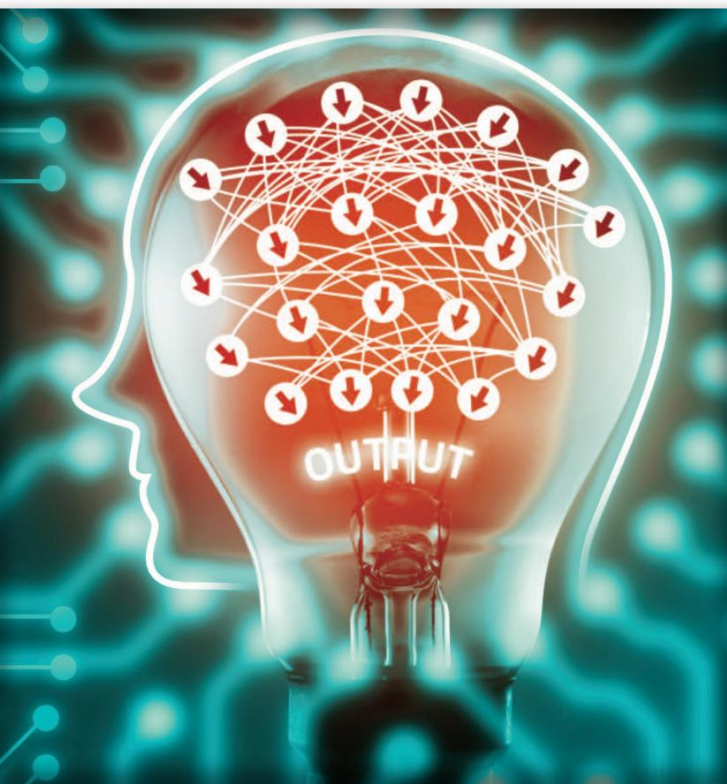


Antonia Creswell, Tom White, Vincent Dumoulin,
Kai Arulkumaran, Biswa Sengupta và Anil A. Bharath

Mạng đối kháng sinh sản

Tổng quan



©Istockphoto.com/zapp2photo

Mạng đối kháng tạo sinh (GAN) cung cấp một cách để

dữ liệu đào tạo. Họ đạt được điều này bằng cách lấy tín hiệu truyền ngược thông qua một quá trình cạnh tranh liên quan đến một cặp mạng. Các biểu diễn có thể được học bởi GAN có thể được sử dụng trong nhiều ứng dụng khác nhau, bao gồm tổng hợp hình ảnh, chỉnh sửa hình ảnh nghệ thuật, chuyển giao kiểu, siêu phân giải hình ảnh và phân loại. Mục đích của bài đánh giá này là cung cấp tổng quan về GAN cho cộng đồng xử lý tín hiệu, dựa trên các phép loại suy và khái niệm quen thuộc khi có thể. Ngoài việc xác định các phương pháp khác nhau để đào tạo và xây dựng GAN, chúng tôi cũng chỉ ra những thách thức còn lại trong lý thuyết và ứng dụng của chúng.

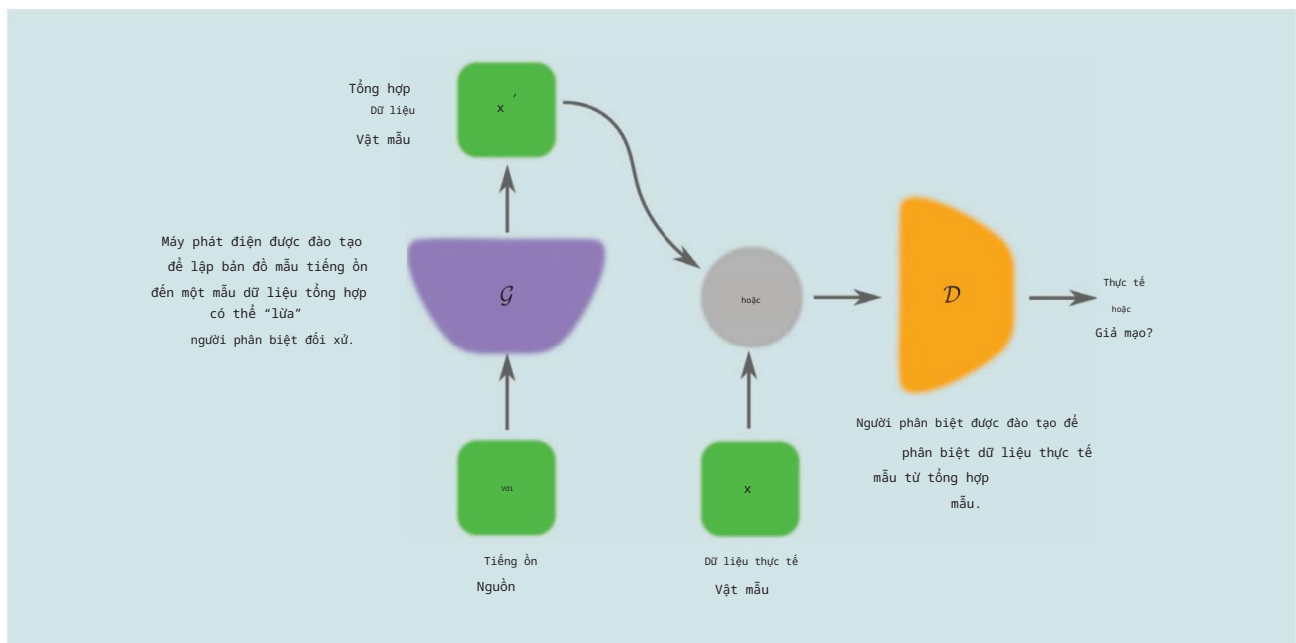
Giới thiệu

GAN là một kỹ thuật mới nổi cho cả học bán giám sát và không giám sát. Chúng đạt được điều này thông qua việc mô hình hóa ngầm các phân phối dữ liệu có chiều cao. Được đề xuất vào năm 2014 [1], chúng có thể được đặc trưng bằng cách đào tạo một cặp mạng cạnh tranh với nhau. Một phép loại suy phổ biến, phù hợp với dữ liệu trực quan, là coi một mạng như một kẻ làm giả nghệ thuật và mạng kia như một chuyên gia nghệ thuật. Kẻ làm giả, được biết đến trong tài liệu GAN là trình tạo, G , tạo ra các tác phẩm giả mạo, với mục đích tạo ra những hình ảnh chân thực. Chuyên gia, được gọi là người phân biệt, D , nhận cả tác phẩm giả mạo và hình ảnh thực (xác thực), và có mục đích phân biệt chúng (xem Hình 1). Cả hai đều được đào tạo đồng thời và cạnh tranh với nhau.

Điều quan trọng là trình tạo không có quyền truy cập trực tiếp vào hình ảnh thực tế—cách duy nhất để nó học là thông qua tương tác với bộ phân biệt. Bộ phân biệt có thể truy cập vào cả mẫu tổng hợp và mẫu được lấy từ chồng hình ảnh thực.

Tín hiệu lỗi cho bộ phân biệt được cung cấp thông qua sự thật cơ bản đơn giản là biết hình ảnh đến từ ngân xếp thực hay từ máy phát. Cùng một tín hiệu lỗi, thông qua bộ phân biệt, có thể được sử dụng để đào tạo máy phát, dẫn đến khả năng tạo ra hàng giả có chất lượng tốt hơn.

Các mạng biểu diễn máy phát điện và bộ phân biệt thường được triển khai bằng các mạng nhiều lớp bao gồm



Hình 1. Hai mô hình được học trong quá trình đào tạo cho GAN là bộ phân biệt \hat{D} và bộ tạo \hat{G} . Chúng thường được triển khai bằng mạng nơ-ron, nhưng chúng có thể được triển khai bằng bất kỳ dạng hệ thống phân biệt nào ánh xạ dữ liệu từ không gian này sang không gian khác; xem văn bản bài viết để biết chi tiết.

của các lớp tích chập và/hoặc kết nối đầy đủ. Các mạng máy phát và phân biệt phải có khả năng phân biệt, mặc dù chúng không nhất thiết phải có khả năng đảo ngược trực tiếp. Nếu ta coi mạng máy phát là ánh xạ từ một không gian biểu diễn nào đó, được gọi là không gian tiềm ẩn, đến không gian dữ liệu (chúng ta sẽ tập trung vào hình ảnh), thì chúng ta có thể biểu thị điều này một cách chính thức hơn là $G: \mathcal{Z} \rightarrow \mathcal{X}$, trong đó \mathcal{Z} là một không gian tiềm ẩn \mathcal{R} , \mathcal{X} là một không gian dữ liệu và z là một hình ảnh và x là một hình ảnh và \cdot biểu thị số của các kích thước.

Trong GAN cơ bản, mạng phân biệt, D , có thể được mô tả tương tự như một hàm ánh xạ từ dữ liệu hình ảnh sang xác suất hình ảnh đến từ phân phối dữ liệu thực, thay vì phân phối máy phát: $D: (\mathcal{X} \cup \mathcal{X}) \rightarrow [0, 1]$. Đối với máy phát cố định, G , bộ phân biệt, D , có thể được đào tạo để phân loại hình ảnh thành hình ảnh từ dữ liệu đào tạo (thực, gần bằng một) hoặc từ máy phát cố định (giả, gần bằng không). Khi bộ phân biệt là tối ưu, nó có thể bị đóng băng và bộ tạo, G , có thể tiếp tục được đào tạo để giảm độ chính xác của bộ phân biệt. Nếu phân phối bộ tạo có thể khớp hoàn hảo với phân phối dữ liệu thực, thì bộ phân biệt sẽ bị nhầm lẫn tối đa, dự đoán 0,5 cho tất cả các đầu vào. Trong thực tế, bộ phân biệt có thể không được đào tạo cho đến khi nó tối ưu; chúng ta sẽ khám phá quá trình đào tạo sâu hơn trong phần "Đào tạo GAN".

Ngoài các vấn đề học thuật thú vị liên quan đến việc đào tạo và xây dựng GAN, động lực đằng sau việc đào tạo GAN có thể không nhất thiết là trình tạo hoặc trình phân biệt: các biểu diễn được thể hiện bởi bất kỳ cặp mạng nào có thể được sử dụng trong nhiều nhiệm vụ tiếp theo. Chúng tôi khám phá các ứng dụng của các biểu diễn này trong phần "Ứng dụng của GAN".

Chuẩn bị

Thuật ngữ

Các mô hình sinh học cách nắm bắt phân phối thống kê của dữ liệu đào tạo, cho phép chúng ta tổng hợp các mẫu từ phân phối đã học. Ngoài việc tổng hợp các mẫu dữ liệu mới, có thể được sử dụng cho các tác vụ hạ nguồn như chỉnh sửa hình ảnh ngữ nghĩa [2], tăng cường dữ liệu [3] và chuyển giao phong cách [4], chúng tôi cũng quan tâm đến việc sử dụng các biểu diễn mà các mô hình như vậy học được cho các tác vụ như phân loại [5] và truy xuất hình ảnh [6].

Đôi khi chúng tôi đề cập đến các lớp tích chập và kết nối hoàn toàn của các mạng sâu; đây là những khái quát của perceptron hoặc các ngân hàng bộ lọc không gian với hậu xử lý phi tuyến tính. Trong mọi trường hợp, trọng số mạng được học thông qua truyền ngược [7].

Ký hiệu

Tài liệu GAN thường đề cập đến các vectơ đa chiều và thường biểu diễn các vectơ trong không gian xác suất bằng chữ in nghiêng (ví dụ, không gian tiềm ẩn là z). Trong lĩnh vực xử lý tín hiệu, người ta thường biểu diễn các vectơ bằng các ký hiệu in đậm, thường và chúng tôi áp dụng quy ước này để nhấn mạnh bản chất đa chiều của các biến. Theo đó, chúng tôi thường gọi $p_{data}(x)$ là biểu diễn hàm mật độ xác suất trên \mathcal{X} . Chúng tôi sử dụng x để biểu thị một vectơ ngẫu nhiên x nằm trong \mathcal{X} phân phối của các x . Chúng tôi sẽ sử dụng $p_G(x)$ để biểu thị vectơ được tạo ra bởi mạng lưới máy phát của GAN. Chúng tôi sử dụng các ký hiệu thư pháp G và D .

Để biểu thị mạng máy phát và mạng phân biệt, tương ứng. Cả hai mạng đều có các bộ tham số (trọng số), θ_G và θ_D , được học thông qua quá trình tối ưu hóa trong quá trình đào tạo.

