Trong ánh xạ miền không giám sát (unsupervised domain mapping), người học được cung cấp hai tập dữ liệu chưa có A và B. Mục tiêu là tìm hiểu ánh xạ biến một mẫu trong A sang mẫu tương tự trong B.

Gần đây CycleGAN và DiscoGAN đã chỉ ra rằng khi học cùng một ánh xạ từ A đến B và ánh xạ ngược từ B đến A, ta sẽ thu được các ánh xạ có độ chính xác cao. Trong bài báo này, chúng ta sẽ tìm hiểu một phương pháp học lập ánh xạ từ A đến B mà không có ánh xạ nghịch đảo (từ B đến A). Điều này được thực hiện bằng cách học một ánh xạ duy trì khoảng cách giữa một cặp mẫu A và B.

Bài báo nhóm em tìm hiểu chỉ trình bày các kết quả thử nghiệm rằng phương pháp mới không chỉ cho phép học tập lập ánh xạ một chiều mà còn dẫn đến các kết quả số thích hợp hơn so với ràng buộc dựa trên vòng tròn hiện tại.

1. Giới thiệu

Sự ra đời của công nghệ Generative Adversarial Network (GAN) đã cho phép tạo ra các hình ảnh thực tế giống một tập huấn luyện nhất định bằng cách ghi lại chính xác những gì nằm trong lớp đã cho và phân biệt cái gì là giả. Trong số nhiều công việc được thực hiện bởi GAN, việc ánh xạ hình ảnh trong miền nguồn tới hình ảnh tương tự trong miền đích là sự quan tâm đặc biệt.

Các giải pháp đề xuất cho vấn đề này thường được phân tách bằng số lượng giám sát bắt buộc. Trên cùng một phương pháp, các phương pháp giám sát sử dụng các cặp mẫu phù hợp, một trong mỗi miền, để tìm hiểu ánh xạ. Giám sát trực tiếp ít hơn đã được chứng minh bằng cách sử dụng ánh xạ vào một không gian ngữ nghĩa và yêu cầu mẫu ban đầu và mẫu tương tự trong miền đích (target domain) cùng chia sẻ một biểu diễn ngữ nghĩa.

Nếu hai miền có liên quan cao, nó đã được chứng minh rằng chỉ bằng cách chia sẻ trọng số giữa các mạng hoạt động trên hai miền và không cần giám sát thêm, người ta có thể ánh xạ mẫu giữa hai miền. Đối với các miền xa hơn, nó đã được chứng minh gần đây rằng bằng ánh xạ đối xứng nghiêng theo cả hai hướng, ta sẽ thu được điều tương tự như trên. Điều này được thực hiện bằng cách yêu cầu tính tròn, tức là ánh xạ mẫu từ miền này sang miền khác và sau đó quay lại, tạo mẫu gốc.

Trong bài báo này, chúng ta tiến thêm một bước nữa và cho thấy rằng có thể tìm hiểu ánh xạ giữa miền nguồn và miền đích theo cách không giám sát một chiều (one-sided unsupervised), bằng cách thực thi tương quan chéo giữa các khoảng cách tương ứng được tính toán trong mỗi miền. Ràng buộc mới cho phép lập ánh xạ một chiều, trong các thí nghiệm của tác giả, kết quả số tốt hơn so với tính tròn. Kết hợp cả hai ràng buộc này với nhau thường dẫn đến những cải tiến hơn nữa.

Việc tìm hiểu ràng buộc mới đòi hỏi phải so sánh các cặp mẫu

* 1. Công việc liên quan

Style transfer:

Những phương thức này thường nhận được dưới dạng đầu vào một kiểu hình ảnh với một nội dung hình ảnh và tạo ra một hình ảnh mới có kiểu đầu tiên và nội dung của hình ảnh thứ hai. Vấn đề nằm ở chuyển hình ảnh giữa các miền khác nhau vì khi ánh xạ giữa các miền, một phần của nội dung được thay thế bằng nội dung mới khớp với tên miền đích và không chỉ là kiểu. Tuy nhiên, sự khác biệt là không sắc nét, và nhiều ví dụ lập ánh xạ miền chéo trong tài liệu gần như có thể được xem như là style transfer. Ví dụ, một con ngựa vằn và một con ngựa, con ngựa được ánh xạ từ ngựa vằn, thay đổi làn da của ngựa thành làn da ngựa vằn.

Generative Adversarial Networks:

Các phương thức GAN đào tạo một mạng sinh G (a generator network G) để tổng hợp các mẫu từ một phân phối đích (target distribution), cho các vectơ nhiễu, cùng nhau đào tạo mạng thứ hai D. Trong ánh xạ hình ảnh, hình ảnh được tạo dựa trên hình ảnh đầu vào chứ không dựa trên dữ liệu nhiễu ngẫu nhiên.

GAN được sinh ra với kỳ vọng tạo ra được những hệ thống có độ chính xác cao mà cần ít hoạt động của con người trong khâu huấn luyện.

