

MÔ PHỎNG LÃO HÓA KHUÔN MẶT NGƯỜI VỚI GENERATIVE ADVERSARIAL NETWORKS (GANs)

*Face Aging with Generative Adversarial Networks (GANs)

1st Trần Tuấn Vũ

*Student. Industrial University Of HCMC Student. Industrial University Of HCMC Computer vision project
Ho Chi Minh, Vietnam
tuanvufit@gmail.com*

2nd Đào Duy Trường

*Student. Industrial University Of HCMC Student. Industrial University Of HCMC Computer vision project
Ho Chi Minh, Vietnam
aotruong123@gmail.com*

3rd Phan Lê hoàng Việt

*Student. Industrial University Of HCMC Computer vision project
Ho Chi Minh, Vietnam
phanlehoangviet1309@gmail.com*

4th Huỳnh Anh Tú

*Student. Industrial University Of HCMC Computer vision project
Ho Chi Minh, Vietnam
nnt12092001@gmail.com*

Tóm tắt nội dung—Mạng học sâu tự sinh đối nghịch (GAN) đang nổi lên là một trong những mô hình học sâu sinh ra ảnh được áp dụng trong nhiều bài toán thực tế. Chúng tôi đề xuất việc áp dụng các mô hình GAN cho bài toán thể hiện sự lão hóa khuôn mặt của con người. Trái ngược với công việc trước đây sử dụng GAN để thay đổi các đặc trưng trên khuôn mặt, trong bài toán này chúng tôi đặc biệt nhấn mạnh vào việc bảo toàn danh tính, cụ thể là trong quá trình sinh ảnh của mô hình nhưng vẫn giữ được những đặc trưng của khuôn mặt. Để đạt được mục tiêu này, chúng tôi sẽ sử dụng các biến thể GAN. Sự khác biệt giữa những hình ảnh khuôn mặt được sinh ra từ mô hình được dựa trên các nhóm tuổi chính.

I. INTRODUCTION

Lão hóa khuôn mặt người là công việc mô phỏng khuôn mặt của một người nào đó theo một độ tuổi nhất định. Nó đang thu hút ngày càng nhiều sự chú ý của các nhà nghiên cứu vì các ứng dụng khác nhau của nó trong nhận dạng khuôn mặt và giải trí ở mọi lứa tuổi. Ví dụ, nó có thể được áp dụng trong việc bảo mật hoặc để dự đoán ai đó sẽ trông như thế nào trong tương lai. Nhiều nghiên cứu đã được thực hiện về lão hóa khuôn mặt [1] [2] [3].

Tuy nhiên để mô phỏng trọn vẹn lão hóa khuôn mặt của người nào đó cần rất nhiều thời gian [4][5][6][7] khiên cho việc thu thập dữ liệu là một nhiệm vụ cực kỳ khó khăn. Các phương pháp lão hóa khuôn mặt truyền thống có thể được phân loại thành các phương pháp dựa trên nguyên mẫu [8] và phương pháp dựa trên mô hình vật lý [9].

Các phương pháp tiếp cận dựa trên nguyên mẫu thường tính toán khuôn mặt trung bình trong một nhóm và sự khác biệt

giữa các khuôn mặt trung bình khác nhau từ các nhóm tuổi khác nhau sẽ được coi là mô hình lão hóa sẽ được sử dụng để mô phỏng khuôn mặt bị lão hóa [2]. Do đó, thông tin cá nhân cụ thể của từng người sẽ bị mất, dẫn đến khuôn mặt được tổng hợp trông không thực tế. Ngược lại, các phương pháp tiếp cận dựa trên mô hình vật lý mô hình hóa hình dạng và kết cấu thay đổi theo độ tuổi về màu tóc, cơ bắp và nếp nhăn, v.v. với mô hình tham số, thường yêu cầu nhiều dữ liệu dành cho việc huấn luyện và rất tốn kém về mặt tính toán.

Gần đây, các phương pháp tiếp cận dựa trên Generative Adversarial Networks (GAN) đã được chứng minh là thành công trong việc tạo ra hình ảnh chất lượng cao [10] [11] [12]. Đối với bài báo này, các mô hình chúng tôi trong bài báo này là CGANs, IPCGANs, Cycle GANs, trong đó Conditional Generative Adversarial Networks (CGAN) [10] [11] [9] lấy thông tin trước trong quá trình tạo ảnh và làm cho ảnh được tạo có thuộc tính mong muốn nhất định. Lấy cảm hứng từ CGAN, chúng tôi đề xuất thêm mô hình IPCGAN cho quá trình lão hóa khuôn mặt. Ngoài ra vì quá trình biến đổi với mục tiêu bảo toàn danh tính, nên chúng tôi đề xuất mô hình CycleGan - một mô hình tiêu biểu trong bài toán image-to-image translation

Những đóng góp của bài báo này được tóm tắt như sau:

- 1) Chúng tôi đảm bảo các khuôn mặt được tạo phù hợp với độ tuổi mục tiêu và bảo toàn các đặc trưng của đối tượng đầu vào. Các thử nghiệm mở rộng xác nhận tính hiệu quả của cả hai thuật ngữ để bảo toàn thông tin nhận dạng và làm cho hiệu ứng lão hóa trên khuôn mặt trở nên rõ ràng.