Như với tất cả các hệ thống học sâu, việc đào tạo đòi hỏi chúng ta phải có một số hàm mục tiêu rõ ràng. Theo ký hiệu thông thường, chúng ta sử dụng $J_G(\theta_G; H, D)$ và $J_D(\theta_D; H, G)$ để tham chiếu đến hàm mục tiêu của bộ tạo và bộ phân biệt tương ứng. Việc lựa chọn ký hiệu nhắc nhở chúng ta rằng theo một nghĩa nào đó, hai hàm mục tiêu phụ thuộc lẫn nhau vào các tập tham số đang phát triển H và D của các mạng khi chúng được cập nhật theo từng lần lặp. Chúng ta sẽ khám phá thêm về điều này trong phần "Đào tạo GAN". Cuối cùng, lưu ý rằng các gradient đa chiều được sử dụng trong các bản cập nhật; chúng ta sử dụng DHG để biểu thị toán tử gradient đối với các trọng số của các tham số bộ tạo và DHD

Để biểu thị toán tử gradient liên quan đến trọng số của bộ phân biệt. Các gradient dự kiến được biểu thị bằng ký hiệu E_D .

Thu thập dữ liệu phân phối

Một vấn đề cốt lõi của xử lý tín hiệu và thống kê là ước tính mật độ: thu được một biểu diễn-ngầm định hoặc rõ ràng, tham số hoặc không tham số-của dữ liệu trong thế giới thực. Đây là động lực chính đằng sau GAN. Trong tài liệu về GAN, thuật ngữ phân phối tạo dữ liệu thường được sử dụng để chỉ mật độ xác suất cơ bản hoặc hàm khối lượng xác suất của dữ liệu quan sát. GAN học thông qua việc tính toán ngầm một số loại điểm tương đồng giữa phân phối của mô hình ứng viên và phân phối tương ứng với dữ liệu thực (xem Hình 2).

Tại sao lại bận tâm đến ước tính mật độ? Câu trả lời nằm ở trung tâm của-có thể tranh luận-nhiều vấn đề về suy luận trực quan, bao gồm phân loại hình ảnh, phát hiện và nhận dạng đối tượng trực quan, theo dõi đối tượng và đăng ký ký đối tượng. Về nguyên tắc, thông qua định lý Bayes, tất cả các vấn đề suy luận của thị giác máy tính đều có thể được giải quyết thông qua việc ước tính các hàm mật độ có điều kiện, có thể gián tiếp dưới dạng một mô hình học phân phối chung của các biến quan tâm và dữ liệu quan sát được. Khó khăn mà chúng ta gặp phải là các hàm khả năng cho dữ liệu hình ảnh thực tế, có chiều cao rất khó xây dựng. Mặc dù GAN không cung cấp rõ ràng cách đánh giá các hàm mật độ, nhưng đối với một bộ phân biệt tạo

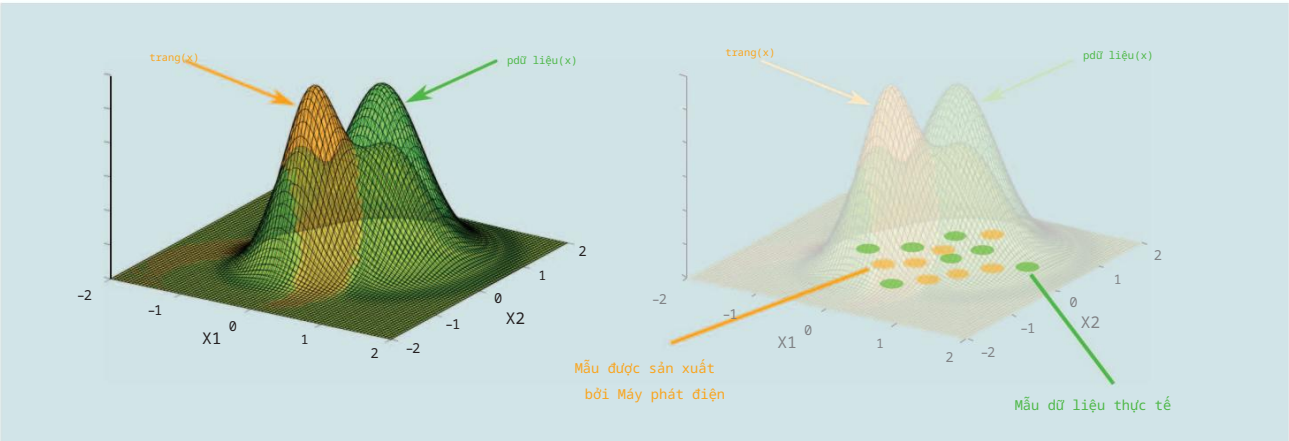
cấp dung lượng phù hợp, máy phát điện sẽ ngấm nắm bắt sự phân phối của dữ liệu.

Công việc liên quan

Người ta có thể xem các nguyên tắc của mô hình sinh bằng cách so sánh với các kỹ thuật chuẩn trong xử lý tín hiệu và phân tích dữ liệu. Ví dụ, xử lý tín hiệu sử dụng rộng rãi ý tưởng biểu diễn tín hiệu dưới dạng tổ hợp có trọng số của các hàm cơ sở. Các hàm cơ sở cố định là cơ sở của các kỹ thuật chuẩn như biểu diễn dựa trên Fourier và biểu diễn wavelet. Các phương pháp tiếp cận dựa trên dữ liệu để xây dựng các hàm cơ sở có thể bắt nguồn từ phép biến đổi Hotelling [8], bắt nguồn từ quan sát của Pearson rằng các thành phần chính giảm thiểu lỗi tái tạo theo tiêu chí lỗi bình phương tối thiểu. Mặc dù được sử dụng rộng rãi, phân tích thành phần chính chuẩn (PCA) không có mô hình thống kê rõ ràng cho dữ liệu quan sát được, mặc dù đã chỉ ra rằng các cơ sở của PCA có thể được suy ra như một bài toán ước tính tham số có khả năng xảy ra tối đa.

Mặc dù được áp dụng rộng rãi, PCA vẫn có hạn chế - các hàm cơ sở xuất hiện dưới dạng các vectơ riêng của ma trận hiệp phương sai trên các quan sát của dữ liệu đầu vào và phép ánh xạ từ không gian biểu diễn trở lại không gian tín hiệu hoặc hình ảnh là tuyến tính. Vì vậy, chúng ta có cả ánh xạ nông và ánh xạ tuyến tính, hạn chế độ phức tạp của mô hình và do đó, của dữ liệu có thể được biểu diễn.

Phân tích thành phần độc lập (ICA) cung cấp một cấp độ tinh vi hơn nữa, trong đó các thành phần tín hiệu không còn cần phải trực giao nữa; các hệ số trộn được sử dụng để trộn các thành phần với nhau để xây dựng các ví dụ về dữ liệu chỉ được coi là độc lập về mặt thống kê. ICA có nhiều công thức khác nhau về các hàm mục tiêu được sử dụng trong quá trình ước tính các thành phần tín hiệu hoặc trong mô hình sinh ra thể hiện cách các tín hiệu hoặc hình ảnh được tạo ra từ các thành phần đó. Một cải tiến gần đây được khám phá thông qua ICA là ước tính tương phản nhiễu (NCE); điều này có thể được coi là tiếp cận tinh thần của GAN [9]: hàm mục tiêu để học các thành phần độc lập so sánh một thống kê được áp dụng



Hình 2. Trong quá trình đào tạo GAN, trình tạo được khuyến khích tạo ra một phân phối mẫu, $p_G(\mathbf{x})$ để khớp với dữ liệu thực, $p_D(\mathbf{x})$. Đối với GAN được tham số hóa và đào tạo phù hợp, các phân phối này sẽ gần như giống hệt nhau. Các biểu diễn được thể hiện bởi GAN được ghi lại trong các tham số đã học (trọng số) của mạng trình tạo và phân biệt.

để làm nhiều với nhiều do mô hình tạo ứng viên tạo ra [10]. Phương pháp NCE ban đầu không bao gồm các bản cập nhật cho trình tạo.

Có thể thực hiện những so sánh nào khác giữa GAN và các công cụ chuẩn của xử lý tín hiệu? Đối với các biểu diễn PCA, ICA, Fou-rier và wavelet, không gian tiềm ẩn của GAN, theo phép tương tự, là không gian hệ số của cái mà chúng ta thường gọi là không gian biến đổi. Điều khiển GAN khác biệt so với các công cụ chuẩn của xử lý tín hiệu này là mức độ phức tạp của các mô hình ánh xạ các vectơ từ không gian tiềm ẩn sang không gian ảnh. Vì các mạng máy phát chứa các tính phi tuyến tính và có thể có độ sâu gần như tùy ý, nên việc ánh xạ này—cũng như nhiều phương pháp học sâu khác—có thể cực kỳ phức tạp.

Đối với các mô hình dựa trên hình ảnh sâu, các phương pháp tiếp cận hiện đại đối với mô hình hóa hình ảnh tạo ra có thể được nhóm thành các mô hình mật độ rõ ràng và ngầm định. Các mô hình mật độ rõ ràng có thể xử lý được (các mô hình thay đổi biến, các mô hình tự hồi quy) hoặc không thể xử lý được (các mô hình có hướng được đào tạo bằng suy luận biến phân, các mô hình không có hướng được đào tạo bằng chuỗi Markov). Ngầm

mô hình mật độ nắm bắt phân phối thống kê của dữ liệu thông qua một quá trình tạo ra sử dụng lấy mẫu tổ tiên [11] hoặc lấy mẫu dựa trên chuỗi Markov. GAN thuộc loại mô hình ngầm có hướng. Tổng quan chi tiết hơn và các bài báo có liên quan có thể được tìm thấy trong [12].

Kiến trúc GAN

GAN được kết nối đầy đủ

Kiến trúc GAN đầu tiên sử dụng mạng nơ-ron được kết nối đầy đủ cho cả bộ tạo và bộ phân biệt [1]. Kiểu kiến trúc này được áp dụng cho các tập dữ liệu hình ảnh tương đối đơn giản: MNIST (chữ số viết tay), CIFAR-10 (hình ảnh tự nhiên) và Tập dữ liệu khuôn mặt Toronto (TFD).

GAN tích chập

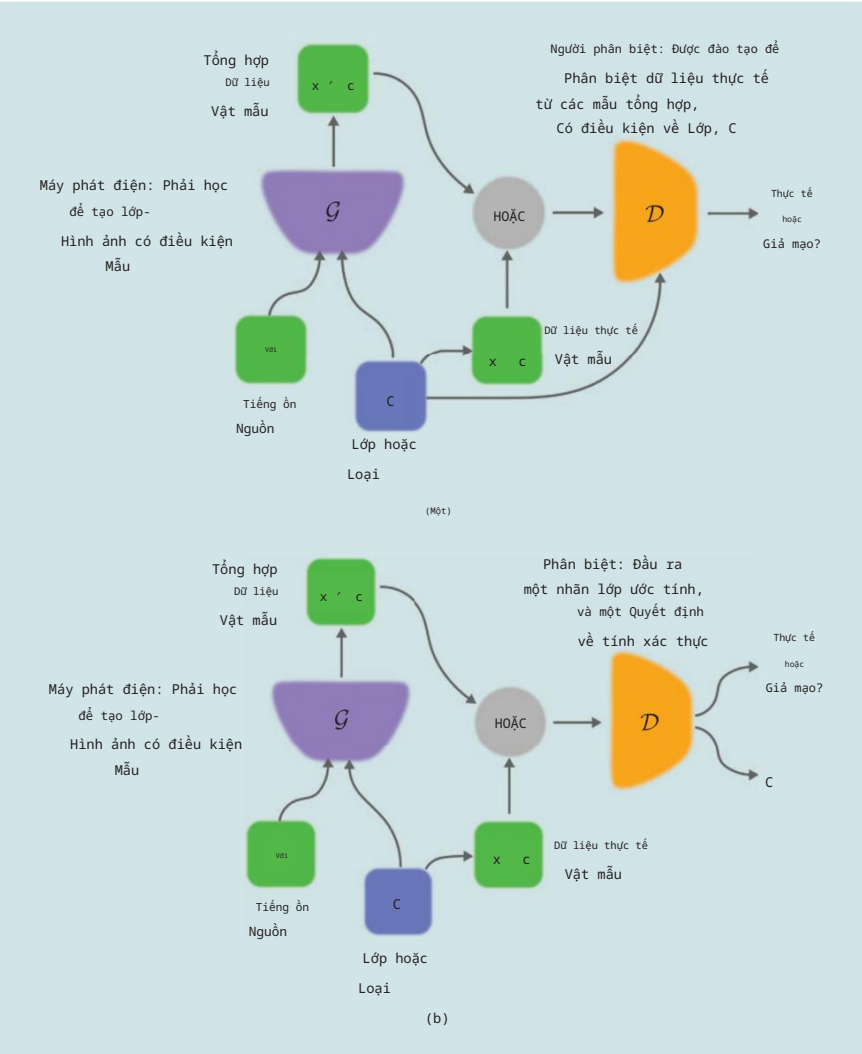
Chuyển từ mạng nơ-ron tích chập (CNN) hoàn toàn kết nối là một sự mở rộng tự nhiên, vì CNN cực kỳ phù hợp với dữ liệu hình ảnh. Các thí nghiệm ban đầu được tiến hành trên CIFAR-10 cho thấy rằng việc đào tạo mạng máy phát và mạng phân biệt khó khăn hơn khi sử dụng CNN có cùng mức dung lượng và sức mạnh biểu diễn như những mạng được sử dụng cho học có giám sát.

