

GIẤY KHẢO SÁT

Truy cập mở



Một cuộc khảo sát về Tăng cường dữ liệu hình ảnh cho Deep Learning

Connor rút ngắn*  và Taghi M. Khoshgoftaar

* Thư tín:
cshorten2015@fau.edu
Khoa máy tính
và Kỹ thuật Điện và
Khoa học Máy tính,
Đại học Florida Atlantic,
Boca Raton, Hoa Kỳ

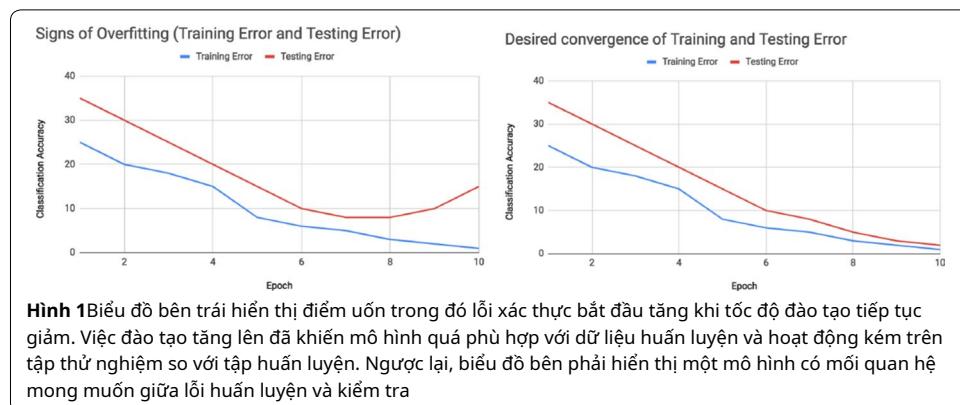
trùu tượng

Mạng nơ-ron tích chập sâu đã hoạt động rất tốt trong nhiều nhiệm vụ Thị giác Máy tính. Tuy nhiên, các mạng này phụ thuộc rất nhiều vào dữ liệu lớn để tránh tình trạng bị quá mức. Quá khứ đề cập đến hiện tượng khi mạng học một hàm có phương sai rất cao, chẳng hạn như mô hình hóa dữ liệu huấn luyện một cách hoàn hảo. Thật không may, nhiều miến ứng dụng không có quyền truy cập vào dữ liệu lớn, chẳng hạn như phân tích hình ảnh y tế. Khảo sát này tập trung vào Tăng cường dữ liệu, một giải pháp không gian dữ liệu cho vấn đề hạn chế dữ liệu. Tăng cường dữ liệu bao gồm một bộ kỹ thuật giúp nâng cao quy mô và chất lượng của các tập dữ liệu đào tạo để có thể xây dựng các mô hình Deep Learning tốt hơn bằng cách sử dụng chúng. Các thuật toán tăng cường hình ảnh được thảo luận trong khảo sát này bao gồm các phép biến đổi hình học, tăng cường không gian màu, bộ lọc hạt nhân, trộn hình ảnh, xóa ngẫu nhiên, có tính năng tăng cường không gian, đào tạo đối thủ, mạng lưới đối thủ tổng hợp, chuyển giao phong cách thần kinh và siêu học tập. Việc áp dụng các phương pháp tăng cường dựa trên GAN được đề cập rất nhiều trong cuộc khảo sát này. Ngoài các kỹ thuật tăng cường, bài viết này sẽ thảo luận ngắn gọn về các đặc điểm khác của Tăng cường dữ liệu như tăng thời gian thử nghiệm, tác động của độ phân giải, kích thước tập dữ liệu cuối cùng và chương trình học. Khảo sát này sẽ trình bày các phương pháp hiện có để Tăng cường dữ liệu, những phát triển đây hứa hẹn và các quyết định ở cấp độ tổng hợp để triển khai Tăng cường dữ liệu. Người đọc sẽ hiểu cách Tăng cường dữ liệu có thể cải thiện hiệu suất của các mô hình của họ và mở rộng các tập dữ liệu hạn chế để tận dụng các khả năng của dữ liệu lớn, và siêu học tập. Việc áp dụng các phương pháp tăng cường dựa trên GAN được đề cập rất nhiều trong cuộc khảo sát này. Ngoài các kỹ thuật tăng cường, bài viết này sẽ thảo luận ngắn gọn về các đặc điểm khác của Tăng cường dữ liệu như tăng thời gian thử nghiệm, tác động của độ phân giải, kích thước tập dữ liệu cuối cùng và chương trình học. Khảo sát này sẽ trình bày các phương pháp hiện có để Tăng cường dữ liệu, những phát triển đây hứa hẹn và các quyết định ở cấp độ tổng hợp để triển khai Tăng cường dữ liệu. Người đọc sẽ hiểu cách Tăng cường dữ liệu có thể cải thiện hiệu suất của các mô hình của họ và mở rộng các tập dữ liệu hạn chế để tận dụng các khả năng của dữ liệu lớn. Bài viết này sẽ thảo luận ngắn gọn về các đặc điểm khác của Tăng cường dữ liệu chẳng hạn như tăng thời gian thử nghiệm, tác động của độ phân giải, kích thước tập dữ liệu cuối cùng và việc học chương trình giảng dạy. Khảo sát này sẽ trình bày các phương pháp hiện có để Tăng cường dữ liệu, những phát triển đây hứa hẹn và các quyết định ở cấp độ tổng hợp để triển khai Tăng cường dữ liệu. Người đọc sẽ hiểu cách Tăng cường dữ liệu có thể cải thiện hiệu suất của các mô hình của họ và mở rộng các tập dữ liệu hạn chế để tận dụng các khả năng của dữ liệu lớn.

Từ khóa: Tăng cường dữ liệu, Dữ liệu lớn, Dữ liệu hình ảnh, Học sâu, GAN

Giới thiệu

Các mô hình Deep Learning đã đạt được tiến bộ đáng kinh ngạc trong các nhiệm vụ phân biệt đối xử. Điều này được thúc đẩy bởi sự tiến bộ của kiến trúc mạng sâu, khả năng tính toán mạnh mẽ và quyền truy cập vào dữ liệu lớn. Mạng lưới thần kinh sâu đã được áp dụng thành công cho các tác vụ Thị giác máy tính như phân loại hình ảnh, phát hiện đối tượng và phân đoạn hình ảnh nhờ sự phát triển của mạng lưới thần kinh tích chập (CNN). Các mạng thần kinh này sử dụng các hạt nhân được tham số hóa, kết nối thưa thớt để bảo toàn các đặc điểm không gian của hình ảnh. Các lớp tích chập tuần tự giảm mẫu độ phân giải không gian của hình ảnh đồng thời mở rộng độ sâu của bản đồ đặc trưng của chúng. Chuỗi phép biến đổi tích chập này có thể tạo ra các biểu diễn hình ảnh có chiều thấp hơn và hữu ích hơn nhiều so với những gì có thể được tạo ra bằng tay.



Hình 1 Biểu đồ bên trái hiển thị điểm uốn trong đó lỗi xác thực bắt đầu tăng khi tốc độ đào tạo tiếp tục giảm. Việc đào tạo tăng lên đã khiến mô hình quá phù hợp với dữ liệu huấn luyện và hoạt động kém trên tập thử nghiệm so với tập huấn luyện. Ngược lại, biểu đồ bên phải hiển thị một mô hình có mối quan hệ mong muốn giữa lỗi huấn luyện và kiểm tra

Có nhiều nhánh nghiên cứu hy vọng cải thiện các điểm chuẩn hiện tại bằng cách áp dụng mạng tích chập sâu cho các nhiệm vụ Thị giác Máy tính. Việc nâng cao khả năng khái quát hóa của các mô hình này là một trong những thách thức khó khăn nhất. Tính tổng quát đề cập đến sự khác biệt về hiệu suất của một mô hình khi được đánh giá trên dữ liệu đã thấy trước đó (dữ liệu huấn luyện) so với dữ liệu mà nó chưa từng thấy trước đó (dữ liệu thử nghiệm). Các mô hình có khả năng khái quát hóa kém đã trang bị quá mức dữ liệu huấn luyện. Một cách để phát hiện hiện tượng trang bị quá mức là vẽ sơ đồ độ chính xác trong quá trình huấn luyện và xác nhận ở mỗi giai đoạn trong quá trình huấn luyện. Biểu đồ bên dưới mô tả việc trang bị quá mức có thể trông như thế nào khi hình dung những độ chính xác này qua các giai đoạn huấn luyện (Hình 2).¹)

Để xây dựng các mô hình Deep Learning hữu ích, lỗi xác thực phải tiếp tục giảm cùng với lỗi đào tạo. Tăng cường dữ liệu là một phương pháp rất mạnh mẽ để đạt được điều này. Dữ liệu tăng cường sẽ thể hiện một tập hợp toàn diện hơn các điểm dữ liệu có thể có, do đó giảm thiểu khoảng cách giữa tập huấn luyện và tập xác thực, cũng như bất kỳ tập kiểm tra nào trong tương lai.

Tăng cường dữ liệu, trọng tâm của cuộc khảo sát này, không phải là kỹ thuật duy nhất được phát triển để giảm tình trạng trang bị quá mức. Một số đoạn văn sau đây sẽ giới thiệu các giải pháp khác có sẵn để tránh trang bị quá mức trong các mô hình Deep Learning. Danh sách này nhằm mục đích cung cấp cho người đọc sự hiểu biết rộng hơn về bối cảnh Tăng cường dữ liệu.

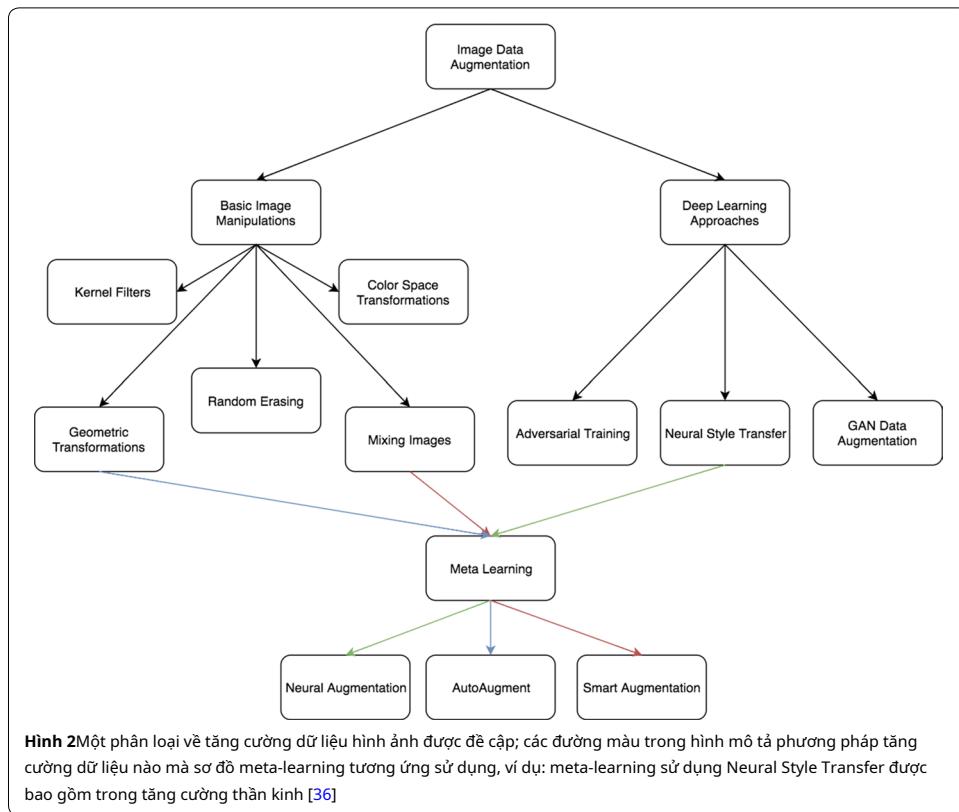
Nhiều chiến lược khác để tăng hiệu suất tổng quát hóa tập trung vào chính kiến trúc của mô hình. Điều này đã dẫn đến một chuỗi các kiến trúc ngày càng phức tạp hơn từ AlexNet [1] đến VGG-16 [2], ResNet [3], Khởi đầu-V3 [4] và DenseNet [5]. Các giải pháp chức năng như chính quy hóa bỏ học, chuẩn hóa hàng loạt, học chuyển và đào tạo trước đã được phát triển để cố gắng mở rộng Deep Learning cho ứng dụng trên các tập dữ liệu nhỏ hơn. Dưới đây là mô tả ngắn gọn về các giải pháp trang bị quá mức này. Một cuộc khảo sát đầy đủ về các phương pháp chính quy hóa trong Deep Learning đã được Kukacka et al. [6]. Kiến thức về các giải pháp trang bị quá mức này sẽ cung cấp cho người đọc về các công cụ hiện có khác, từ đó định hình bối cảnh cấp cao của Tăng cường dữ liệu và Học sâu.

- Rời ra ngoài [7] là một kỹ thuật chính quy hóa nhằm loại bỏ các giá trị kích hoạt của các nơ-ron được chọn ngẫu nhiên trong quá trình huấn luyện. Hạn chế này buộc mạng phải tìm hiểu các tính năng mạnh mẽ hơn thay vì dựa vào khả năng dự đoán của một tập hợp nhỏ các nơ-ron trong mạng. Thompson và cộng sự. [số 8] đã mở rộng ý tưởng này sang tích chập

mạng có Spatial Dropout, loại bỏ toàn bộ bản đồ đặc trưng thay vì từng nơ-ron riêng lẻ.

- Chuẩn hóa hàng loạt [9] là một kỹ thuật chính quy hóa khác giúp chuẩn hóa tập hợp kích hoạt trong một lớp. Quá trình chuẩn hóa hoạt động bằng cách trừ đi giá trị trung bình của lô từ mỗi lần kích hoạt và chia cho độ lệch chuẩn của lô. Kỹ thuật chuẩn hóa này, cùng với tiêu chuẩn hóa, là một kỹ thuật tiêu chuẩn trong quá trình tiền xử lý các giá trị pixel.
- Chuyển tiếp học tập [10,11] là một mô hình thú vị khác để ngăn chặn việc trang bị quá mức. Transfer Learning hoạt động bằng cách đào tạo mạng trên một tập dữ liệu lớn như ImageNet [12] và sau đó sử dụng các trọng số đó làm trọng số ban đầu trong nhiệm vụ phân loại mới. Thông thường, chỉ sao chép trọng số trong các lớp chập chứ không phải toàn bộ mạng bao gồm các lớp được kết nối đầy đủ. Điều này rất hiệu quả vì nhiều bộ dữ liệu hình ảnh có chung các đặc điểm không gian thấp nên được học tốt hơn với dữ liệu lớn. Tìm hiểu mối quan hệ giữa các miền dữ liệu được truyền là một nhiệm vụ nghiên cứu đang được tiến hành [13]. Yosinski và cộng sự. [14] thấy rằng khả năng chuyển đổi bị ảnh hưởng tiêu cực chủ yếu bởi sự chuyên môn hóa của các nơ-ron lớp cao hơn và những khó khăn trong việc phân tách các nơ-ron đồng thích nghi.
- Tập huấn trước [15] về mặt khái niệm rất giống với việc học chuyển giao. Trong Huấn luyện trước, kiến trúc mạng được xác định và sau đó được huấn luyện trên một tập dữ liệu lớn như ImageNet [12]. Điều này khác với Học chuyển giao vì trong Học chuyển giao, kiến trúc mạng như VGG-16 [2] hoặc ResNet [3] phải được chuyển giao cũng như trọng lượng. Huấn luyện trước cho phép khởi tạo trọng số bằng cách sử dụng bộ dữ liệu lớn, đồng thời vẫn mang lại sự linh hoạt trong thiết kế kiến trúc mạng.
- Học một lần và không cần học [16,17] Các thuật toán đại diện cho một mô hình khác để xây dựng các mô hình với dữ liệu cực kỳ hạn chế. Học một lần thường được sử dụng trong các ứng dụng nhận dạng khuôn mặt [18]. Một cách tiếp cận để học một lần là sử dụng mạng Xiêm [19] học một hàm khoảng cách sao cho có thể phân loại hình ảnh ngay cả khi mạng chỉ được huấn luyện trên một hoặc một vài trường hợp. Một cách tiếp cận rất phổ biến khác đối với phương pháp học một lần là sử dụng mạng tăng cường trí nhớ [20]. Học không bắn là một mô hình cực đoan hơn trong đó mạng sử dụng các phần nhúng vector đầu vào và đầu ra như Word2Vec [21] hoặc GloVe [22] để phân loại hình ảnh dựa trên thuộc tính mô tả.

Ngược lại với các kỹ thuật được đề cập ở trên, Tăng cường dữ liệu tiếp cận quá mức từ gốc rễ của vấn đề, tập dữ liệu huấn luyện. Điều này được thực hiện với giả định rằng có thể trích xuất nhiều thông tin hơn từ tập dữ liệu gốc thông qua các phần bổ sung. Những sự gia tăng này làm tăng kích thước tập dữ liệu huấn luyện một cách giả tạo bằng cách làm cong vênh dữ liệu hoặc lấy mẫu quá mức. Tăng cường làm cong vênh dữ liệu biến đổi các hình ảnh hiện có sao cho nhãn của chúng được giữ nguyên. Điều này bao gồm các tính năng bổ sung như biến đổi hình học và màu sắc, xóa ngẫu nhiên, huấn luyện đối nghịch và chuyển kiểu thần kinh. Tăng cường lấy mẫu quá mức tạo ra các phiên bản tổng hợp và thêm chúng vào tập huấn luyện. Điều này bao gồm trộn hình ảnh, tăng cường không gian đặc trưng và mạng đối thủ tổng hợp (GAN). Việc tăng cường lấy mẫu quá mức và làm cong vênh dữ liệu không tạo thành sự phân loại đôi lận nhau. Ví dụ: các mẫu GAN có thể được xếp chồng lên nhau bằng cách cắt ngẫu nhiên để tăng thêm tập dữ liệu. Các quyết định xung quanh kích thước tập dữ liệu cuối cùng,



việc tăng thời gian kiểm tra, học tập theo chương trình giảng dạy và tác động của việc giải quyết được đề cập trong khảo sát này dưới mục "[Cân nhắc thiết kế cho việc tăng cường dữ liệu hình ảnh](#)" phần. Mô tả các kỹ thuật tăng cường riêng lẻ sẽ được liệt kê trong "[Kỹ thuật tăng cường dữ liệu hình ảnh](#)" phần. Phân loại nhanh về Tăng cường dữ liệu được mô tả bên dưới trong [Hình.2](#).

Trước khi thảo luận về các kỹ thuật tăng cường hình ảnh, sẽ rất hữu ích khi xác định bối cảnh của vấn đề và xem xét điều gì khiến việc nhận dạng hình ảnh trở thành một nhiệm vụ khó khăn ngay từ đầu. Trong các ví dụ phân biệt đối xử cổ điển như mèo và chó, phần mềm nhận dạng hình ảnh phải khắc phục các vấn đề về góc nhìn, ánh sáng, độ che khuất, nền, tỷ lệ, v.v. Nhiệm vụ của Tăng cường dữ liệu là đưa các biến tịnh tiến này vào tập dữ liệu sao cho các mô hình kết quả sẽ hoạt động tốt chấp những thách thức này.

Người ta thường chấp nhận quan điểm rằng các bộ dữ liệu lớn hơn sẽ tạo ra các mô hình Deep Learning tốt hơn [23,24]. Tuy nhiên, việc tập hợp các bộ dữ liệu khổng lồ có thể là một nhiệm vụ rất khó khăn do nỗ lực thu thập và dán nhãn dữ liệu thủ công. Bộ dữ liệu hạn chế là một thách thức đặc biệt phổ biến trong phân tích hình ảnh y tế. Với dữ liệu lớn, mạng tích chập sâu đã được chứng minh là rất mạnh mẽ cho các nhiệm vụ phân tích hình ảnh y tế như phân loại tổn thương da như Esteva et al. [25]. Điều này đã truyền cảm hứng cho việc sử dụng CNN trong các nhiệm vụ phân tích hình ảnh y tế [26] chẳng hạn như phân loại tổn thương gan, phân tích quét não, tiếp tục nghiên cứu về phân loại tổn thương da, v.v. Nhiều hình ảnh được nghiên cứu có nguồn gốc từ chụp cắt lớp vi tính (CT) và chụp cộng hưởng từ (MRI), cả hai đều tổn kém và tổn nhiều công sức để thu thập. Việc xây dựng bộ dữ liệu hình ảnh y tế lớn đặc biệt khó khăn do bệnh hiếm gặp, bệnh nhân

quyền riêng tư, yêu cầu của các chuyên gia y tế về ghi nhãn cũng như chi phí và nỗ lực thủ công cần thiết để thực hiện các quy trình chụp ảnh y tế. Những trở ngại này đã dẫn đến nhiều nghiên cứu về Tăng cường dữ liệu hình ảnh, đặc biệt là lấy mẫu quá mức dựa trên GAN, từ góc độ ứng dụng phân loại hình ảnh y tế.

Nhiều nghiên cứu về tính hiệu quả của Tăng cường dữ liệu sử dụng các bộ dữ liệu hình ảnh học thuật phổ biến để đánh giá kết quả. Các bộ dữ liệu này bao gồm nhận dạng chữ số viết tay MNIST, CIFAR-10/100, ImageNet, tiny-imagenet-200, SVHN (số nhà ở chế độ xem phô), Caltech-101/256, địa điểm MIT, tập dữ liệu MIT-Adobe 5K, Pascal VOC và Xe Stanford. Các bộ dữ liệu được thảo luận thường xuyên nhất là CIFAR-10, CIFAR-100 và ImageNet. Việc mở rộng các bộ dữ liệu nguồn mở đã mang lại cho các nhà nghiên cứu nhiều trường hợp khác nhau để so sánh kết quả hiệu suất của các kỹ thuật Tăng cường dữ liệu. Hầu hết các bộ dữ liệu như ImageNet sẽ được phân loại là dữ liệu lớn. Nhiều thử nghiệm tự giới hạn mình vào một tập hợp con của tập dữ liệu để mô phỏng các vấn đề về dữ liệu hạn chế.

Ngoài việc tập trung vào các tập dữ liệu hạn chế, chúng tôi cũng sẽ xem xét vấn đề mất cân bằng lớp và cách Tăng cường dữ liệu có thể là một giải pháp lấy mẫu quá mức hữu ích. Sự mất cân bằng lớp mô tả một tập dữ liệu có tỷ lệ sai lệch giữa mẫu đa số và mẫu thiểu số. Leevy và cộng sự. [27] mô tả nhiều giải pháp hiện có cho sự mất cân bằng cấp cao giữa các loại dữ liệu. Khảo sát của chúng tôi sẽ cho thấy cách lấy mẫu quá mức cân bằng lớp trong dữ liệu hình ảnh có thể được thực hiện bằng Tăng cường dữ liệu.

Nhiều khía cạnh của mô hình Deep Learning và mạng lưới thần kinh được so sánh với trí thông minh của con người. Ví dụ, một giai thoại về trí tuệ con người về học chuyển giao được minh họa trong việc học nhạc. Nếu hai người đang cố gắng học cách chơi guitar và một người đã biết chơi piano thì có vẻ như người chơi piano sẽ học chơi guitar nhanh hơn. Tương tự như việc học âm nhạc, một mô hình có thể phân loại hình ảnh ImageNet có thể sẽ hoạt động tốt hơn trên hình ảnh CIFAR-10 so với mô hình có trọng số ngẫu nhiên.

Tăng cường dữ liệu tương tự như trí tưởng tượng hoặc giấc mơ. Con người tưởng tượng ra các kịch bản khác nhau dựa trên kinh nghiệm. Trí tưởng tượng giúp chúng ta hiểu rõ hơn về thế giới của chúng ta. Các phương pháp Tăng cường dữ liệu như GAN và Truyền kiểu thần kinh có thể 'tưởng tượng' những thay đổi đối với hình ảnh để chúng hiểu rõ hơn về chúng. Phần còn lại của bài viết được tổ chức như sau: Tóm tắt "[Lý lịch](#)" được cung cấp để cung cấp cho độc giả bối cảnh lịch sử của Tăng cường dữ liệu và Học sâu. "[Kỹ thuật tăng cường dữ liệu hình ảnh](#)" thảo luận chi tiết từng kỹ thuật tăng cường hình ảnh cùng với kết quả thử nghiệm. "[Cân nhắc thiết kế cho việc tăng cường dữ liệu hình ảnh](#)" thảo luận về các đặc điểm bổ sung của việc tăng cường chẵng hạn như tăng thời gian thử nghiệm và tác động của độ phân giải hình ảnh. Bài viết kết thúc bằng một "[Cuộc thảo luận](#)" của tài liệu được trình bày, các lĩnh vực "[Công việc tương lai](#)", Và "[Phản kết luận](#)".

Lý lịch

Có thể tìm thấy tính năng tăng cường hình ảnh dưới dạng cong vênh dữ liệu trong LeNet-5 [28]. Đây là một trong những ứng dụng đầu tiên của CNN về phân loại chữ số viết tay. Tăng cường dữ liệu cũng đã được nghiên cứu trong các ứng dụng lấy mẫu quá mức. Lấy mẫu quá mức là một kỹ thuật được sử dụng để lấy mẫu lại các phân phối lớp không cân bằng sao cho mô hình không quá thiên về việc ghi nhãn các trường hợp là loại lớp đa số. Lấy mẫu ngẫu nhiên (ROS) là một phương pháp đơn giản, sao chép hình ảnh một cách ngẫu nhiên từ

lớp thiểu số cho đến khi đạt được tỷ lệ lớp mong muốn. Kỹ thuật lấy mẫu quá mức thông minh có từ SMOTE (Kỹ thuật lấy mẫu quá mức cho thiểu số tổng hợp), được phát triển bởi Chawla et al. [29]. SMOTE và phần mở rộng của Borderline-SMOTE [30] tạo các phiên bản mới bằng cách nội suy các điểm mới từ các phiên bản hiện có thông qua k-Nearest Neighbors. Trọng tâm chính của kỹ thuật này là giảm bớt các vấn đề do mất cân bằng lớp và SMOTE chủ yếu được sử dụng cho dữ liệu dạng bảng và vectơ.

Kiến trúc AlexNet CNN được phát triển bởi Krizhevsky et al. [1] đã cách mạng hóa việc phân loại hình ảnh bằng cách áp dụng mạng chập cho bộ dữ liệu ImageNet. Tăng cường dữ liệu được sử dụng trong các thử nghiệm của họ để tăng kích thước tập dữ liệu lên gấp 2048. Điều này được thực hiện bằng cách cắt ngẫu nhiên 224×224 từ hình ảnh gốc, lật chúng theo chiều ngang và thay đổi cường độ của các kênh RGB bằng cách sử dụng tính năng tăng màu PCA. Tính năng Tăng cường dữ liệu này đã giúp giảm tình trạng bị quá mức khi đào tạo mạng lưới thần kinh sâu. Các tác giả cho rằng sự tăng cường của họ đã làm giảm tỷ lệ lỗi của mô hình hơn 1%.

Kể từ đó, GAN được giới thiệu vào năm 2014 [31], Chuyển giao phong cách thần kinh [32] vào năm 2015 và Tìm kiếm Kiến trúc Thần kinh (NAS) [33] vào năm 2017. Nhiều công trình khác nhau về các tiện ích mở rộng GAN như DCGAN, CycleGAN và GAN đang phát triển dần dần [34] được xuất bản lần lượt vào năm 2015, 2017 và 2017. Sự chuyển giao phong cách thần kinh được đẩy nhanh cùng với sự phát triển của Những tổn thất về nhận thức của Johnson và cộng sự. [35] vào năm 2016. Việc áp dụng các khái niệm metalearning từ NAS vào Tăng cường dữ liệu ngày càng trở nên phổ biến với các công trình như Tăng cường thần kinh [36], Tăng cường thông minh [37] và Tự động tăng cường [38] được xuất bản lần lượt vào năm 2017, 2017 và 2018.

Áp dụng Deep Learning vào hình ảnh y tế đã là một ứng dụng phổ biến của CNN kể từ khi chúng trở nên phổ biến vào năm 2012. Deep Learning và hình ảnh y tế ngày càng trở nên phổ biến với sự trình diễn khả năng phát hiện ung thư da ở cấp độ bác sĩ da liễu của Esteva et al. [25] vào năm 2017.