- 2) Ngoài việc đánh giá định lượng chất lượng của các khuôn mặt được tổng hợp, chúng tôi cũng đề xuất tiến hành xác minh khuôn mặt và phân loại tuổi khuôn mặt cho các khuôn mặt già được tạo bằng phương pháp nghiên cứu người dùng. Thử nghiệm tăng cường dữ liệu được đề xuất của chúng tôi cũng xác nhận tính hiệu quả của 3 mô hình GANs được sử dụng.
- 3) Tiến hành thử nghiệm bài toán trên 3 mô hình: CGANs, IPCGANs, Cycle GANs. Để có thể so sánh và cải tiến.

II. CÁC NGHIÊN CỨU LIÊN QUAN:

A. Mô phỏng lão hóa khuôn mặt:

Các phương pháp lão hóa khuôn mặt truyền thống có thể được phân loại thành phương pháp dựa trên nguyên mẫu và phương pháp dựa trên mô hình vật lý. Chúng tôi giới thiệu độc giả đến để có một cuộc khảo sát toàn diện về các phương pháp này. Cụ thể, các phương pháp tiếp cận dựa trên mô hình vật lý thường tập trung vào sự thay đổi cấu trúc giải phẫu của da, thay đổi cơ mặt và một số phép đo vật lý khác để điều chỉnh khuôn mặt theo tuổi tác. Các mô hình này thường rất phức tạp và yêu cầu nhiều dữ liệu huấn luyện. Phương pháp tiếp cận dựa trên nguyên mẫu tận dụng sự khác biệt giữa các khuôn mặt trung bình của các nhóm tuổi khác nhau để chuyển đổi mẫu tuổi. Tuy nhiên, chiến lược như vậy bỏ qua sự khác biệt giữa những người khác nhau, điều này làm cho các khuôn mặt được tạo ra trông không thực tế. Hơn nữa, một số đặc trưng quan trọng về tuổi tác, chẳng hạn như nếp nhăn, có thể được tính trung bình. Để tránh điều này, trong các phương pháp tiếp cận dựa trên phản hồi rời rạc đối tượng đã được áp dụng để mô hình hóa các thuộc tính khuôn mặt của một người cụ thể để tổng hợp các khuôn lão hóa. Mặc dù thông tin nhận dạng có thể được bảo toàn ở một mức độ nào đó bằng các phương pháp này, sự thay đổi của các khuôn mặt được tổng hợp giữa các nhóm tuổi lân cận diễn ra suôn sẻ hơn, nhưng thông tin nhận dạng không được giải thích rõ ràng trong bài báo này. Quá trình đào tạo gồm ba giai đoạn. Phương pháp này không hiệu quả tại thời điểm suy luận vì chúng phải giải quyết vấn đề về thời gian hoạt động LBFGS cho mỗi hình ảnh. Để lưu giữ thông tin nhận dạng tốt hơn, họ đề xuất phương pháp tiếp cận Local Manifold Adaptation trong. Kết hợp với [2], họ tăng cường xác minh khuôn mặt theo độ tuổi thông qua chuẩn hóa độ tuổi. Tương tự như chúng tôi, đã đề xuất một GAN có điều kiện mã hóa tự động để mã hóa hình ảnh đầu vào thành một đa tạp và sau đó tái tạo lại các hình ảnh cũ. Tuy nhiên, khuôn mặt bị lão hóa của họ dường như ít thay đổi do các điều kiện tuổi tác khác nhau.

B. Generative Adversarial Networks

GAN là mạng để sinh dữ liệu mới giống với dữ liệu trong dataset có sẵn và có 2 mạng trong GAN là Generator và Discriminator. GAN cấu tạo gồm 2 mạng là Generator và Discriminator. Trong khi Generator sinh ra các dữ liệu giống như thật thì Discriminator cố gắng phân biệt đâu là dữ liệu được sinh ra từ Generator và đâu là dữ liệu thật có. Ví dụ bài toán dùng GAN để generate ra tiền giả mà có thể dùng để chi tiêu được. Dữ liệu có là tiền thật. Generator giống như người



Hình 1. Face Aging

làm tiền giả còn Discriminator giống như cảnh sát. Người làm tiền giả sẽ cố gắng làm ra tiền giả mà cảnh sát cũng không phân biệt được. Còn cảnh sát sẽ phân biệt đâu là tiền thật và đâu là tiền giả. Mục tiêu cuối cùng là người làm tiền giả sẽ làm ra tiền giống với tiền thật. Trong quá trình train GAN thì cảnh sát có 2 việc: 1 là học cách phân biệt tiền nào là thật, tiền nào là giả, 2 là nói cho người làm tiền giả biết là tiền nó làm ra vẫn chưa qua mắt được và cần cải thiện hơn. Dần dần thì người làm tiền giả sẽ làm tiền giống tiền thật hơn và cảnh sát cũng thành thạo việc phân biệt tiền giả và tiền thật. Và mong đợi là tiền giả từ GAN sẽ đánh lừa được cảnh sát. Những kiến trúc sau vẫn dựa trên ý tưởng chủ đạo của model GAN đầu tiên nhưng có sự cải tiến đầu vào, phương pháp huấn luyện, hàm loss function để kết quả học được tốt hơn [17]

Tóm lại nguyên lí hoạt động của GAN như sau

- **Generator:** Học cách sinh ra dữ liệu giả để lừa mô hình Discriminator. Để có thể đánh lừa được Discriminator thì đòi hỏi mô hình sinh ra output phải thực sự tốt. Do đó chất lượng ảnh phải càng tốt càng tốt.
- **Discriminator:** Học cách phân biệt giữa dữ liệu giả được sinh từ mô hình Generator với dữ liệu thật. Discriminator như một giáo viên chấm điểm cho Generator biết cách nó sinh dữ liệu đã đủ tinh xảo để qua mặt được Discriminator chưa và nếu chưa thì Generator cần tiếp tục phải học để tạo ra ảnh thật hơn. Đồng thời Discriminator cũng phải cải thiện khả năng phân biệt của mình vì chất lượng ảnh được tạo ra từ Generator càng ngày càng giống thật hơn. Thông qua quá trình huấn luyện thì cả Generator và Discriminator cùng cải thiện được khả năng của mình.

C. Style Transfer:

Mục tiêu tổng hợp khuôn mặt với độ tuổi mục tiêu cũng liên quan đến công việc thay đổi phong cách. Đưa ra một hình ảnh đầu vào (được chuyển với một số phong cách nghệ thuật) và một hình ảnh theo phong cách nghệ thuật, mục tiêu của việc chuyển đổi phong cách là tạo ra một hình ảnh có nội dung được lấy từ ảnh đầu vào trước đó trong khi phong cách là từ cái sau. Để đạt được mục tiêu này, việc mất nội dung và mất kiểu trong không gian đặc trưng được tối ưu hóa cùng nhau. Cụ thể, cả mất nội dung và mất kiểu đều được gọi là mất trí giác vì chúng phụ thuộc vào các tính năng được trích xuất từ mạng mô hình được đào tạo trước. Một mạng lưới thần kinh trích xuất các tính năng ý nghĩa trừu tượng và trực quan hơn

so với các tính năng pixel thô. Mặc dù [6] có thể tạo ra hình ảnh chất lượng cao nhưng giai đoạn thử nghiệm lại chậm vì suy luận cần giải quyết vấn đề tối ưu hóa LBFGS. Để tránh điều này, trong [10], một mạng chuyển tiếp nguồn cấp dữ liệu được thông qua. Khác với chuyển phong cách, chuyển phong cách của một hình ảnh này sang hình ảnh khác, trong quá trình lão hóa khuôn mặt, mong muốn chuyển kiểu tuổi trong nhóm tuổi mục tiêu sang một khuôn mặt. Do đó, chuyển giao phong cách không thể được áp dụng trực tiếp cho lão hóa khuôn mặt.



Hình 2. Style Transfer

D. Unpair Image-to-Image Translation:

Dịch hình ảnh sang hình ảnh là một quá trình chuyển đổi hình ảnh từ miền này sang miền khác, trong đó mục tiêu là học ánh xạ giữa các hình ảnh đầu vào và hình ảnh đầu ra. Những ý tưởng tương tự cũng được áp dụng trong các tác vụ như tô ảnh truyện tranh, tạo ảnh từ phác thảo. Quá trình này thường được thực hiện bằng cách sử dụng các tập dữ liệu huấn luyện được cẩn chỉnh cặp với nhau ví dụ như mô hình Pix2Pix. Gần đây các mô hình GAN phát triển, dịch hình ảnh sang hình ảnh mà không cần dữ liệu ghép cặp với nhau, mục tiêu của tác vụ này là liên kết hai miền dữ liệu: X - Y. Lấy ý tưởng từ từ lĩnh vực ngôn ngữ, cải thiện các bản dịch bằng cách "dịch ngược" và "đổi chiều", tương tự với hình ảnh được liên tục biến đổi và kiểm tra từ đó cải thiện chất lượng hình ảnh.

III. CONDITION GAN - IDENTITY-PRESERVED CONDITIONAL - CYCLE GAN

A. KIẾN TRÚC MẠNG GAN:

Để tạo mô hình tổng quát của dữ liệu thông qua việc học một phép biến đổi từ các điểm thuộc phân phối đơn giản trước ($z \sim P_z$) sang các điểm thuộc phân phối dữ liệu ($z \sim P_{data}$). Một mô hình GAN điển hình bao gồm hai mô-đun chơi trò chơi đối nghịch: một bộ phân biệt đối xử và một bộ tạo. Trong khi trình tạo học cách tạo các mẫu giả không thể phân biệt được với các mẫu thực, thì bộ phân biệt học cách phân biệt giữa các mẫu giả này $G(z) \sim P_G$ và các mẫu thực $x \sim P_{data}$, do đó đưa ra đầu ra vô hướng $y = \{0, 1\}$. Mục tiêu của bộ tạo là đánh lừa bộ phân biệt bằng cách tạo ra các mẫu ảnh thực tế giống với các mẫu từ dữ liệu thực trong khi mục tiêu của

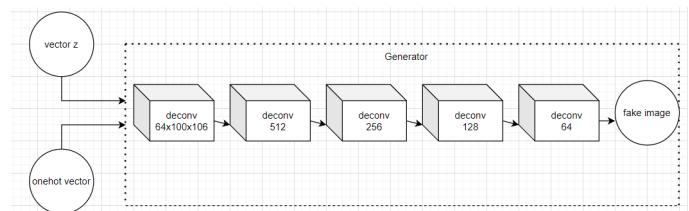
bộ phân biệt là phân biệt chính xác giữa dữ liệu thực và dữ liệu được tạo. Hai mô hình, thường được thiết kế dưới dạng mạng thần kinh, chơi trò chơi tối thiểu tối đa với hàm mục tiêu như trong biểu thức.

Mô hình GAN sẽ huấn luyện đồng thời cả hai model là generator và discriminator. Đây là một trò chơi zero-sum game trong lý thuyết trò chơi mà được xem như là hai người chơi đối nghịch lợi ích. Mô hình generator sẽ tạo ra ảnh fake chất lượng tốt nhất để đánh lừa discriminator và discriminator sẽ tìm cách phân loại ảnh real và ảnh fake. Hàm loss function của GAN là kết hợp giữa loss function của generator và discriminator:

$$\min_G \max_D V(D, G) = \underbrace{\mathbb{E}_{x \sim p_{data}(x)} [\log D(x)]}_{\text{log-probability that D predict x is real}} + \underbrace{\mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]}_{\text{log-probability D predicts G(z) is fake}} (1)$$

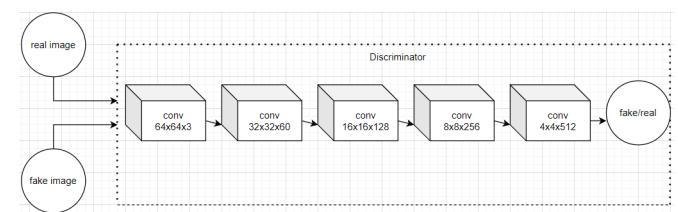
B. KIẾN TRÚC MẠNG CGAN:

Trong bài toán mô phỏng lão hóa khuôn mặt người theo các nhóm tuổi chúng tôi đã tiến hành triển khai đầu tiên trên mô hình CGans (conditional generative adversarial networks) áp dụng với tập dữ liệu IMDB . Đầu vào đối với generator bao gồm một vector z được chúng tôi khởi tạo ngẫu nhiên cùng với đó là nhóm tuổi giả được chúng tôi mã hóa để đưa về vector kết hợp với vector z để làm đầu vào cho generator thông qua các lớp deconv để tạo ra khuôn mặt (hình 3) .



Hình 3. Generator CGans

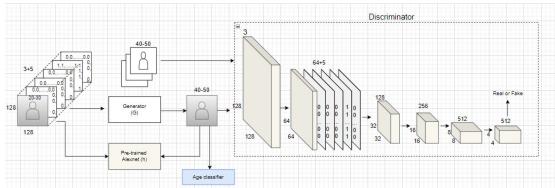
Từ hình ảnh được tạo ra bởi generator, hình ảnh này sẽ được đưa vào discriminator, nhiệm vụ của discriminator là phân biệt hình ảnh được tạo ra bởi generator và hình ảnh thật đã được chuẩn hóa có từ tập dữ liệu IMDB để phân biệt thật giả (hình 4)



Hình 4. Discriminator CGans

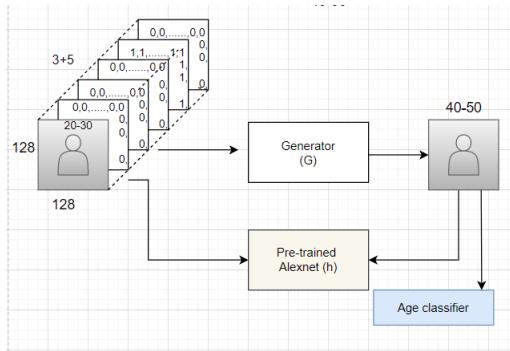
Với việc dùng hàm loss BCE (1) cả 2 generator và discriminator sẽ cùng nhau training và cho ra kết quả ảnh mô phỏng lão hóa khuôn mặt theo độ tuổi .

C. KIẾN TRÚC MẠNG IPCGAN:



Hình 5. Mô phỏng Đa góc nhìn đến khuôn mặt

Hình trên là kiến trúc của mô hình IPCGAN(hình 5) . Ta có thể thấy IPCGAN cũng cấu thành từ hai mô đun chính là Generator và Discriminator

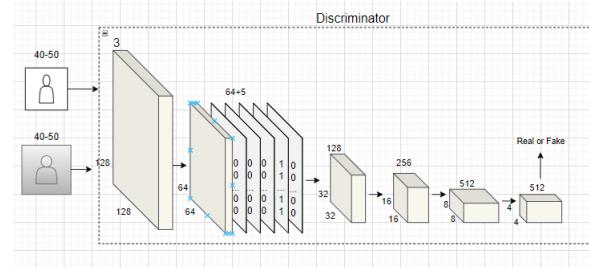


Hình 6. Generator IPCGAN

Từ hình 6 đã có thể thấy ngay sự khác biệt của Generator của IPCGAN và CGAN Khác với đầu vào là một vector z được khởi tạo ngẫu nhiên và một vector mã hóa (hình 3), đầu vào của Generator mô hình IPCGAN là một bức ảnh kèm theo đó là một vector mã hóa thông qua Generator tạo ra một ảnh giả. Thay vì đầu vào là một vector z Generator của CGan sẽ tạo ra một hình ảnh nguêch ngoạc và cần phải có đủ thời gian huấn luyện mô hình mới có thể cho ra hình ảnh chính chu, việc cho dữ liệu đầu vào là hình ảnh sẽ giúp Generator nhanh chóng tạo ra ảnh giả chính chu và gần giống với ảnh đầu vào. Việc đó càng thêm độ tin cậy khi ảnh giả và ảnh đầu vào sẽ cùng được đưa vào mô hình Alexnet để tính toán độ giống nhau từ đó nâng cao chất lượng hình ảnh, thông tin nhận dạng hình ảnh ít bị thiếu đi. Thông thường ảnh giả được tạo ra từ Generator sẽ lấy lại nhãn giả đầu vào để đưa vào Discriminator và sẽ dẫn tới trường hợp ảnh giả tạo ra bị lệch so với nhãn giả đầu vào và rất dễ làm mô hình rơi bao tình trạng overfit và IPCGAN đã được cải tiến điều đó, để tăng tính chính xác cho việc gán nhãn ảnh giả được tạo ra từ Generator thì IPCGAN còn có thêm mô hình Age classification với mục đích là sau khi ảnh được tạo ra sẽ được đưa vào mô hình Age classification đã được train từ tập dữ liệu gốc để cho ra nhãn phù hợp với ảnh giả được tạo ra. Sau khi ảnh giả được tạo ra từ Generator tiếp tục được đưa vào Discriminator với vai trò là ảnh giả kèo theo ảnh gốc để tiến hành phân biệt thật giả (hình 7)

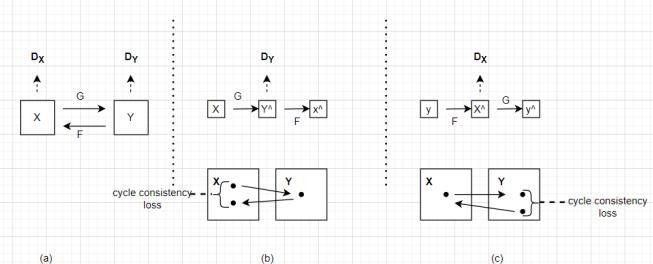
D. KIẾN TRÚC MẠNG CYCLE GAN:

Với bài toán lão hoá khuôn mặt, việc chuyển đổi ảnh từ độ tuổi này sang độ tuổi khác chúng ta có thể quy về dạng



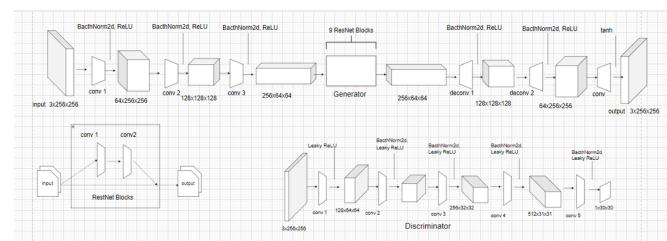
Hình 7. Discriminator IPCGans

image-to-image translation. Dạng bài toán này thường cần một lượng lớn dữ liệu được cặp với nhau - những bộ dữ liệu này rất khó và tốn kém để chuyển bị và trong một số trường hợp không thể chuyển đổi nét riêng như tranh ảnh của các họa sĩ đã mất từ lâu. Vì vậy chúng tôi chọn CycleGan, là một kĩ thuật huấn luyện image-to-image translation mà không cần dữ liệu ghép. Mô hình được huấn luyện không giám sát sử dụng dữ liệu ảnh không ghép và không liên quan đến nhau.



Hình 8. Kiến trúc CycleGan

Kiến trúc của CycleGan bao gồm hai bộ generator $G(X)$, $F(Y)$ và hai bộ discriminators D_x , D_y . Mô hình huấn luyện miền X với G , D_y và miền Y được huấn luyện với F , D_x . Hình ảnh được chuyển đổi từ miền X sang miền Y (hoặc ngược lại) phải nhất quán, điều kiện này được gọi là tính nhất quán của chu kỳ (Cycle Consistency). Phần hình ảnh $G(X)$ được tạo (nằm trong miền Y) được đưa vào Generator $F(Y)$ và được chuyển đổi trở lại thành hình ảnh của miền X, quy trình tương tự đối với hình ảnh được tạo bởi $F(Y)$.



Hình 9. Cấu trúc CycleGan Network

Chúng tôi sử dụng CycleGan gốc được công bố ở bài báo. Phần Generator có Encoder (với các lớp conv), tiếp

theo là 9 blocks Resnet và cuối cùng là Decoder (cùng các lớp conv tương tự Encoder). Phần Discriminator bao gồm các lớp Downsampling với Batchnorm2d và hàm LeakyRelu xen kẽ.

Hàm Loss CycleGan bao gồm hai phần chính: hàm loss của Adversarial Loss của cặp G-D và hàm Cycle Consistency Loss

$$L_{GAN}(F, D_y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)]$$

$$+ \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))]$$

$$L_{GAN}(G, D_x, X, Y) = \mathbb{E}_{x \sim p_{data}(x)} [\log D_X(x)]$$

$$+ \mathbb{E}_{y \sim p_{data}(y)} [\log(1 - D_X(F(y)))]$$

$$L_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\| F(G(x)) - x \|_1]$$

$$+ \mathbb{E}_{y \sim p_{data}(y)} [\| G(F(y)) - y \|_1]$$

Hàm Cycle Consistency Loss để ngăng G và F trái ngược với nhau. Trong bài toán lão hoá khuôn mặt, CCL đảm bảo bảo toàn danh tính trong quá trình lão hoá. Cuối cùng hàm mục tiêu của CycleGan là tổng hợp các hàm loss trên

$$L(G, X, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y)$$

$$+ L_{GAN}(F, D_X, Y, X)$$

$$+ \lambda L_{cyc}(G, F)$$

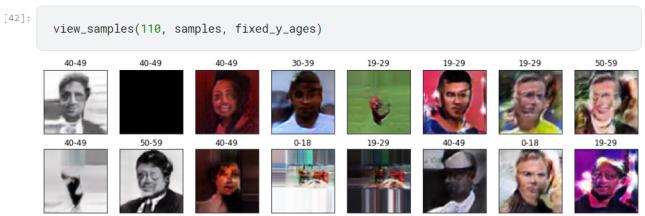
IV. TẬP DỮ LIỆU

Dữ liệu được chúng tôi chọn sử dụng trong bài toán này bao gồm hai tập dữ liệu lần lượt là IMDB (wikidatasets)[19] và CARC2000 (cross-age reference coding)[18]. Đổi với tập dữ liệu IMDB bao gồm 120k ảnh và được chia thành 5 nhóm tuổi bao gồm lần lượt các khoảng tuổi là 0-18, 19-29, 30-39, 40-49, 50-59. Tuy nhiên về tính trung thực của bộ dữ liệu IMDB chưa thực sự được cao khi hầu hết các tấm ảnh ở các giá trị khoảng tuổi đều bị đánh sai tuổi khá là nhiều đặc biệt là độ tuổi 0-18 hầu như các hình ảnh đều là người lớn. Các hình ảnh trong tập dữ liệu hầu hết là ảnh màu tuy nhiên vẫn có khá nhiều ảnh xám, đen. Khác với tập dữ liệu IMDB thì toàn bộ hình ảnh trong tập dữ liệu CARC2000 là ảnh màu và được đánh nhãn với số tuổi có độ trung thực cao. Số nhóm tuổi của tập dữ liệu CARC2000 bằng với số nhóm tuổi của tập dữ liệu IMDB tuy nhiên về khoảng tuổi của các nhóm tuổi có 1 chút sự khác biệt khi CARC2000 chia nhóm tuổi theo các khoảng như sau <20, 20-29, 30-39, 40-49, >50.

V. QUÁ TRÌNH HUẤN LUYỆN VÀ KẾT QUẢ

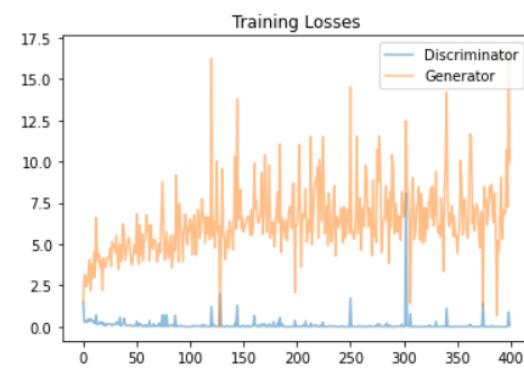
A. Train mô hình CGan

Như đã nói ở phần kiến trúc của cgans , sau khi xây dựng mô hình và có được bộ dữ liệu IMDB, chúng tôi tiến hành huấn luyện mô hình CGans. Do hạn chế về cấu hình cũng như bộ dữ liệu IMDB có số lượng ảnh quá lớn, chúng tôi quyết định sử dụng 12000 ảnh trong tập dữ liệu IMDB kết hợp với nền tảng Kaggle để tiến hành train mô hình cgans. Chúng tôi train thử nghiệm mô hình với 100 epoch đầu tiên trong vòng 5 giờ cùng với các thông số tối ưu hóa learning rate = 0.0001, các hệ số beta1 beta2 lần lượt là 0.5 và 0.999 kết quả chưa được tốt (hình 10) Nhận thấy được kết quả của mô hình



Hình 10. Kết quả dự đoán mô hình CGan 100 epoch

CGans sau 5 giờ huấn luyện chưa thực sự tốt, chúng tốt đã cố gắng huấn luyện mô hình lâu hơn. Trong lần huấn luyện này chúng tôi đã tăng thêm 100 epoch so với lần huấn luyện đầu và các thông số khác như learning rate, beta1,beta2 chúng tôi vẫn giữ nguyên. Tổng thời gian huấn luyện mô hình lần này hết 11 giờ, tuy nhiên kết quả cũng không có khả quan hơn so với lần trước. Cụ thể là mô hình cgans sau 200 epoch có dấu hiệu bị overfit, độ lỗi có dấu hiệu càng ngày càng tăng (hình 11)



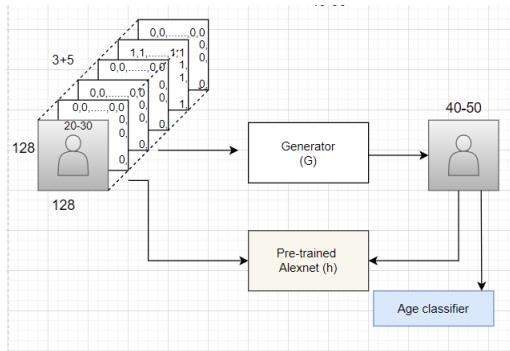
Hình 11. Loss của Generator và Discriminator

Và kết quả generator của mô hình CGans cũng chưa được tốt (hình 12)

B. Train mô hình IPCGan

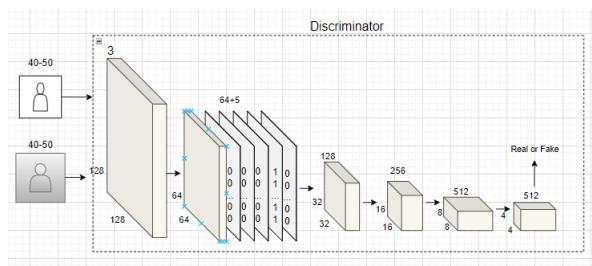
Với kết quả chưa tốt từ việc huấn luyện mô hình CGans ở trên chúng tôi quyết định sử dụng mô hình IPCGAN để áp dụng đối với bài toán và tập dữ liệu CARC2000 được sử dụng

thay thế cho tập dữ liệu IMDB. Khác với đầu vào của generator của mô hình CGans là vector z được khởi tạo ngẫu nhiên cùng với vector onehot đóng vai trò kiểm soát đầu ra thì generator của mô hình IPCGANS với đầu vào là các hình ảnh theo kèm là các vector one hot cùng kích cỡ, việc này sẽ giúp tối ưu thời gian học cho generator và cho ra kết quả sát so với ảnh đầu vào. Để đảm bảo rằng ảnh được tạo ra từ generator giống khuôn mặt, cụ thể là gần giống với ảnh đầu vào thì chúng tôi tiến hành chúng tôi cho xây dựng mạng alexnet dựa trên mô hình có sẵn (pretrain) để tiến hành kiểm tra sự sai lệch giữa ảnh tạo ra và ảnh gốc. IPCGAN ưu việt hơn mô hình cgans ở việc trong khi ảnh được tạo ra bởi generator của mô hình cgans được gán lại vector onehot đầu vào làm nhãn dẫn tới việc ảnh được tạo nhiều khi không phù hợp nhãn ban đầu và làm cho mô hình rơi vào tình trạng overfitting. Nhận thấy điều này thì IPCGAN đã cải tiến hơn, để tăng tính chính xác việc gán nhãn cho ảnh được tạo ra là hợp lý thì IPCGAN còn có mô hình phân loại nhóm tuổi (age classification). Ảnh được tạo ra từ mô hình generator sẽ được đưa vào mô hình phân loại nhóm tuổi được huấn luyện từ tập dữ liệu CARC2000 để cho ra nhóm tuổi phù hợp với ảnh được tạo (hình 13).



Hình 12. Generator IPCGAN

Sau khi qua các tiền trình trên , ảnh được tạo bởi generator tiếp sẽ được đưa vào discriminator cùng với ảnh từ bộ dữ liệu sẽ tiến hành phân loại ảnh thật ảnh giả để phản hồi thông tin cho generator có thông tin để tạo ra ảnh sau cho kết quả tốt hơn . Nhìn chung chức năng của discriminator của IPCGANS và CGAN là giống như nhau , tuy nhiên thì về kiến trúc thì discriminator của IPCGAN có phần phức tạp hơn nhằm mục đích bóc tách , đảm bảo thông tin đầy đủ (hình 14).



Hình 13. Discriminator của IPCGAN

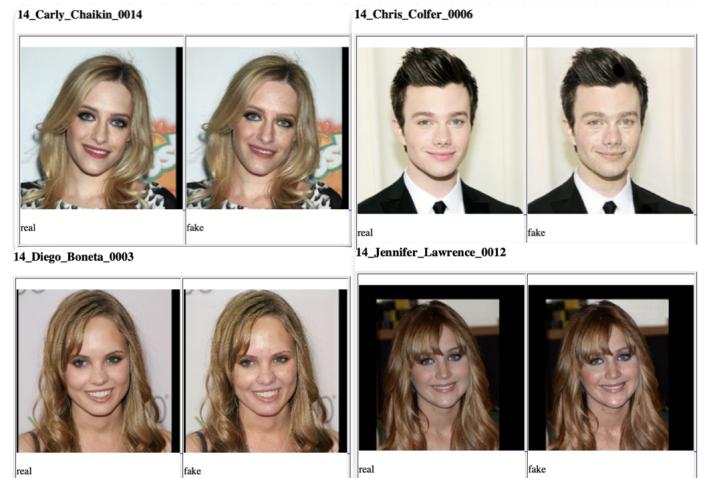
Việc có thêm một số thay đổi về kiến trúc đã giúp IPCGAN có kết quả tốt hơn rất nhiều so với kết quả của CGans , tuy nhiên do việc huấn luyện mô hình trên local còn nhiều hạn chế về phần cứng nên việc bổ xung thêm dữ liệu để huấn luyện mô hình còn gặp khó khăn , dưới đây là các kết quả dự đoán của mô hình IPCGAN cho bài toán của chúng tôi



Hình 14. Kết quả dự đoán của mô hình IPCGAN

C. Train mô hình Cycle-Gan

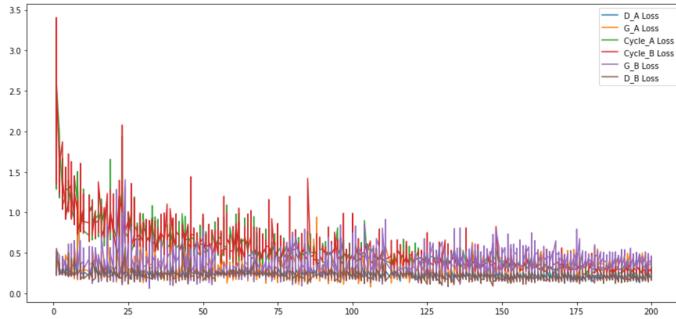
Một trong những vấn đề lớn của CycleGAN là việc training mất khá nhiều thời gian để nó thể hội tụ ở mức tương đối tốt. Chúng tôi training 4 network với 28M tham số (11.4M với mỗi generator và 2.8M cho mỗi discriminator). Vì vậy chúng tôi tăng kích cỡ batch size là 8 và huấn luyện dual GPU T4, áp dụng thêm kĩ thuật transfer learning với 4500 ảnh cho việc huấn luyện chúng tôi chỉ mất 612s cho một epoch và sấp sỉ 34 giờ để hoàn thành training



Hình 15. Kết quả predict CycleGAN giữa ảnh thật và giả

Trong quá trình huấn luyện, chúng tôi nhận thấy rằng quá trình làm biến đổi khuôn mặt giữa các mốc tuổi không thật sự rõ rệt đồng thời việc huấn luyện trên các mốc tuổi cũng tốn nhiều thời gian vì tính chất của mô hình (mỗi lần huấn luyện là một mốc tuổi, chúng tôi cần phải huấn luyện $5 \times 34 = 170$ giờ) Vì vậy chúng tôi chọn huấn luyện chuyển đổi từ mốc tuổi 20 sang 50+ tuổi để thể hiện rõ quá trình lão hóa của khuôn mặt.

Dựa vào ảnh trên ta có thể thấy Cycle Loss A, Cycle Loss B nhanh chóng giảm dần. Điều này có thể giải thích rằng



Hình 16. Log loss CycleGan

CycleGan đang cố gắng điều chỉnh tính nhất quán chu kì của hình ảnh. Trong 100 epoch đầu tiên cả Generator lẫn Discriminator không thực sự ổn định. Tuy nhiên theo thời gian có thể thấy xu hướng đang giảm dần và ổn định sau 200 epoch. Kết hợp với thực tế hình ảnh được tạo ra rõ nét hơn và biến đổi lão hóa được thể hiện rõ trên ảnh hơn.

VI. SO SÁNH VÀ KẾT LUẬN

Sau thời gian huấn luyện 3 mô hình CGans, IPCGAN, CycleGan cho ra các kết quả, chúng tôi thấy kết quả của 3 mô hình có phần lệch nhau khác nhiều, cụ thể như sau:

Kết quả của mô hình CGan: mô hình có dấu hiệu overfit khi càng huấn luyện độ lỗi càng tăng. Kết quả ảnh generator tạo ra với độ hoàn thiện còn thấp một phần là do bộ dữ liệu IMDB sai lệch trong việc gán nhãn cho các ảnh, trong tập dữ liệu còn nhiều ảnh nhiễu như ảnh xám, ảnh đen. Số lượng ảnh của từng nhóm tuổi nhiều kéo theo độ đa dạng, đặc trưng nhiều CGans chưa thể học kịp. Ngoài ra CGans chúng tôi sử dụng ở mức cơ bản chưa có các phương pháp giúp tăng tính chính xác khâu generator ảnh giả.

Kết quả của mô hình IPCGan: với kiến trúc cũng như sự tăng cường về khâu tăng tính chính xác của việc tạo ảnh giả đã mang lại kết quả tốt hơn so với mô hình CGans. Ảnh được đưa vào để lão hóa theo các nhóm tuổi vẫn giữ được các đặc trưng của khuôn mặt tuy nhiên vì các nhóm tuổi cách nhau không xa dẫn tới việc lão hóa mặt không có quá nhiều khác biệt giữa các nhóm tuổi liền kề nhau.

Kết quả của mô hình CycleGan: với việc giảm đi số nhóm tuổi và tạo ra khoảng cách lớn giữa các nhóm tuổi đã giúp CycleGan chiến thắng trong việc lão hóa mặt theo nhóm tuổi so với 2 mô hình CGan và IPCGan. Khác biệt giữa ảnh gốc và ảnh được tạo ra là rõ rệt. Ảnh được lão hóa vẫn có những nét giống so với ảnh gốc.

Tổng kết lại : mô hình Cycle Gans cho kết quả tốt nhất , sau đó là mô hình IPCGANs và cuối cùng là mô hình Cgans

- [2] I. Kemelmacher-Shlizerman, S. Suwananakorn, and S. M. Seitz. Illumination-aware age progression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3334–3341, 2014.
- [3] Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. IEEE transactions on pattern analysis and machine intelligence, 32(11):1955–1976, 2010.
- [4] G. Panis and A. Lanitis. An overview of research activities in facial age estimation using the fg-net aging database. In European Conference on Computer Vision, pages 737–750. Springer, 2014.
- [5] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, pages 341–345. IEEE, 2006.
- [6] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In Proceedings of the European Conference on Computer Vision (ECCV), 2014.
- [7] R. Rothe, R. Timofte, and L. V. Gool. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision (IJCV), July 2016.
- [8] I. Kemelmacher-Shlizerman, S. Suwananakorn, and S. M. Seitz. Illumination-aware age progression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3334–3341, 2014.
- [9] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(3):385–401, 2010.
- [10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672–2680, 2014.
- [11] M. Mirza and S. Osindero. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784, 2014.
- [12] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028, 2017.
- [13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image to-image translation with conditional adversarial networks. arXiv preprint arXiv:1611.07004, 2016.
- [14] Goodfellow, I.J., et al.: Generative adversarial nets. In: NIPS (2014).
- [15] Yin, X., Yu, X., Sohn, K., Liu, X., Chandraker, M.: Towards Large-pose face frontalizationin the wild. In: ICCV (2017).
- [16] Tian, Y., Peng, X., Zhao, L., Zhang, S., Metaxas, D.N.: CR-GAN: learning completerepresentations for multi-view generation (2018).
- [17] <https://phamdinhkhanh.github.io/2020/07/13/GAN.html31-nguy>
- [18] <https://bcsiriuschen.github.io/CARC>
- [19] <https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki>

TÀI LIỆU

- [1] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe. Recurrent face aging. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2378–2386, 2016.