Kim tự tháp Laplacian của mạng đối nghịch (LAPGAN) [13] đưa ra một giải pháp cho vấn đề này, bằng cách phân tích quá trình tạo ra bằng nhiều thang đo: một hình ảnh chân thực được phân tích thành một pyr-amid Laplacian và một GAN tích chập có điều kiện được đào tạo để tạo ra từng lớp dựa trên lớp ở trên.

Ngoài ra, Radford et al. [5] đã đề xuất một họ kiến trúc mạng được gọi là GAN tích chập sâu

(DCGAN), cho phép đào tạo một cặp mạng máy phát tích chập sâu và mạng phân biệt. DCGAN sử dụng tích chập có bước và có bước phân đoạn, cho phép học các toán tử hạ mẫu không gian và tăng mẫu trong quá trình đào tạo. Các toán tử này xử lý sự thay đổi về tốc độ lấy mẫu và vị trí, một yêu cầu chính trong việc ánh xạ từ không gian hình ảnh sang không gian tiềm ẩn có thể có chiều thấp hơn và từ không gian hình ảnh sang bộ phân biệt. Các chi tiết hơn về kiến trúc DCGAN và đào tạo được trình bày trong phần “Các mẹo đào tạo”.

Là một phần mở rộng để tổng hợp hình ảnh trong hai chiều, Wu et al. [14] đã trình bày GAN có khả năng



Hình 3. (a) GAN có điều kiện, do Mirza và cộng sự [15] đề xuất, thực hiện tổng hợp hình ảnh có điều kiện theo lớp; bộ phân biệt thực hiện phân biệt có điều kiện theo lớp giữa hình ảnh thật và hình ảnh giả. (b) Một khác, InfoGAN [16] có mạng phân biệt cũng ước tính nhãn lớp.

tổng hợp các mẫu dữ liệu ba chiều (3-D) bằng cách sử dụng phép tích chập thể tích. Wu et al. [14] đã tổng hợp các vật thể mới bao gồm ghế, bàn và ô tô; ngoài ra, họ cũng trình bày một phương pháp để ánh xạ từ hình ảnh hai chiều (2-D) sang phiên bản 3-D của các vật thể được mô tả trong các hình ảnh đó.

GAN có điều kiện

Mirza et al. [15] đã mở rộng khuôn khổ GAN (2-D) thành bối cảnh có điều kiện bằng cách biến cả mạng máy phát và mạng phân biệt thành có điều kiện lớp (Hình 3). GAN có điều kiện có lợi thế là có thể cung cấp các biểu diễn tốt hơn cho việc tạo dữ liệu đa phương thức. Có thể rút ra sự tương đồng giữa GAN có điều kiện và InfoGAN [16], phân tích nguồn nhiễu thành một nguồn không nén được và một "mã tiềm ẩn", cố gắng khám phá các yếu tố biến đổi tiềm ẩn bằng cách tối đa hóa thông tin chung giữa mã tiềm ẩn và đầu ra của máy phát. Mã tiềm ẩn này có thể được sử dụng để khám phá các lớp đối tượng theo cách hoàn toàn không có giám sát, mặc dù không nhất thiết mã tiềm ẩn phải là mã phân loại. Các biểu diễn mà InfoGAN học được có vẻ có ý nghĩa về mặt ngữ nghĩa, xử lý các yếu tố phức tạp đan xen trong hình ảnh, bao gồm các biến thể về tư thế, ánh sáng và nội dung cảm xúc của hình ảnh khuôn mặt [16].

GAN với các mô hình suy luận

Trong công thức ban đầu của mình, GAN không có cách nào để ánh xạ một quan sát nhất định, x , thành một vectơ trong không gian tiềm ẩn—trong tài liệu về GAN, điều này thường được gọi là cơ chế suy luận. Một số kỹ thuật đã được đề xuất để đảo ngược bộ tạo GAN được đào tạo trước [17], [18]. Các đề xuất độc lập

suy luận học được theo hướng đối nghịch (ALI) [19] và GAN song hướng (BiGAN) [20] cung cấp các phần mở rộng đơn giản nhưng hiệu quả, giới thiệu một mạng suy luận trong đó các bộ phân biệt kiểm tra các cặp chung (dữ liệu, tiềm ẩn).

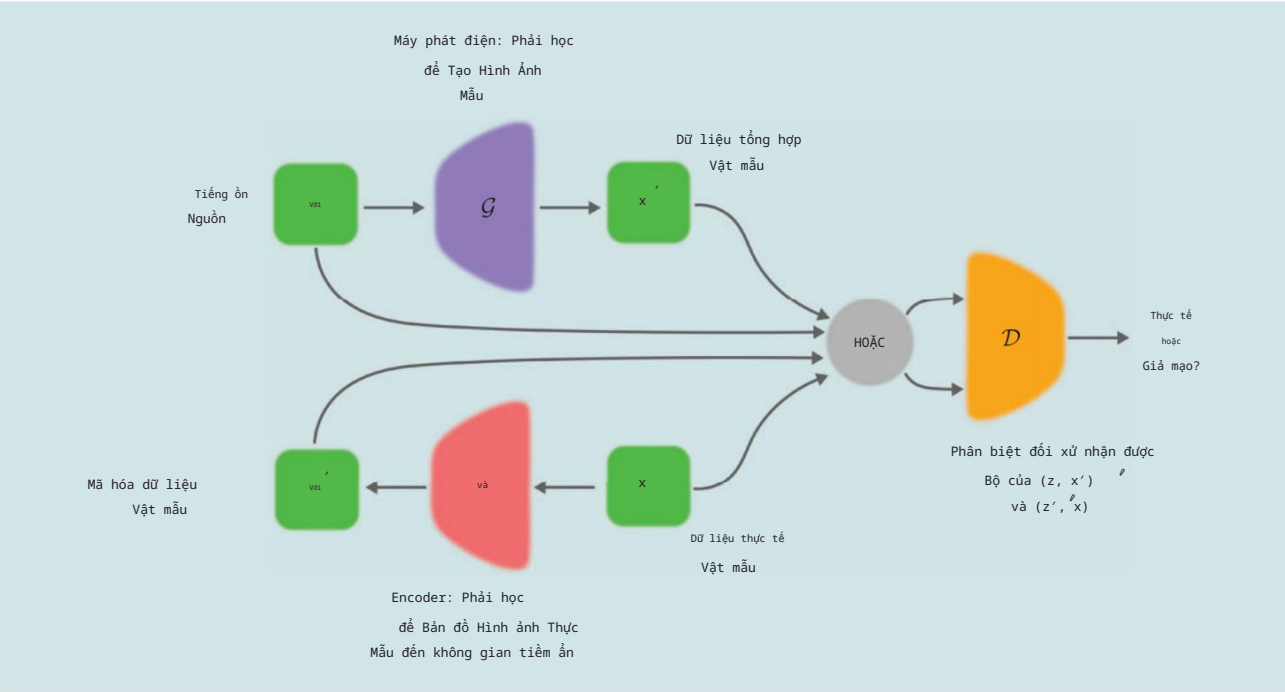
Trong công thức này, bộ tạo bao gồm hai mạng: "bộ mã hóa" (mạng suy luận) và "bộ giải mã". Chúng được đào tạo chung để đánh lừa bộ phân biệt. Bản thân bộ phân biệt nhận được các cặp vectơ (z, x_z) (xem Hình 4) và phải xác định cặp nào tạo thành một bộ chính hãng bao gồm mẫu ảnh thực và mã hóa của nó, hoặc một mẫu ảnh giả và đầu vào không gian tiềm ẩn tương ứng cho bộ tạo.

Về mặt lý tưởng, trong mô hình mã hóa-giải mã, đầu ra, được gọi là tái tạo, phải tương tự như đầu vào. Thông thường, độ trung thực của các mẫu dữ liệu tái tạo được tổng hợp bằng ALI/BiGAN là kém. Độ trung thực của các mẫu có thể được cải thiện bằng chi phí đối nghịch bổ sung về phân phối các mẫu dữ liệu và tái tạo của chúng [21].

Bộ mã hóa tự động đối nghịch

Autoencoders là mạng lưới, bao gồm một bộ mã hóa và giải mã, học cách ánh xạ dữ liệu thành một biểu diễn tiềm ẩn bên trong và ra ngoài một lần nữa. Nghĩa là, chúng học một phép ánh xạ xác định (thông qua bộ mã hóa) từ một không gian dữ liệu, ví dụ, hình ảnh, thành một không gian tiềm ẩn hoặc biểu diễn, và một phép ánh xạ (thông qua bộ giải mã) từ không gian tiềm ẩn trở lại không gian dữ liệu. Sự kết hợp của hai phép ánh xạ này dẫn đến một sự tái tạo, và hai phép ánh xạ được đào tạo sao cho một hình ảnh được tái tạo càng gần với hình ảnh gốc càng tốt.

Bộ mã hóa tự động gợi nhớ đến các bộ lọc tái tạo hoàn hảo được sử dụng rộng rãi trong xử lý hình ảnh và tín hiệu.



Hình 4. Cấu trúc ALI/BiGAN [19], [20] bao gồm ba mạng. Một trong số này đóng vai trò là bộ phân biệt, một mạng khác ánh xạ các vectơ nhiễu từ không gian tiềm ẩn sang không gian hình ảnh (bộ giải mã, được mô tả là bộ tạo G trong hình), với mạng cuối cùng (bộ mã hóa, được mô tả là E) ánh xạ từ không gian hình ảnh sang không gian tiềm ẩn.

Tuy nhiên, các bộ mã hóa tự động thường học các ánh xạ phi tuyến tính theo cả hai hướng. Hơn nữa, khi được triển khai với các mạng lưới sâu, các kiến trúc có thể được sử dụng để triển khai các bộ mã hóa tự động có tính linh hoạt đáng kể. Việc đào tạo có thể không được giám sát, với việc truyền ngược được áp dụng giữa hình ảnh được tái tạo và hình ảnh gốc để học các tham số của cả bộ mã hóa và bộ giải mã.

Như đã đề xuất trước đó, người ta thường muốn không gian tiềm ẩn có một tổ chức hữu ích. Ngoài ra, người ta có thể muốn thực hiện lấy mẫu tổ tiên, truyền thẳng [11] từ một bộ mã hóa tự động. Đào tạo đối nghịch cung cấp một lộ trình để đạt được hai mục tiêu này. Cụ thể, đào tạo đối nghịch có thể được áp dụng giữa không gian tiềm ẩn và phân phối trước mong muốn trên không gian tiềm ẩn (GAN không gian tiềm ẩn). Điều này dẫn đến một hàm mất kết hợp [22] phản ánh cả lỗi tái tạo và thước đo mức độ khác biệt của phân phối trước so với phân phối do mạng mã hóa ứng viên tạo ra. Cách tiếp cận này tương tự như một bộ mã hóa tự động biến thiên (VAE) [23] trong đó GAN không gian tiềm ẩn đóng vai trò là Kullback-

Thuật ngữ phân kỳ Leibler (KL) của hàm mất mát. Mescheder et al. [24] đã thống nhất VAE với đào tạo đối nghịch dưới dạng khuôn khổ Bayes biến thiên đối nghịch (AVB). Những ý tưởng tương tự đã được trình bày trong [12]. AVB cố gắng tối ưu hóa cùng một tiêu chí như của VAE, nhưng sử dụng mục tiêu đào tạo đối nghịch thay vì sự phân kỳ KL.

Đào tạo GAN

Giới thiệu

Quá trình đào tạo GAN bao gồm cả việc tìm các tham số của bộ phân biệt để tối đa hóa độ chính xác phân loại của nó và tìm các tham số của bộ tạo để tối đa hóa độ chính xác phân loại.

hợp nhất bộ phân biệt. Quá trình đào tạo này được tóm tắt trong Hình 5.