Việc sử dụng GAN trong hình ảnh y tế đã được ghi lại rõ ràng trong một cuộc khảo sát của Yi et al. [39]. Khảo sát này đề cập đến việc sử dụng GAN trong tái thiết như khử nhiễu CT [40], chụp ảnh cộng hưởng từ tăng tốc [41], khử nhiễu PET [42] và ứng dụng GAN siêu phân giải trong phân đoạn mạch máu võng mạc [43]. Ngoài ra, Yi và cộng sự. [39] đề cập đến việc sử dụng tổng hợp hình ảnh GAN trong các ứng dụng hình ảnh y tế như tổng hợp MRI não [44,45], chẩn đoán ung thư phổi [46], tổng hợp tổn thương da có độ phân giải cao [47], và phân loại bất thường trên X-quang ngực [48]. Tổng hợp hình ảnh dựa trên GAN Tăng cường dữ liệu được sử dụng bởi Frid-Adar et al. [49] vào năm 2018 để phân loại tổn thương gan. Điều này đã cải thiện hiệu suất phân loại từ độ nhạy 78,6% và độ đặc hiệu 88,4% bằng cách sử dụng các mức tăng cổ điển lên độ nhạy 85,7% và độ đặc hiệu 92,4% bằng cách sử dụng Tăng cường dữ liệu dựa trên GAN.

Hầu hết các cải tiến được đề cập đều trung vào việc cải thiện các mô hình Nhận dạng Hình ảnh. Nhận dạng hình ảnh là khi một mô hình dự đoán nhãn đầu ra, chẳng hạn như 'chó' hoặc 'mèo' cho hình ảnh đầu vào.

Tuy nhiên, có thể mở rộng kết quả từ nhận dạng hình ảnh sang các tác vụ Thị giác máy tính khác như Phát hiện đối tượng do thuật toán YOLO [50], R-CNN [51], R-CNN nhanh [52] và R-CNN nhanh hơn [53] hoặc Phân đoạn theo ngữ nghĩa [54] bao gồm các thuật toán như U-Net [55].

Kỹ thuật tăng cường dữ liệu hình ảnh

Các minh chứng sớm nhất cho thấy tính hiệu quả của Tăng cường dữ liệu đến từ các phép biến đổi đơn giản như lật ngang, tăng cường không gian màu và cắt xén ngẫu nhiên.

Những phép biến đổi này mã hóa nhiều bất biến đã thảo luận trước đó, đặt ra những thách thức đối với các nhiệm vụ nhận dạng hình ảnh. Các phần mở rộng được liệt kê trong khảo sát này là các phép biến đổi hình học, biến đổi không gian màu, bộ lọc hạt nhân, trộn hình ảnh, xóa ngẫu nhiên, tăng không gian tính năng, đào tạo đối thủ, tăng cường dựa trên GAN, chuyển kiểu thần kinh và sơ đồ siêu học. Phần này sẽ giải thích cách hoạt động của từng thuật toán tăng cường, báo cáo kết quả thử nghiệm và thảo luận về những nhược điểm của kỹ thuật tăng cường.

Tăng cường dữ liệu dựa trên các thao tác hình ảnh cơ bản

Các phép biến đổi hình học

Phần này mô tả các phép tăng cường khác nhau dựa trên các phép biến đổi hình học và nhiều chức năng xử lý ảnh khác. Lớp tăng cường được thảo luận dưới đây có thể được đặc trưng bởi tính dễ thực hiện của chúng. Việc hiểu những chuyển đổi này sẽ cung cấp cơ sở hữu ích để nghiên cứu sâu hơn về các kỹ thuật Tăng cường dữ liệu.

Chúng tôi cũng mô tả các phép tăng cường hình học khác nhau trong bối cảnh ứng dụng 'an toàn' của chúng. Tính an toàn của phương pháp Tăng cường dữ liệu đề cập đến khả năng duy trì nhãn sau chuyển đổi. Ví dụ: xoay và lật thường an toàn trong các thử thách ImageNet chẳng hạn như mèo so với chó, nhưng không an toàn đối với các nhiệm vụ nhận dạng chữ số như 6 so với 9. Một phép biến đổi bảo toàn không có nhãn có thể có khả năng tăng cường khả năng của mô hình trong việc đưa ra phản hồi chỉ ra rằng nó không tự tin về dự đoán của mình. Tuy nhiên, để đạt được điều này đòi hỏi phải có nhãn hiệu tinh tế [56] sau tăng cường. Nếu nhãn của hình ảnh sau khi chuyển đổi bảo toàn không nhãn có giá trị như [0,5 0,5], thì mô hình có thể tìm hiểu các dự đoán về độ tin cậy mạnh mẽ hơn. Tuy nhiên, việc xây dựng các nhãn được tinh chỉnh cho mỗi lần Tăng cường dữ liệu không an toàn là một quá trình tốn kém về mặt tính toán.

Do thách thức trong việc xây dựng các nhãn được tinh chỉnh cho dữ liệu sau tăng cường, điều quan trọng là phải xem xét tính 'an toàn' của việc tăng cường. Điều này phần nào phụ thuộc vào miền, tạo ra thách thức cho việc phát triển các chính sách tăng cường tổng quát, (xem [Tự động tăng cường](#) [38] để khám phá thêm về việc tìm kiếm các phần mở rộng có thể khai quát hóa). Không có chức năng xử lý hình ảnh nào không thể dẫn đến chuyển đổi nhãn thay đổi ở một mức độ méo nào đó. Điều này thể hiện thiết kế tăng cường theo dữ liệu cụ thể và thách thức trong việc phát triển các chính sách tăng cường có thể khai quát hóa. Đây là một cân nhắc quan trọng đối với các phép tăng cường hình học được liệt kê dưới đây.

Lật

Việc lật trực ngang phổ biến hơn nhiều so với việc lật trực dọc. Phần mở rộng này là một trong những cách dễ thực hiện nhất và đã được chứng minh là hữu ích trên các bộ dữ liệu như CIFAR-10 và ImageNet. Trên các tập dữ liệu liên quan đến nhận dạng văn bản như MNIST hoặc SVHN, đây không phải là phép chuyển đổi bảo toàn nhãn.

Không gian màu

Dữ liệu hình ảnh kỹ thuật số thường được mã hóa dưới dạng tensor của kích thước (chiều cao×chiều rộng×kênh màu). Thực hiện tăng cường trong không gian kênh màu là một chiến lược khác rất thiết thực để thực hiện. Việc tăng màu rất đơn giản bao gồm cách ly một kênh màu duy nhất như R, G hoặc B. Một hình ảnh có thể được chuyển đổi nhanh chóng thành biểu diễn của nó trong một kênh màu bằng cách tách ma trận đó và thêm 2 ma trận 0 từ các kênh màu khác. Ngoài ra, các giá trị RGB có thể được thao tác dễ dàng bằng các thao tác ma trận đơn giản để tăng hoặc giảm độ sáng của hình ảnh. Việc tăng cường màu sắc nâng cao hơn đến từ việc lấy biểu đồ màu mô tả hình ảnh. Việc thay đổi giá trị cường độ trong các biểu đồ này dẫn đến thay đổi ánh sáng chẳng hạn như ánh sáng được sử dụng trong các ứng dụng chỉnh sửa ảnh.

Cắt xén

Cắt xén hình ảnh có thể được sử dụng như một bước xử lý thực tế đối với dữ liệu hình ảnh có kích thước chiều cao và chiều rộng hỗn hợp bằng cách cắt xén một phần trung tâm của mỗi hình ảnh. Ngoài ra, việc cắt xén ngẫu nhiên cũng có thể được sử dụng để mang lại hiệu ứng rất giống với bản dịch. Sự tương phản giữa cắt ngẫu nhiên và dịch là việc cắt xén sẽ làm giảm kích thước của đầu vào như $(256,256) \rightarrow (224, 224)$, trong khi các bản dịch giữ nguyên kích thước không gian của hình ảnh. Tùy thuộc vào ngữ cảnh giảm được chọn để cắt xén, đây có thể không phải là phép chuyển đổi bảo toàn nhãn.

Vòng xoay

Tăng cường xoay được thực hiện bằng cách xoay hình ảnh sang phải hoặc sang trái trên một trục trong khoảng từ 1° đến 359° . Sự an toàn của việc tăng cường xoay được xác định chủ yếu bởi tham số độ quay. Xoay nhẹ chẳng hạn như từ 1 đến 20 hoặc -1 đến -20 có thể hữu ích trong các tác vụ nhận dạng chữ số như MNIST, nhưng khi mức độ xoay tăng lên, nhãn của dữ liệu không còn được giữ nguyên sau khi chuyển đổi.

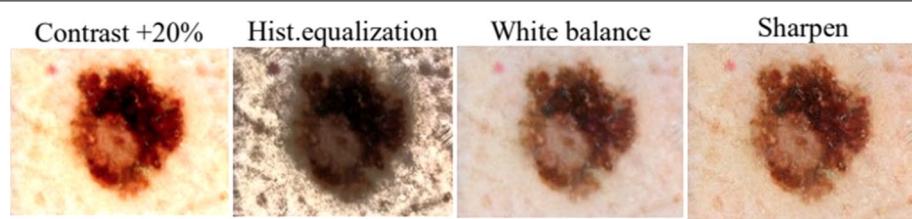
Dịch

Dịch chuyển hình ảnh sang trái, phải, lên hoặc xuống có thể là một phép chuyển đổi rất hữu ích để tránh sai lệch vị trí trong dữ liệu. Ví dụ: nếu tất cả các hình ảnh trong tập dữ liệu đều được căn giữa, điều này thường xảy ra trong các tập dữ liệu nhận dạng khuôn mặt, thì điều này cũng yêu cầu mô hình phải được kiểm tra trên các hình ảnh được căn giữa hoàn hảo. Khi hình ảnh gốc được dịch theo một hướng, không gian còn lại có thể được lấp đầy bằng một giá trị không đổi như 0 s hoặc 255 s hoặc có thể được lấp đầy bằng nhiều ngẫu nhiên hoặc nhiều Gauss. Phần đệm này bảo toàn các kích thước không gian của hình ảnh sau khi tăng cường.

Tiêm tiếng ồn

Việc chèn nhiễu bao gồm việc đưa vào một ma trận các giá trị ngẫu nhiên thường được rút ra từ phân bố Gaussian. Việc giảm tiếng ồn được thử nghiệm bởi Moreno-Barea et al. [57] trên chín bộ dữ liệu từ kho lưu trữ UCI [58]. Việc thêm nhiễu vào hình ảnh có thể giúp CNN tìm hiểu các tính năng mạnh mẽ hơn.

Các phép biến đổi hình học là giải pháp rất tốt cho các sai lệch vị trí có trong dữ liệu huấn luyện. Có nhiều nguồn sai lệch tiềm tàng có thể tách biệt các



Hình 3Ví dụ về Tăng cường màu sắc do Mikolajczyk và Grochowski cung cấp [72] trong lĩnh vực phân loại khối u ác tính

phân phối dữ liệu huấn luyện từ dữ liệu kiểm tra. Nếu có sự thiên vị về vị trí, chẳng hạn như trong tập dữ liệu nhận dạng khuôn mặt trong đó mọi khuôn mặt đều được căn giữa hoàn hảo trong khung, thì các phép biến đổi hình học là một giải pháp tuyệt vời. Ngoài khả năng mạnh mẽ trong việc khắc phục các sai lệch vị trí, các phép biến đổi hình học cũng rất hữu ích vì chúng dễ thực hiện. Có nhiều thư viện xử lý hình ảnh giúp bạn bắt đầu thực hiện các thao tác như lật ngang và xoay dễ dàng. Một số nhược điểm của phép biến đổi hình học bao gồm bộ nhớ bổ sung, chi phí tính toán phép biến đổi và thời gian đào tạo bổ sung. Một số phép biến đổi hình học như dịch chuyển hoặc cắt xén ngẫu nhiên phải được quan sát thủ công để đảm bảo chúng không làm thay đổi nhãn của hình ảnh. Cuối cùng, trong nhiều lĩnh vực ứng dụng được đề cập, chẳng hạn như phân tích hình ảnh y tế, các sai lệch làm lệch dữ liệu huấn luyện khỏi dữ liệu thử nghiệm phức tạp hơn các phương sai vị trí và tịnh tiến. Do đó, phạm vi áp dụng các phép biến đổi hình học ở đâu và khi nào là tương đối hạn chế.

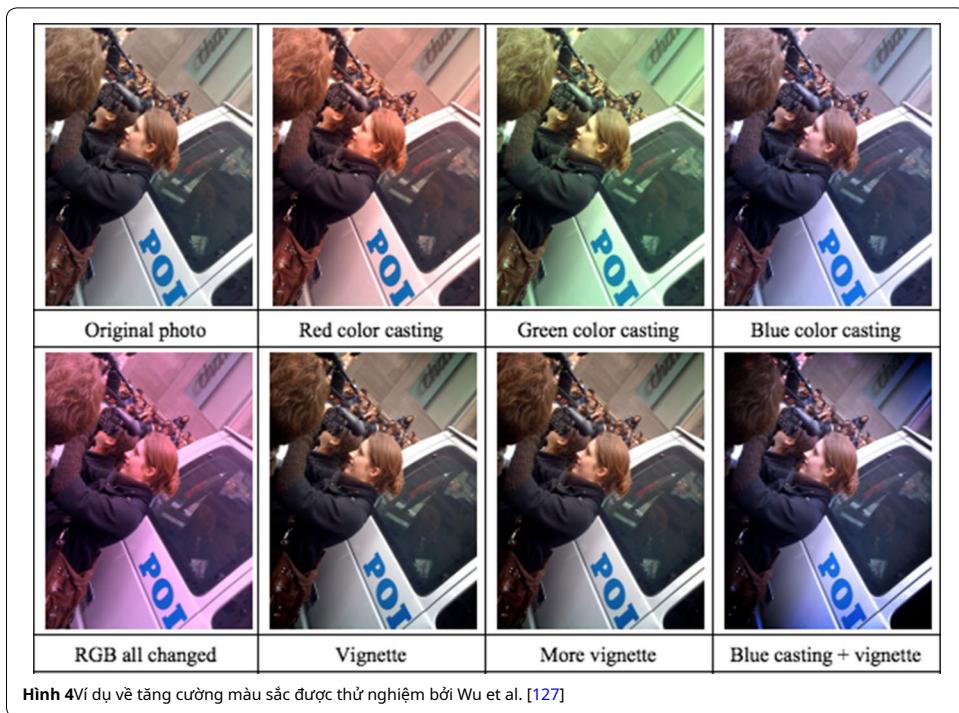
Chuyển đổi không gian màu

Dữ liệu hình ảnh được mã hóa thành 3 ma trận xếp chồng lên nhau, mỗi ma trận có kích thước chiều cao×chiều rộng. Các ma trận này biểu thị các giá trị pixel cho một giá trị màu RGB riêng lẻ. Độ lệch ánh sáng là một trong những thách thức xảy ra thường xuyên nhất đối với các vấn đề nhận dạng hình ảnh. Do đó, hiệu quả của các phép biến đổi không gian màu, còn được gọi là các phép biến đổi trắc quang, khá trực quan để khái niệm hóa. Cách khắc phục nhanh nhất đối với hình ảnh quá sáng hoặc quá tối là lặp qua các hình ảnh và giảm hoặc tăng giá trị pixel theo một giá trị không đổi. Một thao tác không gian màu nhanh khác là tách các ma trận màu RGB riêng lẻ. Một phép biến đổi khác bao gồm việc giới hạn các giá trị pixel ở một giá trị tối thiểu hoặc tối đa nhất định. Sự thể hiện nội tại của màu sắc trong hình ảnh kỹ thuật số phù hợp với nhiều chiến lược tăng cường.

Các phép biến đổi không gian màu cũng có thể được bắt nguồn từ các ứng dụng chỉnh sửa hình ảnh. Các giá trị pixel của hình ảnh trong mỗi kênh màu RGB được tổng hợp để tạo thành biểu đồ màu. Biểu đồ này có thể được thao tác để áp dụng các bộ lọc làm thay đổi đặc điểm không gian màu của hình ảnh.

Có rất nhiều sự tự do để sáng tạo với việc tăng cường không gian màu. Việc thay đổi sự phân bổ màu sắc của hình ảnh có thể là một giải pháp tuyệt vời cho những thách thức về ánh sáng mà dữ liệu thử nghiệm gặp phải (Hình 2).[3,4](#).

Bộ dữ liệu hình ảnh có thể được đơn giản hóa trong việc biểu diễn bằng cách chuyển đổi ma trận RGB thành một hình ảnh thang độ xám duy nhất. Điều này dẫn đến hình ảnh nhỏ hơn, chiều cao×chiều rộng×1,



Hình 4 Ví dụ về tăng cường màu sắc được thử nghiệm bởi Wu et al. [127]

dẫn đến tính toán nhanh hơn. Tuy nhiên, điều này đã được chứng minh là làm giảm độ chính xác của hiệu suất. Chatfield và cộng sự. [59] nhận thấy độ chính xác phân loại giảm ~ 3% giữa hình ảnh thang độ xám và hình ảnh RGB với các thử nghiệm của họ trên ImageNet [12] và PASCAL [60] Tập dữ liệu VOC. Ngoài hình ảnh RGB so với thang độ xám, còn có nhiều cách khác để thể hiện màu kỹ thuật số như HSV (Hue, Saturation và Value). Jurio và cộng sự. [61] khám phá hiệu suất của Phân đoạn hình ảnh trên nhiều cách biểu diễn không gian màu khác nhau từ RGB đến YUV, CMY và HSV.

Tương tự như các phép biến đổi hình học, nhược điểm của phép biến đổi không gian màu là tăng bộ nhớ, chi phí chuyển đổi và thời gian huấn luyện. Ngoài ra, các phép biến đổi màu có thể loại bỏ thông tin màu quan trọng và do đó không phải lúc nào cũng là phép biến đổi bảo toàn nhãn. Ví dụ: khi giảm giá trị pixel của hình ảnh để mô phỏng môi trường tối hơn, có thể không thể nhìn thấy các đối tượng trong hình ảnh. Một ví dụ giàn tiếp khác về chuyển đổi màu không bảo toàn nhãn là trong Phân tích cảm tính hình ảnh [62]. Trong ứng dụng này, CNN cố gắng dự đoán trực quan điểm cảm xúc của một hình ảnh, chẳng hạn như: rất tiêu cực, tiêu cực, trung tính, tích cực hoặc rất tích cực. Một dấu hiệu của hình ảnh tiêu cực/rất tiêu cực là sự hiện diện của máu. Màu đỏ sẫm của máu là thành phần quan trọng để phân biệt máu với nước hoặc sơn. Nếu các phép biến đổi không gian màu liên tục thay đổi không gian màu sao cho mô hình không thể nhận ra máu đỏ từ sơn xanh lục thì mô hình sẽ hoạt động kém trong Phân tích cảm tính hình ảnh. Trên thực tế, các phép biến đổi không gian màu sẽ loại bỏ các sai lệch màu có trong tập dữ liệu theo hướng có lợi cho các đặc điểm không gian. Tuy nhiên, đối với một số nhiệm vụ, màu sắc là đặc điểm phân biệt rất quan trọng.

Bảng 1 Kết quả thí nghiệm Tăng cường Dữ liệu của Taylor và Nitschke trên Caltech101 [63]

	Độ chính xác cao nhất (%)	Độ chính xác của top 5 (%)
Đường cơ sở	48,13±0,42	64,50±0,65
lật	49,73±1.13	67,36±138
Xoay	50,80±0,63	69,41±0,48
Cắt xén	61,95±1,01	79,10±0,80
Hiện tượng giật màu	49,57±0,53	67,18±0,42
Mở rộng đường biên	49,29 + 1,16	66,49 + 0,84
PCA ưa thích	49,41±0,84	67,54±1,01

Kết quả của họ cho thấy phép biến đổi hình học cắt xén mang lại kết quả phân loại chính xác nhất. Giá trị in nghiêng biểu thị hiệu suất cao theo các số liệu so sánh

Các phép biến đổi hình học và trắc quang

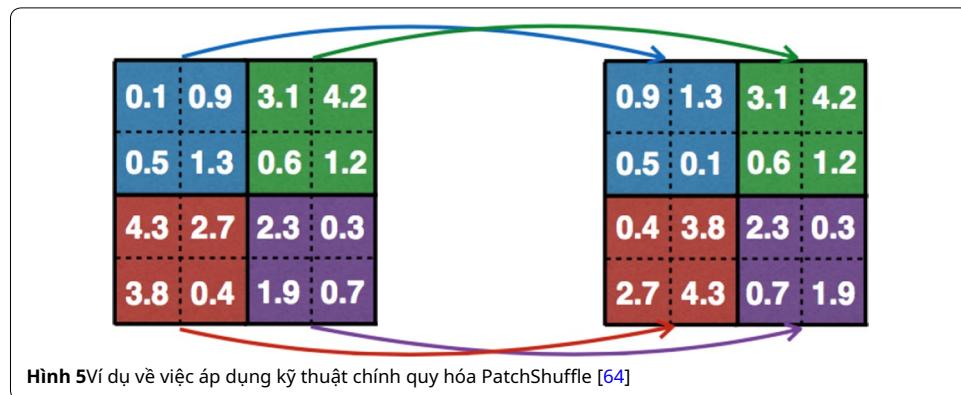
Taylor và Nitschke [63] cung cấp một nghiên cứu so sánh về hiệu quả của các phép biến đổi hình học và trắc quang (không gian màu). Các phép biến đổi hình học được nghiên cứu là lật, xoay -30° đến 30° và cắt xén. Các phép biến đổi không gian màu được nghiên cứu là hiện tượng biến đổi màu, (thao tác màu ngẫu nhiên), tăng cường cạnh và PCA. Họ đã thử nghiệm những cải tiến này bằng cách thực chéo 4 lần trên tập dữ liệu Caltech101 được lọc thành 8421 hình ảnh có kích thước 256×256 (Bảng1).

Bộ lọc hạt nhân

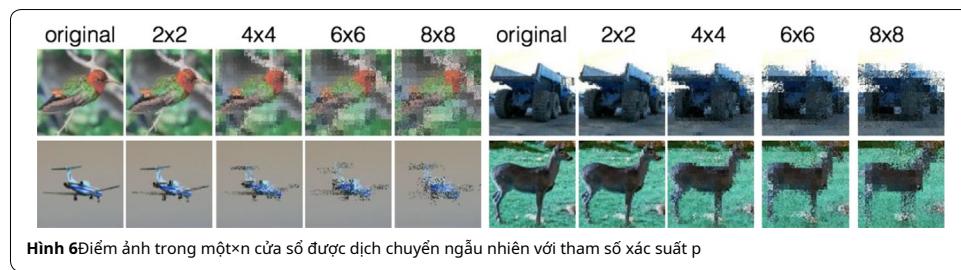
Bộ lọc hạt nhân là một kỹ thuật rất phổ biến trong xử lý ảnh để làm sắc nét và làm mờ hình ảnh. Các bộ lọc này hoạt động bằng cách trượt một $N \times N$ ma trận trên một hình ảnh bằng bộ lọc làm mờ Gaussian, sẽ dẫn đến hình ảnh mờ hơn hoặc bộ lọc cạnh dọc hoặc ngang có độ tương phản cao sẽ mang lại hình ảnh sắc nét hơn dọc theo các cạnh. Theo trực giác, việc làm mờ hình ảnh để Tăng cường dữ liệu có thể dẫn đến khả năng chống mờ chuyển động cao hơn trong quá trình thử nghiệm. Ngoài ra, việc làm sắc nét hình ảnh để Tăng cường dữ liệu có thể giúp gói gọn nhiều chi tiết hơn về các đối tượng quan tâm.

Làm sắc nét và làm mờ là một số cách cổ điển để áp dụng bộ lọc hạt nhân cho hình ảnh. Kang và cộng sự. [64] thử nghiệm một bộ lọc hạt nhân duy nhất có khả năng hoán đổi ngẫu nhiên các giá trị pixel trong một $N \times N$ cửa sổ trượt. Họ gọi kỹ thuật tăng cường này là PatchShuffle Regularization. Thử nghiệm trên các kích thước bộ lọc khác nhau và xác suất xáo trộn các pixel ở mỗi bước, họ chứng minh tính hiệu quả của việc này bằng cách đạt được tỷ lệ lỗi 5,66% trên CIFAR-10 so với tỷ lệ lỗi đạt được là 6,33% khi không sử dụng Chính quy PatchShuffle. Cài đặt siêu tham số đạt được điều này bao gồm 2×2 bộ lọc và xác suất hoán đổi là 0,05. Những thử nghiệm này được thực hiện bằng ResNet [3] Kiến trúc CNN (Hình.5,6).

Bộ lọc hạt nhân là một lĩnh vực tương đối chưa được khám phá để Tăng cường dữ liệu. Nhược điểm của kỹ thuật này là nó rất giống với cơ chế bên trong của CNN. CNN có các hạt nhân tham số học cách tối ưu để thể hiện hình ảnh theo từng lớp. Ví dụ: một cái gì đó như PatchShuffle Regularization có thể được triển khai với lớp chập. Điều này có thể đạt được bằng cách sửa đổi các tham số lớp tích chập tiêu chuẩn sao cho các tham số đệm duy trì độ phân giải không gian và lớp kích hoạt tiếp theo giữ các giá trị pixel trong khoảng từ 0 đến 255, trái ngược với thứ gì đó giống như kích hoạt sigmoid ánh xạ các pixel thành các giá trị trong khoảng từ 0 đến 1. Do đó hạt nhân



Hình 5Ví dụ về việc áp dụng kỹ thuật chính quy hóa PatchShuffle [64]



Hình 6Điểm ảnh trong một khung cửa sổ được dịch chuyển ngẫu nhiên với tham số xác suất p

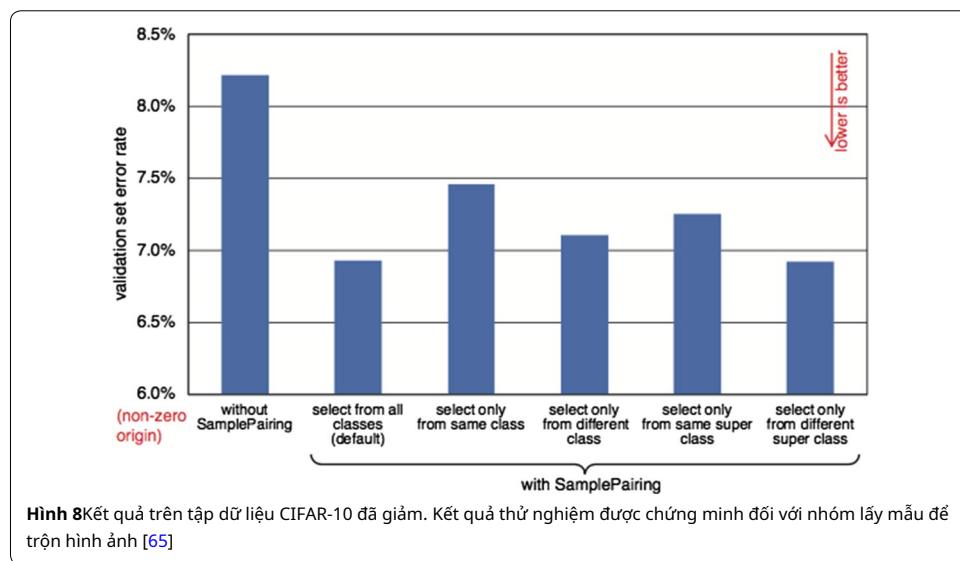
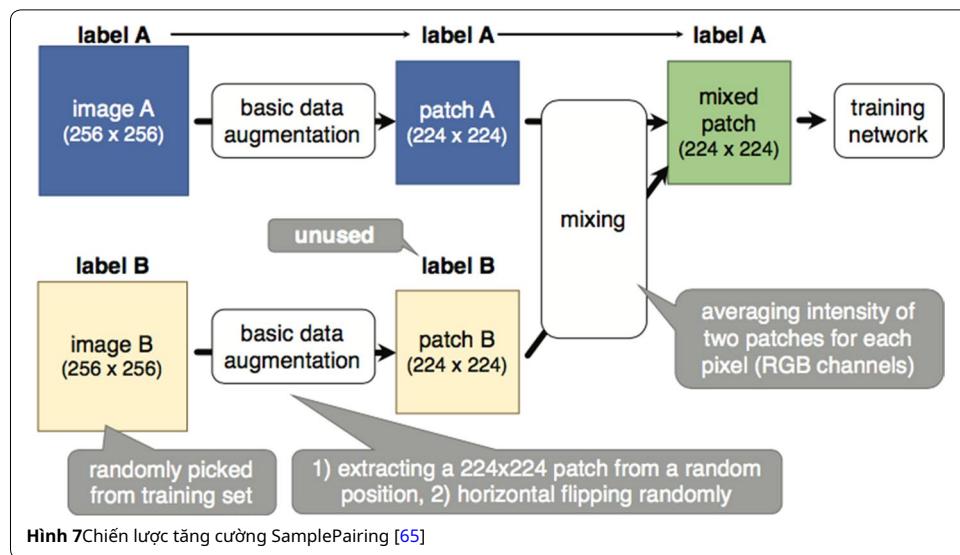
các bộ lọc có thể được triển khai tốt hơn dưới dạng một lớp của mạng thay vì dưới dạng bổ sung cho tập dữ liệu thông qua Tăng cường dữ liệu.