Generative Adversarial Network - tạm dịch Mạng Chống đối Tạo sinh - hình thành trên ý tưởng về sự cạnh tranh của hai mạng neural network:

* Discriminative network (mạng phân biệt): có nhiệm vụ cố gắng phân biệt giữa data thật và data giả mạo.
* Generative network (mạng sinh): sinh ra data giả, và mục tiêu là làm sao sinh ra được các data giống với data thật nhất, làm Discriminator network (mạng phân biệt) không thể phân biệt đâu là data thật đâu là data giả. Các sample được sinh ra dựa trên các latent code (mã ngầm) z

Unsupervised Mapping

Đây là việc liên quan nhiều nhất đến công việc của nhóm nghiên cứu trong bài báo, không sử dụng giám sát ngoại trừ các hình ảnh mẫu từ hai miền. Điều này đã được thực hiện rất gần đây trong hình ảnh để dịch hình ảnh và hơi sớm hơn để dịch giữa các ngôn ngữ tự nhiên. Lưu ý rằng đề xuất phương pháp "GAN với tổn thất tái thiết", áp dụng hạn chế chu kỳ ở một bên và chỉ đào tạo một GAN. Tuy nhiên, không giống như phương thức của chúng ta, phương thức này yêu cầu sự phục hồi của cả hai ánh xạ và được thực hiện tốt hơn bằng phương thức hai chiều đầy đủ.

Phương thức CoGAN, tìm hiểu ánh xạ từ một vector đầu vào ngẫu nhiên đến các mẫu phù hợp từ hai miền. Nó đã được chứng minh rằng phương pháp có thể được sửa đổi để thực hiện dịch miền.

Weakly Supervised Mapping

Sự kết hợp giữa miền nguồn và miền đích được thực hiện bằng cách kết hợp một ánh xạ cố định được huấn luyện trước f và yêu cầu f không thay đổi, tức là các hoạt động của f là giống nhau đối với các mẫu đầu vào và cho các mẫu được ánh xạ.

Supervised Mapping

Khi được cung cấp các cặp kết hợp (hình ảnh đầu vào, hình ảnh đầu ra), việc giám sát có thể được thực hiện trực tiếp. Một ví dụ về phương thức như vậy cũng sử dụng GAN, trong đó người phân biệt D nhận được một cặp hình ảnh có một hình ảnh nguồn và hình ảnh kia là hình ảnh đích (cặp "thực") hoặc một hình ảnh được tạo ra (cặp "giả") ). Việc liên kết giữa nguồn và ảnh đích được tăng cường hơn nữa bằng cách sử dụng kiến trúc U-net.

Domain adaptation

Trong cài đặt này, chúng tôi thường được cung cấp hai miền, một miền có giám sát ở dạng nhãn phù hợp, trong khi thứ hai có ít hoặc không có giám sát. Mục tiêu là tìm hiểu cách gắn nhãn các mẫu từ miền thứ hai. Điều gì là phổ biến cho cả hai lĩnh vực và những gì là riêng biệt được tách ra để cải thiện trên các mô hình hiện có. Một phép biến đổi được học ở cấp độ pixel, từ một tên miền này sang miền khác, bằng cách sử dụng GAN. Một phương pháp tiếp cận đối lập không giám sát đối với phân đoạn ngữ nghĩa, sử dụng cả kỹ thuật thích ứng miền cụ thể và toàn cầu, được đề xuất.

1. Preliminaries (Sơ bộ)

i.i.d: independent and identically distributed random variables

(Tạm dịch: Biến ngẫu nhiên phân tán độc lập và phân biệt)

Trong bài toán không được giám sát, thuật toán được cung cấp với các tập dữ liệu không có nhãn từ hai miền, A và B. Tập dữ liệu thứ nhất bao gồm các mẫu i.i.d từ phân phối pA và tập dữ liệu thứ hai bao gồm các mẫu i.i.d từ phân phối pB. Chính thức, được đưa ra



mục tiêu của nhóm nghiên cứu là tìm hiểu hàm GAB, ánh xạ các mẫu trong miền A đến các mẫu tương tự trong miền B (xem ví dụ bên dưới). Cần phải đồng thời khôi phục hàm GBA thứ hai, tương tự ánh xạ các mẫu trong miền B tới các mẫu tương tự trong miền A.