Chi phí đào tạo được đánh giá bằng cách sử dụng hàm giá trị, V^h h GD, phụ thuộc vào cả bộ tạo và bộ phân biệt. Đào tạo liên quan đến việc giải quyết

$$\max_{\theta_G} \min_{\theta_D} V^h(\theta_G, \theta_D)$$

ở đâu

$$V^h(\theta_G, \theta_D) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))]$$

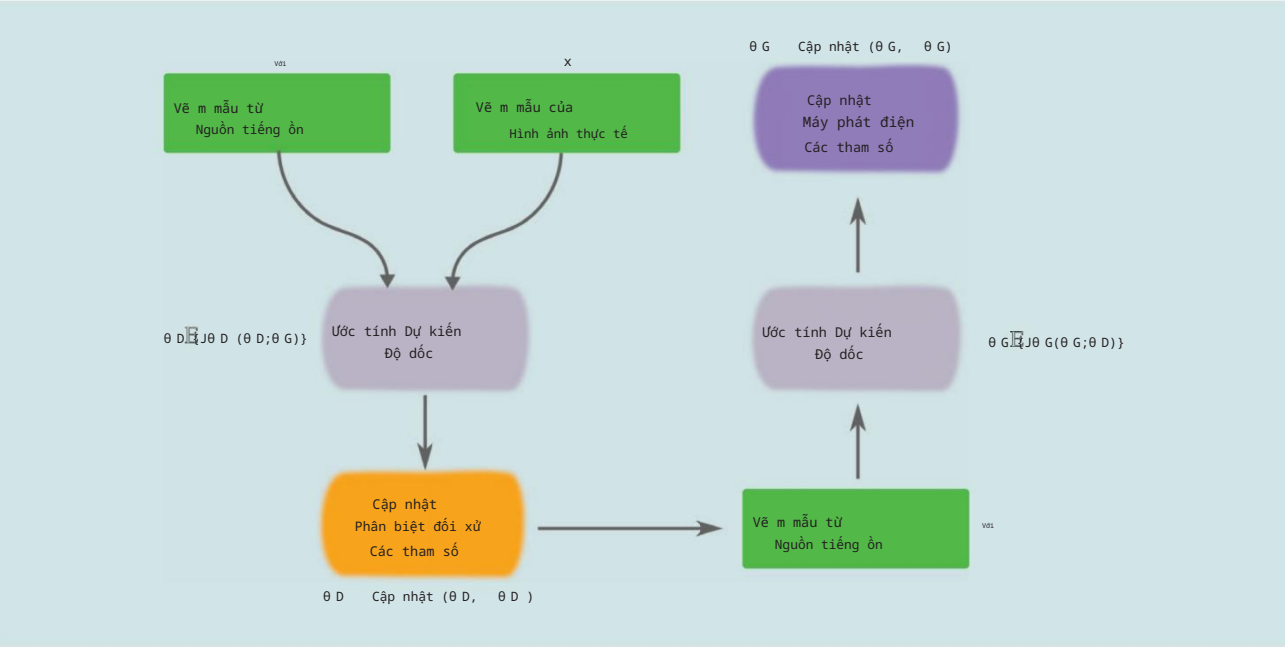
Trong quá trình đào tạo, các tham số của mô hình được cập nhật, trong khi các tham số của mô hình kia được cố định. Goodfellow et al. [1] chỉ ra rằng, đối với một máy phát điện cố định, có một bộ phân biệt tối ưu D^* duy nhất, $D^*(x) = \frac{p(x)}{p(x) + p_G(x)}$. x p dữ liệu Ho cũng chỉ ra rằng máy phát điện, G , là tối ưu khi dữ liệu (x), D^* = $\frac{p(x)}{p(x) + p_G(x)}$ / tương đương với bộ phân biệt tối

$$p_G(x) = \frac{p(x)}{2}$$

nator dự đoán 0,5 cho tất cả các mẫu được lấy từ x . Nói cách khác, máy phát điện là tối ưu khi bộ phân biệt, D , bị nhầm lẫn tối đa và không thể phân biệt được các mẫu thật với các mẫu giả.

Về mặt lý tưởng, bộ phân biệt được đào tạo cho đến khi tối ưu đối với máy phát điện hiện tại; sau đó máy phát điện được cập nhật lại. Tuy nhiên, trên thực tế, bộ phân biệt có thể không được đào tạo cho đến khi tối ưu mà chỉ có thể được đào tạo cho một số lượng nhỏ các lần lặp lại và máy phát điện được cập nhật đồng thời với bộ phân biệt. Hơn nữa, một tiêu chí đào tạo thay thế, không bão hòa thường được sử dụng cho máy phát điện, sử dụng $\max(1 - G \circ D(G(z)))$ thay vì $\min(G \log(1 - D(G(z))))$.

Mặc dù có sự tồn tại về mặt lý thuyết của các giải pháp duy nhất, việc đào tạo GAN vẫn còn nhiều thách thức và thường không ổn định vì một số lý do [5], [25], [26]. Một cách tiếp cận để cải thiện đào tạo GAN là



Hình 5. Vòng lặp chính của quá trình đào tạo GAN. Các mẫu dữ liệu mới, x_1 , có thể được rút ra bằng cách truyền các mẫu ngẫu nhiên, z , qua mạng máy phát. Độ dốc của bộ phân biệt có thể được cập nhật k lần trước khi cập nhật máy phát.

đánh giá các “triệu chứng” thực nghiệm có thể gặp phải trong quá trình đào tạo. Các triệu chứng này bao gồm:

- khó khăn trong việc hội tụ cấp mô hình [5]
- mô hình sinh sản “sụp đổ” để tạo ra các mẫu rất giống nhau cho các đầu vào khác nhau [25]
- tổn thất phân biệt hội tụ nhanh chóng về 0 [26], không cung cấp đường dẫn đáng tin cậy nào để cập nhật độ dốc cho máy phát điện.

Một số tác giả đã đề xuất các phương pháp tiếp cận theo kinh nghiệm để giải quyết những vấn đề này [1], [25]; những vấn đề này sẽ được thảo luận trong phần tiếp theo.

Những nỗ lực ban đầu để giải thích lý do tại sao đào tạo GAN không ổn định được đề xuất bởi Goodfellow và Salimans et al. [1], [25], những người đã quan sát thấy rằng các phương pháp giảm dần độ dốc thường được sử dụng để cập nhật cả tham số của bộ tạo và bộ phân biệt là không phù hợp khi giải pháp cho vấn đề tối ưu hóa do đào tạo GAN đặt ra thực sự cấu thành một điểm yên ngựa. Salimans et al. đã cung cấp một ví dụ đơn giản cho thấy điều này [25]. Tuy nhiên, giảm dần độ dốc ngẫu nhiên thường được sử dụng để cập nhật mạng nơ-ron và có các môi trường lập trình máy học được phát triển tốt giúp dễ dàng xây dựng và cập nhật mạng bằng cách sử dụng giảm dần độ dốc ngẫu nhiên.

Mặc dù một phương pháp xử lý lý thuyết ban đầu [1] cho thấy dữ liệu máy phát điện là tối ưu khi $p_g(x) = p_r(x)$, một kết quả rất gọn gàng với trực giác cơ bản mạnh mẽ, các mẫu dữ liệu thực nằm trên một đa tạp nằm trong không gian có nhiều chiều của các biểu diễn có thể. Ví dụ, nếu các mẫu ảnh màu là \mathbb{R}^3 , kích thước NN $\# \# 3$ giá trị pixel $[0, 255]^3$, không gian có thể được biểu diễn—mà chúng ta có thể gọi là X —có chiều với mỗi chiều lấy các giá trị từ 0 đến 255^3 , cường độ pixel có thể đo được lớn nhất. Tuy nhiên, các mẫu dữ liệu trong hỗ trợ của pdata tạo thành đa tạp của dữ liệu thực liên quan đến một số vấn đề cụ thể, thường chiếm một phần rất nhỏ trong tổng không gian, X . Tương tự như vậy, các mẫu do máy phát điện tạo ra cũng chỉ nên chiếm một phần nhỏ của X .

Arjovsky et al. [26] đã chỉ ra rằng hỗ trợ $p_g(x)$ và $p_{data}(x)$ nằm trong không gian có chiều thấp hơn không gian tương ứng với X . Hậu quả của điều này là $p_g(x)$ và $p_{data}(x)$ có thể không có sự chồng chéo, và do đó tồn tại một bộ phân biệt gần như tầm thường có khả năng phân biệt các mẫu thật, $p_{data}(x) \neq p_g(x)$ với các mẫu giả, $x \sim p_g(x)$ với độ chính xác 100%. Trong trường hợp này, lỗi phân biệt nhanh chóng hội tụ về 0. Các tham số của máy phát điện chỉ có thể được cập nhật thông qua bộ phân biệt, vì vậy khi điều này xảy ra, các gradient được sử dụng để cập nhật các tham số của máy phát điện cũng hội tụ về 0 và có thể không còn hữu ích cho các bản cập nhật cho máy phát điện nữa. Các giải thích của Arjovsky et al. giải thích cho một số triệu chứng liên quan đến đào tạo GAN [26].

Goodfellow et al. [1] cũng chỉ ra rằng khi D là tối ưu, việc đào tạo G tương đương với việc giảm thiểu sự phân kỳ Jensen-Shannon (JS) giữa $p_g(x)$ và $p_{data}(x)$. Nếu D không tối ưu mal, bản cập nhật có thể ít có ý nghĩa hoặc không chính xác. Điều này ý tưởng thách thức

hiểu biết lý thuyết đã thúc đẩy nghiên cứu về các hàm chi phí dựa trên khoảng cách thay thế. Một số trong số này được khám phá trong phần “Công thức thay thế”.

Mẹo huấn luyện

Một trong những cải tiến lớn đầu tiên trong quá trình đào tạo GAN để tạo hình ảnh là kiến trúc DCGAN do Radford và cộng sự đề xuất [5]. Công trình này là kết quả của quá trình khám phá sâu rộng các kiến trúc CNN trước đây được sử dụng trong thị giác máy tính và tạo ra một bộ hướng dẫn để xây dựng và đào tạo cả bộ tạo và bộ phân biệt. Trong phần “GAN tích chập”, chúng tôi đã ám chỉ đến tầm quan trọng của tích chập có bước và có bước phân số [27], đây là các thành phần chính của thiết kế kiến trúc. Điều này cho phép cả bộ tạo và bộ phân biệt học các hoạt động lấy mẫu lên và lấy mẫu xuống tốt, có thể góp phần cải thiện chất lượng tổng hợp hình ảnh. Cụ thể hơn đối với đào tạo, chuẩn hóa theo lô [28] được khuyến nghị sử dụng trong cả hai mạng để ổn định quá trình đào tạo trong các mô hình sâu hơn.

Một đề xuất khác là giảm thiểu số lượng các lớp được kết nối hoàn toàn được sử dụng để tăng tính khả thi của việc đào tạo các mô hình sâu hơn. Cuối cùng, Radford et al. [5] đã chỉ ra rằng việc sử dụng các hàm kích hoạt đơn vị tuyến tính chính lưu rò rỉ (ReLU) giữa các lớp trung gian của bộ phân biệt mang lại hiệu suất vượt trội hơn so với việc sử dụng ReLU thông thường.

Sau đó, Salimans et al. [25] đề xuất thêm các phương pháp tiếp cận theo phương pháp heuristic để ổn định quá trình đào tạo GAN. Đầu tiên, phương pháp so khớp đặc điểm, thay đổi mục tiêu của trình tạo một chút để tăng lượng thông tin có sẵn. Cụ thể, bộ phân biệt vẫn được đào tạo để phân biệt giữa các mẫu thật và giả, nhưng trình tạo hiện được đào tạo để so khớp các kích hoạt trung gian dự kiến (các đặc điểm) của mẫu giả của bộ phân biệt với các kích hoạt trung gian dự kiến của các mẫu thật. Thứ hai, phương pháp phân biệt minibatch, thêm một đầu vào bổ sung cho bộ phân biệt, đây là một đặc điểm mã hóa khoảng cách giữa một mẫu nhất định trong một minibatch và các mẫu khác. Điều này nhằm mục đích ngăn chặn sự sụp đổ chế độ, vì bộ phân biệt có thể dễ dàng biết được liệu trình tạo có tạo ra cùng một đầu ra hay không.

Một thủ thuật thứ ba, trung bình hóa theo phương pháp heuristic, phạt các tham số mạng nếu chúng lệch khỏi giá trị trung bình đang chạy của các giá trị trước đó, có thể giúp hội tụ đến trạng thái cân bằng. Thủ thuật thứ tư, chuẩn hóa lô ảo, làm giảm sự phụ thuộc của một mẫu vào các mẫu khác trong lô nhỏ bằng cách tính toán số liệu thống kê lô để chuẩn hóa với mẫu được đặt trong lô nhỏ tham chiếu được cố định khi bắt đầu đào tạo.

Cuối cùng, việc làm mịn nhấn một chiều làm cho mục tiêu của bộ phân biệt là 0,9 thay vì một, làm mịn ranh giới phân loại của bộ phân biệt, do đó ngăn chặn bộ phân biệt quá tự tin sẽ cung cấp các gradient yếu cho bộ tạo. Sønderby et al. [29] đã đưa ra

bộ phân biệt bằng cách thêm nhiễu vào các mẫu trước khi đưa chúng vào bộ phân biệt. Sønderby et al. [29] lập luận rằng việc làm mịn nhân một chiều làm sai lệch bộ phân biệt tối ưu, trong khi kỹ thuật của họ, nhiễu trường hợp, di chuyển các đa tạp của các mẫu thực và giả lại gần nhau hơn, đồng thời ngăn cản bộ phân biệt dễ dàng tìm thấy ranh giới phân biệt tách biệt hoàn toàn các mẫu thực và giả. Trong thực tế, điều này có thể được thực hiện bằng cách thêm nhiễu Gauss vào cả hình ảnh tổng hợp và thực, ù độ lệch chuẩn theo thời gian. Quá trình tương tự đã được Arjovsky et al. [26] đề xuất độc lập.