Trộn hình ảnh

Trộn các hình ảnh lại với nhau bằng cách lấy trung bình các giá trị pixel của chúng là một cách tiếp cận rất phản trực giác đối với Tăng cường dữ liệu. Những hình ảnh được tạo ra bằng cách này sẽ không giống như một sự biến đổi hữu ích đối với người quan sát. Tuy nhiên, Ionue [65] đã chứng minh làm thế nào việc ghép các mẫu có thể được phát triển thành một chiến lược tăng cường hiệu quả. Trong thí nghiệm này, hai hình ảnh được cắt ngẫu nhiên từ 256×256 đến 224×224 và ngẫu nhiên lật ngang. Những hình ảnh này sau đó được trộn bằng cách lấy trung bình các giá trị pixel cho từng kênh RGB. Điều này dẫn đến một hình ảnh hỗn hợp được sử dụng để huấn luyện mô hình phân loại. Nhãn được gán cho hình ảnh mới giống với hình ảnh được chọn ngẫu nhiên đầu tiên (Hình 2).[7](#).

Trên tập dữ liệu CIFAR-10, Ionue đã báo cáo tỷ lệ lỗi giảm từ 8,22 xuống 6,93% khi sử dụng kỹ thuật Tăng cường dữ liệu SamplePairing. Nhà nghiên cứu thậm chí còn tìm thấy kết quả tốt hơn khi thử nghiệm tập dữ liệu có kích thước giảm, giảm tổng số CIFAR-10 xuống còn 1000 mẫu với 100 mẫu trong mỗi lớp. Với tập dữ liệu có kích thước giảm, SamplePairing giúp giảm tỷ lệ lỗi từ 43,1 xuống 31,0%. Kết quả CIFAR-10 chứng tỏ tính hữu ích của kỹ thuật SamplePairing trong các ứng dụng dữ liệu hạn chế (Hình 2).[8](#).

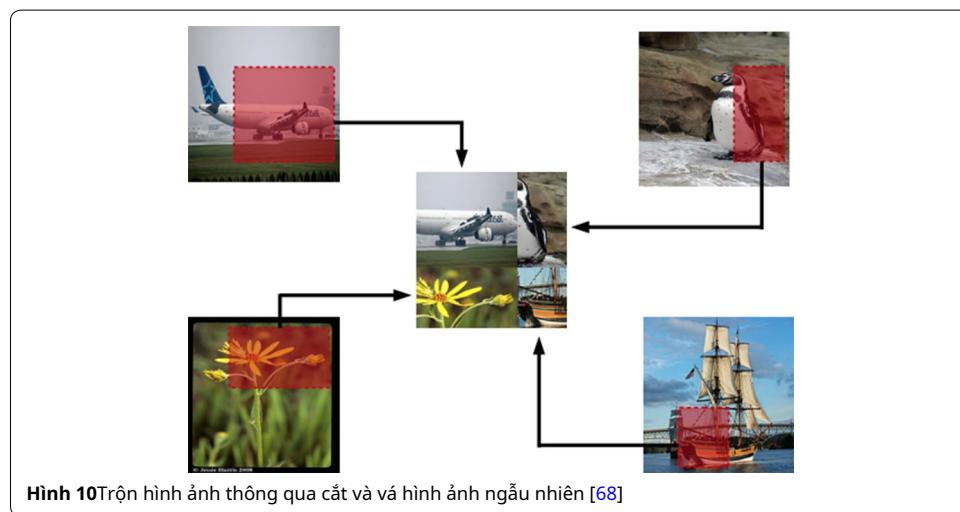
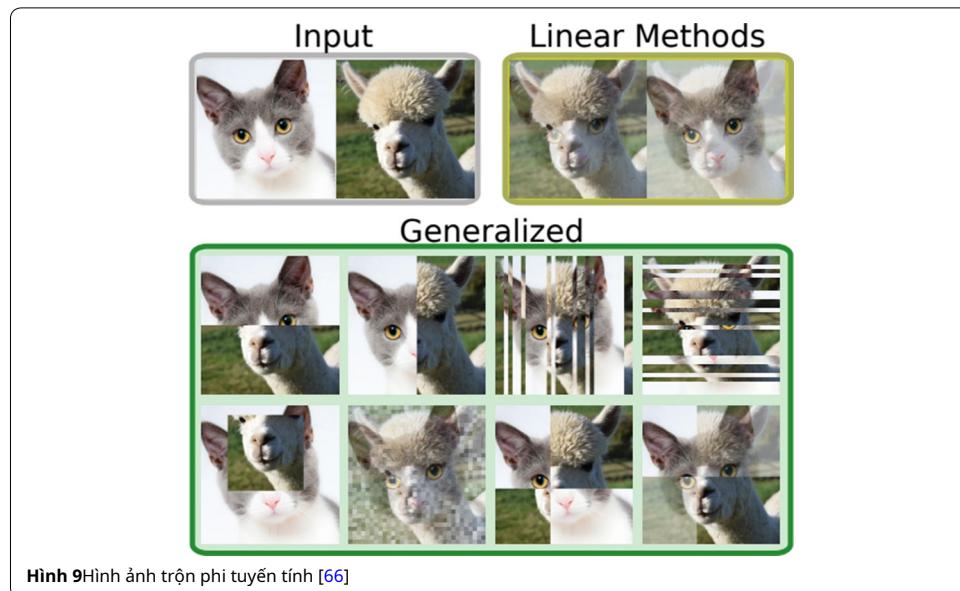
Một chi tiết khác được tìm thấy trong nghiên cứu là kết quả tốt hơn thu được khi trộn hình ảnh từ toàn bộ tập huấn luyện thay vì từ các trường hợp chỉ thuộc cùng một lớp. Bắt đầu từ tập huấn luyện có kích thước N , SamplePairing tạo ra tập dữ liệu có kích thước N_2+N . Ngoài ra, Ghép mẫu có thể được xếp chồng lên trên các kỹ thuật tăng cường khác. Ví dụ: nếu sử dụng các phép tăng cường được thể hiện trong



bài báo AlexNet của Krizhevsky et al. [1], năm 2048×việc tăng dữ liệu có thể được mở rộng hơn nữa đến (2048×N)₂.

Khái niệm trộn hình ảnh theo cách không trực quan đã được Summers và Dinneen nghiên cứu sâu hơn [66]. Họ đã xem xét việc sử dụng các phương pháp phi tuyến tính để kết hợp hình ảnh vào các phiên bản đào tạo mới. Tất cả các phương pháp họ sử dụng đều mang lại hiệu suất tốt hơn so với các mô hình cơ bản (Hình 2).⁹.

Trong số các mức tăng phi tuyến tính được thử nghiệm này, kỹ thuật tốt nhất đã giúp giảm sai số từ 5,4 xuống 3,8% trên CIFAR-10 và 23,6% xuống 19,7% trên CIFAR-100. Tương tự như vậy, Liang et al. [67] đã sử dụng GAN để tạo ra các hình ảnh hỗn hợp. Họ nhận thấy rằng việc đưa các hình ảnh hỗn hợp vào dữ liệu huấn luyện giúp giảm thời gian huấn luyện và tăng tính đa dạng của các mẫu GAN. Takahashi và Matsubara [68] cuộc thí nghiệm



bằng một cách tiếp cận khác để trộn các hình ảnh cắt ngẫu nhiên các hình ảnh và ghép các phần cắt lại với nhau để tạo thành các hình ảnh mới như mô tả bên dưới. Kết quả kỹ thuật của họ, cũng như khả năng tăng cường ghép mẫu và trộn lẫn, chứng minh tính hiệu quả đối khi không hợp lý của dữ liệu lớn với các mô hình Deep Learning (Hình 2).10).

Một nhược điểm rõ ràng của kỹ thuật này là nó ít có ý nghĩa từ góc độ con người. Việc tăng hiệu suất được tìm thấy từ việc trộn hình ảnh là rất khó hiểu hoặc giải thích. Một lời giải thích khả dĩ cho điều này là kích thước tập dữ liệu tăng lên dẫn đến sự thể hiện mạnh mẽ hơn các đặc điểm cấp thấp như đường và cạnh. Việc kiểm tra hiệu suất của điều này so với các phương pháp học chuyển tiếp và đào tạo trước là một lĩnh vực thú vị cho công việc trong tương lai. Học chuyển giao và đào tạo trước là những kỹ thuật khác giúp tìm hiểu các đặc điểm cấp thấp trong CNN. Ngoài ra, nó sẽ được



Thật thú vị khi xem hiệu suất thay đổi như thế nào nếu chúng ta phân vùng dữ liệu huấn luyện sao cho 100 kỷ nguyên đầu tiên được huấn luyện với hình ảnh gốc và hỗn hợp và 50 kỷ nguyên cuối cùng chỉ với hình ảnh gốc. Những loại chiến lược này sẽ được thảo luận sâu hơn trong Cân nhắc thiết kế về tăng cường dữ liệu liên quan đến việc học trong chương trình giảng dạy [69]. Ngoài ra, bài viết sẽ đề cập đến một kỹ thuật siêu học được phát triển bởi Lemley và cộng sự. [37] sử dụng mạng lưới thần kinh để tìm hiểu cách trộn hình ảnh tối ưu.

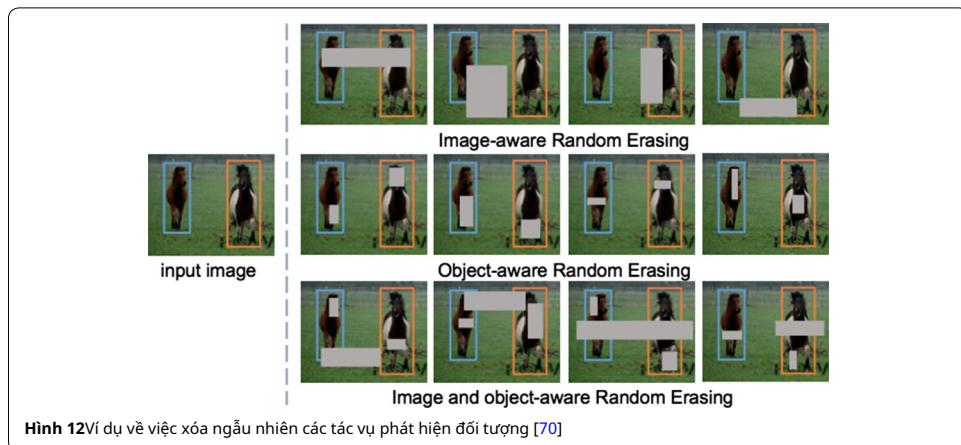
Xóa ngẫu nhiên

Xóa ngẫu nhiên [70] là một kỹ thuật Tăng cường dữ liệu thú vị khác được phát triển bởi Zhong et al. Lấy cảm hứng từ các cơ chế chính quy hóa bỏ đi, việc xóa ngẫu nhiên có thể được coi là tương tự như bỏ đi ngoại trừ trong không gian dữ liệu đầu vào thay vì được nhúng vào kiến trúc mạng. Kỹ thuật này được thiết kế đặc biệt để giải quyết các thách thức về nhận dạng hình ảnh do tắc nghẽn. Tắc nghẽn đề cập đến khi một số phần của đối tượng không rõ ràng. Việc xóa ngẫu nhiên sẽ ngăn chặn điều này bằng cách buộc mô hình tìm hiểu thêm các đặc điểm mô tả về một hình ảnh, ngăn không cho nó khớp quá mức với một đặc điểm hình ảnh nhất định trong hình ảnh. Đặc biệt, bên cạnh thách thức trực quan về sự tắc nghẽn, việc xóa ngẫu nhiên là một kỹ thuật đầy hứa hẹn để đảm bảo chú ý đến toàn bộ hình ảnh, thay vì chỉ một tập hợp con của nó.

Xóa ngẫu nhiên hoạt động bằng cách chọn ngẫu nhiên một $n \times m$ của một hình ảnh và che nó bằng 0 giây, 255 giây, giá trị pixel trung bình hoặc giá trị ngẫu nhiên. Trên tập dữ liệu CIFAR-10, điều này giúp giảm tỷ lệ lỗi từ 5,17 xuống 4,31%. Phương pháp điền bản vá tốt nhất được tìm thấy là các giá trị ngẫu nhiên. Phương pháp tô màu và kích thước của mặt nạ là các thông số duy nhất cần được thiết kế bằng tay trong quá trình thực hiện (Hình 2).[11, 12](#).

Xóa ngẫu nhiên là một phương pháp Tăng cường dữ liệu nhằm tìm cách trực tiếp ngăn chặn việc trang bị quá mức bằng cách thay đổi không gian đầu vào. Bằng cách loại bỏ một số bản vá đầu vào nhất định, mô hình buộc phải tìm các đặc điểm mô tả khác. Phương pháp tăng cường này cũng có thể được xếp chồng lên trên các kỹ thuật tăng cường khác như lật ngang hoặc bộ lọc màu. Việc xóa ngẫu nhiên tạo ra một trong những độ chính xác cao nhất trên bộ dữ liệu CIFAR-10. DeVries và Taylor [71] đã thực hiện một nghiên cứu tương tự có tên là Cutout Regularization. Giống như nghiên cứu xóa ngẫu nhiên, họ đã thử nghiệm các vùng che giấu ngẫu nhiên của hình ảnh (Bảng 2).

Mikolajczyk và Grochowski [72] đã trình bày một ý tưởng thú vị để kết hợp việc xóa ngẫu nhiên với GAN được thiết kế để vẽ hình ảnh. Hình ảnh inpainting mô tả nhiệm vụ của

**Hình 12**Ví dụ về việc xóa ngẫu nhiên các tác vụ phát hiện đối tượng [70]**Bảng 2**Kết quả của việc cắt bỏ chính quy [104], cộng biểu thị bằng cách sử dụng các phương pháp tăng cường truyền thống, lật ngang và cắt xén

Phương pháp	C10	C10+	C100	C100+	SVHN
ResNet18 [5]	$10,63 \pm 0,26$	$4,72 \pm 0,21$	$36,68 \pm 0,57$	$22,46 \pm 0,31$	-
ResNet18 + phần cắt bỏ	$9,31 \pm 0,18$	$3,99 \pm 0,13$	$34,98 \pm 0,29$	$21,96 \pm 0,24$	-
WideResNet [21]	$6,97 \pm 0,22$	$3,87 \pm 0,08$	$26,06 \pm 0,22$	$18,8 \pm 0,08$	$1,60 \pm 0,05$
WideResNet + phần cắt bỏ	$5,54 \pm 0,08$	$3,08 \pm 0,16$	$23,94 \pm 0,15$	$18,41 \pm 0,27$	$1h30 \pm 0,03$
Chính quy hóa lắc-lắc [4] Chính	-	2,86	-	15,85	-
quy hóa lắc + cắt bỏ	-	$2,56 \pm 0,07$	-	$15h20 \pm 0,21$	-

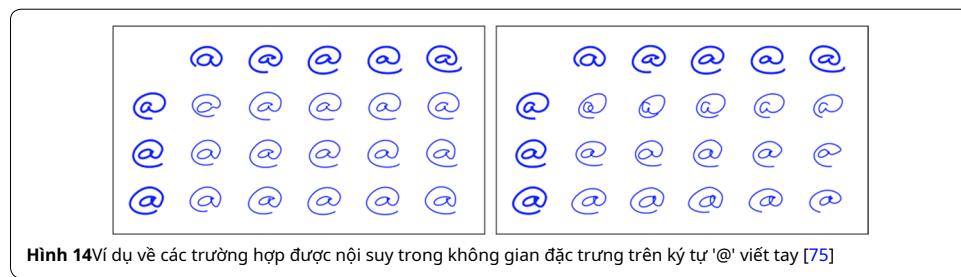
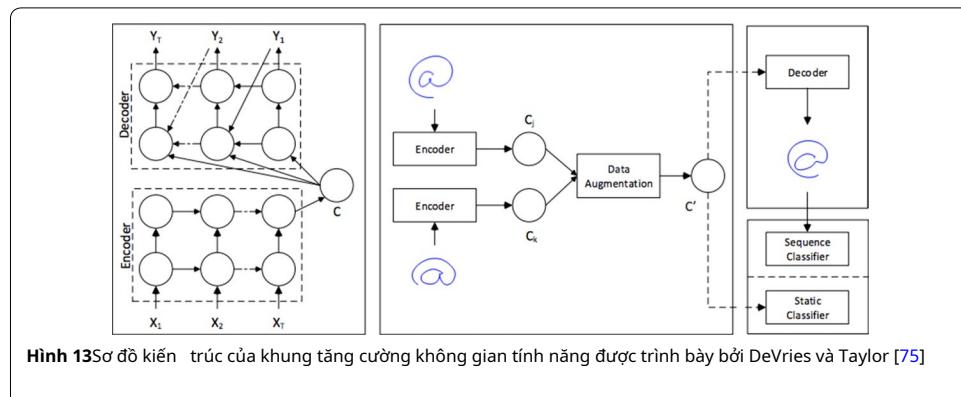
Tỷ lệ lỗi đạt được là 2,56% trên CIFAR-10 bằng cách sử dụng các phương pháp cắt bỏ và tăng cường truyền thống. Giá trị in nghiêng biểu thị hiệu suất cao theo các số liệu so sánh

điền vào phần còn thiếu của hình ảnh. Bằng cách sử dụng một bộ sưu tập đa dạng các công cụ vẽ GAN, việc tăng cường xóa ngẫu nhiên có thể tạo ra những phép ngoại suy rất thú vị. Sẽ rất thú vị để xem liệu có thể đạt được kết quả tốt hơn bằng cách xóa các mảng có hình dạng khác nhau như hình tròn thay vì xóa hay không. *Nó* là hình chữ nhật. Phần mở rộng của việc này sẽ là tham số hóa hình học của các bản vá bị xóa ngẫu nhiên và tìm hiểu cấu hình xóa tối ưu.

Một bất lợi của việc xóa ngẫu nhiên là nó không phải lúc nào cũng là một phép biến đổi bảo toàn nhãn. Trong nhận dạng chữ số viết tay, nếu phần trên cùng của số '8' bị cắt ngẫu nhiên thì nó không khác gì số '6'. Trong nhiều nhiệm vụ chi tiết như bộ dữ liệu Stanford Cars [73], việc xóa ngẫu nhiên các phần của hình ảnh (logo, v.v.) có thể khiến thương hiệu ô tô không thể nhận dạng được. Do đó, một số can thiệp thủ công có thể cần thiết tùy thuộc vào tập dữ liệu và nhiệm vụ.

Lưu ý khi kết hợp tăng cường

Trong số các phép tăng cường được thảo luận, các phép biến đổi hình học, các phép biến đổi không gian màu, bộ lọc hạt nhân, trộn hình ảnh và xóa ngẫu nhiên, gần như tất cả các phép biến đổi này đều đi kèm với tham số cường độ biến dạng liên quan. Tham số này mã hóa sự khác biệt về độ méo giữa góc quay 45° và góc quay 30° . Với một danh sách lớn các phép tăng cường tiềm năng và không gian độ lớn gần như liên tục, thật dễ dàng để khái niệm hóa kích thước khổng lồ của không gian tìm kiếm tăng cường. Việc kết hợp các tính năng bổ sung như cắt xén, lật, thay đổi màu sắc và xóa ngẫu nhiên có thể dẫn đến



kích thước tập dữ liệu tăng cao ồ ạt. Tuy nhiên, điều này không đảm bảo sẽ có lợi. Trong các miền có dữ liệu rất hạn chế, điều này có thể dẫn đến tình trạng trang bị quá mức. Do đó, điều quan trọng là phải xem xét các thuật toán tìm kiếm để lấy được một tập hợp con dữ liệu tăng cường tối ưu để huấn luyện các mô hình Deep Learning. Thông tin thêm về chủ đề này sẽ được thảo luận trong phần Cân nhắc thiết kế về tăng cường dữ liệu.

Tăng cường dữ liệu dựa trên Deep Learning *Tăng cường không gian tính năng*

Tất cả các phương pháp tăng cường được thảo luận ở trên đều được áp dụng cho hình ảnh trong không gian đầu vào. Mạng lưới thần kinh cực kỳ mạnh mẽ trong việc ánh xạ các đầu vào có chiều cao thành các biểu diễn chiều thấp hơn. Các mạng này có thể ánh xạ hình ảnh tới các lớp nhị phân hoặc $N \times 1$ vectơ trong các lớp phẳng. Việc xử lý tuần tự của các mạng nơ-ron có thể được điều khiển sao cho các biểu diễn trung gian có thể được tách ra khỏi toàn bộ mạng. Các biểu diễn chiều thấp hơn của dữ liệu hình ảnh trong các lớp được kết nối đầy đủ có thể được trích xuất và tách biệt. Konno và Iwazume [74] tìm thấy mức tăng hiệu suất trên CIFAR-100 từ độ chính xác 66 lên 73% bằng cách điều khiển tính mô-jun của mạng thần kinh để cô lập và tinh chỉnh các lớp riêng lẻ sau khi đào tạo. Các biểu diễn chiều thấp hơn được tìm thấy trong các lớp cấp cao của CNN được gọi là không gian đặc trưng. DeVries và Taylor [75] đã trình bày một bài viết thú vị thảo luận về việc tăng cường trong không gian đặc trưng này. Điều này mở ra cơ hội cho nhiều thao tác vector cho Data Augmentation.

SMOTE là một tính năng tăng cường phổ biến được sử dụng để giảm bớt các vấn đề về mất cân bằng lớp học. Kỹ thuật này được áp dụng cho không gian đặc trưng bằng cách nối các hàng xóm gần nhất để hình thành các thể hiện mới. DeVries và Taylor thảo luận về việc thêm nhiễu, nội suy và ngoại suy như các hình thức tăng cường không gian đặc trưng phổ biến (Hình 2).^{13,14}

Bảng 3 Kết quả hiệu suất của thử nghiệm với phép ngoại suy tính năng so với không gian đầu vào trên MNIST và CIFAR-10 [75]

Người mẫu	MNIST	CIFAR-10
Đường cơ sở	$1.093 \pm 0,057$	$30,65 \pm 0,27$
Phép biến đổi affine đường cơ sở + không gian đầu vào	$1.477 \pm 0,068$	–
Đường cơ sở + ngoại suy không gian đầu vào Đường cơ	$1.010 \pm 0,065$	–
sở + ngoại suy không gian đặc trưng	$0,950 \pm 0,036$	$29,24 \pm 0,27$

Giá trị in nghiêng biểu thị hiệu suất cao theo số liệu so sánh

Việc sử dụng bộ mã hóa tự động đặc biệt hữu ích để thực hiện tăng cường không gian đặc trưng trên dữ liệu. Bộ mã hóa tự động hoạt động bằng cách có một nửa mạng, bộ mã hóa, ánh xạ hình ảnh thành các biểu diễn vectơ có chiều thấp sao cho nửa còn lại của mạng, bộ giải mã, có thể tái tạo lại các vectơ này thành hình ảnh gốc. Biểu diễn được mã hóa này được sử dụng để tăng cường không gian đặc trưng.

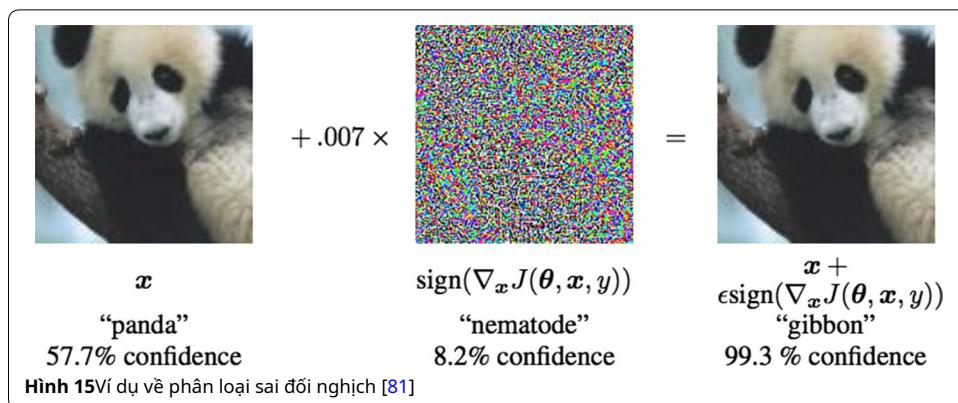
DeVries và Taylor [75] đã thử nghiệm kỹ thuật tăng cường không gian đặc trưng của họ bằng cách ngoại suy giữa 3 lân cận gần nhất trên mỗi mẫu để tạo dữ liệu mới và so sánh kết quả của chúng với phép ngoại suy trong không gian đầu vào và sử dụng các phép biến đổi affine trong không gian đầu vào (Bảng 3).

Việc tăng cường không gian đối tượng có thể được triển khai bằng bộ mã hóa tự động nếu cần thiết để tái tạo lại các phiên bản mới trở lại không gian đầu vào. Cũng có thể thực hiện tăng cường không gian đối tượng chỉ bằng cách tách các biểu diễn vectơ khỏi CNN. Điều này được thực hiện bằng cách cắt bỏ lớp đầu ra của mạng, sao cho đầu ra là một vectơ chiều thấp chứ không phải là nhãn lớp. Sau đó, các biểu diễn vectơ được tìm thấy bằng cách huấn luyện CNN và sau đó chuyển tập huấn luyện qua CNN bị cắt bớt. Các biểu diễn vectơ này có thể được sử dụng để huấn luyện bất kỳ mô hình học máy nào từ Naive Bayes, Support Vector Machine hoặc quay lại mạng đa lớp được kết nối đầy đủ. Hiệu quả của kỹ thuật này là một chủ đề cho công việc trong tương lai.

Một nhược điểm của việc tăng cường không gian đặc trưng là rất khó diễn giải dữ liệu vectơ. Có thể khôi phục các vectơ mới thành hình ảnh bằng mạng mã hóa tự động; tuy nhiên, điều này đòi hỏi phải sao chép toàn bộ phần mã hóa của CNN đang được huấn luyện. Đối với các CNN chuyên sâu, điều này dẫn đến việc tạo ra các bộ mã hóa tự động khổng lồ, rất khó đào tạo và tốn thời gian. Cuối cùng, Wong và cộng sự. [76] thấy rằng khi có thể chuyển đổi hình ảnh trong không gian dữ liệu, việc tăng cường không gian dữ liệu sẽ tốt hơn việc tăng cường không gian đối tượng.

Huấn luyện đối nghịch

Một trong những giải pháp để tìm kiếm không gian có thể tăng cường là đào tạo đối nghịch. Huấn luyện đối nghịch là một khuôn khổ để sử dụng hai hoặc nhiều mạng với các mục tiêu tương phản được mã hóa trong các hàm mất mát của chúng. Phần này sẽ thảo luận về việc sử dụng huấn luyện đối nghịch làm thuật toán tìm kiếm cũng như hiện tượng tấn công đối nghịch. Tấn công đối nghịch bao gồm một mạng đối thủ học cách tăng cường hình ảnh dẫn đến phân loại sai trong mạng phân loại đối thủ của nó. Những cuộc tấn công đối nghịch này, bị hạn chế bởi việc chèn tiếng ồn, đã thành công một cách đáng ngạc nhiên từ góc độ của mạng đối nghịch. Điều này thật đáng ngạc nhiên vì nó hoàn toàn thách thức trực giác về cách các mô hình này thể hiện hình ảnh. Đối thủ



các cuộc tấn công chứng minh rằng việc thể hiện hình ảnh kém hiệu quả hơn nhiều so với những gì có thể mong đợi. Điều này được chứng minh rõ ràng bởi Moosavi-Dezfooli et al. [77] sử dụng DeepFool, một mạng tìm ra mức nhiễu tối thiểu có thể cần thiết để gây ra sự phân loại sai với độ tin cậy cao. Su và cộng sự. [78] cho thấy 70,97% hình ảnh có thể bị phân loại sai nếu chỉ thay đổi một pixel. Zajac và cộng sự. [79] gây ra sự phân loại sai với các cuộc tấn công đối nghịch giới hạn ở đường viền của hình ảnh. Sự thành công của các cuộc tấn công đối nghịch đặc biệt được phóng đại khi độ phân giải của hình ảnh tăng lên.