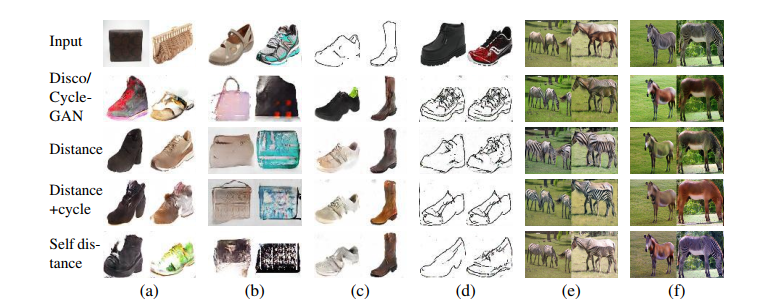
Justification

Để cho phép học không giám sát của một ánh xạ có hướng, chúng tôi giới thiệu ràng buộc rằng các cặp đầu vào x, x', ở khoảng cách nhất định với nhau, được ánh xạ tới cặp đầu ra GAB (x), GAB (x') với khoảng cách tương tự, tức là khoảng cách || x - x' || và ||GAB(x) - GAB(x')|| có tương quan cao. Như chúng tôi trình bày dưới đây, có thể giả định rằng ràng buộc này xấp xỉ trong nhiều tình huống được chứng minh bởi công việc trước đây về dịch miền. Mặc dù là ước chừng, nó vẫn có khả năng, việc ánh xạ giữa các miền chỉ yêu cầu sự giám sát ít hơn của yêu cầu ở trên để phân phối đầu ra của việc ánh xạ sẽ trùng khớp với phân phối đích (phân phối mục tiêu).

Mỗi bộ ba hiển thị gồm (Nguồn: túi xách), (đích: giày) được tạo bởi ánh xạ GAB của CycleGAN và kết quả xấp xỉ của GAB bằng một phép biến đổi tuyến tính phi định tính T, thu được mỗi điểm ảnh đầu ra dưới dạng kết hợp tuyến tính của các pixel đầu vào. Việc chuyển đổi tuyến tính nắm bắt được bản chất của GAB cho thấy rằng nhiều ánh xạ được thực hiện bằng một phép biến đổi không gian cố định. (Xem hình 1 bên dưới)



Xem xét, ví dụ, trường hợp ánh xạ giày tới các cạnh, như được trình bày trong Hình 4. Trong trường hợp này, các điểm cạnh chỉ đơn giản là một tập con của tọa độ ảnh, được chọn bởi tiêu chí hình ảnh cục bộ. Nếu hình ảnh x là trực quan tương tự như hình ảnh x ', có khả năng là các ánh xạ cạnh của chúng tương tự nhau. Trong thực tế, sự tương tự này nhấn mạnh việc sử dụng thông tin gradient trong tài liệu thị giác máy tính cổ điển. Vì vậy, trong khi khoảng cách được dự kiến sẽ khác nhau trong hai miền, người ta có thể mong đợi một mối tương quan cao. (Xem hình 4 bên dưới)



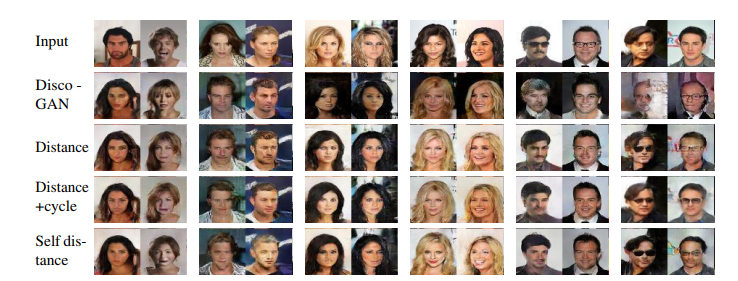
Tiếp theo, hãy xem xét trường hợp của túi xách để lập ánh xạ giày (Hình 4). Tương tự có xu hướng có cùng phân phối màu sắc hình ảnh trong các hình ảnh khác nhau. Giả sử rằng các vị trí pixel không gian của túi xách theo phân phối chặt chẽ (tức là, tập hợp hình ảnh túi xách có cùng hình dạng) và cùng lưu giữ giày, sau đó tồn tại một thiết lập các trường hoán vị kinh điển chuyển đổi túi xách thành giày. Nếu có một hoán vị, đó sẽ là một hoán vị cố định các vị trí pixel, khoảng cách sẽ được giữ nguyên. Trong thực tế, việc chuyển đổi hình ảnh phức tạp hơn.

Để nghiên cứu xem mô hình hoán vị là một xấp xỉ hợp lệ, chúng ta đã học được một phép biến đổi tuyến tính không âm tính T ∈ R+(64^2 × 64^2) ánh xạ, một hướng tại một thời điểm, hình ảnh túi xách có kích thước 64 × 64 × 3 cho hình ảnh giày đầu ra có cùng kích thước được đưa ra bởi phương pháp CycleGAN. Các cột của T có thể được hiểu là trọng số xác định sự lan truyền của khối lượng trong hình ảnh đầu ra cho mỗi vị trí pixel trong hình ảnh đầu vào. Nó được ước tính bằng cách giảm thiểu sai số bình phương của ánh xạ mọi hướng (R, G hoặc B) của một hình ảnh túi xách đến cùng một hướng trong việc trùng khớp với giày. Việc tối ưu hóa được thực hiện bằng độ dốc gốc với phép chiếu tới không gian của các ma trận không âm, nghĩa là, đánh số các phần tử âm của T tại mