏ qua nhóm tối đa để điều chỉnh tích chập mạng nơ-ron. Trong Biên bản Hội nghị quốc tế lần thứ 22 về Xử lý thông tin nơ-ron (trang 46–53). Berlin: Springer.

Wu, J., Yu, Y., Huang, C., & Yu, K. (2015). Học sâu nhiều trường hợp cho phân loại hình ảnh và chú thích tự động. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 3460–3469). Red Hook, NY: Curran.

Xie, L., Wang, J., Wei, Z., Wang, M., & Tian, Q. (2016). DisturbLabel: Đang điều chỉnh CNN trên lớp mất mát. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng Mẫu (trang 4753–4762). Red Hook, NY: Curran.

Xie, S., Girshick, R., Dollar, P., Tu, Z., & He, K. (2016). Biến đổi dư thừa tổng hợp- cho mạng nơ-ron sâu. Bản in trước arXiv:1611.05431.

Xu, B., Wang, N., Chen, T., & Li, M. (2015). Đánh giá thực nghiệm về các hoạt động đường cong chỉnh sửa trong mạng tích chập. Bản in trước của ArXiv:1505.00853v2.

Yadan, O., Adams, K., Taigman, Y., & Ranzato, M. (2014). Đào tạo đa GPU cho Con-Vnets. Bản in trước của ArXiv: 1312. 5853v4.

Yang, J., Yu, K., Gong, Y., & Huang, T. (2009). Ghép hình kim tự tháp không gian tuyến tính sử dụng mã hóa thưa thớt để phân loại hình ảnh. Trong Biên bản Hội nghị IEEE về Tầm nhìn máy tính và nhận dạng mẫu (trang 1794–1801). Red Hook, NY: Curran.

Yu, D., Wang, H., Chen, P., & Wei, Z. (2014). Phân nhóm hỗn hợp cho mạng nơ-ron tích chập. Trong Biên bản báo cáo Hội nghị quốc tế lần thứ 9 về Tập thô và Công nghệ tri thức (trang 364–375). Berlin: Springer.

Yu, F., & Koltun, V. (2015). Tổng hợp ngữ cảnh đa thang đo bằng phép tích chập giãn nở. arXiv bản in trước:1511.07122.

Yu, W., Yang, K., Bai, Y., Yao, H., & Rui, Y. (2014a). Luồng DNN: Tính năng DNN pyra-mid dựa trên hình ảnh khớp. Trong Biên bản Hội nghị về thị giác máy tính của Anh (trang 1–10). Durham, Vương quốc Anh: BMVA Press.

Yu, W., Yang, K., Bai, Y., Yao, H., & Rui, Y. (2014b). Hình dung và so sánh các tích chập mạng nơ-ron nhân tạo. Bản in trước của ArXiv: 1412.6631.

Zagoruyko, S., & Komodakis, N. (2017). Mạng lưới dư thừa rộng. Bản in trước của ArXiv: 1605.07146v3.

Zeiler, MD (2012). ADADELTA: Một phương pháp tốc độ học tập thích ứng. Bản in trước của ArXiv: 1212.5701.

Zeiler, MD, & Fergus, R. (2013). Phân nhóm ngẫu nhiên để điều chỉnh tích chập sâu mạng nơ-ron nhân tạo. Bản in trước của ArXiv:1301.3557.

Zeiler, MD, & Fergus, R. (2014). Hình dung và hiểu các mạng lưới tích chập. Trong Biên bản báo cáo của Hội nghị châu Âu về thị giác máy tính (trang 818–833). Berlin: Springer.

Zeiler, MD, Taylor, GW, & Fergus, R. (2011). Mạng giải tích thích ứng cho việc học tính năng ở mức trung bình và cao. Trong Biên bản báo cáo của IEEE International Hội nghị về Tầm nhìn máy tính (trang 2018–2025). Red Hook, NY: Curran.

Zhai, S., Cheng, Y., & Zhang, ZM (2016). Mạng nơ-ron tích chập kép. Trong D. Lee, M. Sugiyama, UV Luxburg, I. Guyon, & R. Garnett (Biên tập viên), Tiến bộ trong hệ thống xử lý thông tin thần kinh, 29 (trang 1–9). Np: Tiền biên bản.

Zhang, H., Berg, AC, Maire, M., & Malik, J. (2006). SVM-KNN: Phân loại hàng xóm gần nhất phân biệt để nhận dạng danh mục trực quan. Trong Biên bản báo cáo của Hội nghị IEEE về Tầm nhìn máy tính và Nhận dạng mẫu (trang 2126–2136). Đỏ Hook, NY: Hiện tại.

Zhao, Q., & Griffin, LD (2016). Ức chế sự bất thường: Hướng tới CNN mạnh mẽ bằng cách sử dụng hàm kích hoạt đối xứng. Bản in trước của ArXiv:1603.05145v1.

Zhou, E., Cao, Z., & Yin, Q. (2015). Nhận dạng khuôn mặt ngây thơ-sâu sắc: Chạm đến giới hạn của LFW chuẩn mực hay không? Bản in trước của ArXiv:1501.04690.

Zhu, Z., Luo, P., Wang, X., & Tang, X. (2014). Phục hồi khuôn mặt dạng xem chuẩn trong tự nhiên với mạng lưới nơ-ron sâu. Bản in trước của ArXiv:1404.3543.

Zhuang, Y., Chin, W., Juan, Y., & Lin, C. (2013). Một SGD song song nhanh để phân tích ma trận trong các hệ thống bộ nhớ chia sẻ. Biên bản Hội nghị ACM lần thứ 7 về Hệ thống đề xuất (trang 249–256). New York: ACM.

Zinkevich, M., Weimer, M., Li, L., & Smola, AJ (2010). Giảm dần gradient ngẫu nhiên song song. Trong JD Lafferty, CKI Williams, J. Shawe-Taylor, RS Zemel, & A. Culotta (Biên tập viên), Những tiến bộ trong hệ thống xử lý thông tin thần kinh, 23 (trang 2595–2603). Red Hook, NY: Hiện tại.

Nhận ngày 15 tháng 1 năm 2017; chấp nhận ngày 25 tháng 3 năm 2017.