Cuộc tấn công đối nghịch có thể được nhắm mục tiêu hoặc không nhắm mục tiêu, đề cập đến sự cân nhắc trong đó mạng đối thủ đang cố gắng gây ra sự phân loại sai. Các cuộc tấn công đối nghịch có thể giúp minh họa ranh giới quyết định yếu tố tốt hơn các số liệu phân loại tiêu chuẩn có thể.

Ngoài vai trò là thước đo đánh giá, khả năng phòng thủ trước các cuộc tấn công của đối thủ, huấn luyện đối nghịch có thể là một phương pháp hiệu quả để tìm kiếm các biện pháp tăng cường.

Bằng cách hạn chế tập hợp các phép tăng cường và biến dạng có sẵn cho mạng đối nghịch, nó có thể học cách tạo ra các phép tăng cường dẫn đến phân loại sai, từ đó tạo thành một thuật toán tìm kiếm hiệu quả. Những sự tăng cường này rất có giá trị trong việc cung cấp những điểm yếu trong mô hình phân loại. Do đó, đào tạo đối nghịch có thể là một kỹ thuật tìm kiếm hiệu quả để Tăng cường dữ liệu. Điều này trái ngược hoàn toàn với các kỹ thuật tăng cường truyền thống được mô tả trước đây. Sự gia tăng đối nghịch có thể không đại diện cho các ví dụ có khả năng xảy ra trong tập kiểm tra, nhưng chúng có thể cải thiện các điểm yếu trong ranh giới quyết định đã học.

Engstrom và cộng sự. [80] đã chỉ ra rằng các phép biến đổi đơn giản như phép quay và phép dịch có thể dễ dàng gây ra sự phân loại sai của các mô hình CNN sâu. Các phép biến đổi ngẫu nhiên tồi tệ nhất đã làm giảm độ chính xác của MNIST xuống 26%, CIFAR10 xuống 72% và ImageNet (Top 1) xuống 28%. Goodfellow và cộng sự. [81] tạo ra các ví dụ đối nghịch để cải thiện hiệu suất trong nhiệm vụ phân loại MNIST. Sử dụng kỹ thuật tạo các ví dụ đối nghịch được gọi là “phương pháp ký hiệu gradient nhanh”, mạng tối đa [82] đã phân loại sai 89,4% ví dụ đối nghịch với độ tin cậy trung bình là 97,6%. Thử nghiệm này được thực hiện trên tập dữ liệu MNIST. Với quá trình huấn luyện đối nghịch, tỷ lệ lỗi của các ví dụ đối nghịch đã giảm từ 89,4% xuống 17,9% (Hình 2).¹⁵

Li và cộng sự. [83] thử nghiệm phương pháp đào tạo đối thủ mới và so sánh hiệu suất trên dữ liệu thử nghiệm ban đầu và các ví dụ đối nghịch. Kết quả hiển thị

Bảng 4 Độ chính xác của kiểm tra cho thấy tác động của quá trình đào tạo đối thủ, sạch để cập đến dữ liệu thử nghiệm ban đầu, FGSM để cập đến các ví dụ đối thủ bắt nguồn từ Phương pháp Dấu hiệu Độ dốc Nhanh và PGD để cập đến các ví dụ đối nghịch bắt nguồn từ Giảm dần Độ dốc Dự kiến [83]

Người mẫu	MNIST			CIFAR-10		
	Lau dọn	FGSM	PGD	Lau dọn	FGSM	PGD
Tiêu chuẩn	0,9939	0,0922	0	0,9306	0,5524	0,0256
Được đào tạo đối nghịch	0,9932	0,9492	0,0612	0,8755	0,8526	0,1043
Phương pháp của chúng tôi	0,9903	0,9713	0,9171	0,8714	0,6514	0,3440

dưới đây cho thấy việc dự đoán các cuộc tấn công đối nghịch trong quá trình huấn luyện có thể làm giảm đáng kể sự thành công của các cuộc tấn công như thế nào.

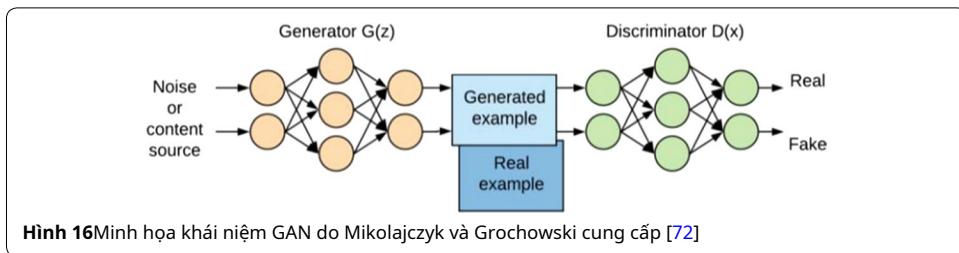
Như thể hiện trong bảng4, việc huấn luyện đối nghịch trong thí nghiệm của họ không cải thiện độ chính xác của bài kiểm tra. Tuy nhiên, nó cải thiện đáng kể độ chính xác kiểm tra của các ví dụ đối nghịch. Phòng thủ đối nghịch là một chủ đề rất thú vị để đánh giá tính bảo mật và độ mạnh mẽ của các mô hình Deep Learning. Cải tiến Phương pháp ký hiệu gradient nhanh, DeepFool, được phát triển bởi Moosavi-Dezfooli et al. [77], sử dụng mạng lưới thần kinh để tìm ra nhiễu loạn nhỏ nhất có thể gây ra sự phân loại sai.

Một khuôn khổ thú vị khác có thể được sử dụng trong bối cảnh đào tạo đối nghịch là yêu cầu đối thủ thay đổi nhãn của dữ liệu huấn luyện. Xie và cộng sự. [84] đã trình bày DisturbLabel, một kỹ thuật chính quy hóa thay thế ngẫu nhiên các nhãn ở mỗi lần lặp. Đây là một ví dụ hiếm hoi về việc thêm nhiễu vào lớp mất mát, trong khi hầu hết các phương pháp tăng cường khác được thảo luận đều thêm nhiễu vào lớp biểu diễn đầu vào hoặc lớp biểu diễn ẩn. Trên tập dữ liệu MNIST với LeNet [28] Kiến trúc CNN, DisturbLabel tạo ra tỷ lệ lỗi 0,32% so với tỷ lệ lỗi cơ bản là 0,39%. DisturbLabel kết hợp với Dropout Regularization tạo ra tỷ lệ lỗi 0,28% so với mức cơ bản 0,39%. Để chuyển điều này sang bối cảnh đào tạo đối nghịch, một mạng lấy dữ liệu huấn luyện của bộ phân loại làm đầu vào và tìm hiểu xem nhãn nào cần lật để tối đa hóa tỷ lệ lỗi của mạng phân loại.

Hiệu quả của việc huấn luyện đối nghịch dưới hình thức tìm kiếm nhiều hoặc tăng cường vẫn là một khái niệm tương đối mới chưa được thử nghiệm và hiểu biết rộng rãi. Tìm kiếm đối nghịch để thêm nhiễu đã được chứng minh là cải thiện hiệu suất trên các mẫu đối nghịch, nhưng không rõ liệu điều này có hữu ích cho mục tiêu giảm quá mức hay không. Công việc trong tương lai tìm cách mở rộng mối quan hệ giữa khả năng chống lại các cuộc tấn công đối nghịch và hiệu suất thực tế trên các tập dữ liệu thử nghiệm.

Tăng cường dữ liệu dựa trên GAN

Một chiến lược thú vị khác để Tăng cường dữ liệu là mô hình hóa tổng quát. Mô hình tổng quát đề cập đến thực tiễn tạo các phiên bản nhân tạo từ tập dữ liệu sao cho chúng giữ được các đặc điểm tương tự với tập hợp ban đầu. Các nguyên tắc đào tạo đối nghịch được thảo luận ở trên đã dẫn đến khung mô hình tổng quát rất thú vị và phổ biến rộng rãi được gọi là GAN. Bowles và cộng sự. [85] mô tả GAN như một cách để “mở khóa” thông tin bổ sung từ tập dữ liệu. GAN không phải là thế hệ duy nhất



kỹ thuật mô hình hóa hiện có; tuy nhiên họ đang dẫn đầu về tốc độ tính toán và chất lượng kết quả.

Một chiến lược hữu ích khác cho mô hình tổng quát đáng được đề cập là bộ mã hóa tự động biến thiên. Khung GAN có thể được mở rộng để cải thiện chất lượng của các mẫu được tạo bằng bộ mã hóa tự động biến thiên [86]. Bộ mã hóa tự động biến thiên hiểu cách biểu diễn các điểm dữ liệu theo chiều thấp. Trong miền hình ảnh, điều này dịch một tensor hình ảnh có kích thước $chiều cao \times chiều rộng \times màu sắc \times kinh$ xuống thành một vectơ có kích thước $N \times 1$, giống hệt với những gì đã được thảo luận liên quan đến việc tăng cường không gian đặc trưng. Các ràng buộc chiều thấp trong biểu diễn vectơ sẽ dẫn đến biểu diễn kém hơn, mặc dù các ràng buộc này tốt hơn cho việc trực quan hóa bằng các phương pháp như t-SNE [87]. Hãy tưởng tượng một biểu diễn vector có kích thước 5×1 được tạo bởi bộ mã hóa tự động. Các bộ mã hóa tự động này có thể thực hiện phân phối dữ liệu được gắn nhãn và ánh xạ chúng vào không gian này. Các lớp này có thể bao gồm 'đầu quay sang trái', 'đầu quay ở giữa' và 'đầu quay sang phải'. Bộ mã hóa tự động tìm hiểu cách biểu diễn theo chiều thấp của các điểm dữ liệu này sao cho các phép toán vectơ như cộng và trừ có thể được sử dụng để mô phỏng chế độ xoay 3D ở chế độ xem trước của một phiên bản mới. Các đầu ra của bộ mã hóa tự động biến thiên có thể được cải thiện hơn nữa bằng cách nhập chúng vào GAN [31]. Ngoài ra, một quy trình thao tác vectơ tương tự có thể được thực hiện trên các đầu vào vectơ nhiều tới GAN thông qua việc sử dụng GAN hai chiều [88].

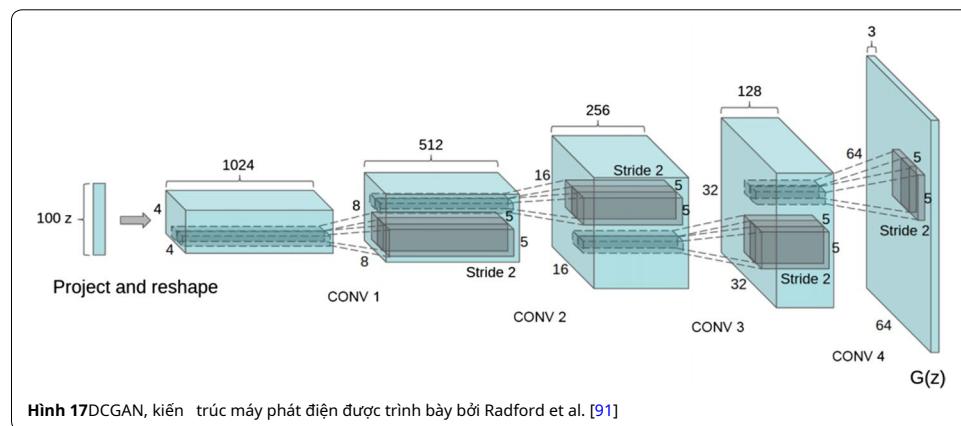
Hiệu suất ấn tượng của GAN đã dẫn đến sự chú ý ngày càng tăng về cách chúng có thể được áp dụng cho nhiệm vụ Tăng cường dữ liệu. Các mạng này có khả năng tạo dữ liệu đào tạo mới giúp mô hình phân loại hoạt động tốt hơn. Kiến trúc GAN lần đầu tiên được đề xuất bởi Ian Goodfellow [31] là một khuôn khổ cho mô hình tổng quát thông qua đào tạo đối nghịch. Giai thoại hay nhất để hiểu GAN là sự tương tự giữa cảnh sát và kẻ làm hàng giả. Kẻ giả mạo (mạng máy phát điện) nhận một số dạng đầu vào. Đây có thể là một vectơ ngẫu nhiên, một hình ảnh, văn bản khác và nhiều thứ khác. Kẻ làm giả học cách tạo ra tiền sao cho cảnh sát (mạng lưới phân biệt đối xử) không thể biết tiền đó là thật hay giả. Sự phân đôi thật hay giả tương tự như việc phiên bản được tạo có phải từ tập huấn luyện hay không hoặc nó được tạo bởi mạng máy phát điện (Hình 2).¹⁶

Sự tương tự giữa kẻ giả mạo và kẻ cướp là cầu nối liền mạch để hiểu GAN trong bối cảnh phát hiện xâm nhập mạng. Lin và cộng sự. [89] sử dụng mạng máy phát điện để tìm hiểu cách đánh lừa hệ thống phát hiện hộp đen. Điều này nêu bật một trong những đặc điểm thú vị nhất của GAN. Các công cụ phân tích bắt nguồn từ lý thuyết trò chơi như chiến lược minimax và Cân bằng Nash [90] gợi ý rằng trình tạo cuối cùng sẽ đánh lừa người phân biệt đối xử. Sự thành công của trình tạo trong việc vượt qua trình phân biệt đối xử khiến nó trở nên rất mạnh mẽ cho mô hình tổng quát. GAN là kỹ thuật tạo mô hình tổng quát hứa hẹn nhất để sử dụng trong Tăng cường dữ liệu.

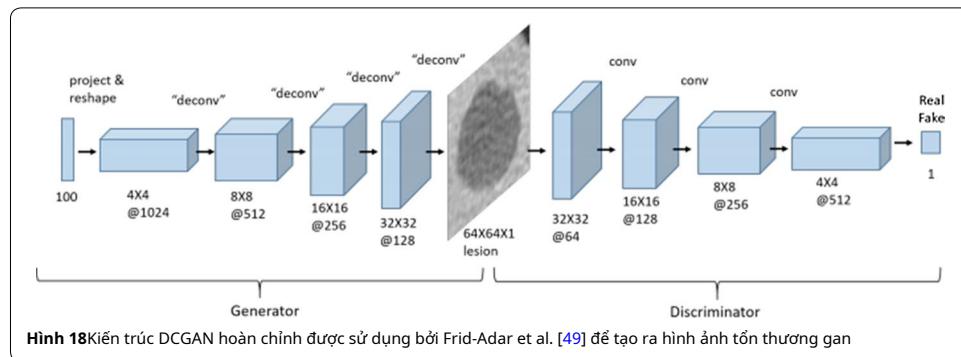
Kiến trúc vanilla GAN sử dụng mạng perceptron đa lớp trong mạng tạo và mạng phân biệt đối xử. Điều này có thể tạo ra các hình ảnh có thể chấp nhận được trên một tập dữ liệu hình ảnh đơn giản, chẳng hạn như các chữ số viết tay MNIST. Tuy nhiên, nó không tạo ra kết quả chất lượng cho các bộ dữ liệu có độ phân giải cao hơn, phức tạp hơn. Trong tập dữ liệu MNIST, mỗi hình ảnh chỉ có $28 \times 28 \times 1$ với tổng số 784 pixel. GAN áp dụng cho dữ liệu MNIST có thể mang lại kết quả thuyết phục. Tuy nhiên, hình ảnh MNIST ít thách thức hơn nhiều so với các bộ dữ liệu hình ảnh khác do phương sai và độ phân giải trong cùng lớp thấp, đó là một số điểm khác biệt của nhiều hình ảnh. Điều này trái ngược hoàn toàn với các bộ dữ liệu khác được nghiên cứu trong hầu hết các bài báo về Thị giác máy tính mang tính học thuật như ImageNet hoặc CIFAR-10. Để tham khảo ngay, hình ảnh ImageNet có độ phân giải $256 \times 256 \times 3$, tổng cộng 196.608 pixel, $250 \times$ tăng số lượng pixel so với MNIST.

Nhiều tài liệu nghiên cứu đã được xuất bản nhằm sửa đổi khung GAN thông qua các kiến trúc mạng, hàm mất mát, phương pháp tiến hóa khác nhau, v.v. Nghiên cứu này đã cải thiện đáng kể chất lượng của các mẫu được tạo bởi GAN. Đã có nhiều kiến trúc mới được đề xuất để mở rộng khái niệm về GAN và tạo ra hình ảnh đầu ra có độ phân giải cao hơn, nhiều trong số đó nằm ngoài phạm vi của bài viết này. Trong số các kiến trúc mới này, DCGAN, GAN tăng dần, CycleGAN và GAN có điều kiện dường như có tiềm năng ứng dụng nhất trong Tăng cường dữ liệu.

DCGAN [91] kiến trúc đã được đề xuất để mở rộng dựa trên độ phức tạp bên trong của mạng tạo và mạng phân biệt đối xử. Kiến trúc này sử dụng CNN cho các mạng tạo và phân biệt đối xử thay vì các perceptron đa lớp. DCGAN đã được thử nghiệm để tạo ra kết quả trên bộ dữ liệu hình ảnh nội thất phòng ngủ LSUN, mỗi hình ảnh là $64 \times 64 \times 3$, với tổng số 12.288 pixel, (so với 784 trong MNIST). Ý tưởng đằng sau DCGAN là tăng độ phức tạp của mạng máy phát để chiếu đầu vào thành một tensor có chiều cao và sau đó thêm các lớp giải mã để chuyển từ tensor được chiếu sang hình ảnh đầu ra. Các lớp giải mã này sẽ mở rộng theo các chiều không gian, ví dụ, từ $14 \times 14 \times 6$ đến $28 \times 28 \times 1$, trong khi lớp tích chập sẽ làm giảm kích thước không gian chẳng hạn như đi từ $14 \times 14 \times 32$ đến $7 \times 7 \times 64$. Kiến trúc DCGAN trình bày chiến lược sử dụng các lớp tích chập trong khung GAN để tạo ra hình ảnh có độ phân giải cao hơn (Hình.17,18).



Hình 17 DCGAN, kiến trúc máy phát điện được trình bày bởi Radford et al. [91]



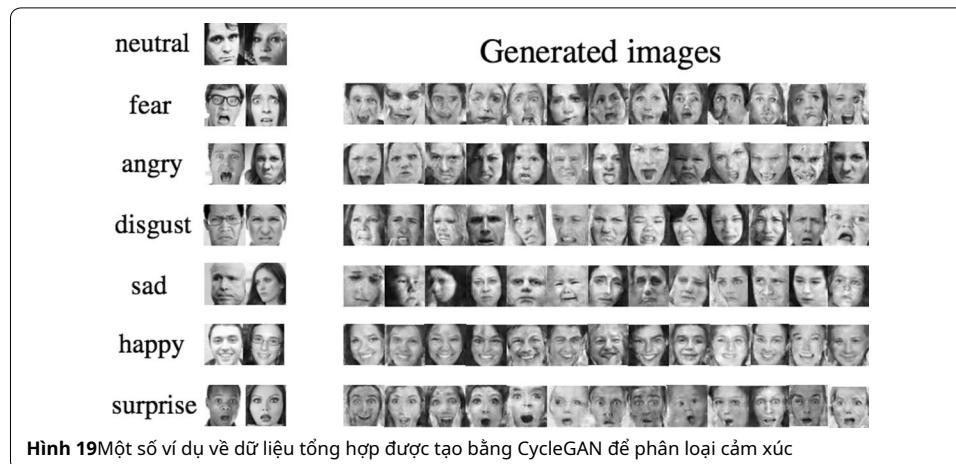
Frid-Adar và cộng sự. [49] đã thử nghiệm tính hiệu quả của việc sử dụng DCGAN để tạo ra hình ảnh y tế về tổn thương gan. Họ sử dụng kiến trúc trong hình trên để tạo ra $64 \times 64 \times 1$ size chụp CT tổn thương gan. Tập dữ liệu ban đầu của họ chứa 182 bản quét CT, (53 U nang, 64 Di căn và 65 U mạch máu). Sau khi sử dụng các phương pháp tăng cường cổ điển để đạt được độ nhạy 78,6% và độ đặc hiệu 88,4%, họ nhận thấy độ nhạy tăng lên 85,7% và độ đặc hiệu 92,4% sau khi thêm các mẫu do DCGAN tạo ra.

Một kiến trúc đáng quan tâm khác được gọi là GAN tăng dần [34]. Kiến trúc này đào tạo một loạt các mạng với độ phức tạp có độ phân giải tăng dần. Các độ phân giải này dao động từ 4×4 đến 8×8 , v.v. cho đến khi đầu ra có kích thước 1024×1024 đã đạt được. Điều này được xây dựng dựa trên khái niệm rằng GAN có thể chấp nhận hình ảnh làm đầu vào cũng như các vectơ ngẫu nhiên. Do đó, chuỗi GAN hoạt động bằng cách chuyển các mẫu từ GAN có độ phân giải thấp hơn đến GAN có độ phân giải cao hơn. Điều này đã tạo ra kết quả rất tuyệt vời trên hình ảnh khuôn mặt.

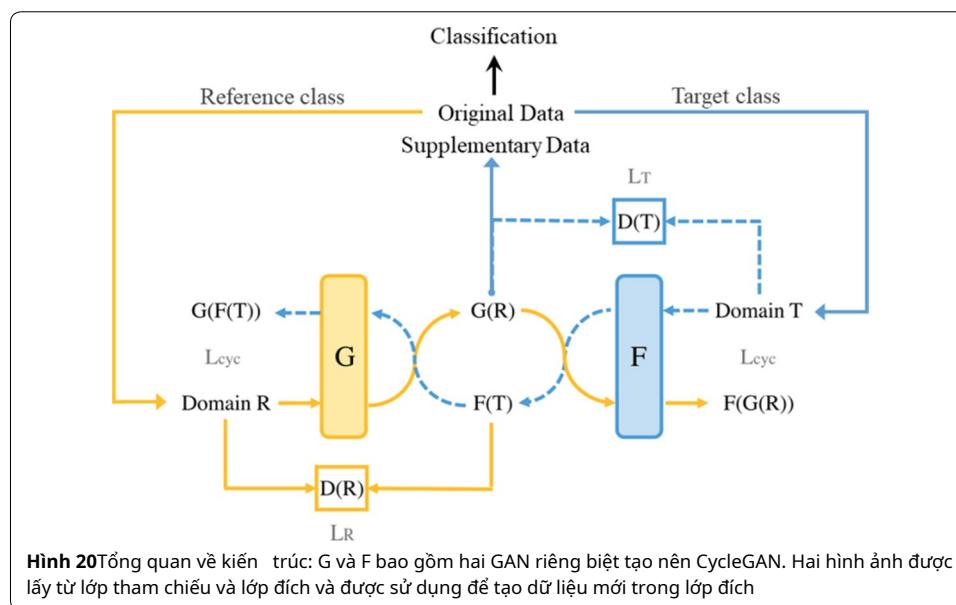
Ngoài việc cải thiện kích thước độ phân giải của GAN, một kiến trúc thú vị khác giúp tăng chất lượng đầu ra là CycleGAN [92] được đề xuất bởi Zhu et al. Cycle-GAN giới thiệu chức năng mất tính nhất quán theo chu kỳ bổ sung để giúp ổn định quá trình đào tạo GAN. Điều này được áp dụng cho việc dịch từ hình ảnh sang hình ảnh. Chuyển giao kiểu thần kinh [32], được thảo luận sâu hơn trong phần bên dưới, tìm hiểu cách dịch từ một hình ảnh sang một hình ảnh. Tuy nhiên, CycleGAN học cách dịch từ miền hình ảnh sang miền khác, chẳng hạn như ngựa sang ngựa vằn. Điều này được thực hiện thông qua các hàm mất tính nhất quán tiến và lùi. Một máy phát điện lấy hình ảnh của những con ngựa và học cách ánh xạ chúng tới ngựa vằn để người phân biệt đối xử không thể biết liệu ban đầu chúng có phải là một phần của bộ ngựa vằn hay không, như đã thảo luận ở trên. Sau đó, những con ngựa vằn được tạo ra từ hình ảnh ngựa sẽ được chuyển qua một mạng để chuyển chúng trở lại thành ngựa. Người phân biệt thứ hai xác định xem hình ảnh được dịch lại này có thuộc về bộ ngựa hay không. Cả hai tổn thất phân biệt đối xử này được tổng hợp để tạo thành tổn thất nhất quán theo chu kỳ.

Việc sử dụng CycleGAN đã được thử nghiệm bởi Zhu et al. [93] trong nhiệm vụ Phân loại Cảm xúc. Sử dụng bộ dữ liệu nhận dạng cảm xúc, FER2013 [94], Cơ sở dữ liệu nhận dạng biểu cảm khuôn mặt, họ xây dựng bộ phân loại CNN để nhận biết 7 cảm xúc khác nhau: tức giận, ghê tởm, sợ hãi, vui, buồn, bất ngờ và trung tính. Các lớp này không cân bằng và CycleGAN được sử dụng như một phương pháp lấy mẫu quá mức thông minh.

CycleGAN đã học cách dịch từ hình ảnh sang hình ảnh không ghép đôi giữa các miền. Một ví dụ về các miền trong bài toán này là trung tính đến mức ghê tởm. CycleGAN học cách dịch một hình ảnh đại diện cho một hình ảnh trung tính thành một hình ảnh đại diện cho cảm xúc ghê tởm (Hình 2).^{19,20}



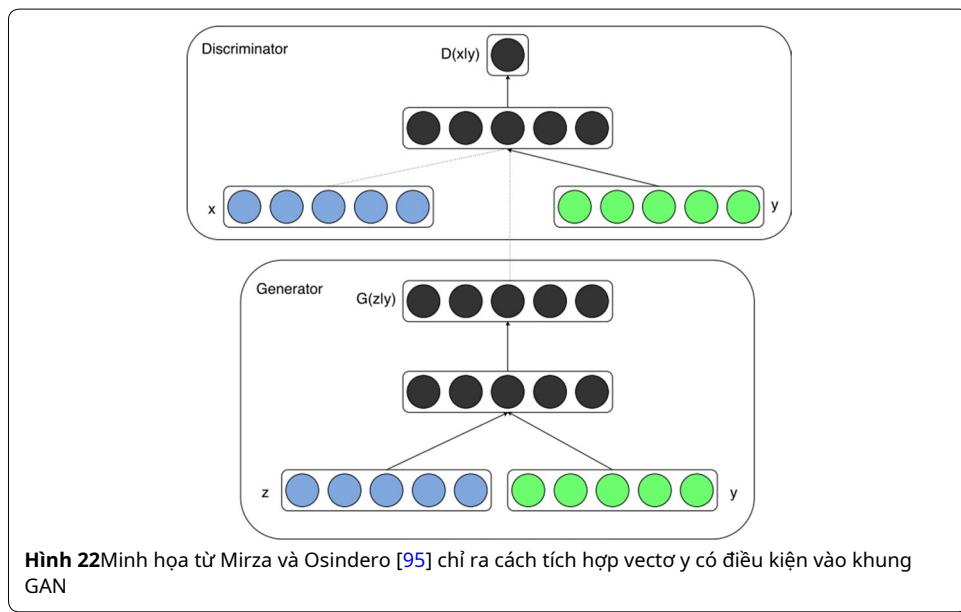
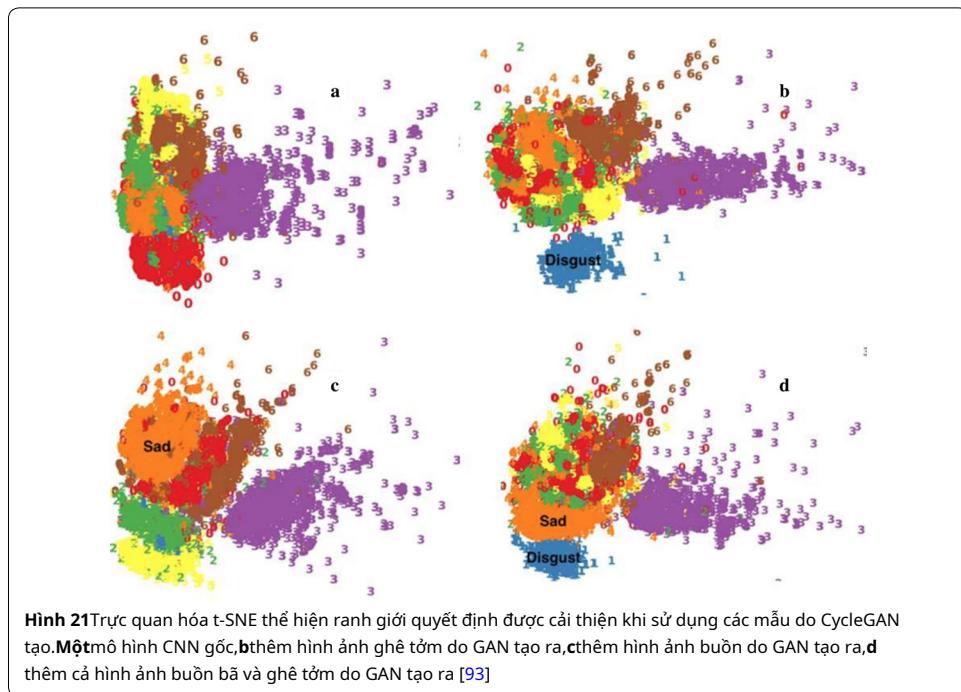
Hình 19 Một số ví dụ về dữ liệu tổng hợp được tạo bằng CycleGAN để phân loại cảm xúc



Hình 20 Tổng quan về kiến trúc: G và F bao gồm hai GAN riêng biệt tạo nên CycleGAN. Hai hình ảnh được lấy từ lớp tham chiếu và lớp đích và được sử dụng để tạo dữ liệu mới trong lớp đích

Việc sử dụng CycleGAN để dịch hình ảnh từ 7 lớp còn lại sang các lớp thiểu số rất hiệu quả trong việc cải thiện hiệu suất của mô hình CNN về nhận dạng cảm xúc. Sử dụng các kỹ thuật này, độ chính xác được cải thiện 5–10%. Để hiểu rõ hơn về tính hiệu quả của việc thêm các phiên bản do GAN tạo, trực quan hóa t-SNE được sử dụng. t-SNE [87] là một kỹ thuật trực quan học cách ánh xạ giữa các vectơ chiều cao vào không gian chiều thấp để tạo điều kiện thuận lợi cho việc trực quan hóa các ranh giới quyết định (Hình 2).²¹

Một kiến trúc GAN thú vị khác để sử dụng trong Tăng cường dữ liệu là GAN có điều kiện [95]. GAN có điều kiện thêm một vectơ điều kiện vào cả bộ tạo và bộ phân biệt đối xử để giảm bớt các vấn đề về thu gọn chế độ. Ngoài việc nhập một vectơ ngẫu nhiên z tới trình tạo, GAN có điều kiện cũng nhập một vectơ có thể giống như nhãn lớp được mã hóa một lần, ví dụ: [0 0 0 1 0]. Nhãn lớp này nhằm mục tiêu một lớp cụ thể cho trình tạo và trình phân biệt đối xử (Hình 2).²²



Lucic và cộng sự. [96] đã tìm cách so sánh các hàm mất GAN mới được phát triển. Họ đã tiến hành một loạt thử nghiệm để xác định rằng hầu hết các hàm mất mát có thể đạt được điểm số tương tự khi tối ưu hóa đủ siêu tham số và khởi động lại ngẫu nhiên. Điều này cho thấy rằng sức mạnh tính toán tăng lên là một lĩnh vực trọng tâm hứa hẹn hơn so với những thay đổi thuật toán trong hàm tạo và mất phân biệt đối xử.

Hầu hết các nghiên cứu được thực hiện trong việc áp dụng GAN vào Tăng cường dữ liệu và báo cáo hiệu suất phân loại thu được đã được thực hiện trong phân tích hình ảnh y sinh

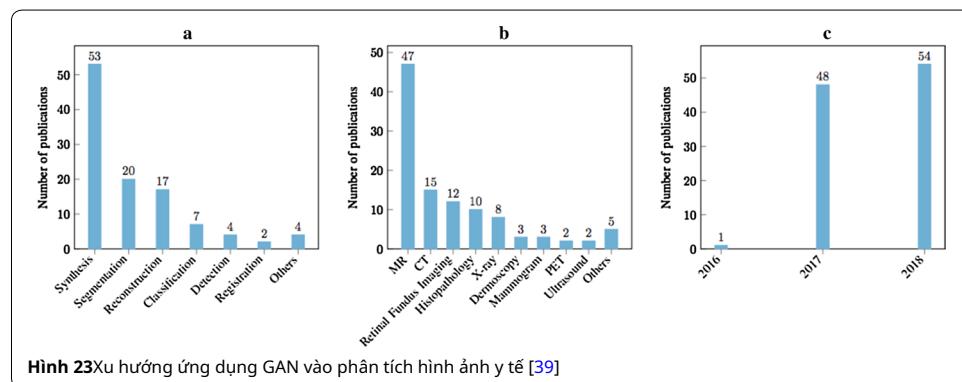
[39]. Các bài viết này đã cho thấy ranh giới phân loại được cải thiện nhờ quá trình đào tạo với dữ liệu thực và được tạo từ các mô hình GAN. Ngoài ra, một số bài báo đo lường chất lượng đầu ra của GAN bằng bài kiểm tra Turing trực quan. Trong các thử nghiệm này, nghiên cứu yêu cầu hai chuyên gia phân biệt giữa hình ảnh thật và hình ảnh nhân tạo trong các nhiệm vụ hình ảnh y tế như phân loại tổn thương da và phát hiện ung thư gan. Bảng 5 cho thấy các chuyên gia thứ nhất và thứ hai chỉ có thể gán nhãn chính xác 62,5% và 58,6% hình ảnh tổn thương gan do GAN tạo ra là giả. Việc gắn nhãn hình ảnh là giả có nghĩa là nguồn gốc của chúng đến từ máy tạo chứ không phải hình ảnh tổn thương gan thực tế (Bảng 6; Quá sun. 23).

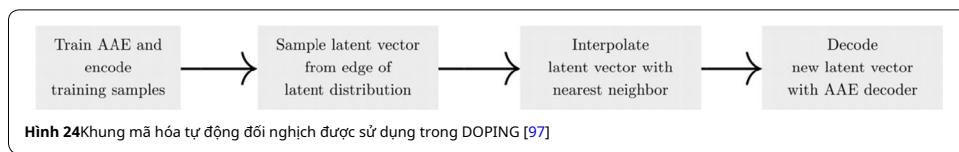
Bảng 5 Kết quả 'Thử nghiệm Turing Trực quan' trên các hình ảnh tổn thương gan do DCGAN tạo ra được trình bày bởi Frid-Adar et al. [139]

	Độ chính xác phân loại			ROI có thật không?
	Thực tế (%)	Tổng hợp (%)	Tổng điểm	
Chuyên gia 1	78	77,5	235\302=77,8%	189\302=62,5%
Chuyên gia 2	69,2	69,2	209\302=69,2%	177\302=58,6%

Bảng 6 Kết quả 'Thử nghiệm Turing trực quan' trên DCGAN- và WGAN khác nhau [104]—hình ảnh MRI khối u não được tạo ra bởi Han và cộng sự. [140]

	Sự chính xác (%)	Thực tế đã chọn như thật	Thực tế như tổng hợp	Synt như thật	Cú pháp như cú pháp
T1 (DCGAN, 128×128) Tlc	70	26	24	6	44
(DCGAN, 128×128) T2	71	24	26	3	47
(DCGAN, 128×128) FLAIR	64	22	28	số 8	42
(DCGAN, 128×128) Concat	54	12	38	số 8	42
(DCGAN, 128×128) Concat	77	34	16	7	43
(DCGAN, 64×64) T1 (WGAN,	54	13	37	9	41
128×128) Tlc (WGAN, 128×	64	20	30	6	44
128) T2 (WGAN, 128×128)	55	13	37	số 8	42
FLAIR (WGAN, 128×128)	58	19	31	11	39
Concat (WGAN, 128×128)	62	16	34	4	46
Concat (WGAN, 64×64)	66	31	19	15	35
	53	18	32	15	35





Các mẫu GAN có thể được sử dụng như một kỹ thuật lấy mẫu quá mức để giải quyết các vấn đề về mất cân bằng lớp. Lim và cộng sự. [97] chỉ ra cách sử dụng mẫu GAN để phát hiện sự bất thường không có giám sát. Bằng cách lấy mẫu quá mức các mẫu bình thường hiếm gặp, là những mẫu xảy ra với xác suất nhỏ, GAN có thể giảm tỷ lệ dương tính giả khi phát hiện bất thường. Họ thực hiện việc này bằng cách sử dụng khung Bộ mã hóa tự động đối nghịch do Makhzani và cộng sự đề xuất. [98] (Quả sung.24).

Dù tiềm năng của GAN rất thú vị nhưng rất khó để có được đầu ra có độ phân giải cao từ các kiến trúc tiên tiến hiện tại. Việc tăng kích thước đầu ra của hình ảnh do trình tạo tạo ra có thể sẽ gây ra sự mất ổn định trong quá trình huấn luyện và không hội tụ. Một nhược điểm khác của GAN là chúng yêu cầu một lượng dữ liệu đáng kể để đào tạo. Do đó, tùy thuộc vào mức độ hạn chế của tập dữ liệu ban đầu, GAN có thể không phải là giải pháp thực tế. Salimans và cộng sự. [99] cung cấp một mô tả đầy đủ hơn về các vấn đề với GAN đào tạo.

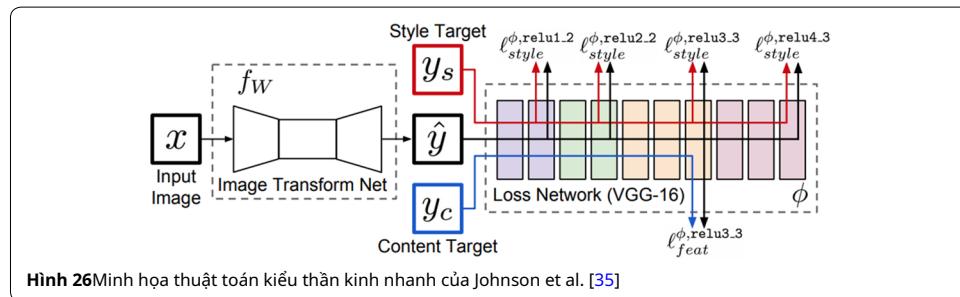
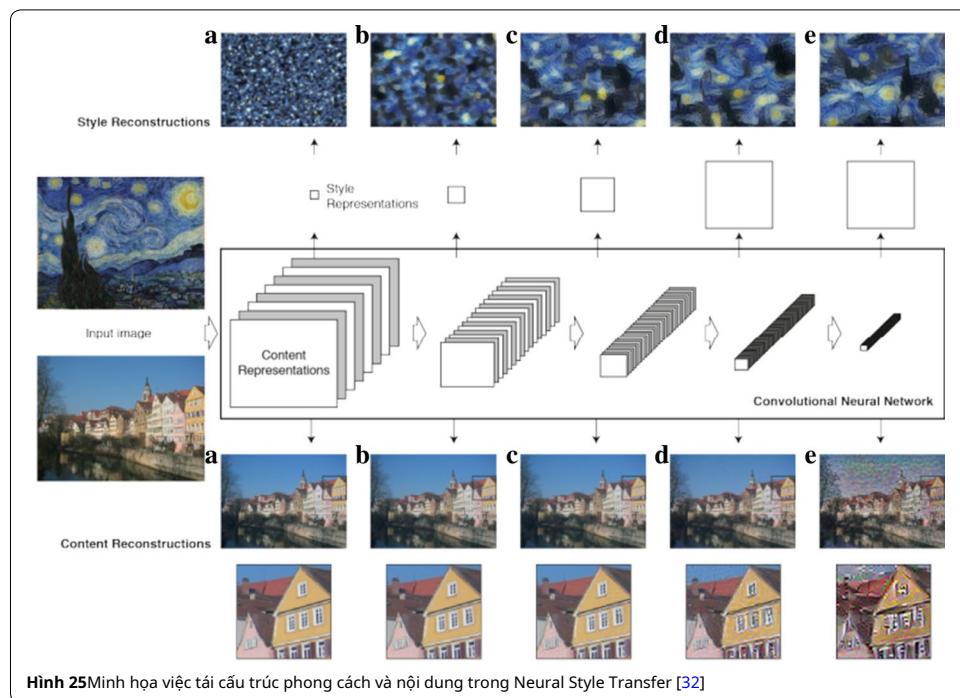
Chuyển giao phong cách thần kinh

Chuyển giao kiểu thần kinh [32] là một trong những minh chứng hào nhoáng nhất về khả năng Deep Learning. Ý tưởng chung là thao tác các biểu diễn hình ảnh được tạo trong CNN. Chuyển kiểu thần kinh có lẽ được biết đến nhiều nhất nhờ các ứng dụng nghệ thuật, nhưng nó cũng đóng vai trò như một công cụ tuyệt vời để Tăng cường dữ liệu. Thuật toán hoạt động bằng cách điều khiển các biểu diễn tuần tự trên CNN sao cho kiểu của một hình ảnh có thể được chuyển sang hình ảnh khác trong khi vẫn giữ nguyên nội dung gốc của nó. Li và cộng sự có thể tìm thấy lời giải thích chi tiết hơn về hoạt động ma trận gram hỗ trợ Chuyển giao kiểu thần kinh. [100] (Quả sung.25).

Điều quan trọng là cũng phải nhận ra sự tiến bộ của thuật toán gốc từ Gatys et al. được gọi là Chuyển kiểu nhanh [35]. Thuật toán này mở rộng hàm mờ mịt từ mờ mỗi pixel đến mờ cảm nhận và sử dụng mạng chuyển tiếp nguồn cấp dữ liệu để tạo kiểu cho hình ảnh. Sự mờ mịt về nhận thức này được giải thích thông qua việc sử dụng một mạng lưới được huấn luyện trước khác. Việc sử dụng sự mờ mịt cảm nhận so với sự mờ mịt trên mỗi pixel cũng cho thấy nhiều hứa hẹn trong việc áp dụng độ phân giải siêu cao [101] cũng như chuyển giao phong cách. Việc cải tiến chức năng mờ mịt này cho phép chuyển kiểu chạy nhanh hơn nhiều, tăng sự quan tâm đến các ứng dụng thực tế. Ngoài ra, Ulyanov và cộng sự. [102] nhận thấy rằng việc thay thế chuẩn hóa hàng loạt bằng chuẩn hóa phiên bản sẽ mang lại sự cải thiện đáng kể cho việc tạo kiểu nhanh (Hình 2).26).

Với mục đích Tăng cường dữ liệu, điều này hơi giống với các phép biến đổi ánh sáng không gian màu. Neural Style Transfer mở rộng các biến thể ánh sáng và cho phép mã hóa các kết cấu cũng như phong cách nghệ thuật khác nhau. Điều này khiến những người thực hành Tăng cường dữ liệu phải quyết định lấy mẫu kiểu nào khi lấy hình ảnh mới thông qua Chuyển kiểu thần kinh.

Việc chọn phong cách để lấy mẫu có thể là một nhiệm vụ đầy thách thức. Đối với các ứng dụng như ô tô tự lái, việc chuyển đổi đào tạo là khá trực quan



dữ liệu theo thang đo từ đêm sang ngày, từ mùa đông đến mùa hè hoặc từ mưa đến nắng. Tuy nhiên, trong các miền ứng dụng khác, tập hợp các kiểu dáng để chuyển vào không quá rõ ràng. Để dễ thực hiện, việc tăng cường dữ liệu thông qua Truyền kiểu thần kinh có thể được thực hiện bằng cách chọn một bộ phong cách và áp dụng chúng cho tất cả các hình ảnh trong tập huấn luyện. Công việc của Tăng cường Phong cách [103], tránh đưa một dạng thiên vị phong cách mới vào tập dữ liệu bằng cách lấy các kiểu ngẫu nhiên từ việc phân phối 79.433 hình ảnh nghệ thuật. Việc chuyển phong cách trong dữ liệu đào tạo đã được thử nghiệm khi chuyển đổi từ môi trường mô phỏng sang thế giới thực. Điều này rất hữu ích cho các tác vụ thao tác robot bằng cách sử dụng Học tăng cường vì có thể gây hư hỏng phần cứng khi đào tạo trong thế giới thực. Nhiều hạn chế như camera có độ chính xác thấp khiến các mô hình này có tính khái quát kém khi được đào tạo về mô phỏng vật lý và triển khai trong thế giới thực.

Tobin và cộng sự. [104] khám phá tính hiệu quả của việc sử dụng các phong cách khác nhau trong mô phỏng đào tạo và đạt được độ chính xác trong vòng 1,5 cm trong thế giới thực khi thực hiện nhiệm vụ định vị đối tượng. Thí nghiệm của họ ngẫu nhiên hóa vị trí và kết cấu của các vật thể được phát hiện



Hình 27Ví dụ về các phong cách khác nhau được mô phỏng bởi Tobin et al. [104]

trên bàn trong mô phỏng cũng như kết cấu, ánh sáng, số lượng đèn và tiếng ồn ngẫu nhiên trong nền. Họ nhận thấy rằng với đủ sự thay đổi trong kiểu dữ liệu huấn luyện, thế giới thực chỉ đơn giản xuất hiện dưới dạng một biến thể khác của mô hình. Điều thú vị là họ nhận thấy rằng sự đa dạng trong phong cách có hiệu quả hơn là mô phỏng trong một môi trường thực tế nhất có thể. Điều này trái ngược với công việc của Shrivastava et al. [105] những người đã sử dụng GAN để làm cho dữ liệu mô phỏng của họ trở nên thực tế nhất có thể (Hình 2).²⁷⁾

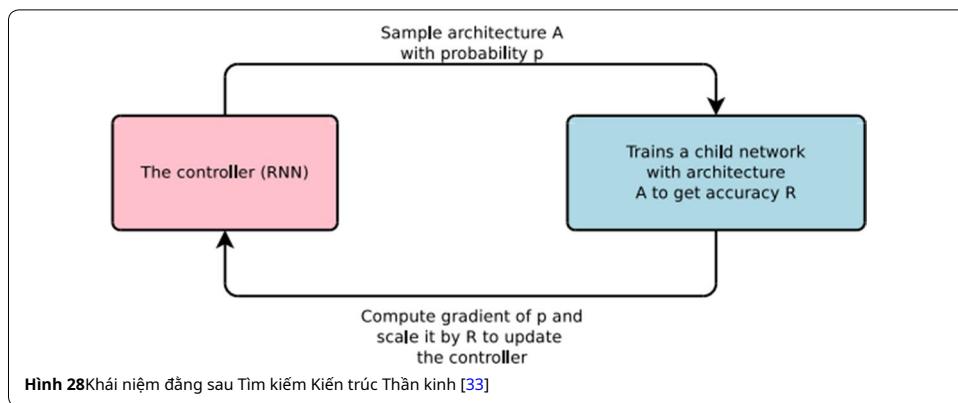
Việc sử dụng dữ liệu mô phỏng để xây dựng các mô hình Thị giác Máy tính đã được nghiên cứu kỹ lưỡng. Một ví dụ về điều này là từ Richter et al. [106]. Họ sử dụng đồ họa máy tính từ các trò chơi thế giới mở hiện đại như Grand Theft Auto để tạo ra các tập dữ liệu phân đoạn theo ngữ nghĩa. Các tác giả nêu bật những giao thoại về chi phí chủ thích thủ công cần thiết để xây dựng các bộ dữ liệu cấp pixel này. Họ đề cập đến tập dữ liệu CamVid [107] yêu cầu 60 phút cho mỗi hình ảnh để chú thích thủ công và bộ dữ liệu Cityscapes [108] yêu cầu 90 phút cho mỗi hình ảnh. Chi phí nhân công và thời gian cao này thúc đẩy việc sử dụng và phát triển các bộ dữ liệu tổng hợp. Chuyển kiểu thần kinh là một chiến lược rất thú vị để cải thiện khả năng khái quát hóa của các bộ dữ liệu mô phỏng.

Một nhược điểm của Tăng cường dữ liệu truyền kiểu thần kinh là nỗ lực cần thiết để chọn kiểu để truyền hình ảnh vào. Nếu tập kiểu dáng quá nhỏ, các sai lệch tiếp theo có thể được đưa vào tập dữ liệu. Cố gắng tái tạo các thí nghiệm của Tobin et al. [104] sẽ cần một lượng lớn bộ nhớ và khả năng tính toán bổ sung để chuyển đổi và lưu trữ 79.433 hình ảnh mới từ mỗi hình ảnh.

Thuật toán ban đầu được đề xuất bởi Gatys et al. [32] có thời gian chạy rất chậm và do đó không thực tế cho việc Tăng cường dữ liệu. Thuật toán được phát triển bởi Johnson et al. [35] nhanh hơn nhiều, nhưng hạn chế chuyển sang một tập hợp các kiểu được đào tạo trước.

Tăng cường dữ liệu học tập meta

Khái niệm meta-learning trong nghiên cứu Deep Learning thường đề cập đến khái niệm tối ưu hóa mạng lưới thần kinh bằng mạng lưới thần kinh. Cách tiếp cận này đã trở nên rất phổ biến kể từ khi xuất bản NAS [33] từ Zoph và Lê. Thực và cộng sự. [109,110] Mà còn



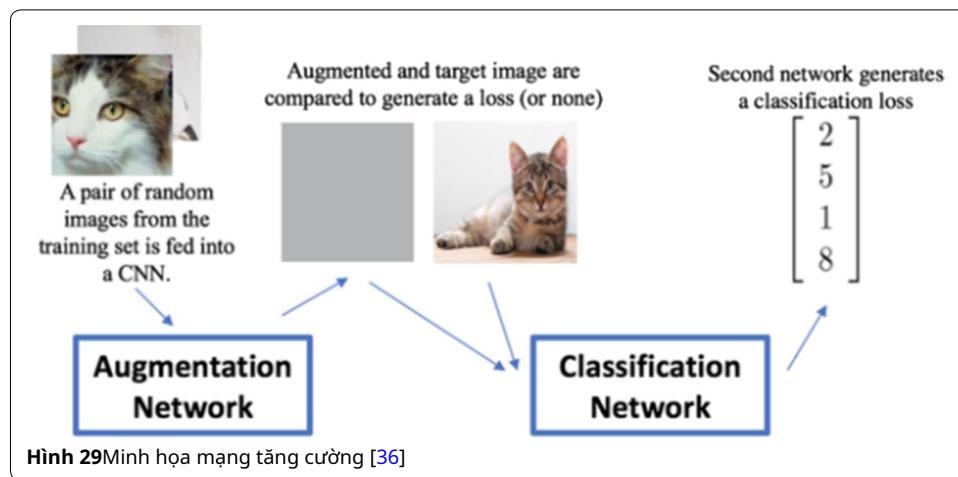
cho thấy tính hiệu quả của các thuật toán tiến hóa để tìm kiếm kiến trúc. Salimans và cộng sự. [111] so sánh trực tiếp các chiến lược tiến hóa với Học tăng cường. Một lựa chọn thú vị khác cho Học tăng cường là tìm kiếm ngẫu nhiên đơn giản [112]. Việc sử dụng các thuật toán tìm kiếm ngẫu nhiên và tiến hóa là một lĩnh vực thú vị trong công việc trong tương lai, nhưng các sơ đồ siêu học tập được xem xét trong khảo sát này đều là mạng thần kinh, dựa trên độ dốc.

Lịch sử tiến bộ của Deep Learning từ kỹ thuật tính năng như SIFT [113] và HOG [114]] tới thiết kế kiến trúc như AlexNet [1], VGGNet [2] và Inception-V3 [4], gợi ý rằng thiết kế siêu kiến trúc là sự thay đổi mô hình tiếp theo. NAS thực hiện một cách tiếp cận mới đối với kiến trúc meta-learning bằng cách sử dụng mạng lặp lại được đào tạo bằng Học tăng cường để thiết kế kiến trúc mang lại độ chính xác tốt nhất. Trên tập dữ liệu CIFAR-10, điều này đạt được tỷ lệ lỗi là 3,65 (Hình 2).28).

Phần này sẽ giới thiệu ba thử nghiệm sử dụng meta-learning để Tăng cường dữ liệu. Các phương pháp này sử dụng mạng thần kinh được chuẩn bị trước để tìm hiểu Tăng cường dữ liệu thông qua việc trộn hình ảnh, Chuyển kiểu thần kinh và các phép biến đổi hình học.

Tăng cường thần kinh Thuật toán Chuyển kiểu thần kinh yêu cầu hai tham số cho trọng số của kiểu và mất nội dung. Perez và Vương [36] đã trình bày một thuật toán để học tổng hợp chiến lược Chuyển giao kiểu thần kinh được gọi là Tăng cường thần kinh. Phương pháp Tăng cường thần kinh lấy hai hình ảnh ngẫu nhiên từ cùng một lớp. Mạng tăng cường được chuẩn bị trước ánh xạ chúng thành một hình ảnh mới thông qua CNN với 5 lớp, mỗi lớp có 16 kênh, 3×3 bộ lọc và chức năng kích hoạt ReLU. Sau đó, hình ảnh xuất ra từ quá trình tăng cường sẽ được chuyển đổi bằng một hình ảnh ngẫu nhiên khác thông qua Chuyển kiểu thần kinh. Việc chuyển kiểu này được thực hiện thông qua CycleGAN [92] phần mở rộng của GAN [31] khuôn khổ. Những hình ảnh này sau đó được đưa vào mô hình phân loại và lỗi từ mô hình phân loại sẽ được truyền ngược để cập nhật mạng Tăng cường thần kinh. Mạng Tăng cường thần kinh sử dụng lỗi này để tìm hiểu trọng số tối ưu cho nội dung và kiểu hình ảnh giữa các hình ảnh khác nhau cũng như ánh xạ giữa các hình ảnh trong CNN (Hình 2).29).

Perez và Wang đã thử nghiệm thuật toán của họ trên bộ dữ liệu MNIST và Tiny-imagenet-200 về các nhiệm vụ phân loại nhị phân như mèo so với chó. Bộ dữ liệu Tiny-imagenet-200 được sử dụng để mô phỏng dữ liệu hạn chế. Bộ dữ liệu Tiny-imagenet-200 chỉ chứa 500 hình ảnh trong mỗi lớp, trong đó 100 hình ảnh được dành riêng để xác thực. Vấn đề này



giới hạn tập dữ liệu này ở 2 lớp. Như vậy chỉ có 800 hình ảnh để huấn luyện. Mỗi hình ảnh Tiny-imagenet-200 là $64 \times 64 \times 3$ và hình ảnh MNIST là $28 \times 28 \times 1$. Thí nghiệm so sánh Tăng cường thần kinh được đề xuất của họ [36] với các kỹ thuật tăng cường truyền thống như cắt xén và xoay, cũng như với cách tiếp cận chuyển kiểu với một tập hợp các kiểu được xác định trước như Đêm/Ngày và Mùa đông/Mùa hè.

Nghiên cứu cơ bản truyền thống đã biến đổi hình ảnh bằng cách chọn mức tăng cường từ một tập hợp (dịch chuyển, phóng to/thu nhỏ, xoay, lật, làm méo hoặc tô bóng bằng màu sắc). Điều này được lặp lại để tăng kích thước tập dữ liệu từ N lên 2 N. Đường cơ sở chuyển kiểu GAN sử dụng 6 kiểu khác nhau để chuyển đổi hình ảnh (Cezanne, Enhance, Monet, Ukiyoe, Van Gogh và Winter). Các kỹ thuật Tăng cường thần kinh được thử nghiệm bao gồm ba cấp độ dựa trên thiết kế của hàm mất mát cho mạng tăng cường (Mất nội dung, Mất kiểu qua ma trận gram và không mất máy tính ở lớp này). Tất cả các thử nghiệm đều được thử nghiệm với một mạng tích chập bao gồm 3 lớp tích chập, mỗi lớp tiếp theo là tổng hợp tối đa và chuẩn hóa hàng loạt, tiếp theo là 2 lớp được kết nối đầy đủ. Mỗi thử nghiệm chạy trong 40 kỷ nguyên với tốc độ học tập là 0.7).

Kết quả thí nghiệm rất hứa hẹn. Kỹ thuật Tăng cường thần kinh hoạt động tốt hơn đáng kể trong nghiên cứu Chó so với Cá vàng và chỉ kém hơn một chút trong nghiên cứu Chó so với Mèo. Kỹ thuật này không có bất kỳ tác động nào tới bài toán MNIST. Bài viết gợi ý rằng chiến lược tốt nhất có thể là kết hợp các phương pháp tăng cường truyền thống và Tăng cường thần kinh.

Tăng cường thông minh Sự tăng cường thông minh [37] sử dụng một khái niệm tương tự như kỹ thuật Tăng cường thần kinh được trình bày ở trên. Tuy nhiên, sự kết hợp của các hình ảnh chỉ được lấy từ các tham số đã học của CNN được thêm vào trước, thay vì sử dụng thuật toán Chuyển kiểu thần kinh.

Tăng cường thông minh là một cách tiếp cận khác để tăng cường siêu học tập. Điều này được thực hiện bằng cách có hai mạng, *Mạng-MÔI* và *Mạng-B*. *Mạng-MÔI* là một mạng tăng cường nhận hai hoặc nhiều hình ảnh đầu vào và ánh xạ chúng thành một hoặc nhiều hình ảnh mới để huấn luyện *Mạng-B*. Sự thay đổi tỷ lệ lỗi trong *Mạng-B* vậy thì

Bảng 7 Kết quả so sánh các phép tăng thêm [36]**Kết quả định lượng trên chó so với cá vàng****Chó vs cá vàng**

Tăng cường	Val. acc.
Không có	0,855
Truyền thống	0,890
GAN	0,865
Thần kinh + không mất Thần	0,915
kinh + mất nội dung Thần kinh +	<u>0,900</u>
phong cách	<u>0,890</u>
Điều khiển	0,840

Kết quả định lượng về chó vs mèo**Chó vs mèo**

Tăng cường	Val. acc.
Không có	0,705
Truyền thống	0,775
GAN	0,720
Thần kinh + không mất Thần	0,765
kinh + mất nội dung Thần kinh +	<u>0,770</u>
phong cách	<u>0,740</u>
Điều khiển	0,710

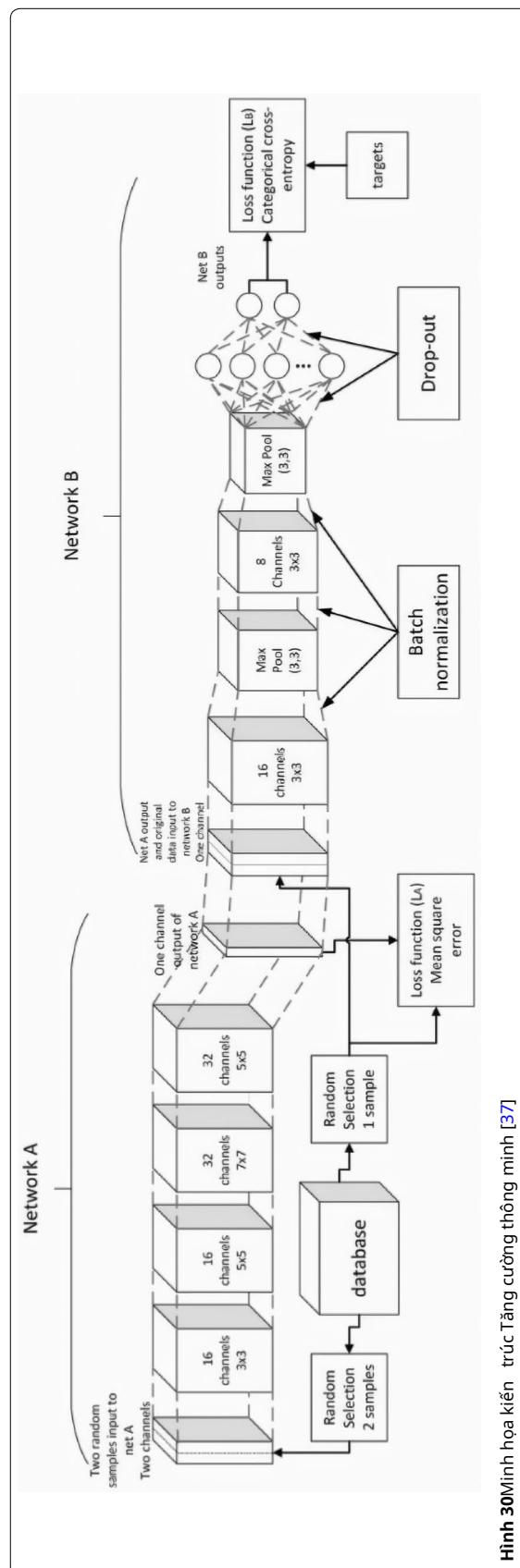
MINIST 0's và 8's

Tăng cường	Val. acc.
Không có	0,972
Thần kinh + không mất Thần	0,975
kinh + mất nội dung	<u>0,968</u>

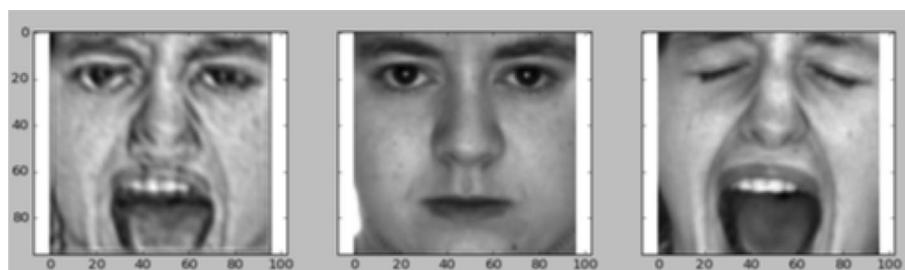
được truyền ngược để cập nhật *Mạng-MÔT*. Ngoài ra, một hàm mất mát khác được tích hợp vào *Mạng-MÔT* để đảm bảo rằng đầu ra của nó tương tự với các đầu ra khác trong lớp. *Mạng-MÔT* sử dụng một loạt các lớp chập để tạo ra hình ảnh tăng cường. Khung khái niệm của *Mạng-MÔT* có thể được mở rộng để sử dụng một số Mạng được đào tạo song song. Nhiều *Mạng-BẮNG* có thể rất hữu ích cho việc học các phần mở rộng dành riêng cho từng lớp thông qua học tập tổng hợp (Hình 2).[30](#).

Tăng cường thông minh tương tự như SamplePairing [65] hoặc các ví dụ hỗn hợp theo nghĩa là sự kết hợp của các ví dụ hiện có sẽ tạo ra các ví dụ mới. Tuy nhiên, cơ chế Tăng cường thông minh phức tạp hơn nhiều, sử dụng CNN thích ứng để tạo ra hình ảnh mới thay vì lấy điểm ảnh trung bình hoặc kết hợp hình ảnh được thiết kế thủ công.

Kỹ thuật Tăng cường thông minh đã được thử nghiệm với nhiệm vụ nhận dạng giới tính. Trên tập dữ liệu Feret, độ chính xác được cải thiện từ 83,52 lên 88,46%. Tập dữ liệu về khán giả đã phản hồi với mức cải thiện từ 70,02% lên 76,06%. Điều thú vị nhất là kết quả từ một tập dữ liệu khuôn mặt khác đã tăng từ 88,15 lên 95,66%. Điều này được so sánh với các kỹ thuật tăng cường truyền thống giúp tăng độ chính xác từ 88,15 lên 89,08%. Ngoài ra, thí nghiệm này đạt được độ chính xác tương tự khi sử dụng hai *Mạng-BẮNG* trong khung tăng cường như đã được tìm thấy với một *Mạng-MÔT*. Thí nghiệm này chứng tỏ



Hình 30 Minh họa kiến trúc Tăng cường thông minh [37]



Hình 31 Trong nhiệm vụ nhận dạng giới tính, hình ảnh bên trái là ví dụ về phiên bản do Mạng-A tạo ra trong Tăng cường thông minh với các hình ảnh bên phải làm đầu vào [37]

hiệu suất tăng lên đáng kể với chiến lược siêu học tập Tăng cường thông minh (Hình 2).[\[31\]](#).

Tự động tăng cường Tự động tăng cường [\[38\]](#), được phát triển bởi Cubuk và cộng sự, là một cách tiếp cận siêu học khác nhiều so với Tăng cường thần kinh hoặc Tăng cường thông minh. Tự động tăng cường là một thuật toán Học tăng cường [\[115\]](#) tìm kiếm một chính sách tăng cường tối ưu giữa một tập hợp các phép biến đổi hình học bị ràng buộc với các mức độ biến dạng khác nhau. Ví dụ: 'translateX 20 pixel' có thể là một trong những phép biến đổi trong không gian tìm kiếm (Bảng [số 8](#)).

Trong thuật toán Học tăng cường, một chính sách tương tự như chiến lược của thuật toán học. Chính sách này xác định những hành động cần thực hiện ở các trạng thái nhất định để đạt được mục tiêu nào đó. Phương pháp AutoAugment tìm hiểu một chính sách bao gồm nhiều chính sách phụ, mỗi chính sách phụ bao gồm một chuyển đổi hình ảnh và mức độ chuyển đổi. Do đó, Học tăng cường được sử dụng như một thuật toán tìm kiếm riêng biệt của các phần mở rộng. Các tác giả cũng gợi ý rằng các thuật toán tiến hóa hoặc tìm kiếm ngẫu nhiên cũng sẽ là những thuật toán tìm kiếm hiệu quả.

AutoAugment đã tìm thấy các chính sách đạt tỷ lệ lỗi 1,48% trên CIFAR-10. Tính năng Tự động tăng cường cũng đạt được độ chính xác Top 1 83,54% trên tập dữ liệu ImageNet. Cũng rất thú vị, các chính sách học được trên bộ dữ liệu ImageNet đã thành công khi được chuyển sang các nhiệm vụ nhận dạng hình ảnh của Stanford Cars và FGVC Aircraft. Trong trường hợp này, chính sách ImageNet áp dụng cho các bộ dữ liệu khác này đã giảm tỷ lệ lỗi lần lượt là 1,16% và 1,76%.

Geng và cộng sự. [\[116\]](#) đã mở rộng trên AutoAugment bằng cách thay thế thuật toán tìm kiếm Học tăng cường bằng Tìm kiếm ngẫu nhiên tăng cường (ARS) [\[112\]](#). Các tác giả chỉ ra rằng các chính sách phụ học được từ AutoAugment vốn có nhiều thiếu sót do không gian tìm kiếm rỗng rạc. Họ chuyển đổi xác suất và mức độ gia tăng thành một không gian liên tục và tìm kiếm các chính sách phụ bằng ARS. Với điều này, họ đạt được tỷ lệ lỗi thấp hơn trên CIFAR-10, CIFAR-100 và ImageNet (Bảng [9](#)).

Minh và cộng sự. [\[117\]](#) cũng đã thử nghiệm sử dụng Học tăng cường [\[115\]](#) để tìm kiếm Tăng cường dữ liệu. Họ khám phá thêm về tính hiệu quả của việc chuyển đổi học tập đối với từng trường hợp riêng lẻ thay vì toàn bộ tập dữ liệu. Họ nhận thấy sự khác biệt về độ chính xác trong phân loại là 70,18% so với 74,42% trên tập dữ liệu CIFAR-10 và 74,61% so với 80,35% trong vấn đề phân loại chó và mèo. Hơn nữa, họ khám phá sự mạnh mẽ của các bộ phân loại liên quan đến việc tăng thời gian thử nghiệm và thấy rằng mô hình

Bảng 8 Chính sách tăng cường AutoAugment được tìm thấy trên tập dữ liệu CIFAR-10 đã rút gọn [38]

	Hoạt động 1	Hoạt động 2
Chính sách phụ 0	(Đảo ngược, 0,1,7)	(Đảo tương phản, 0,2,6)
Chính sách phụ 1	(Xoay, 0,7,2)	(DịchX,0,3,9)
Chính sách phụ 2	(Độ sắc nét, 0,8,1)	(Độ sắc nét, 0,9,3)
Chính sách phụ 3	(CắtY,0,5,8)	(DịchY,0,7,9)
Chính sách phụ 4	(Tự động tương phản, 0,5,8)	(Cân bằng, 0,9,2)
Chính sách phụ 5	(CắtY,0,2,7)	(Áp phích, 0,3,7)
Chính sách phụ 6	(Màu, 0,4,3)	(Độ sáng, 0,6,7)
Chính sách phụ 7	(Độ sắc nét, 0,3,9)	(Độ sáng, 0,7,9)
Chính sách phụ 8	(Cân bằng, 0,6,5)	(Cân bằng, 0,5,1)
Chính sách phụ 9	(Tương phản, 0,6,7)	(Độ sắc nét, 0,6,5)
Chính sách phụ 10	(Màu, 0,7,7)	(DịchX,0,5,8)
Chính sách phụ 11	(Cân bằng, 0,3,7)	(Tự động tương phản, 0,4,8)
Chính sách phụ 12	(DịchY,0,4,3)	(Độ sắc nét, 0,2,6)
Chính sách phụ 13	(Độ sáng, 0,9,6)	(Màu sắc, 0,2,8)
Chính sách phụ 14	(Solarize, 0,5,2)	(Đảo ngược, 0,0,3)
Chính sách phụ 15	(Cân bằng, 0,2,0)	(Tự động tương phản, 0,6,0)
Chính sách phụ 16	(Cân bằng, 0,2,8)	(Cân bằng, 0,6,4)
Chính sách phụ 17	(Màu, 0,9,9)	(Cân bằng, 0,6,6)
Chính sách phụ 18	(Tự động tương phản, 0,8,4)	(Solarize, 0,2,8)
Chính sách phụ 19	(Độ sáng, 0,1,3)	(Màu, 0,7,0)
Chính sách phụ 20	(Solarize, 0,4,5)	(Tự động tương phản, 0,9,3)
Chính sách phụ 21	(DịchY,0,9,9)	(DịchY,0,7,9)
Chính sách phụ 22	(Tự động tương phản, 0,9,2)	(Solarize, 0,8,3)
Chính sách phụ 23	(Cân bằng, 0,8,8)	(Đảo ngược, 0,1,3)
Chính sách phụ 24	(DịchY,0,7,9)	(Tự động tương phản, 0,9,1)

Bảng 9 Hiệu suất của ARS trên không gian liên tục so với AutoAugment trên không gian rác [116]

Người mẫu	Tự động tăng cường	ARS-Tháng 8
Wide-ResNet-28-10	2,68	2,33
Lắc-Lắc (26 2x32 ngày) Lắc-	2,47	2,14
Lắc (26 2x96 ngày) Lắc-Lắc	1,99	1,68
(26 2x112 ngày)	1,89	1,59
AmoebaNet-B (6.128)	1,75	1,49
PyramidNet + ShakeDrop	1,48	1,26

được đào tạo với Tìm kiếm tăng cường Học tập tăng cường hoạt động tốt hơn nhiều. Trên tập dữ liệu CIFAR-10, kết quả này mang lại độ chính xác 50,99% so với 70,06% khi các mô hình được đánh giá trên dữ liệu thử nghiệm tăng cường.

Một bất lợi của meta-learning là nó là một khái niệm tương đối mới và chưa được thử nghiệm nhiều. Ngoài ra, các chương trình siêu học tập có thể khó thực hiện và tốn thời gian. Những người thực hành siêu học sẽ phải giải quyết các vấn đề chủ yếu với độ dốc biến mất [118], trong số những người khác, để đào tạo các mạng này.

So sánh các phần mở rộng

Như được thể hiện xuyên suốt “[Cân nhắc thiết kế cho việc tăng cường dữ liệu hình ảnh](#)” phần, khả năng Tăng cường dữ liệu. Tuy nhiên, không có nhiều nghiên cứu so sánh cho thấy sự khác biệt về hiệu suất của các cách tăng cường khác nhau này. Một nghiên cứu như vậy được thực hiện bởi Shijie et al. [119] so sánh GAN, WGAN, lật, cắt xén, dịch chuyển, hiện tượng jitter PCA, jitter màu, thêm nhiễu, xoay và một số kết hợp trên bộ dữ liệu CIFAR-10 và ImageNet. Ngoài ra, nghiên cứu so sánh có phạm vi kích thước tập dữ liệu khác nhau với tập nhỏ bao gồm 2 k mẫu với 200 mẫu trong mỗi lớp, tập trung bình bao gồm 10 k mẫu với 1 k trong mỗi lớp và tập lớn bao gồm 50 k mẫu với 5 mẫu. k ở mỗi lớp. Họ cũng đã thử nghiệm với 3 cấp độ tăng cường, không tăng cường, mẫu gốc cộng với cùng kích thước của mẫu được tạo và mẫu gốc cộng với kích thước gấp đôi của mẫu được tạo. Họ nhận thấy rằng việc cắt, lật, WGAN và xoay thường hoạt động tốt hơn những cách khác. Sự kết hợp giữa lật+cắt và lật+GAN về tổng thể là tốt nhất, cải thiện hiệu suất phân loại trên CIFAR-10 thêm +3% và +3,5%,

Cân nhắc thiết kế cho việc tăng cường dữ liệu hình ảnh

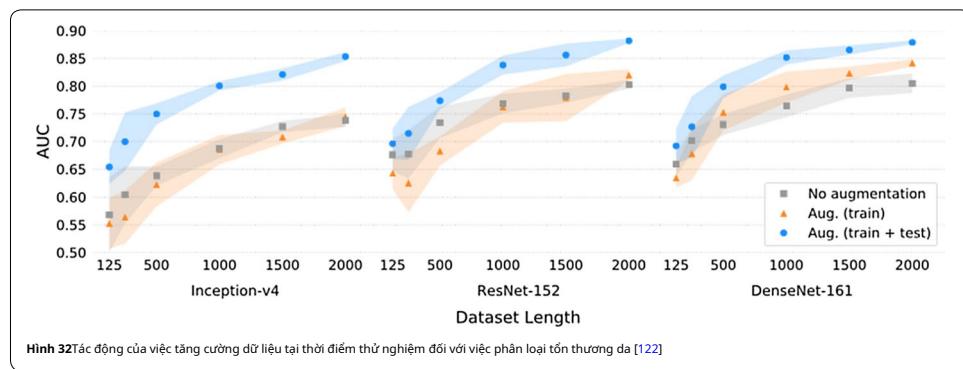
Phần này sẽ mô tả ngắn gọn một số quyết định thiết kế bổ sung liên quan đến kỹ thuật Tăng cường dữ liệu trên dữ liệu hình ảnh.

Tăng thời gian thử nghiệm

Ngoài việc tăng cường dữ liệu đào tạo, nhiều báo cáo nghiên cứu cũng cho thấy tính hiệu quả của việc tăng cường dữ liệu tại thời điểm thử nghiệm. Điều này có thể được coi là tương tự với các kỹ thuật học tập tổng hợp trong không gian dữ liệu. Bằng cách lấy một hình ảnh thử nghiệm và tăng cường nó theo cách tương tự như các hình ảnh huấn luyện, có thể rút ra một dự đoán chắc chắn hơn. Điều này đi kèm với chi phí tính toán tùy thuộc vào số lần tăng cường được thực hiện và nó có thể hạn chế tốc độ của mô hình. Đây có thể là một nút thắt cổ chai rất tốn kém trong các mô hình yêu cầu dự đoán theo thời gian thực. Tuy nhiên, việc tăng thời gian thử nghiệm là một phương pháp đầy hứa hẹn cho các ứng dụng như chẩn đoán hình ảnh y tế. Radosavovic và cộng sự. [120] biểu thị việc tăng thời gian thử nghiệm dưới dạng chất lọc dữ liệu để mô tả việc sử dụng các dự đoán tổng hợp để có được hình ảnh thể hiện tốt hơn.

Vương và cộng sự. [121] đã tìm cách phát triển một khung toán học để xây dựng phương pháp tăng thời gian thử nghiệm. Khi thử nghiệm sơ đồ tăng cường thời gian thử nghiệm trên phân đoạn hình ảnh y tế, họ nhận thấy rằng sơ đồ này hoạt động tốt hơn so với sơ đồ dự đoán cơ sở đơn lẻ và nhiều dự đoán dựa trên tình trạng bồi học. Họ cũng nhận thấy ước tính độ không đảm bảo tốt hơn khi sử dụng tính năng tăng thời gian thử nghiệm, giảm các dự đoán có độ tin cậy cao nhưng không chính xác. Phương pháp tăng cường thời gian thử nghiệm của họ sử dụng mô phỏng Monte Carlo để thu được các tham số cho các mức tăng khác nhau như lật, chia tỷ lệ, xoay và dịch, cũng như chèn tiếng ồn.

Việc tăng thời gian thử nghiệm có thể được tìm thấy trong bài báo Alexnet [1], áp dụng CNN cho tập dữ liệu ImageNet. Trong các thí nghiệm của mình, họ tính trung bình các dự đoán trên mười mảnh đất được cắt ngẫu nhiên. Các bản vá này bao gồm một bản được trích xuất từ trung tâm, bốn phần cắt ở góc và các vùng tương đương trên các hình ảnh được lật theo chiều ngang. Những dự đoán này được tính trung bình để tạo thành đầu ra cuối cùng. Anh ấy và cộng sự. [3] sử dụng quy trình thử nghiệm 10 vụ tương tự để đánh giá kiến trúc ResNet CNN của họ (Hình 2).[\[32\]](#).



Perez và cộng sự. [122] trình bày nghiên cứu về hiệu quả của việc tăng cường thời gian thử nghiệm với nhiều kỹ thuật tăng cường. Những cải tiến được thử nghiệm này bao gồm tăng màu, xoay, cắt, chia tỷ lệ, lật, cắt ngẫu nhiên, xóa ngẫu nhiên, đàn hồi, trộn và kết hợp giữa các kỹ thuật. Bàn 9 cho thấy hiệu suất cao hơn đạt được khi tăng cường hình ảnh thử nghiệm cũng như hình ảnh huấn luyện. Matsunaga và cộng sự. [123] cũng chứng minh tính hiệu quả của việc tăng thời gian thử nghiệm đối với việc phân loại tổn thương da, sử dụng các phép biến đổi hình học như xoay, dịch, chia tỷ lệ và lật.

Tác động của việc tăng thời gian thử nghiệm lên độ chính xác của phân loại là một cơ chế khác để đo lường độ tin cậy của bộ phân loại. Do đó, một bộ phân loại mạnh mẽ được định nghĩa là có phương sai thấp trong các dự đoán qua các lần tăng cường. Ví dụ: dự đoán về một hình ảnh sẽ không khác nhiều khi hình ảnh đó được xoay 20°. Trong các thử nghiệm tìm kiếm sự tăng cường bằng Học tăng cường, Minh và cộng sự. [117] đo lường độ tin cậy bằng cách làm biến dạng các hình ảnh thử nghiệm với xác suất 50% và đổi chiều độ chính xác trên dữ liệu chưa tăng cường với dữ liệu được tăng cường. Trong nghiên cứu này, hiệu suất của mô hình cơ sở giảm từ 74,61 xuống 66,87% khi đánh giá trên các hình ảnh thử nghiệm tăng cường.

Một số mô hình phân loại chưa chắc chắn về mức độ cần thiết của chúng đối với tốc độ. Điều này cho thấy hứa hẹn trong việc phát triển các phương pháp nâng cao dần độ tin cậy của dự đoán. Điều này có thể được thực hiện bằng cách trước tiên đưa ra một dự đoán với mức tăng thời gian thử nghiệm rất ít hoặc không có, sau đó thêm tăng dần thời gian thử nghiệm để tăng độ tin cậy của dự đoán. Các tác vụ Thị giác Máy tính khác nhau yêu cầu những ràng buộc nhất định về việc tăng cường thời gian thử nghiệm có thể được sử dụng. Ví dụ: nhận dạng hình ảnh có thể dễ dàng tổng hợp các dự đoán trên các hình ảnh bị biến dạng. Tuy nhiên, rất khó để tổng hợp các dự đoán về hình ảnh biến đổi hình học trong phát hiện đối tượng và phân đoạn ngữ nghĩa.

Giáo trình học

Bên cạnh nghiên cứu về Tăng cường dữ liệu, nhiều nhà nghiên cứu còn quan tâm đến việc cố gắng tìm ra chiến lược chọn dữ liệu huấn luyện vượt qua lựa chọn ngẫu nhiên. Trong bối cảnh Tăng cường dữ liệu, nghiên cứu đã được xuất bản để điều tra mối quan hệ giữa dữ liệu gốc và dữ liệu tăng cường qua các kỹ nguyên đào tạo. Vài nghiên cứu

Bảng 10 So sánh độ phân giải của ba bộ dữ liệu hình ảnh nguồn mở rất phổ biến

Tập dữ liệu	Nghị quyết
Chữ số viết tay MNIST	28×28×1
CIFAR-10/100	32×32×3
ImageNet	256×256×3

gợi ý rằng tốt nhất ban đầu chỉ huấn luyện với dữ liệu gốc và sau đó kết thúc huấn luyện với dữ liệu gốc và dữ liệu tăng cường, mặc dù không có sự đồng thuận rõ ràng.

Trong SamplePairing [65] nghiên cứu, một kỹ nguyên trên ImageNet và 100 kỹ nguyên trên các tập dữ liệu khác được hoàn thành mà không cần SamplePairing trước khi thêm dữ liệu hình ảnh hỗn hợp vào quá trình đào tạo. Sau khi các hình ảnh SamplePairing được thêm vào tập huấn luyện, chúng sẽ chạy theo chu kỳ giữa các kỹ nguyên 8:2, 8 kỹ nguyên có hình ảnh SamplePairing, 2 không có. Jaderberg và cộng sự. [124] huấn luyện độc quyền bằng dữ liệu tổng hợp để nhận dạng văn bản trong cảnh tự nhiên. Dữ liệu tổng hợp tạo ra dữ liệu huấn luyện bằng cách liệt kê thông qua các phông chữ và phần bổ sung khác nhau. Điều này tạo ra các bộ hình ảnh huấn luyện cho từ vựng có kích thước 50 k và 90 k. Mikolajczyk và Gochowski [72] rút ra những so sánh từ việc học chuyển tiếp. Họ đề xuất rằng việc đào tạo về dữ liệu tăng cường để tìm hiểu các trọng số ban đầu của mạng tích chập sâu cũng tương tự như việc truyền các trọng số được đào tạo trên các bộ dữ liệu khác như ImageNet. Các trọng số này sau đó chỉ được tinh chỉnh với dữ liệu huấn luyện ban đầu.

Các quyết định học tập trong chương trình giảng dạy đặc biệt quan trọng đối với các hệ thống Học một lần như FaceNet, được trình bày bởi Schroff et al. [125]. Điều quan trọng là tìm các khuôn mặt có phần giống với khuôn mặt mới sao cho hàm khoảng cách đã học thực sự hữu ích. Theo nghĩa này, khái niệm học theo chương trình giảng dạy có nhiều điểm tương đồng với các thuật toán tìm kiếm đối nghịch hoặc chỉ học trên các ví dụ khó.

Học tập theo chương trình giảng dạy, một thuật ngữ ban đầu được đặt ra bởi Bengio et al. [126], là một khái niệm có thể áp dụng cho tất cả các mô hình Deep Learning, không chỉ những mô hình bị hạn chế về dữ liệu. Việc vạch ra độ chính xác của quá trình đào tạo theo thời gian trên các tập hợp con đào tạo ban đầu khác nhau có thể giúp tiết lộ các mẫu trong dữ liệu giúp tăng tốc đáng kể thời gian đào tạo. Tăng cường dữ liệu xây dựng quá trình đào tạo tăng cao ô ạt từ các kết hợp như lật, dịch và xóa ngẫu nhiên. Rất có khả năng tồn tại một tập hợp con trong tập hợp này để việc đào tạo sẽ nhanh hơn và chính xác hơn.

Tác động của độ phân giải

Một cuộc thảo luận thú vị khác về Tăng cường dữ liệu trong hình ảnh là tác động của độ phân giải. Hình ảnh có độ phân giải cao hơn như HD (1920×1080×3) hoặc 4 K (3840×2160×3) yêu cầu xử lý và bộ nhớ nhiều hơn để đào tạo các CNN sâu. Tuy nhiên, có vẻ trực quan rằng các mô hình thế hệ tiếp theo sẽ được đào tạo về hình ảnh có độ phân giải cao hơn. Nhiều mô hình hiện tại lấy mẫu hình ảnh xuống từ độ phân giải ban đầu của chúng để làm cho vấn đề phân loại trở nên khả thi hơn về mặt tính toán. Tuy nhiên, đôi khi việc lấy mẫu xuống này làm mất thông tin trong ảnh, khiến việc nhận dạng ảnh trở nên khó khăn hơn (Bảng 10).