Các công thức thay thế

Phần đầu tiên của phần này xem xét các diễn giải và khái quát hóa thông tin-lý thuyết khác của GAN. Phần thứ hai xem xét các hàm chi phí thay thế nhằm mục đích giải quyết trực tiếp vấn đề về độ dốc biến mất.

Tổng quát hóa của hàm chi phí GAN

Nowozin et al. [30] đã chỉ ra rằng đào tạo GAN có thể được tổng quát hóa để giảm thiểu không chỉ sự phân kỳ JS mà còn là ước tính của f-divergences; chúng được gọi là f-GANs. Các f-diver-gences bao gồm các biện pháp phân kỳ nổi tiếng như KL-divergence. Nowozin và cộng sự đã chỉ ra rằng f-divergence có thể được xấp xỉ bằng cách áp dụng các liên hợp Fenchel của f-divergence mong muốn cho các mẫu được rút ra từ phân phối của các mẫu được tạo ra, sau khi truyền các mẫu đó qua một bộ phân biệt [30]. Họ cung cấp một danh sách các liên hợp Fenchel cho các f-divergences thường được sử dụng, cũng như các hàm kích hoạt có thể được sử dụng trong lớp cuối cùng của mạng máy phát, tùy thuộc vào sự lựa chọn f-divergence. Sau khi suy ra các hàm chi phí tổng quát để đào tạo máy phát và bộ phân biệt của f-GAN, Nowozin và cộng sự [30] nhận thấy rằng, ở dạng thô, việc tối đa hóa mục tiêu của máy phát điện có thể dẫn đến các gradient yếu, đặc biệt là khi bắt đầu đào tạo và đề xuất một hàm chi phí thay thế để cập nhật máy phát điện, ít có khả năng bão hòa hơn khi bắt đầu đào tạo.

Nowozin và cộng sự đề xuất rằng khi bộ phân biệt được đào tạo, đạo hàm của f-divergence theo tỷ lệ phân phối dữ liệu thực và giả sẽ được ước tính, trong khi khi bộ tạo được đào tạo thì chỉ có ước tính của f-divergence được giảm thiểu. Uehara et al. [31] mở rộng f-GAN hơn nữa, trong đó ở bước phân biệt, tỷ lệ phân phối dữ liệu thực và giả được dự đoán và ở bước tạo ra, độ phân kỳ f được giảm thiểu trực tiếp. Các phương án thay thế cho độ phân kỳ JS cũng được Goodfellow [12] đề cập.

Các hàm chi phí thay thế để ngăn chặn sự biến mất của gradient Arjovsky et al. [32] đã đề xuất Wasserstein GAN (WGAN), một GAN có hàm chi phí thay thế được lấy từ phép xấp xỉ khoảng cách Wasserstein. Không giống như hàm chi phí GAN ban đầu, WGAN có nhiều khả năng cung cấp các gradient hữu ích cho việc cập nhật trình tạo. Chi phí

Điều gì làm cho GANs khác biệt với các công cụ xử lý tín hiệu tiêu chuẩn này là mức độ phức tạp của các mô hình ánh xạ các vectơ từ không gian tiềm ẩn sang không gian ảnh.

hàm được suy ra cho WGAN dựa vào bộ phân biệt, mà họ gọi là bộ phê bình, là một hàm liên tục k-Lipschitz; trên thực tế, điều này có thể được thực hiện bằng cách chỉ cần cắt các tham số của bộ phân biệt. Tuy nhiên, nghiên cứu gần đây hơn [33] cho thấy rằng việc cắt trọng số làm giảm đáng kể khả năng của mô hình bộ phân biệt, buộc nó phải học các hàm đơn giản hơn. Gulrajani et al. [33] đã đề xuất một phương pháp cải tiến để đào tạo bộ phân biệt cho WGAN, bằng cách phạt chuẩn mực của các gradient bộ phân biệt đối với các mẫu dữ liệu trong quá trình đào tạo, thay vì thực hiện cắt tham số.

So sánh ngắn gọn các biến thể GAN

GAN cho phép chúng ta tổng hợp các mẫu dữ liệu mới từ nhiều ngẫu nhiên, nhưng chúng được coi là khó đào tạo một phần do gradient biến mất. Tất cả các mô hình GAN mà chúng ta đã thảo luận trong bài viết này đều yêu cầu điều chỉnh siêu tham số cẩn thận và lựa chọn mô hình để đào tạo. Tuy nhiên, có lẽ các mô hình dễ đào tạo hơn là bộ mã hóa tự động đối nghịch (AAE) và WGAN.

AAE tương đối dễ đào tạo vì tổn thất đối nghịch được áp dụng cho một phân phối khá đơn giản ở các chiều thấp hơn (so với dữ liệu hình ảnh). WGAN [33] được thiết kế để dễ đào tạo hơn, sử dụng một công thức khác của mục tiêu đào tạo không gặp phải vấn đề về độ dốc biến mất. WGAN cũng có thể được đào tạo thành công ngay cả khi không có chuẩn hóa theo lô; nó cũng ít nhạy cảm hơn với sự lựa chọn các giá trị phi tuyến tính được sử dụng giữa các lớp tích chập.

Các mẫu được tổng hợp bằng GAN hoặc WGAN có thể thuộc bất kỳ lớp nào có trong dữ liệu đào tạo. GAN có điều kiện cung cấp một phương pháp tổng hợp các mẫu có nội dung do người dùng chỉ định.

Rõ ràng từ nhiều kỹ thuật trực quan hóa khác nhau (Hình 6) rằng tổ chức của không gian tiềm ẩn ẩn chứa một số ý nghĩa, nhưng GAN gốc không cung cấp mô hình suy luận để cho phép các mẫu dữ liệu được ánh xạ thành các biểu diễn tiềm ẩn. Cả BiGAN và ALI đều cung cấp một cơ chế để ánh xạ dữ liệu hình ảnh vào không gian tiềm ẩn (suy luận), tuy nhiên, chất lượng tái tạo cho thấy rằng chúng không nhất thiết phải mã hóa và giải mã mẫu một cách trung thực. Một sự phát triển rất gần đây cho thấy ALI có thể khôi phục các mẫu dữ liệu được mã hóa một cách trung thực [21]. Tuy nhiên, mô hình này có nhiều điểm chung với AVB và AAE. Đây là các bộ mã hóa tự động, tương tự như VAE, trong đó không gian tiềm ẩn được chính quy hóa bằng cách sử dụng đào tạo đối nghịch thay vì sự phân kỳ KL giữa các mẫu được mã hóa và trước đó.

Cấu trúc của không gian tiềm ẩn

GAN xây dựng các biểu diễn riêng của chúng về dữ liệu mà chúng được đào tạo và khi làm như vậy, tạo ra các không gian vectơ hình học có cấu trúc cho các miền khác nhau. Đây là một chất lượng được chia sẻ với các mô hình mạng nơ-ron khác, bao gồm VAE [23], cũng như các mô hình ngôn ngữ như word2vec [34]. Nhìn chung, miền dữ liệu được mô hình hóa được ánh xạ vào một không gian vectơ,

có ít chiều hơn không gian dữ liệu, buộc mô hình phải khám phá cấu trúc thú vị trong dữ liệu và biểu diễn nó một cách hiệu quả. Không gian tiềm ẩn này nằm ở đầu “gốc” của mạng máy phát, và dữ liệu ở cấp độ biểu diễn này (không gian tiềm ẩn) có thể được cấu trúc cao và có thể hỗ trợ các hoạt động ngữ nghĩa cấp cao [5]. Các ví dụ bao gồm phép quay khuôn mặt từ quỹ đạo qua không gian tiềm ẩn, cũng như các phép loại suy hình ảnh có tác dụng thêm các thuộc tính trực quan như kính mắt vào khuôn mặt “trần trụi”.

Tất cả các mô hình GAN (vanilla) đều có một trình tạo ánh xạ dữ liệu từ không gian tiềm ẩn vào không gian cần mô hình hóa, nhưng nhiều mô hình GAN có bộ mã hóa hỗ trợ thêm ánh xạ ngược [19], [20]. Đây trở thành một phương pháp mạnh mẽ để khám phá và sử dụng không gian tiềm ẩn có cấu trúc của mạng GAN. Với bộ mã hóa, các tập hợp hình ảnh được gắn nhãn có thể được ánh xạ vào không gian tiềm ẩn và được phân tích để khám phá “các vectơ khái niệm” biểu thị các thuộc tính cấp cao như “cười” hoặc “đội mũ”. Các vectơ này có thể được áp dụng ở các độ lệch tỷ lệ trong không gian tiềm ẩn để ảnh hưởng đến hành vi của trình tạo (Hình 6). Tương tự như việc sử dụng quy trình mã hóa để mô hình hóa phân phối các mẫu tiềm ẩn, Gurumurthy et al. [35] đề xuất mô hình hóa không gian tiềm ẩn dưới dạng hỗn hợp các Gauss-ian và tìm hiểu các thành phần hỗn hợp giúp tối đa hóa khả năng tạo ra các mẫu dữ liệu theo phân phối tạo dữ liệu.

Ứng dụng của GAN

Khám phá các ứng dụng mới cho việc đào tạo đối kháng của mạng lưới sâu là một lĩnh vực nghiên cứu đang được triển khai. Chúng tôi xem xét một số ứng dụng thị giác máy tính đã xuất hiện trong tài liệu và sau đó được tinh chỉnh. Các ứng dụng này được chọn để làm nổi bật một số cách tiếp cận khác nhau khi sử dụng các biểu diễn dựa trên GAN để thao tác, phân tích hoặc mô tả hình ảnh và không phản ánh đầy đủ phạm vi ứng dụng tiềm năng của GAN.

Việc sử dụng GAN để phân loại hình ảnh đặt chúng vào bối cảnh rộng hơn của máy học và cung cấp đánh giá định lượng hữu ích về các tính năng được trích xuất trong không siêu

GANs xây dựng riêng của họ biểu diễn dữ liệu mà chúng được đào tạo và bằng cách đó tạo ra các không gian vectơ hình học có cấu trúc cho các miền khác nhau.

học vised. Tổng hợp hình ảnh vẫn là một khả năng cốt lõi của GAN và đặc biệt hữu ích khi hình ảnh được tạo ra có thể phải tuân theo các ràng buộc đã tồn tại từ trước. Siêu phân giải [36]–[38] đưa ra một ví dụ về cách tiếp cận hiện có có thể được bổ sung bằng thành phần mất mát đối nghịch để đạt được kết quả chất lượng cao hơn. Cuối cùng, phép dịch hình ảnh sang hình ảnh chứng minh cách GAN cung cấp giải pháp mục đích chung cho một nhóm tác vụ yêu cầu tự động chuyển đổi hình ảnh đầu vào thành hình ảnh đầu ra.

Phân loại và hồi quy

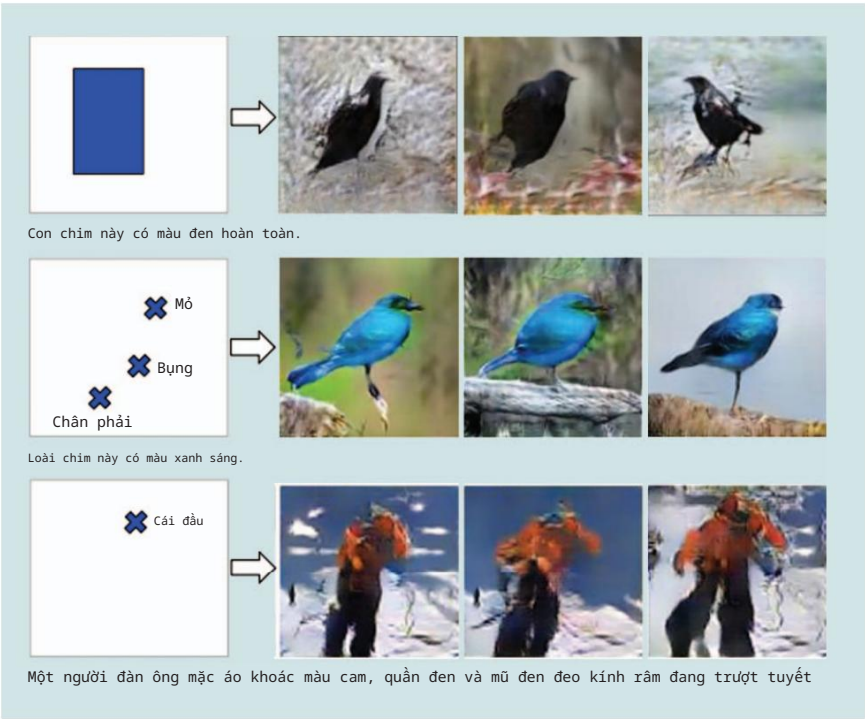
Sau khi hoàn tất quá trình đào tạo GAN, mạng nơ-ron có thể được sử dụng lại cho các tác vụ hạ lưu khác. Ví dụ, đầu ra của các lớp tích chập của bộ phân biệt có thể được sử dụng làm bộ trích xuất tính năng, với các mô hình tuyến tính đơn giản được lắp trên các tính năng này bằng cách sử dụng một lượng nhỏ các cặp (hình ảnh, nhãn) [5], [25]. Chất lượng của các biểu diễn không giám sát trong mạng DCGAN đã được đánh giá bằng cách áp dụng bộ phân loại L2-SVM chính quy cho một vectơ tính năng được trích xuất từ bộ phân biệt (đã đào tạo) [5]. Điểm phân loại tốt đã đạt được khi sử dụng phương pháp này trên cả tập dữ liệu có giám sát và bán giám sát, ngay cả những tập dữ liệu không giao nhau với dữ liệu đào tạo ban đầu.

Chất lượng biểu diễn dữ liệu có thể được cải thiện khi đào tạo đối nghịch bao gồm việc học chung một cơ chế suy luận như với ALI [19]. Một vectơ biểu diễn được xây dựng bằng cách sử dụng ba lớp ẩn cuối cùng của bộ mã hóa ALI, một bộ phân loại L2-SVM tương tự, nhưng đạt được tỷ lệ phân loại sai thấp hơn đáng kể so với DCGAN [19]. Ngoài ra, ALI đã đạt được kết quả phân loại tiên tiến khi thông tin nhận được đưa vào quy trình đào tạo.

Khi dữ liệu đào tạo được gắn nhãn có hạn, đào tạo đối nghịch cũng có thể được sử dụng để tổng hợp thêm các mẫu đào tạo. Shrivastava et al. [39] sử dụng GAN để tinh chỉnh hình ảnh tổng hợp trong khi vẫn duy trì thông tin chú thích của chúng. Bằng cách chỉ đào tạo các mô hình trên hình ảnh tổng hợp được tinh chỉnh bằng GAN (tức là không có dữ liệu đào tạo thực tế), Shrivastava et al. [39] đã đạt được hiệu suất tiên tiến nhất trong các tác vụ ước tính tư thế và ánh mắt. Tương tự như vậy, đã thu được kết quả tốt cho ước tính và dự đoán ánh mắt



Hình 6. Một ví dụ về việc áp dụng “vectơ nụ cười” với mô hình ALI [19]. Hình ảnh đầu tiên là ví dụ về một người phụ nữ không cười và hình ảnh cuối cùng là ví dụ về một người phụ nữ đang cười. Giá trị z cho hình ảnh đầu tiên được suy ra là z_1 và cho hình ảnh cuối cùng là z_2 . Nội suy dọc theo một vectơ kết nối z_1 và z_2 , sẽ cung cấp các giá trị z có thể được truyền qua một trình tạo để tổng hợp các mẫu mới. Lưu ý hàm ý: một vectơ dịch chuyển trong không gian tiềm ẩn đi qua “cường độ” nụ cười trong không gian hình ảnh. (Hình ảnh được sử dụng với sự cho phép của Tom White.)



Hình 7. Ví dụ về tổng hợp hình ảnh sử dụng GAWWN. Trong GAWWN, hình ảnh được điều kiện hóa trên cả mô tả văn bản và vị trí hình ảnh được chỉ định là điểm chính hoặc hộp giới hạn. (Hình được sao chép từ [44] với sự cho phép.)

sử dụng kiến trúc GAN không gian thời gian [40]. Trong một số trường hợp, các mô hình được đào tạo trên dữ liệu tổng hợp không khái quát hóa tốt khi áp dụng vào dữ liệu thực [3]. Bousmalis et al. [3] đề xuất giải quyết vấn đề này bằng cách điều chỉnh các mẫu tổng hợp từ miền nguồn để khớp với miền mục tiêu bằng cách sử dụng đào tạo đối nghịch. Ngoài ra, Liu et al. [41] đề xuất sử dụng nhiều GAN—một cho mỗi miền—với trọng số được liên kết để tổng hợp các cặp mẫu hình ảnh tương ứng từ các miền khác nhau. Vì chất lượng của các mẫu được tạo ra khó có thể đánh giá định lượng trên các mô hình, nên các tác vụ phân loại có khả năng vẫn là một công cụ định lượng quan trọng để đánh giá hiệu suất của GAN, ngay cả khi các ứng dụng mới và đa dạng trong thị giác máy tính đang được khám phá.

Tổng hợp hình ảnh

Phần lớn nghiên cứu GAN gần đây tập trung vào việc cải thiện chất lượng và tiện ích của khả năng tạo hình ảnh. Mô hình LAPGAN đã giới thiệu một chuỗi các mạng tích chập trong khuôn khổ kim tự tháp Laplacian để tạo hình ảnh theo cách từ thô đến tinh [13]. Một cách tiếp cận tương tự được sử dụng bởi Huang et al. [42] với GAN hoạt động trên các biểu diễn trung gian thay vì hình ảnh có độ phân giải thấp hơn.

LAPGAN cũng mở rộng phiên bản có điều kiện của mô hình GAN trong đó cả mạng G và D đều nhận thông tin nhân bổ sung làm đầu vào; kỹ thuật này đã chứng minh là hữu ích và hiện là một thông lệ phổ biến để cải thiện chất lượng hình ảnh. Ý tưởng về điều kiện hóa GAN này sau đó đã được mở rộng để kết hợp ngôn ngữ tự nhiên. Ví dụ, Reed et al. [43] đã sử dụng kiến trúc GAN để tổng hợp hình ảnh từ mô tả văn bản, trong đó

có thể mô tả như là chú thích ngược. Ví dụ, với chú thích văn bản về một con chim như "màu trắng với một ít màu đen trên đầu và cánh cùng cái mỏ dài màu cam", GAN đã được đào tạo có thể tạo ra một số hình ảnh hợp lý khớp với mô tả.

Ngoài việc điều kiện hóa trên các mô tả văn bản, mạng lưới đối nghịch sinh ra cái gì-ở-đầu (GAWWN) điều kiện hóa trên vị trí hình ảnh [44]. Hệ thống GAWWN hỗ trợ một giao diện tương tác trong đó các hình ảnh lớn có thể được xây dựng gia tăng với các mô tả văn bản về các bộ phận và hộp giới hạn do người dùng cung cấp (Hình 7).

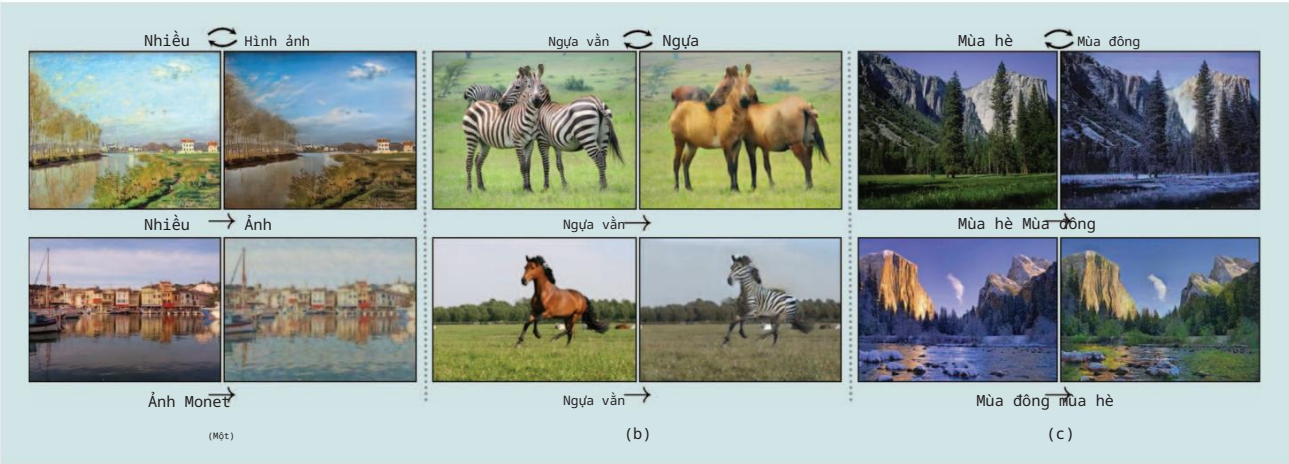
GAN có điều kiện không chỉ cho phép chúng ta tổng hợp các mẫu mới với các thuộc tính cụ thể mà còn cho phép chúng ta phát triển các công cụ để chỉnh sửa hình ảnh trực quan; ví dụ, thay đổi kiểu tóc của một người trong hình ảnh, bất họ đeo kính hoặc chỉnh sửa hình ảnh để họ trông trẻ hơn [35]. Các ứng dụng bổ sung của GAN vào việc chỉnh sửa hình ảnh bao gồm công trình của Zhu và Brock et al. [2], [45].

Bản dịch từ hình ảnh sang hình ảnh
Mạng đối nghịch có điều kiện rất phù hợp để dịch một hình ảnh đầu vào thành một hình ảnh đầu ra, đây là một chủ đề thường gặp trong đồ họa máy tính, xử lý hình ảnh và thị giác máy tính. Mô hình pix2pix cung cấp một giải pháp chung cho nhóm các vấn đề này [46]. Ngoài việc học cách ánh xạ từ hình ảnh đầu vào sang hình ảnh đầu ra, pix2pix

mô hình cũng xây dựng một hàm mất mát để đào tạo phép ánh xạ này. Mô hình này đã chứng minh kết quả hiệu quả cho các vấn đề khác nhau của thị giác máy tính mà trước đây đòi hỏi máy móc riêng biệt, bao gồm phân đoạn ngữ nghĩa, tạo bản đồ từ ảnh chụp trên không và tô màu cho ảnh đen trắng. Wang và cộng sự trình bày một ý tưởng tương tự, sử dụng GAN để tổng hợp bản đồ bề mặt-pháp tuyến (tương tự như bản đồ độ sâu) và sau đó ánh xạ những hình ảnh này vào các cảnh tự nhiên.

CycleGAN [4] mở rộng công trình này bằng cách giới thiệu một chu kỳ mất tính nhất quán cố gắng bảo toàn hình ảnh gốc sau một chu kỳ dịch chuyển và dịch chuyển ngược. Trong công thức này, các cặp hình ảnh khớp nhau không còn cần thiết để đào tạo nữa. Điều này làm cho việc chuẩn bị dữ liệu đơn giản hơn nhiều và mở ra kỹ thuật này cho một họ ứng dụng lớn hơn. Ví dụ, chuyển giao phong cách nghệ thuật [47] tạo ra hình ảnh tự nhiên theo phong cách của các nghệ sĩ, chẳng hạn như Picasso hoặc Monet, chỉ bằng cách được đào tạo trên một bộ sưu tập tranh và hình ảnh tự nhiên không ghép đôi (Hình 8).

Siêu phân giải
Siêu phân giải cho phép tạo ra hình ảnh có độ phân giải cao từ hình ảnh có độ phân giải thấp hơn, với mô hình được đào tạo suy ra các chi tiết thực tế như ảnh trong khi lấy mẫu. SRGAN



Hình 8. Mô hình CycleGAN học cách dịch chuyển hình ảnh sang hình ảnh giữa hai bộ sưu tập hình ảnh không có thứ tự. Dưới đây là các ví dụ về ánh xạ hình ảnh hai chiều: (a) Tranh của Monet sang ảnh phong cảnh, (b) ngựa vằn sang ngựa và (c) ảnh mùa hè sang mùa đông ở Công viên quốc gia Yosemite. (Hình ảnh được sao chép từ [4] với sự cho phép.)

mô hình [36] mở rộng những nỗ lực trước đó bằng cách thêm một thành phần mất mát đối nghịch, hạn chế hình ảnh nằm trên nhiều hình ảnh tự nhiên.

Máy phát SRGAN được điều chỉnh trên hình ảnh có độ phân giải thấp và suy ra hình ảnh tự nhiên chân thực như ảnh với 4 # các yếu tố nâng cấp. Không giống như hầu hết các ứng dụng GAN, tổn thất đối nghịch là một thành phần của hàm tổn thất lớn hơn, hàm này cũng bao gồm tổn thất nhận thức từ bộ phân loại được đào tạo trước và tổn thất chính quy khuyến khích hình ảnh mạch lạc về mặt không gian. Trong bối cảnh này, sự mất mát đối nghịch hạn chế giải pháp tổng thể thành nhiều hình ảnh tự nhiên, tạo ra các giải pháp có sức thuyết phục hơn về mặt nhận thức.

Việc tùy chỉnh các ứng dụng học sâu thường bị cản trở do thiếu các tập dữ liệu đào tạo được quản lý có liên quan. Tuy nhiên, SRGAN dễ dàng tùy chỉnh theo các miền cụ thể, vì các cặp ảnh đào tạo mới có thể dễ dàng được xây dựng bằng cách hạ mẫu một tập hợp các ảnh có độ phân giải cao. Đây là một cân nhắc quan trọng trong thực tế, vì các chi tiết ảnh-thực tế suy ra mà GAN tạo ra sẽ khác nhau tùy thuộc vào miền ảnh được sử dụng trong tập đào tạo.

Cuộc thảo luận

Câu hỏi mở

GAN đã thu hút được sự chú ý đáng kể do khả năng tận dụng lượng lớn dữ liệu chưa được gắn nhãn. Mặc dù đã có nhiều tiến bộ để giảm bớt một số thách thức liên quan đến việc đào tạo và đánh giá GAN, nhưng vẫn còn một số thách thức chưa được giải quyết.

Thu gọn chế độ

Như đã nêu trong phần “Đào tạo GAN”, một vấn đề phổ biến của GAN liên quan đến việc máy phát điện bị sụp đổ để tạo ra một họ nhỏ các mẫu tương tự (sụp đổ một phần) và trong trường hợp xấu nhất, chỉ tạo ra một mẫu duy nhất (sụp đổ hoàn toàn) [26], [48]. Sự đa dạng trong máy phát điện có thể được tăng lên bằng các thủ thuật thực tế để cân bằng sự phân phối các mẫu được tạo ra

bằng bộ phân biệt đối với các lô thực và giả, hoặc bằng cách sử dụng nhiều GAN để bao phủ các chế độ khác nhau của phân phối xác suất [49]. Một giải pháp khác để giảm thiểu sự sụp đổ chế độ là thay đổi phép đo khoảng cách được sử dụng để so sánh các phân phối thống kê. Arjovsky [32] đã đề xuất so sánh các phân phối dựa trên khoảng cách Wasserstein thay vì độ phân kỳ dựa trên KL (DCGAN [5]) hoặc khoảng cách biến thiên tổng thể (GAN dựa trên năng lượng [50]). Metz và cộng sự [51] đã đề xuất mở cuộn bộ phân biệt trong một số bước, tức là, để nó tính toán các bản cập nhật của nó trên máy phát điện hiện tại trong một số bước, sau đó sử dụng các bộ phân biệt “mở cuộn” để cập nhật máy phát điện bằng cách sử dụng mục tiêu minimax bình thường. Như bình thường, bộ phân biệt chỉ đào tạo về bản cập nhật của nó từ một bước, nhưng máy phát điện hiện có quyền truy cập vào cách bộ phân biệt sẽ tự cập nhật. Với mục tiêu máy phát điện một bước thông thường, bộ phân biệt sẽ chỉ cần gán một xác suất thấp cho các đầu ra trước đó của máy phát điện, buộc máy phát điện phải di chuyển, dẫn đến sự hội tụ hoặc chu kỳ nhảy chế độ vô tận. Tuy nhiên, với mục tiêu không được cuộn, máy phát điện có thể ngăn bộ phân biệt tập trung vào bản cập nhật trước đó và cập nhật các thể hệ của riêng nó với tầm nhìn xa về cách bộ phân biệt sẽ phản ứng.

Sự mất ổn định trong quá trình đào tạo-điểm yên xe

Trong GAN, Hessian của hàm mất mát trở nên không xác định. Do đó, giải pháp tối ưu nằm ở việc tìm một điểm yên ngựa thay vì một điểm cực tiểu cục bộ. Trong học sâu, một số lượng lớn các trình tối ưu hóa chỉ phụ thuộc vào đạo hàm đầu tiên của hàm mất mát; hội tụ đến một điểm yên ngựa cho GAN đòi hỏi phải khởi tạo tốt. Bằng cách áp dụng định lý đa tạp ổn định từ lý thuyết hệ thống phi tuyến tính, Lee và cộng sự [52] đã chỉ ra rằng, nếu chúng ta chọn các điểm ban đầu của một trình tối ưu hóa một cách ngẫu nhiên, thì quá trình giảm dần độ dốc sẽ không hội tụ đến một điểm yên ngựa với xác suất là một (xem thêm [25] và [53]). Ngoài ra, Mescheder và cộng sự [54] đã lập luận rằng sự hội tụ của hàm mục tiêu của GAN bị ảnh hưởng bởi sự hiện diện của một phần thực bằng không của ma trận Jacobian cũng như các giá trị riêng có phần ảo lớn. Điều này làm nản lòng đối với quá trình đào tạo GAN; tuy nhiên, do sự tồn tại của

bộ tối ưu hóa, không phải tất cả hy vọng đều mất. Thật không may, các phương pháp kiểu Newton có độ phức tạp về thời gian tính toán được chia theo cấp số nhân hoặc cấp số nhân với chiều của các tham số. Do đó, một dòng câu hỏi khác nằm ở việc áp dụng và chia tỷ lệ các bộ tối ưu hóa bậc hai cho đào tạo đối nghịch.

Một vấn đề cơ bản hơn là sự tồn tại của trạng thái cân bằng cho GAN. Sử dụng kết quả từ các phép đo phi tham số Bayesian, Arora et al. [48] kết nối sự tồn tại của trạng thái cân bằng với một hỗn hợp hữu hạn các mạng nơ-ron-điều này có nghĩa là, dưới một dung lượng nhất định, không có trạng thái cân bằng nào có thể tồn tại. Trên một lưu ý liên quan chặt chẽ, người ta cũng lập luận rằng, trong khi đào tạo GAN có vẻ như đã hội tụ, thì phân phối đã đào tạo vẫn có thể cách xa phân phối mục tiêu. Để giảm bớt vấn đề này, Arora et al. [48] đề xuất một biện pháp mới gọi là khoảng cách mạng nơ-ron.

Đánh giá các mô hình sinh sản

Làm thế nào để đánh giá độ trung thực của các mẫu được tổng hợp bởi các mô hình sinh sản? Chúng ta có nên sử dụng ước tính khả năng không? Một GAN được đào tạo bằng một phương pháp luận có thể được so sánh với một phương pháp luận khác (so sánh mô hình) không? Đây là những câu hỏi mở không chỉ liên quan đến GAN mà còn liên quan đến các mô hình xác suất nói chung. Theis [55] lập luận rằng việc đánh giá GAN bằng các biện pháp khác nhau có thể dẫn đến các kết luận mâu thuẫn về chất lượng của các mẫu tổng hợp; quyết định chọn một biện pháp này hơn biện pháp khác phụ thuộc vào ứng dụng.

Kết luận

Sự bùng nổ về mối quan tâm đến GAN không chỉ được thúc đẩy bởi tiềm năng của chúng trong việc học các ánh xạ sâu, phi tuyến tính cao từ không gian tiềm ẩn vào không gian dữ liệu và ngược lại mà còn bởi tiềm năng của chúng trong việc sử dụng lượng lớn dữ liệu hình ảnh không có nhãn vẫn đóng đối với việc học biểu diễn sâu. Trong phạm vi tính tế của đào tạo GAN, có nhiều cơ hội để phát triển lý thuyết và thuật toán, và với sức mạnh của mạng lưới sâu, có rất nhiều cơ hội cho các ứng dụng mới.

Lời cảm ơn

Chúng tôi muốn cảm ơn David Warde-Farley vì những phản hồi có giá trị của anh ấy về các bản sửa đổi trước đây của bài viết. Antonia Creswell ghi nhận sự hỗ trợ của Hội đồng nghiên cứu khoa học kỹ thuật và vật lý thông qua học bổng đào tạo tiến sĩ.

Tác giả

Antonia Creswell (ac2211@ic.ac.uk) đã nhận bằng hạng nhất từ Imperial College London chuyên ngành kỹ thuật y sinh vào năm 2011 và hiện đang là nghiên cứu sinh tiến sĩ tại Nhóm nghiên cứu về thị giác máy tính lấy cảm hứng từ sinh học tại Imperial College London. Trọng tâm nghiên cứu tiến sĩ của cô là cải thiện việc đào tạo các mạng đối nghịch tạo sinh và áp dụng chúng vào tìm kiếm trực quan và biểu diễn học tập trong các nguồn dữ liệu hình ảnh không có nhãn.

Tom White (tom@sixdozen.com) đã nhận bằng Cử nhân Toán học của Đại học Georgia và bằng Thạc sĩ Nghệ thuật và Khoa học Truyền thông của Viện Công nghệ Massachusetts. Hiện tại, ông là giảng viên cao cấp tại Khoa Thiết kế của Đại học Victoria Wellington, New Zealand. Nghiên cứu hiện tại của ông tập trung vào việc khám phá việc sử dụng ngày càng tăng của con-

học máy mang tính xây dựng trong thiết kế tính toán và tiềm năng sáng tạo của các nhà thiết kế con người khi làm việc cộng tác với mạng nơ-ron nhân tạo trong quá trình khám phá ý tưởng thiết kế và tạo mẫu.

Vincent Dumoulin (vi.dumoulin@gmail.com) đã nhận bằng Cử nhân Khoa học Vật lý và Khoa học Máy tính từ Đại học Montréal, Canada. Anh là ứng viên tiến sĩ tại Viện Thuật toán Học tập Montréal dưới sự đồng giám sát của Yoshua Bengio và Aaron Courville, làm việc về các phương pháp học sâu để tạo mô hình.

Kai Arulkumaran (kailash.arulkumaran13@imperial.ac.uk) đã nhận bằng Cử nhân Khoa học máy tính của Đại học Cambridge, Vương quốc Anh vào năm 2012 và bằng Thạc sĩ Kỹ thuật Y sinh của Đại học Imperial College London vào năm 2014, nơi anh hiện là ứng viên Tiến sĩ tại Khoa Kỹ thuật Sinh học. Anh là thực tập sinh nghiên cứu tại Twitter Magic Pony và Microsoft Research vào năm 2017. Trọng tâm nghiên cứu của anh là học tăng cường sâu và thị giác máy tính để kiểm soát thị giác vận động.

Biswa Sengupta (biswasengupta@gmail.com) nhận bằng Cử nhân Kỹ thuật (danh dự) về kỹ thuật điện và máy tính năm 2004 và bằng Thạc sĩ Khoa học máy tính lý thuyết năm 2005 từ Đại học York, Vương quốc Anh. Ông nhận bằng Thạc sĩ Khoa học thứ hai về khoa học thần kinh và hành vi năm 2007 từ Viện Max Planck về Điều khiển học Sinh học, Đức và bằng Tiến sĩ về khoa học thần kinh lý thuyết năm 2011 từ Đại học Cambridge, Vương quốc Anh. Ông đã được đào tạo thêm về thống kê Bayes và hình học vi phân tại University College London và Đại học Cambridge trước khi lãnh đạo Cortexica Vision Systems với tư cách là nhà khoa học trưởng. Hiện tại, ông là nhà khoa học thỉnh giảng tại Imperial College London và cũng đang lãnh đạo nghiên cứu máy học tại Phòng thí nghiệm Noah's Ark của Huawei Technologies Vương quốc Anh.

Anil A. Bharath (a.bharath@imperial.ac.uk) đã nhận bằng B. Bằng kỹ sư về kỹ thuật điện tử và điện tử của University College London năm 1988, và bằng tiến sĩ về xử lý tín hiệu của Imperial College London năm 1993, nơi ông hiện là giảng viên tại Khoa Kỹ thuật sinh học. Ông là thành viên học thuật của Viện khoa học dữ liệu của Imperial và là thành viên của Viện Kỹ thuật và Công nghệ. Ông là khách mời học thuật của Nhóm xử lý tín hiệu tại Đại học Cambridge năm 2006. Ông là người đồng sáng lập Cortexica Vision Systems. Ông quan tâm đến nghiên cứu về kiến trúc sâu cho suy luận trực quan.

Tài liệu tham khảo

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville và Y. Bengio, "Mạng đối nghịch sinh sản", trong Proc. Advances Neural Information Processing Systems Conf., 2014, trang 2672-2680.

[2] J.-Y. Zhu, P. Krähenbühl, E. Shechtman và AA Efros, " Thao tác trực quan tạo ra trên đa tạp hình ảnh tự nhiên ," trong Proc. Hội nghị Châu Âu Tầm nhìn máy tính, 2016, tr. 597-613.

[3] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan và D. Krishnan, "Thích ứng miền cấp độ pixel không giám sát với mạng đối nghịch tạo sinh", trong Proc. IEEE Conf. Nhận dạng mẫu thị giác máy tính, 2016, trang 3722-3731.

[4] J.-Y. Zhu, T. Park, P. Isola, và AA Efros. (2017). Dịch chuyển hình ảnh sang hình ảnh không ghép đôi bằng cách sử dụng các mạng đối nghịch nhất quán chu kỳ , Proc. Int. Conf. Computer Vision. [Trực tuyến]. Có sẵn: <https://arxiv.org/abs/1703.10593>

[5] A. Radford, L. Metz và S. Chintala, “Học biểu diễn không giám sát với mạng đối nghịch tạo sinh tích chập sâu,” trong Proc. 5th Int. Conf. Psychol., tập 24, số 6, tr. 417, 1933.
Hội thảo về Biểu diễn học tập, 2016.

[6] A. Creswell và AA Bharath, “Huấn luyện đối kháng để truy xuất bản phác thảo,” trong Proc. European Conf. Computer Vision Workshops, Amsterdam, Hà Lan, 2016, tr. 798-809.

[7] Y. LeCun, Y. Bengio, và G. Hinton, “Học sâu,” Nature, tập 521, số 7553, trang 436-444, 2015.

[8] H. Hotelling, “Phân tích phức hợp các biến thống kê thành các thành phần chính,” J. Educ. Psychol., tập 24, số 6, tr. 417, 1933.
[9] IJ Goodfellow, “Về tiêu chí phân biệt để ước tính các mô hình tạo ra,” trong Proc. Int. Conf. Learning Representations Workshop Track, 2015.

[10] M. Gutmann và A. Hyvärinen, “Ước tính tương phản nhiễu: Một nguyên tắc ước tính mới cho các mô hình thống kê không chuẩn hóa,” Artif. Intell. Statist., tập 1, số 2, trang 6, 2010.

[11] Y. Bengio, L. Yao, G. Alain và P. Vincent, “Bộ mã hóa tự động khử nhiễu tổng quát như các mô hình tạo ra”, trong Proc. Advances Neural Information Processing Systems Conf., 2013, trang 899-907.

[12] I. Goodfellow. (2016). Hướng dẫn NIPS 2016: Mạng đối nghịch tạo sinh, Proc. Neural Information Processing Systems Conf. [Trực tuyến]. Có sẵn: <https://arxiv.org/abs/1701.00160>

[13] EL Denton, S. Chintala, A. Szlam và R. Fergus, “Các mô hình hình ảnh tạo ra sâu sử dụng kim tự tháp Laplacian của các mạng đối nghịch,” trong Proc. Advances Neural Information Processing Systems Conf., 2015, tr. 1486-1494.

[14] J. Wu, C. Zhang, T. Xue, B. Freeman và J. Tenenbaum, “Học không gian tiềm ẩn xác suất của hình dạng vật thể thông qua mô hình đối nghịch-sáng tạo 3D,” trong Proc. Hội nghị về Hệ thống xử lý thông tin thần kinh tiên tiến, 2016, trang 82-90.

[15] M. Mirza và S. Osindero, “Mạng đối nghịch tạo ra có điều kiện”, Bản in trước arXiv, arXiv:1411.1784, 2014.

[16] X. Chen, Y. Duan, R. Houthoof, J. Schulman, I. Sutskever và P. Abbeel, “Infogan: Học biểu diễn có thể diễn giải bằng cách tối đa hóa thông tin các mạng đối nghịch tạo sinh,” trong Proc. Advances Neural Information Processing Systems Conf., 2016, tr. 2172-2180.

[17] A. Creswell và AA Bharath, “Đào ngược máy phát điện của mạng đối kháng-đối lập tạo ra,” trong Proc. Neural Information Processing Systems Workshop Adversarial Training, 2016.

[18] ZC Lipton và S. Tripathi, “Phục hồi chính xác các vectơ tiềm ẩn từ các mạng đối nghịch sinh sản”, trong Proc. Int. Conf. Learning Representations Workshop Track, 2017.

[19] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky và A. Courville, “Suy luận học được theo hướng đối nghịch”, trong Proc. Int. Conf. Learning Representations, 2017.

[20] J. Donahue, P. Krähenbühl và T. Darrell, “Học tính năng đối nghịch,” trong Hội nghị quốc tế về biểu diễn học tập, 2017.

[21] C. Li, H. Liu, C. Chen, Y. Pu, L. Chen, R. Henao và L. Carin, “Hướng tới hiểu biết về học tập đối nghịch để khớp phân phối chung,” trong Proc. Hội nghị về Hệ thống xử lý thông tin thần kinh tiên tiến, 2017.

[22] A. Makhzani, J. Shlens, N. Jaitly, và I. Goodfellow. (2016). Bộ mã hóa tự động đối nghịch, Proc. Int. Conf. Biểu diễn học tập. [Trực tuyến]. Có sẵn: <http://arxiv.org/abs/1511.05644>

[23] DP Kingma và M. Welling, “Tự động mã hóa các phép biến thiên Bayes,” trong Proc. 2nd Int. Conf. Learning Representations, 2014.

[24] LM Mescheder, S. Nowozin và A. Geiger. (2017). Bayes biến thiên đối nghịch: Thống nhất các bộ mã hóa tự động biến thiên và mạng đối nghịch tạo sinh. [Trực tuyến]. Có sẵn: <http://arxiv.org/abs/1701.04722>

[25] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford và X. Chen, “Các kỹ thuật cải tiến để đào tạo GANS,” trong Proc. Advances Neural Information Processing Systems Conf., 2016, tr. 2226-2234.

[26] M. Arjovsky và L. Bottou, “Hướng tới các phương pháp có nguyên tắc để đào tạo các mạng đối kháng sinh sản,” trong Proc. Neural Information Processing Systems Conf. Hội thảo Đào tạo đối kháng, 2016.

[27] E. Shelhamer, J. Long và T. Darrell, “Mạng tích chập hoàn toàn để phân đoạn ngữ nghĩa”, IEEE Trans. Pattern Anal. Mach. Intell., tập 39, số 4, trang 640-651, 2017.

[28] S. Ioffe và C. Szegedy, “Chuẩn hóa theo lô: Tăng tốc đào tạo mạng sâu bằng cách giảm sự dịch chuyển biến phụ thuộc nội bộ,” trong Proc. 32nd Int. Conf. Machine Learning, 2015, trang 448-456.

[29] CK Sønderby, J. Caballero, L. Theis, W. Shi và F. Huszár, “Suy luận bản đồ khẩu hao cho siêu phân giải hình ảnh,” trong Proc. Int. Conf. Learning Representations, 2017.

[30] S. Nowozin, B. Cseke và R. Tomioka, “F-GAN: Đào tạo các bộ lấy mẫu thần kinh tạo ra bằng cách sử dụng phương pháp giảm thiểu sự phân kỳ biến thiên,” trong Proc. Advances Neural Information Processing Systems Conf., 2016, trang 271-279.

[31] M. Uehara, I. Sato, M. Suzuki, K. Nakayama và Y. Matsuo, “Mạng đối nghịch tạo ra từ góc độ ước tính tỷ lệ mật độ,” Bản in trước arXiv, arXiv:1610.02920, 2016.

[32] M. Arjovsky, S. Chintala và L. Bottou, “Wasserstein GAN,” trong Proc. 34th Int. Conf. Machine Learning, 2017, trang 214-223.

[33] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin và A. Courville, “Cải thiện việc đào tạo GANS wasserstein,” trong Proc. Advances Neural Information Processing Systems Conf., 2017.

[34] T. Mikolov, K. Chen, G. Corrado và J. Dean, “Ước tính hiệu quả các biểu diễn từ trong không gian vectơ,” trong Proc. Int. Conf. Learning Representations, 2013.

[35] S. Gurumurthy, R. K. Sarvadevabhatla và V. B. Radhakrishnan, “DeIgan: Mạng đối kháng tạo ra dữ liệu đa dạng và hạn chế,” trong Proc. Hội nghị IEEE Nhận dạng mẫu thị giác máy tính, 2017, trang 166-174.

[36] J. Ledig, L. Theis, F. Huszár, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang và W. Shi, “Siêu phân giải hình ảnh đơn thực tế bằng cách sử dụng mạng đối nghịch tạo sinh”, trong Proc. IEEE Conf. Nhận dạng mẫu thị giác máy tính, 2017, trang 4681-4690.

[37] X. Yu và F. Porikli, “Hình ảnh khuôn mặt có độ phân giải cực cao bằng mạng tạo phân biệt,” trong Proc. European Conf. Computer Vision, 2016, trang 318-333.

[38] X. Yu và F. Porikli, “Tạo ảo giác cho hình ảnh khuôn mặt có độ phân giải rất thấp, không thẳng hàng và nhiễu bằng bộ mã hóa tự động phân biệt biến đổi,” trong Proc. IEEE Conf. Nhận dạng mẫu thị giác máy tính, 2017, trang 3760-3768.

[39] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang và R. Webb, “Học từ hình ảnh mờ phông và không giám sát thông qua đào tạo đối nghịch,” trong Proc. IEEE Conf. Nhận dạng mẫu thị giác máy tính, 2016, tr. 2107-2116.

[40] M. Zhang, KT Ma, JH Lim, Q. Zhao và J. Feng, “Cải nhìn sâu sắc về tương lai: Cải nhìn dự đoán về các video ích kỷ bằng cách sử dụng các mạng đối đầu,” trong Proc. IEEE Conf. Nhận dạng mẫu thị giác máy tính, 2017, trang 4372-4381.

[41] M.-Y. Liu và O. Tuzel, “Mạng đối kháng sinh sản kết hợp,” trong Proc. Hội nghị về Hệ thống xử lý thông tin thần kinh tiên tiến, 2016, trang 469-477.

[42] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft và S. Belongie, “Mạng đối kháng tạo sinh xếp chồng”, trong Proc. IEEE Conf. Nhận dạng mẫu thị giác máy tính, 2016.

[43] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele và H. Lee. (2016). Tổng hợp văn bản đối nghịch thành hình ảnh, Proc. Int. Conf. Học máy. [Trực tuyến]. Có sẵn: <https://arxiv.org/abs/1605.05396>

[44] SE Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele và H. Lee, “Học cách vẽ gì và vẽ ở đâu,” trong Proc. Advances Neural Information Processing Systems Conf., 2016, tr. 217-225.

[45] A. Brock, T. Lim, JM Ritchie và N. Weston, “Chỉnh sửa ảnh thần kinh với mạng lưới đối nghịch hướng nội,” trong Proc. 6th Int. Conf. Learning Representations, 2017.

[46] P. Isola, J.-Y. Zhu, T. Zhou và AA Efros, “Phép dịch hình ảnh sang hình ảnh với mạng đối nghịch có điều kiện,” trong Proc. IEEE Conf. Nhận dạng mẫu thị giác máy tính, 2016, trang 1125-1134.

[47] C. Li và M. Wand, “Tổng hợp kết cấu thời gian thực được tính toán trước với mạng đối nghịch tạo ra Markovian,” trong Proc. European Conf. Computer Vision, 2016, trang 702-716.

[48] S. Arora, R. Ge, Y. Liang, T. Ma và Y. Zhang, “Tổng quát hóa và cân bằng trong mạng đối nghịch sinh sản (GANS),” trong Proc. 34th Int. Conf. Machine Learning, 2017, tr. 224-232.

[49] I. Tolstikhin, S. Gelly, O. Bousquet, C.-J. Simon-Gabriel và B. Schölkopf, “ADAGAN: Tăng cường các mô hình tạo sinh”, Bản in trước arXiv arXiv:1701.02386. [Trực tuyến]. Có sẵn: <https://arxiv.org/abs/1701.02386>

[50] J. Zhao, M. Mathieu, và Y. LeCun. (2017). Mạng đối nghịch tạo ra dựa trên năng lượng, Proc. Int. Conf. Biểu diễn học tập. [Trực tuyến]. Có sẵn: <https://arxiv.org/abs/1609.03126>

[51] L. Metz, B. Poole, D. Pfau và J. Sohl-Dickstein. (2017). Mạng lưới đối kháng sinh thành chưa được triển khai, Proc. Quốc tế Conf. Học cách biểu diễn. [Trực tuyến]. Có sẵn: <https://arxiv.org/abs/1611.02163>

[52] JD Lee, M. Simchowitz, MI Jordan và B. Recht, “Sự giảm dần theo độ dốc chỉ hội tụ đến các bộ tối thiểu hóa,” trong Proc. Conf. Learning Theory, 2016, trang 1246-1257.

[53] R. Pemantle, “Sự không hội tụ đến các điểm không ổn định trong các mô hình urn và các phép tính xấp xỉ ngẫu nhiên,” Ann. Probab., tập 18, số 2, trang 698-712, tháng 4 năm 1990.

[54] LM Mescheder, S. Nowozin, và A. Geiger. (2017). Số học của GANS, Proc. Advances Neural Information Processing Systems Conf. [Trực tuyến]. Có sẵn: <http://arxiv.org/abs/1705.10461>

[55] L. Theis, A. van den Oord và M. Bethge, “Ghi chú về việc đánh giá các mô hình sinh sản,” trong Proc. Int. Conf. Learning Representations.