Thật thú vị khi điều tra bản chất của việc lấy mẫu xuống này và so sánh hiệu suất thu được. Wu và cộng sự. [127] so sánh sự cân bằng giữa độ chính xác và tốc độ

Original image			Low-resolution model			High-resolution model		
Rank	Score	Class	Rank	Score	Class	Rank	Score	Class
1	0.2287	ant	1	0.103	lacewing	1	0.103	lacewing
2	0.0997	damselfly	2	0.074	dragonfly	2	0.074	dragonfly
3	0.057	nematode	3	0.074	damselfly	3	0.074	damselfly
4	0.0546	chainlink fence	4	0.063	walking stick	4	0.063	walking stick
5	0.0522	long-horned	5	0.039	long-horned	5	0.039	long-horned
6	0.0307	walking stick	6	0.027	leafhopper	6	0.027	leafhopper
7	0.0287	dragonfly	7	0.025	nail	7	0.025	nail
8	0.0267	tiger beetle	8	0.023	grasshopper	8	0.023	grasshopper
9	0.0225	doormat	9	0.019	ant	9	0.019	ant
10	0.0198	flute	10	0.015	mantis	10	0.015	mantis
11	0.0198	grey whale	11	0.015	fly	11	0.015	fly
12	0.0178	mantis	12	0.013	hammer	12	0.013	hammer
13	0.0171	lacewing	13	0.012	American	13	0.012	American
14	0.0161	radiator	14	0.012	gar	14	0.012	gar
15	0.0161	scabbard	15	0.011	chainlink	15	0.011	chainlink
16	0.0157	slide rule	16	0.011	padlock	16	0.011	padlock
17	0.0148	fly	17	0.011	tree frog	17	0.011	tree frog
18	0.0129	leafhopper	18	0.011	cicada	18	0.011	cicada
19	0.0101	cucumber	19	0.01	screwdriver	19	0.01	screwdriver
20	0.0094	velvet	20	0.01	harvestman	20	0.01	harvestman

Hình 33 Phân loại Hình ảnh bên phải theo các mô hình độ phân giải khác nhau do Wu và cộng sự đào tạo. [127]

khi giảm tần số lấy mẫu hình ảnh xuống các độ phân giải khác nhau. Các nhà nghiên cứu nhận thấy rằng việc tạo ra một tập hợp các mô hình được đào tạo với hình ảnh có độ phân giải cao và thấp sẽ hoạt động tốt hơn bất kỳ mô hình riêng lẻ nào. Dự đoán tổng thể này được tìm thấy bằng cách lấy trung bình các dự đoán softmax. Các mô hình được đào tạo trên 256×256 hình ảnh và 512×512 hình ảnh đạt tỷ lệ lỗi top 5 lần lượt là 7,96% và 7,42%. Khi tổng hợp lại, chúng đạt được tỷ lệ lỗi trong top 5 thấp hơn là 6,97%. Do đó, các hình ảnh được lấy mẫu xuống khác nhau có thể được xem dưới dạng sơ đồ Tăng cường dữ liệu khác (Hình 2).³³.

Với sự tiến bộ của Mạng thần kinh chuyển đổi siêu phân giải được trình bày bởi Chong et al. [128] hoặc SRGAN, Mạng đối thủ tạo ra độ phân giải siêu cao, được trình bày bởi Ledig et al. [129], thật thú vị khi cân nhắc xem liệu việc nâng cấp hình ảnh lên độ phân giải cao hơn nữa có mang lại mô hình tốt hơn hay không. Lấy mẫu nâng cao chất lượng trên hình ảnh CIFAR-10 từ thậm chí 32×32×3 đến 64×64×3 có thể dẫn tới các bộ phân loại hình ảnh tốt hơn và mạnh mẽ hơn.

Độ phân giải cũng là một chủ đề rất quan trọng với GAN. Việc tạo ra các kết quả đầu ra có độ phân giải cao từ GAN là rất khó do các vấn đề về độ ổn định trong quá trình huấn luyện và sự sụp đổ của chế độ. Nhiều kiến trúc GAN mới hơn như StackGAN [130] và GAN đang phát triển dần dần [34] được thiết kế để tạo ra hình ảnh có độ phân giải cao hơn. Ngoài các kiến trúc này, việc sử dụng các mạng siêu phân giải như SRGAN có thể là một kỹ thuật hiệu quả để cải thiện chất lượng đầu ra từ DCGAN [91] người mẫu. Khi việc tạo ra các đầu ra có độ phân giải cao từ các mẫu GAN là thực tế, những đầu ra này sẽ rất hữu ích cho việc Tăng cường dữ liệu.

Kích thước tập dữ liệu cuối cùng

Một thành phần cần thiết của Tăng cường dữ liệu là xác định kích thước tập dữ liệu cuối cùng. Ví dụ: nếu tất cả hình ảnh được lật theo chiều ngang và được thêm vào tập dữ liệu, kích thước tập dữ liệu thu được sẽ thay đổi từ N thành 2N. Một trong những cân nhắc chính liên quan đến kích thước tập dữ liệu cuối cùng là các ràng buộc về bộ nhớ và tính toán bổ sung liên quan đến việc tăng dữ liệu. Người thực hành có thể lựa chọn giữa việc sử dụng các trình tạo để chuyển đổi dữ liệu nhanh chóng trong quá trình đào tạo hoặc chuyển đổi dữ liệu trước và

lưu trữ nó trong bộ nhớ. Việc chuyển đổi dữ liệu nhanh chóng có thể tiết kiệm bộ nhớ nhưng sẽ dẫn đến quá trình đào tạo chậm hơn. Việc lưu trữ tập dữ liệu trong bộ nhớ có thể cực kỳ khó khăn tùy thuộc vào mức độ tăng kích thước của tập dữ liệu. Việc lưu trữ các tập dữ liệu tăng cường trong bộ nhớ đặc biệt có vấn đề khi tăng cường dữ liệu lớn. Quyết định này thường được phân loại thành tăng cường dữ liệu trực tuyến hoặc ngoại tuyến, (với tăng cường trực tuyến đề cập đến tăng cường nhanh chóng và tăng cường ngoại tuyến đề cập đến việc chỉnh sửa và lưu trữ dữ liệu trên đĩa).

Trong việc thiết kế một hệ thống đào tạo phân tán rộng rãi, Chilimbi et al. [131] tăng cường hình ảnh trước khi đào tạo để tăng tốc độ phân phát hình ảnh. Bằng cách tăng cường hình ảnh trước, hệ thống phân tán có thể yêu cầu và lưu trữ các đợt đào tạo vào bộ đệm. Các phần mở rộng cũng có thể được tích hợp vào biểu đồ tính toán được sử dụng để xây dựng các mô hình Deep Learning và tạo điều kiện phân biệt nhanh chóng. Những sự tăng cường này xử lý hình ảnh ngay sau tensor hình ảnh đầu vào.

Ngoài ra, cũng rất thú vị khi khám phá một tập hợp con dữ liệu tăng cao sẽ mang lại hiệu suất cao hơn hoặc tương tự cho toàn bộ tập huấn luyện. Đây là một khái niệm tương tự như việc học theo chương trình giảng dạy, vì ý tưởng trọng tâm là tìm ra thứ tự tối ưu của dữ liệu huấn luyện. Ý tưởng này cũng liên quan rất nhiều đến kích thước tập dữ liệu cuối cùng cũng như các cân nhắc về tính toán chuyển đổi và bộ nhớ khả dụng để lưu trữ hình ảnh nâng cao.

Giảm bớt sự mất cân bằng lớp bằng Tăng cường dữ liệu

Mất cân bằng lớp là một vấn đề phổ biến trong đó tập dữ liệu chủ yếu bao gồm các ví dụ từ một lớp. Điều này có thể tự biểu hiện trong một vấn đề phân loại nhị phân sao cho có sự phân biệt rõ ràng về giai cấp đa số và thiểu số hoặc trong phân loại nhiều lớp trong đó có một hoặc nhiều lớp đa số và một hoặc nhiều lớp thiểu số. Các bộ dữ liệu không cân bằng có hại vì chúng thiên về các mô hình theo dự đoán của lớp đa số. Các bộ dữ liệu không cân bằng cũng đưa ra độ chính xác như một thước đo hiệu suất sai lệch. Buda và cộng sự. [132] cung cấp một nghiên cứu có hệ thống đặc biệt điều tra tác động của dữ liệu mất cân bằng trong việc xử lý dữ liệu hình ảnh của CNN. Leevy và cộng sự. [27] đề cập đến nhiều giải pháp ở cấp độ Dữ liệu và Thuật toán để giải quyết tình trạng mất cân bằng lớp trong dữ liệu lớn nói chung. Tăng cường dữ liệu nằm trong giải pháp Cấp dữ liệu cho sự mất cân bằng lớp và có nhiều chiến lược khác nhau để triển khai.

Một giải pháp đơn giản để lấy mẫu quá mức bằng Tăng cường dữ liệu sẽ là lấy mẫu quá mức ngẫu nhiên đơn giản với các phép biến đổi hình học nhỏ chẳng hạn như xoay 30° . Các thao tác hình ảnh đơn giản khác như tăng màu, trộn ảnh, bộ lọc hạt nhân và xóa ngẫu nhiên cũng có thể được mở rộng sang dữ liệu mẫu theo cách tương tự như tăng cường hình học. Điều này có thể hữu ích để dễ thực hiện và thử nghiệm nhanh chóng với các tỷ lệ lớp khác nhau. Một vấn đề của việc lấy mẫu quá mức với các phép biến đổi hình ảnh cơ bản là nó có thể gây ra tình trạng quá khớp đối với lớp thiểu số đang bị lấy mẫu quá mức. Những thành kiến hiện diện trong tầng lớp thiểu số phổ biến hơn sau khi lấy mẫu bằng các kỹ thuật này.

Các phương pháp lấy mẫu quá mức dựa trên Học sâu như đào tạo đối nghịch, Chuyển giao kiểu thần kinh, GAN và sơ đồ siêu học cũng có thể được sử dụng như một chiến lược lấy mẫu quá mức thông minh hơn. Chuyển kiểu thần kinh là một cách thú vị để tạo ra hình ảnh mới. Những hình ảnh mới này có thể được tạo thông qua kiểu ngoại suy với kiểu nước ngoài hoặc bằng cách nội suy các kiểu giữa các phiên bản trong tập dữ liệu. Sử dụng GAN

lấy mẫu dữ liệu quá mức có thể là một cách hiệu quả khác để tăng quy mô lớp thiểu số trong khi vẫn duy trì sự phân bố bên ngoài. Việc lấy mẫu quá mức với GAN có thể được thực hiện bằng cách sử dụng toàn bộ lớp thiểu số làm ví dụ "thực" hoặc bằng cách sử dụng các tập hợp con của lớp thiểu số làm đầu vào cho GAN. Việc sử dụng lấy mẫu tiến hóa [133] để tìm các tập hợp con này để nhập vào GAN để lấy mẫu lớp là một lĩnh vực đầy hứa hẹn cho công việc trong tương lai.

Cuộc thảo luận

Các cách thú vị để tăng cường dữ liệu hình ảnh thuộc hai loại chung: làm cong vênh dữ liệu và lấy mẫu quá mức. Nhiều phần mở rộng trong số này làm sáng tỏ cách cải thiện bộ phân loại hình ảnh, trong khi những phần mở rộng khác thì không. Có thể dễ dàng giải thích lợi ích của việc lật ngang hoặc cắt ngẫu nhiên. Tuy nhiên, không rõ tại sao việc trộn các pixel hoặc toàn bộ hình ảnh lại với nhau như trong chính quy hóa PatchShuffle hoặc SamplePairing lại hiệu quả đến vậy. Ngoài ra, rất khó để diễn giải các biểu diễn mà mạng thần kinh học được để tăng cường dựa trên GAN, bộ mã hóa tự động biến thiên và siêu học. Trực quan hóa CNN được dẫn dắt bởi Yosinski et al. [134] bằng phương pháp hình dung sâu sắc của họ. Việc có được sự hiểu biết ở cấp độ con người về các tính năng của mạng tích chập có thể giúp ích rất nhiều cho việc hướng dẫn quá trình tăng cường.

Thao tác sức mạnh đại diện của mạng lưới thần kinh đang được sử dụng theo nhiều cách thú vị để tiếp tục phát triển các kỹ thuật tăng cường. Các kỹ thuật tăng cường thủ công truyền thống như cắt xén, lật và thay đổi không gian màu đang được mở rộng bằng cách sử dụng GAN, Chuyển kiểu thần kinh và thuật toán tìm kiếm siêu học.

Dịch từ hình ảnh sang hình ảnh có nhiều ứng dụng tiềm năng trong Tăng cường dữ liệu. Neural Style Transfer sử dụng các lớp thần kinh để dịch hình ảnh sang các kiểu mới. Kỹ thuật này không chỉ sử dụng các biểu diễn thần kinh để tách 'phong cách' và 'nội dung' khỏi hình ảnh mà còn sử dụng các phép biến đổi thần kinh để chuyển phong cách của hình ảnh này sang hình ảnh khác. Chuyển kiểu thần kinh là một kỹ thuật tăng cường mạnh mẽ hơn nhiều so với tăng cường không gian màu truyền thống, nhưng ngay cả những phương pháp này cũng có thể được kết hợp với nhau.

Một đặc điểm thú vị của các phương pháp tăng cường này là khả năng kết hợp của chúng với nhau. Ví dụ: kỹ thuật xóa ngẫu nhiên có thể được xếp chồng lên trên bất kỳ phương pháp tăng cường nào trong số này. Khung GAN sở hữu một thuộc tính nội tại của đệ quy rất thú vị. Các mẫu lấy từ GAN có thể được tăng cường bằng các phương pháp tăng cường truyền thống như bộ lọc ánh sáng hoặc thậm chí được sử dụng trong các chiến lược tăng cường mạng thần kinh như Tăng cường thông minh hoặc Tăng cường thần kinh để tạo ra nhiều mẫu hơn nữa. Các mẫu này có thể được đưa vào các GAN tiếp theo và tăng đáng kể kích thước của tập dữ liệu gốc. Khả năng mở rộng của khung GAN là một trong nhiều lý do khiến chúng rất thú vị đối với các nhà nghiên cứu Deep Learning.

Việc tăng thời gian thử nghiệm tương tự như việc học tập trong không gian dữ liệu. Thay vì tổng hợp các dự đoán của các thuật toán học tập khác nhau, chúng tôi tổng hợp các dự đoán trên các hình ảnh được tăng cường. Chúng tôi thậm chí có thể mở rộng thuật toán giải pháp để tham số hóa các trọng số dự đoán từ các mức tăng khác nhau. Đây có vẻ như là một giải pháp tốt cho các hệ thống quan tâm đến việc đạt được điểm hiệu suất rất cao, hơn cả khả năng dự đoán

tốc độ. Xác định tính hiệu quả của việc tăng thời gian thử nghiệm bằng cách chủ yếu khám phá các phép biến đổi hình học trong thời gian thử nghiệm và Chuyển giao kiểu thần kinh, là một lĩnh vực công việc trong tương lai.

Một câu hỏi thú vị về Tăng cường dữ liệu thực tế là làm thế nào để xác định kích thước tập dữ liệu sau khi tăng cường. Không có sự đồng thuận về tỷ lệ kích thước tập dữ liệu gốc và cuối cùng sẽ mang lại mô hình hoạt động tốt nhất. Tuy nhiên, hãy tưởng tượng chỉ sử dụng tăng cường màu sắc. Nếu tập dữ liệu huấn luyện ban đầu bao gồm 50 con chó và 50 con mèo và mỗi hình ảnh được tăng cường 100 bộ lọc màu để tạo ra 5000 con chó và 5000 con mèo thì tập dữ liệu này sẽ thiên về đặc điểm không gian của 50 con chó và 50 con mèo ban đầu. Dữ liệu tăng cường màu quá rộng này sẽ khiến một mô hình sâu bị quá khớp, thậm chí còn tệ hơn so với ban đầu. Từ giai thoại này, chúng ta có thể khái niệm hóa sự tồn tại của kích thước tối ưu cho dữ liệu sau tăng cường.

Ngoài ra, không có sự đồng thuận về chiến lược tốt nhất để kết hợp các kỹ thuật lấy mẫu quá mức và làm cong vênh dữ liệu. Một điều quan trọng cần cân nhắc là độ lệch nội tại trong tập dữ liệu ban đầu, có giới hạn. Hiện tại không có kỹ thuật tăng cường nào có thể sửa một tập dữ liệu có tính đa dạng rất kém đối với dữ liệu thử nghiệm. Tất cả các thuật toán tăng cường này hoạt động tốt nhất với giả định rằng dữ liệu huấn luyện và dữ liệu kiểm tra đều được lấy từ cùng một phân phối. Nếu điều này không đúng thì rất khó có khả năng những phương pháp này sẽ hữu ích.

Công việc tương lai

Công việc trong tương lai về Tăng cường dữ liệu sẽ tập trung vào nhiều lĩnh vực khác nhau như thiết lập phân loại kỹ thuật tăng cường, cải thiện chất lượng của mẫu GAN, tìm hiểu các cách mới để kết hợp siêu học tập và Tăng cường dữ liệu, khám phá mối quan hệ giữa Tăng cường dữ liệu và kiến trúc phân loại, và mở rộng các nguyên tắc này sang các loại dữ liệu khác. Chúng tôi quan tâm đến việc xem thành phần chuỗi thời gian trong dữ liệu video tác động như thế nào đến việc sử dụng các kỹ thuật tăng cường hình ảnh tĩnh. Tăng cường dữ liệu không giới hạn ở miền hình ảnh và có thể hữu ích cho văn bản, tin sinh học, bản ghi dạng bảng, v.v.

Công việc trong tương lai của chúng tôi nhằm mục đích khám phá các điểm chuẩn hiệu suất thông qua việc tăng cường không gian hình học và màu sắc trên một số bộ dữ liệu từ các tác vụ nhận dạng hình ảnh khác nhau. Các bộ dữ liệu này sẽ bị hạn chế về kích thước để kiểm tra tính hiệu quả đối với các vấn đề về dữ liệu hạn chế. Zhang và cộng sự. [135] thử nghiệm kỹ thuật tăng cường GAN mới của họ trên tập dữ liệu SVHN trên 50, 80, 100, 200 và 500 trường hợp đào tạo. Tương tự như công việc này, chúng tôi sẽ xem xét thiết lập thêm các tiêu chuẩn cho các cấp độ dữ liệu hạn chế khác nhau.

Cải thiện chất lượng của các mẫu GAN và kiểm tra tính hiệu quả của chúng trên nhiều bộ dữ liệu là một lĩnh vực rất quan trọng khác cho công việc trong tương lai. Chúng tôi muốn khám phá thêm về khả năng tổ hợp của các mẫu GAN bằng các kỹ thuật tăng cường khác, chẳng hạn như áp dụng một loạt các chuyển đổi kiểu cho các mẫu do GAN tạo.

Mạng siêu phân giải thông qua việc sử dụng SRCNN, Mạng thần kinh chuyển đổi siêu phân giải và SRGAN cũng là những lĩnh vực rất thú vị cho công việc Tăng cường dữ liệu trong tương lai. Chúng tôi muốn khám phá sự khác biệt về hiệu suất giữa các kiến trúc với các hình ảnh được lấy mẫu lại, chẳng hạn như mở rộng hình ảnh CIFAR-10 từ 32×32 đến 64×64 đến 128×128 và vân vân. Một trong những khó khăn chính với các mẫu GAN là cố gắng đạt được kết quả đầu ra có độ phân giải cao. Do đó, sẽ rất thú vị khi xem làm thế nào chúng ta có thể sử dụng mạng siêu phân giải để đạt được độ phân giải cao như mẫu DCGAN

được nhập vào SRCNN hoặc SRGAN. Kết quả của chiến lược này sẽ được so sánh với hiệu suất của kiến trúc GAN đang phát triển dần dần.

Việc tăng thời gian thử nghiệm có khả năng tạo ra sự khác biệt lớn về hiệu suất Thị giác Máy tính và chưa được khám phá nhiều. Chúng tôi muốn thiết lập điểm chuẩn cho các nhóm tăng cường thời gian thử nghiệm khác nhau và điều tra các thuật toán giải pháp được sử dụng. Hiện tại, bô phiếu theo đa số dường như là thuật toán giải pháp chiếm ưu thế để tăng thời gian thử nghiệm. Có vẻ như rất có khả năng việc tăng cường thời gian thử nghiệm có thể được cải thiện hơn nữa nếu trọng số của mỗi dự đoán hình ảnh tăng cường được tham số hóa và học hỏi thêm. Ngoài ra, chúng tôi sẽ khám phá tính hiệu quả của việc tăng thời gian thử nghiệm trong việc phát hiện đối tượng, so sánh các mức tăng không gian màu và thuật toán Chuyển kiểu thần kinh.

Kiến trúc GAN siêu học tập là một lĩnh vực thú vị khác được quan tâm. Việc sử dụng các thuật toán Học tăng cường như NAS trên kiến trúc bộ tạo và bộ phân biệt đối xử có vẻ rất hứa hẹn. Một lĩnh vực thú vị khác cần nghiên cứu sâu hơn là sử dụng phương pháp tiến hóa để tăng tốc độ đào tạo GAN thông qua tính toán song song và tính toán cụm.

Một lĩnh vực quan trọng khác của công việc trong tương lai nhằm tích hợp thực tế Tăng cường dữ liệu vào quy trình làm việc của Deep Learning là phát triển các công cụ phần mềm. Tương tự như cách Tensorflow [136] tự động hóa các quy trình back-end của học tập giảm dần độ dốc, các thư viện Tăng cường dữ liệu sẽ tự động hóa các chức năng tiền xử lý. Máy Keras [137] cung cấp một lớp ImageDataGenerator hỗ trợ rất nhiều cho việc thực hiện các phép tăng cường hình học. Buslaev và cộng sự. đã trình bày một công cụ tăng cường khác mà họ gọi là Albulementations [138]. Sự phát triển của Chuyển giao kiểu thần kinh, đào tạo đối nghịch, GAN và API siêu học sẽ giúp các kỹ sư tận dụng sức mạnh hiệu suất của các kỹ thuật Tăng cường dữ liệu tiên tiến nhanh hơn và dễ dàng hơn nhiều.

Phần kết luận

Khảo sát này trình bày một loạt các giải pháp Tăng cường dữ liệu cho vấn đề trang bị quá mức trong các mô hình Deep Learning do dữ liệu hạn chế. Các mô hình Deep Learning dựa vào dữ liệu lớn để tránh trang bị quá mức. Việc tăng cường các bộ dữ liệu một cách giả tạo bằng cách sử dụng các phương pháp được thảo luận trong khảo sát này sẽ mang lại lợi ích cho dữ liệu lớn trong miền dữ liệu hạn chế. Tăng cường dữ liệu là một kỹ thuật rất hữu ích để xây dựng các bộ dữ liệu tốt hơn. Nhiều cải tiến đã được đề xuất mà thường có thể được phân loại là kỹ thuật làm cong vênh dữ liệu hoặc kỹ thuật lấy mẫu quá mức.

Tương lai của Tăng cường dữ liệu rất tươi sáng. Việc sử dụng các thuật toán tìm kiếm kết hợp các phương pháp làm cong vênh dữ liệu và lấy mẫu quá mức có tiềm năng rất lớn. Kiến trúc phân lớp của mạng lưới thần kinh sâu mang lại nhiều cơ hội cho Tăng cường dữ liệu. Hầu hết các phần mở rộng được khảo sát đều hoạt động ở lớp đầu vào. Tuy nhiên, một số có nguồn gốc từ các biểu diễn lớp ẩn và một phương thức, DisturbLabel [28], thậm chí còn được thể hiện ở lớp đầu ra. Không gian của các biểu diễn trung gian và không gian nhãn là các khu vực chưa được khám phá của Tăng cường dữ liệu với các kết quả thú vị. Cuộc khảo sát này tập trung vào các ứng dụng cho dữ liệu hình ảnh, mặc dù nhiều kỹ thuật và khái niệm này có thể được mở rộng sang các miền dữ liệu khác.

Tăng cường dữ liệu không thể khắc phục tất cả các thành kiến có trong một tập dữ liệu nhỏ. Ví dụ: trong nhiệm vụ phân loại giống chó, nếu chỉ có chó bulldog và không có trường hợp nào

chó tha mồi vàng, không có phương pháp tăng cường nào được thảo luận, từ SamplePairing đến AutoAugment đến GAN, sẽ tạo ra chó săn vàng. Tuy nhiên, một số dạng sai lệch như ánh sáng, độ che khuất, tỷ lệ, nền, v.v. có thể ngăn ngừa được hoặc ít nhất là giảm đáng kể nhờ Tăng cường dữ liệu. Việc trang bị quá mức thường không phải là vấn đề lớn đối với việc truy cập vào dữ liệu lớn. Tăng cường dữ liệu ngăn chặn việc trang bị quá mức bằng cách sửa đổi các tập dữ liệu hạn chế để có các đặc điểm của dữ liệu lớn.

Các từ viết tắt

GAN: mạng đối thủ tổng quát; CNN: mạng lưới thần kinh tích chập; DCGAN: mạng đối thủ tạo phức tạp sâu; NAS: tìm kiếm kiến trúc thần kinh; SRCNN: mạng nơ ron tích chập siêu phân giải; SRGAN: mạng đối thủ tạo ra siêu phân giải; CT: chụp cắt lớp vi tính; MRI: chụp cộng hưởng từ; PET: chụp cắt lớp phát xạ positron; ROS: lấy mẫu ngẫu nhiên; SMOTE: kỹ thuật lấy mẫu tổng hợp thiểu số; RGB: đỏ-xanh-xanh; PCA: phân tích thành phần chính; UCI: Đại học California Irvine; MNIST: Viện Tiêu chuẩn và Công nghệ Quốc gia Sứa đổi; CIFAR: Viện Nghiên cứu Cao cấp Canada; t-SNE: nhúng hàng xóm ngẫu nhiên phân phối t.

Sự nhìn nhận

Chúng tôi xin cảm ơn những người đánh giá trong Phòng thí nghiệm khai thác dữ liệu và học máy tại Đại học Florida Atlantic. Ngoài ra, chúng tôi nhận sự hỗ trợ một phần của NSF (CNS-1427536). Các ý kiến, phát hiện, kết luận hoặc khuyến nghị trong bài viết này chỉ thuộc về các tác giả và không phản ánh quan điểm của NSF.

Tác giả đóng góp

CS đã thực hiện việc xem xét và phân tích tài liệu cơ bản cho tác phẩm này, đồng thời soạn thảo bản thảo. TMK, JLL, RAB, RZ, KW, NS và RK đã làm việc với CS để phát triển khuôn khổ và trọng tâm của bài viết. TMK đã giới thiệu chủ đề này với CS và giúp hoàn thiện và hoàn thiện công việc này. Tất cả các tác giả đều đọc và phê duyệt bản thảo cuối cùng.

Kính phí

Không áp dụng được.

Sự sẵn có của dữ liệu và tài liệu

Không áp dụng được.

Lợi ích cạnh tranh

Nhiều tác giả tuyên bố rằng họ không có hứng thú với việc cạnh tranh.

Đồng ý xuất bản

Không áp dụng được.

Phê duyệt đạo đức và đồng ý tham gia

Không áp dụng được.

Đã nhận: ngày 9 tháng 1 năm 2019 Được chấp nhận: ngày 22 tháng 4 năm 2019
Published online: 06 July 2019

Người giới thiệu

1. Krizhevsky A, Sutskever I, Hinton GE. Phân loại ImageNet với mạng lưới thần kinh tích chập sâu. Hệ thống xử lý thông tin thần kinh Adv. 2012;25:1106-14.
2. Karen S, Andrew Z. Mạng tích chập rất sâu để nhận dạng hình ảnh quy mô lớn. Bản in điện tử arXiv. 2014.
3. Kaiming H, Xiangyu Z, Shaoqing R, Jian S. Học phân류 sâu để nhận dạng hình ảnh. Trong: CVPR, 2016.
4. Christian S, Vincent V, Sergey I, Jon S, Zbigniew W. Xem xét lại kiến trúc khởi đầu cho thị giác máy tính. Bản in điện tử arXiv, 2015.
5. Gao H, Zhuang L, Laurens M, Kilian QW. Mạng tích chập được kết nối dày đặc. Bản in trước arXiv. 2016.
6. Jan K, Vladimir G, Daniel C. Chính quy hóa cho học sâu: phân loại. Bản in trước arXiv. 2017.
7. Nitish S, Geoffrey H, Alex K, Ilya S, Ruslan S. Dropout: một cách đơn giản để ngăn chặn tình trạng mạng lưới thần kinh bị trang bị quá mức. J Mach Tim hiểu Res. 2014;15(1):1929-58.
8. Jonathan T, Ross G, Arjun J, Yann L, Christoph B. Bản địa hóa đối tượng hiệu quả bằng cách sử dụng mạng tích chập. Trong: CVPR'15. 2015.
9. Sergey I, Christian S. Chuẩn hóa hàng loạt: tăng tốc đào tạo mạng lưới sâu bằng cách giảm sự dịch chuyển hiệp phương sai nội bộ. Trong: ICML; 2015.
10. Karl W, Taghi MK, DingDing W. Khảo sát về học tập chuyển tiếp. J Dữ liệu lớn. 2016;3:9.
11. Shao L. Chuyển giao học tập để phân loại trực quan: một cuộc khảo sát. Hệ thống học tập mạng thần kinh IEEE Trans. 2015;26(5):1019-34.
12. Jia D, Wei D, Richard S, Li-Jia L, Kai L, Li FF. ImageNet: cơ sở dữ liệu hình ảnh phân cấp quy mô lớn. Trong: CVPR09, 2009.
13. Amir Z, Alexander S, William S, Leonidas G, Jitendra M, Silvio S. Nhiệm vụ: tháo gỡ nhiệm vụ học tập chuyển giao. Trong: CVPR '18. 2018.
14. Yosinski J, Clune J, Bengio Y, Lipson H. Các tính năng có thể chuyển đổi trong mạng lưới thần kinh sâu như thế nào? Hệ thống xử lý thông tin thần kinh Adv. 2014;27:3320-8.

15. Erhan D, Bengio Y, Courville A, Manzagol PA, Vincent P. Tại sao đào tạo trước không có giám sát lại giúp học sâu? J Mach Tùm hiểu Res. 2010;11:625–60.
16. Mark P, Dean P, Geoffrey H, Tom MM. Học tập không cần nỗ lực với mã đầu ra ngữ nghĩa. Trong: NIPS; 2009.
17. Yongqin X, Christoph HL, Bernt S, Zeynep A. Học tập không bắn—một đánh giá toàn diện về cái tốt, cái xấu và cái xấu. bản in trước arXiv. 2018.
18. Yaniv T, Ming Y, Marc' AR, Lior W. DeepFace: thu hẹp khoảng cách về hiệu suất ở cấp độ con người trong xác minh khuôn mặt. Trong: CVPR '14; 2014.
19. Mạng lưới thần kinh Gregory K, Richard Z, Ruslan S. Siamese để nhận dạng dạng hình ảnh một lần. Trong: Hội thảo Học sâu ICML; 2015.
20. Adam S, Sergey B, Matthew B, Dean W, Timothy L. Học một lần với mạng lưới thần kinh tăng cường trí nhớ. bản in trước arXiv. 2016.
21. Tomas M, Ilya S, Kai C, Greg C, Jeffrey D. Các cách biểu diễn phân tán của các từ và cụm từ cũng như thành phần của chúng. Được chấp nhận vào NIPS 2013.
22. Jeffrey P, Richard S, Christopher DM. GloVe: vector toàn cầu để biểu diễn từ. Trong: Ký yếu các phương pháp thực nghiệm trong xử lý ngôn ngữ tự nhiên (EMNLP 2014) 12. 2014.
23. Halevy A, Norvig P, Pereira F. Tính hiệu quả không hợp lý của dữ liệu. Hệ thống IEEE Intell. 2009;24:8–12.
24. Chen S, Abhinav S, Saurabh S, Abhinav G. Phản đối tính hiệu quả phi lý của dữ liệu trong kỹ nguyên học sâu. Trong: ICCV; 2017. tr. 843–52.
25. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Phân loại ung thư da ở cấp độ bác sĩ da liễu với mạng lưới thần kinh sâu. Thiên nhiên. 2017;542:115–8.
26. Geert L, Thijs K, Babak EB, Arnaud AAS, Francesco C, Mohsen G, Jeroen AWM, van Bram G, Clara IS. Một cuộc khảo sát về học sâu trong phân tích hình ảnh y tế. Hình ảnh Med hậu môn. 2017;42:60–88.
27. Joffrey LL, Taghi MK, Richard AB, Naeem S. Một cuộc khảo sát về giải quyết sự mất cân bằng cấp cao trong dữ liệu lớn. Dữ liệu lớn của Springer J. 2018;5:42.
28. LeCun Y, Bottou L, Bengio Y, Haffner P. Học tập dựa trên gradient được áp dụng để nhận dạng tài liệu. Proc IEEE. 1998;86(11):2278–324.
29. Nitesh VC, Kevin WB, Lawrence OH, Kegelmeyer W. SMOTE: kỹ thuật lấy mẫu quá mức tổng hợp thiểu số. J Artif Intellig Res. 2002;16:321–57.
30. Hui H, Wen-Yuan W, Bing-Huan M. Borderline-SMOTE: một phương pháp lấy mẫu quá mức mới trong việc học tập dữ liệu không cân bằng. Trong: Ký yếu của ICIC, tập. 3644, Ghi chú bài giảng về Khoa học Máy tính, New York. 2005, tr. 878–87.
31. Ian JG, Jean PA, Mehdi M, Bing X, David WF, Sherjil O, Aaron C, Yoshua B. Lưới tạo ra đối thủ. NIPS. 2014.
32. Leon AG, Alexander SE, Matthias B. Một thuật toán thần kinh về phong cách nghệ thuật. ArXiv. 2015.
33. Barret Z, Quốc VL. Tìm kiếm kiến trúc thần kinh với học tập tăng cường. Trong: Hội nghị quốc tế về đại diện học tập, 2017.
34. Tero K, Timo A, Samuli L, Jaakko L. Sự phát triển dần dần của GAN để cải thiện chất lượng, tính ổn định và sự biến đổi. CoRR, abs/1710.10196, 2017.
35. Justin J, Alexandre A, Li FF. Tồn thắt về nhận thức khi truyền kiểu thời gian thực và độ phân giải siêu cao. ECCV. 2016;2016:694–711.
36. Luis P, Jason W. Hiệu quả của việc tăng cường dữ liệu trong phân loại hình ảnh bằng cách sử dụng học sâu. Trong: Báo cáo nghiên cứu của Đại học Stanford, 2017.
37. Lemley J, Barzrafkan S, Corcoran P. Tăng cường thông minh học một chiến lược tăng cường dữ liệu tối ưu. Trong: Truy cập IEEE. 2017.
38. Ekin DC, Barret Z, Dandelion M, Vijay V, Quốc VL. AutoAugment: học các chính sách tăng cường từ dữ liệu. bản in trước ArXiv. 2018.
39. Xin Y, Paul SB, Ekta W. Mạng lưới đối thủ sáng tạo trong hình ảnh y tế: đánh giá. bản in trước arXiv. 2018.
40. Jelmer MW, Tim L, Max AV, Ivana I. Mạng đối nghịch sáng tạo để giảm tiếng ồn trong CT liều thấp. Trong: Giao dịch của IEEE về Hình ảnh Y tế. 2017.
41. Ohad S, Tammy RR. Hình ảnh cộng hưởng từ được tăng tốc bằng mạng lưới thần kinh đối nghịch. Trong: DLMIA/ML-CDS@ MICCAI, 2017.
42. Wang Y, Biting Y, Wang L, Chen Z, Lalush DS, Lin W, Xi W, Chu J, Shen D, Chu L. Mạng tạo đối thủ có điều kiện 3D để ước tính hình ảnh PET chất lượng cao ở liều thấp. Hình ảnh thần kinh. 2018;174:550–62.
43. Dwarakanath M, Behzad B. Phân đoạn mạch máu võng mạc bằng cách sử dụng bản đồ độ mặn cục bộ và mạng lưới thần kinh đối thủ tổng hợp để có độ phân giải siêu hình ảnh. bản in trước arXiv. 2017.
44. Francesco C, Aldo M, Claudio S, Giorgio T. Tăng cường dữ liệu y sinh bằng cách sử dụng mạng lưới thần kinh đối lập tổng quát. Trong: Hội nghị quốc tế về mạng lưới thần kinh nhân tạo. Berlin: Springer; 2017. Trang 626–34.
45. Camilo B, Andrew JP, Larry TD, Allen TN, Susan MR, Bennett AL. Học tập đa dạng MRI não tiềm ẩn với học sâu. Int Soc Opt Photonics. 2018;10574:105741.
46. Maria JMC, Sarfaraz H, Jeremy B, Ulas B. Làm thế nào để đánh lừa các bác sĩ X quang bằng mạng lưới tạo ra đối thủ? Một xét nghiệm thí giác để chẩn đoán ung thư phổi. bản in trước arXiv. 2017.
47. Baur C, Albarqouni S, Navab N. MelanoGANs: tổng hợp tổn thương da có độ phân giải cao với GAN. bản in trước arXiv. 2018.
48. Madani A, Moradi M, Karargyris A, Syeda-Mahmood T. Tạo X-quang ngược và tăng cường dữ liệu để phân loại bất thường về tim mạch. Trong: Hình ảnh y tế 2018. Xử lý hình ảnh 2018;10574:105741.
49. Maayan FA, Eyal K, Jacob G, Hayit G. Tăng cường dữ liệu dựa trên GAN để cải thiện việc phân loại tổn thương gan. bản in trước arXiv. 2018.
50. Joseph R, Santosh D, Ross G, Ali F. Bạn chỉ nhìn một lần: phát hiện đối tượng thống nhất, theo thời gian thực. Trong: CVPR'16. 2016.
51. Ross G, Jeff D, Trevor D, Jitendra M. Hệ thống phân cấp tính năng phong phú giúp phát hiện đối tượng chính xác và phân đoạn theo ngữ nghĩa. Trong: CVPR '14. 2014.
52. Ross G. R-CNN nhanh. CoRR, abs/1504.08083. 2015.
53. Shaoqing R, Kaiming H, Ross G, Jian S. R-CNN nhanh hơn: hướng tới phát hiện đối tượng theo thời gian thực với các mạng để xuất khu vực. Trong: NIPS, 2015.
54. Jonathan L, Evan S, Trevor D. Mạng tích chập hoàn toàn để phân đoạn theo ngữ nghĩa. CoRR, abs/1411.4038. 2014.

55. Olaf R, Philipp F, Thomas B. U-Net: mạng tích chập để phân đoạn hình ảnh y sinh. Trong: MICCAI. lò xo; 2015, tr. 234–41.
56. Hessam B, Maxwell H, Mohammad R, Ali F. Nhà máy lọc nhãnh: cải thiện việc phân loại mạng hình ảnh thông qua quá trình phát triển nhãnh. bản in trước arXiv. 2018.
57. Francisco JM-B, Fiammetta S, Jose MJ, Daniel U, Leonardo F. Sơ đồ điều chỉnh nhiễu chuyển tiếp để tăng cường dữ liệu. bản in trước arXiv. 2018.
58. Dua D, Karra TE. Kho lưu trữ máy học UCI [<http://archive.ics.uci.edu/ml>]. Irvine, CA: Đại học California, Trường Khoa học Thông tin và Máy tính; 2017.
59. Ken C, Karen S, Andrea V, Andrew Z. Sự trở lại của ma quỷ trong các chi tiết: đi sâu vào các mạng xoắn. Trong: Ký yếu của BMVC. 2014.
60. Mark E, Luc VG, Christopher KW, John W, Andrew Z. Thủ thách của các lớp đối tượng thi giác pascal (VOC). <http://www.pascal-network.org/challenges/VOC/voc2008/workshop/>. 2008.
61. Aranzazu J, Miguel P, Mikel G, Carlos LM, Daniel P. Một nghiên cứu so sánh về các khống gian màu khác nhau trong phân đoạn hình ảnh dựa trên phân cụm. IPMU; 2010.
62. Quanzeng Y, Jiebo L, Hailin J, Jianchao Y. Phân tích cảm xúc hình ảnh mạnh mẽ bằng cách sử dụng các mạng sâu được đào tạo dần dần và chuyển miền. Trong: AAAI. 2015, tr. 381–8.
63. Luke T, Geoff N. Cải thiện việc học sâu bằng cách tăng cường dữ liệu chung. bản in trước arXiv. 2017.
64. Guoliang K, Yuanyi D, Liang Z, Yi Y. Chính quy hóa PatchShuffle. bản in trước arXiv. 2017.
65. Hiroshi I. Tăng cường dữ liệu bằng cách ghép các mẫu để phân loại hình ảnh. Bản in điện tử ArXiv. 2018.
66. Cecilia S, Michael JD. Cải thiện việc tăng cường dữ liệu ví dụ hỗn hợp. Bản in trước ArXiv. 2018.
67. Daojun L, Feng Y, Tian Z, Peter Y. Tim hiểu các phương pháp luyện tập kết hợp. Trong: Truy cập IEEE. 2018. tr. 1.
68. Ryo T, Takashi M. Tăng cường dữ liệu bằng cách cắt xén hình ảnh ngẫu nhiên và các bản vá cho CNN sâu. bản in trước arXiv. 2018.
69. Yoshua B, Jerome L, Ronan C, Jason W. Chương trình giảng dạy. Trong: Ký yếu hội nghị quốc tế thường niên lần thứ 26 về học máy, ACM. 2009, tr. 41–8.
70. Zhun Z, Liang Z, Guoliang K, Shaozi L, Yi Y. Tăng cường xóa dữ liệu ngẫu nhiên. Bản in điện tử ArXiv. 2017.
71. Terrance V, Graham WT. Cải thiện việc chính quy hóa các mạng thần kinh tích chập có tính năng cắt bỏ. bản in trước arXiv. 2017.
72. Agnieszka M, Michal G. Tăng cường dữ liệu để cải thiện việc học sâu trong bài toán phân loại hình ảnh. Trong: Tiến sĩ liên ngành quốc tế IEEE 2018. Hội thảo, 2018.
73. Jonathan K, Michael S, Jia D, Li FF. Biểu diễn đối tượng 3D để phân loại chi tiết. Trong: Hội thảo lần thứ 4 của IEEE về Biểu diễn và Nhận dạng 3D, tại ICCV 2013 (3dRR-13). Sydney, Úc. Ngày 8 tháng 12 năm 2013.
74. Tomohiko K, Michiaki I. Icing on the cake: một phương pháp học sau dễ dàng và nhanh chóng mà bạn có thể thử sau khi học sâu. bản in trước arXiv. 2018.
75. Terrance V, Graham WT. Tăng cường tập dữ liệu trong không gian đặc trưng. Trong: Ký yếu hội nghị quốc tế về học máy (ICML), hội thảo, 2017.
76. Sebastien CW, Adam G, Victor S, Mark DM. Hiểu cách tăng cường dữ liệu để phân loại: khi nào nên làm cong vênh? CoRR, abs/1609.08764, 2016.
77. Seyed-Mohsen MD, Alhussein F, Pascal F. DeepFool: một phương pháp đơn giản và chính xác để đánh lừa các mạng lưới thần kinh sâu. bản in trước arXiv. 2016.
78. Jiawei S, Danilo VV, Sakurai K. Tấn công một pixel để đánh lừa mạng lưới thần kinh sâu. bản in trước arXiv. 2018.
79. Michal Z, Konrad Z, Negar R, Pedro OP. Đóng khung đối nghịch để phân loại hình ảnh và video. bản in trước arXiv. 2018.
80. Logan E, Brandon T, Dimitris T, Ludwig S, Aleksander M. Một phép quay và một bản dịch là đủ: đánh lừa CNN bằng các phép biến đổi đơn giản. Bản in trước ArXiv. 2018.
81. Goodfellow I, Shlens J, Szegedy C. Giải thích và khai thác các ví dụ đối nghịch. Hội nghị quốc tế về đại diện học tập, 2015.
82. Mang Ian JG, David WF, Mehdi M, Aaron C, Yoshua B. Maxout. bản in trước arXiv. 2013.
83. Shuangtao L, Yuanke C, Yanlin P, Lin B. Học các tính năng mạnh mẽ hơn thông qua luyện tập đối nghịch. Bản in trước của ArXiv. 2018.
84. Lingxi X, Jingdong W, Zhen W, Meng W, Qi T. DisturbLabel: chuẩn hóa CNN trên lớp mất mát. bản in trước arXiv. 2016.
85. Christopher B, Liang C, Ricardo GPB, Roger G, Alexander H, David AD, Maria VH, Joanna W, Daniel R. GAN Tăng cường: tăng cường dữ liệu đào tạo bằng cách sử dụng mạng đối thủ tổng quát. bản in trước arXiv. 2018.
86. Doersch C. Hướng dẫn về Bộ mã hóa tự động biến thiên. Bản in điện tử ArXiv. 2016.
87. Laurens M, Geoffrey H. Thực quan hóa dữ liệu bằng t-SNE. J Mach Tim hiểu Res. 2008;9:2431–56.
88. Jeff D, Philipp K, Trevor D. Học đặc điểm đối nghịch. Trong: CVPR'16. 2016.
89. Lin Z, Shi Y, Xue Z. IDSGAN: Mạng đối thủ sáng tạo để tạo ra cuộc tấn công chống lại phát hiện xâm nhập. bản in trước arXiv; 2018.
90. William F, Mihaela R, Balaji L, Andrew MD, Shakir M, Ian G. Nhiều con đường dẫn đến trạng thái cân bằng: GAN không cần giảm độ phân kỳ ở mỗi bước. Trong: Hội nghị quốc tế về biểu diễn học tập (ICLR); 2017.
91. Alec R, Luke M, Soumith C. Học biểu diễn không giám sát với các mạng đối thủ tạo ra tích chập sâu. ICLR, 2016.
92. Jun-Yan Z, Taesung P, Phillip I, Alexei AE. Dịch thuật từ hình ảnh sang hình ảnh không ghép đôi bằng cách sử dụng mạng đối thủ nhất quán theo chu kỳ. Trong: Hội nghị quốc tế về thị giác máy tính (ICCV), 2017.
93. Xinyue Z, Yifan L, Zengchang Q, Jiahong L. Phân loại cảm xúc bằng cách tăng cường dữ liệu bằng cách sử dụng mạng đối nghịch tổng hợp. CoRR, tập. abs/1711.00648. 2017.
94. Goodfellow IJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Cukierski W, Tang Y, Thaler D, Lee DH, et al. Những thách thức trong học tập biểu diễn: Báo cáo về ba cuộc thi học máy. Trong: NIPS. Berlin: Springer; 2013. tr.117–24.
95. Mehdi M, Simon O. Lưới tạo ra đối thủ có điều kiện. bản in trước arXiv. 2014.
96. Mario L, Karol K, Marcin M, Olivier B, Sylvain G. Các GAN có được tạo ra như nhau không? Một nghiên cứu quy mô lớn. bản in trước arXiv. 2018.

97. Swee KL, Yi L, Ngọc-Trung T, Ngai-Man C, Gemma R, Yuval E. DOPING: tăng cường dữ liệu tổng quát để phát hiện dị thường không giám sát bằng GAN. bản in trước arXiv. 2018.
98. Alireza M, Jonathon S, Navdeep J, Ian G, Brendan F. Bộ mã hóa tự động đối nghịch. bản in trước arXiv. 2015.
99. Tim S, Ian G, Wojciech Z, Vicki C, Alec R, Xi C. Cải tiến kỹ thuật đào tạo GAN. bản in trước arXiv. 2016.
100. Yanghao L, Naiyan W, Jiaying L, Xiaodi H. Chuyển giao phong cách thần kinh làm mờ đi. bản in trước arXiv. 2017.
101. Khizar H. Siêu phân giải thông qua học sâu. bản in trước arXiv. 2017.
102. Dmitry U, Andrea V, Victor L. Chuẩn hóa sơ thẩm: thành phần cần thiêu để cách điệu nhanh. bản in trước arXiv. 2016.
103. Philip TJ, Amir AA, Stephen B, Toby B, Boguslaw O. Tăng cường kiểu dáng: tăng cường dữ liệu thông qua ngẫu nhiên hóa kiểu dáng. bản in điện tử arXiv. 2018.
104. Josh T, Rachel F, Alex R, Jonas S, Wojciech Z, Pieter A. Chọn ngẫu nhiên miền để chuyển mạng lưới thần kinh sâu từ mô phỏng sang thế giới thực. bản in trước arXiv. 2017.
105. Ashish S, Tomas P, Oncel T, Josh S, Wenda W, Russ W. Học từ các hình ảnh mô phỏng và không được giám sát thông qua đào tạo đối nghịch. Trong: Hội thảo về thị giác máy tính và nhận dạng mẫu, 2017.
106. Stephan RR, Vibhav V, Stefan R, Vladlen K. Chơi để lấy dữ liệu: sự thật nền tảng từ trò chơi máy tính. Trong: Hội nghị châu Âu về thị giác máy tính (ECCV); 2016.
107. Brostow Gabriel J, Fauqueur Julien, Cipolla Roberto. Các lớp đối tượng ngữ nghĩa trong video: cơ sở dữ liệu thực tế có độ phân giải cao. Nhận dạng mẫu Lett. 2008;30(2):88–97.
108. Marius C, Mohamed O, Sebastian R, Timo R, Markus E, Rodrigo B, Uwe F, Stefan R, Bernt S. Bộ dữ liệu cảnh quan thành phố để hiểu cảnh quan đô thị theo ngữ nghĩa. Trong: CVPR; 2016.
109. Esteban R, Sherry M, Andrew S, Saurabh S, Yutaka LS, Jie T, Quốc VL, Alexey K. Sự phát triển quy mô lớn của các bộ phân loại hình ảnh. Trong: Kỷ yếu hội nghị quốc tế lần thứ 34 về học máy (ICML '17). 2017.
110. Esteban R, Alok A, Yanping H, Quốc VL. Tiết hóa thường xuyên cho tìm kiếm kiến trúc phân loại hình ảnh. bản in trước arXiv. 2018.
111. Các chiến lược của Tim S, Jonathan H, Xi C, Szymon S, Ilya S. Evolution như một giải pháp thay thế có thể mở rộng cho học tập tăng cường. bản in điện tử arXiv. 2017.
112. Horia M, Aurelia G, Benjamin R. Tìm kiếm ngẫu nhiên đơn giản cung cấp cách tiếp cận mang tính cạnh tranh cho việc học tăng cường. Trong: Những tiến bộ trong hệ thống xử lý thông tin thần kinh (NIPS); 2018.
113. David GL. Các tính năng hình ảnh đặc biệt từ các điểm chính bắt biến theo tỷ lệ. Máy tính Int J Vis. 2004;2004:91–110.
114. Navneet D, Bill T. Biểu đồ độ dốc định hướng để phát hiện con người. Trong: CVPR, 2005.
115. Sutton RS, Tăng cường AG. Học tập: giới thiệu. New York: Nhà xuất bản MIT; 1998.
116. Mingyang G, Kele X, Bo D, Huaimin W, Lei Z. Học các chính sách tăng cường dữ liệu bằng cách sử dụng tìm kiếm ngẫu nhiên tăng cường. bản in trước arXiv. 2018.
117. Trần NM, Mathieu S, Hoàng TL, Martin W. Tiền xử lý dữ liệu ảnh tự động với học tăng cường sâu. bản in trước arXiv. 2018.
118. Hochreiter S. Bài toán biến thiên gradient trong quá trình học mạng nơ-ron hồi quy và lời giải bài toán. Int J Uncertain Fuzzin Know Dýa Syst. 1998;6(02):107–16.
119. Jia S, Wang P, Jia P, Hu S. Nghiên cứu tăng cường dữ liệu để phân loại hình ảnh dựa trên mạng nơ ron tích chập. Trong: Đại hội tự động hóa Trung Quốc (CAC) 2017, 2017. p. 4165–70.
120. Ilija R, Piotr D, Ross G, Georgia G, Kaiming H. Chất lọc dữ liệu: hướng tới học tập được giám sát toàn diện. Trong: CVPR '18; 2018.
121. Guotai W, Michael A, Sebastian O, Wenqi L, Jan D, Tom V. Tăng cường thời gian thử nghiệm với ước tính độ không đảm bảo cho phân đoạn hình ảnh y tế dựa trên học sâu. OpenReview.net. 2018.
122. Fabio P, Christina V, Sandra A, Eduardo V. Tăng cường dữ liệu để phân tích tổn thương da. Trong: Hội thảo và thử thách phân tích hình ảnh da ISIC @ MICCAI 2018. 2018.
123. Karzuhisa M, Akira H, Akane M, Hiroshi K. Phân loại hình ảnh của khối u ác tính, nevus và dày sừng tiết bã theo quần thể mạng lưới thần kinh sâu. Trong: Thủ thách hợp tác hình ảnh da quốc tế (ISIC) 2017 tại hội nghị chuyên đề quốc tế về hình ảnh y sinh (ISBI). 2017.
124. Max J, Karen S, Andrea V, Andrew Z. Dữ liệu tổng hợp và mạng lưới thần kinh nhân tạo để nhận dạng văn bản cảnh tự nhiên. bản in trước arXiv. 2014.
125. Florian S, Dmitry K, James P. FaceNet: một phương pháp nhúng thông nhất để nhận dạng khuôn mặt và phân cụm. Trong: CVPR '15. 2015.
126. Xudong M, Qing L, Haoran X, Raymond YKL, Zhen W, Stephen PS. Mạng đối thủ sinh ra bình phương tối thiểu. Trong: Hội nghị quốc tế về thị giác máy tính (ICCV), 2017.
127. Ren W, Shengen Y, Yi S, Qingqing D, Gang S. Hình ảnh sâu: mở rộng quy mô nhận dạng hình ảnh. CoRR, abs/1501.02876, 2015.
128. Chao D, Chen CL, Kaiming H, Ziaouo T. Tìm hiểu mạng tích chập sâu cho hình ảnh siêu phân giải. Trong: ECCV. Berlin: Springer; 2014., tr. 184–99.
129. Christian L, Lucas T, Ferenc H, Jose C, Andrew C, Alejandro A, Andrew A, Alykhan T, Johannes T, Zehan W, Wenzhe S. Siêu phân giải hình ảnh đơn thực tế bằng cách sử dụng mạng đối lập tổng quát. bản in trước arXiv. 2016.
130. Han Z, Tao X, Hongsheng L, Shaoting Z, Xiaogang W, Xiaolei H, Dimitris M. StackGAN: tổng hợp văn bản thành hình ảnh thực tế với các mạng đối nghịch tổng hợp chồng lên nhau. Trong: ICCV, 2017.
131. Trishul C, Yutaka S, Johnson A, Karthik K. Projekt adam: xây dựng một hệ thống đào tạo deep learning hiệu quả và có thể mở rộng. Trong: Kỷ yếu của OSDI. 2014. Trang 571–82.
132. Buda Mateusz, Maki Atsuto, Mazurowski Maciej A. Một nghiên cứu có hệ thống về văn đề mắt cân bằng lớp trong mạng lưới thần kinh tích chập. Mạng lưới thần kinh. 2018;106:249–59.
133. Drown DJ, Khoshgoftaar TM, Seliya N. Lấy mẫu tiền hóa và mô hình hóa chất lượng phần mềm của các hệ thống có độ đảm bảo cao. Hệ thống chuyển đổi IEEE. 2009;39(5):1097–107.
134. Jason Y, Jeff C, Anh N, Thomas F, Hod L. Tìm hiểu mạng lưới thần kinh thông qua hình dung sâu sắc. Trong: Hội nghị châu Âu về thị giác máy tính (ECCV). Berlin: Springer; 2015. tr. 818–33.
135. Xiaofeng Z, Zhangyang W, Dong L, Qing L. DADA: tăng cường dữ liệu đối nghịch sâu để phân loại chế độ dữ liệu cực thấp. bản in trước arXiv. 2018.
136. Martin A, Paul B, Jianmin C, Zhifeng C, Andy D, Jeffrey D, Matthieu D, Sanjay G, Geoffrey I, Michael I, Manjunath K, Josh L, Rajat M, Sherry M, Derek GM, Benoit S, Paul T, Vijay V, Pete W, Matrin W, Yuan Y, Xiaoqiang Z. TensorFlow:

một hệ thống học máy quy mô lớn. Trong: Kỷ yếu hội nghị chuyên đề USENIX lần thứ 12 về thiết kế và triển khai hệ điều hành (OSDI '16), 2016.

137. Máy ảnh <https://keras.io/>. 2015.
138. Alexander B, Alex P, Eugene K, Vladimir II, Alexandr AK. Albumentations: tăng cường hình ảnh nhanh chóng và linh hoạt. Bản in trước của ArXiv. 2018.
139. Maayan FA, Idit D, Eyal K, Michal A, Jacob G, Hayit G. Tăng cường hình ảnh y tế tổng hợp dựa trên GAN cho tăng hiệu suất CNN trong phân loại tổn thương gan. bản in trước arXiv. 2018.
140. Changhee H, Hideaki H, Leonardo R, Ryosuke A, Wataru S, Shinichi M, Yujiro F, Giancarlo M, Hideki N. Tạo hình ảnh mr não tổng hợp dựa trên GAN. Trong: Hội nghị chuyên đề quốc tế lần thứ 15 của IEEE về hình ảnh y sinh năm 2018 (ISBI 2018). IEEE, 2011. P. 734–8.

Ghi chú của nhà xuất bản

Springer Nature vẫn trung lập đối với các yêu cầu về quyền tài phán trong các bản đồ được xuất bản và các tổ chức liên kết.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com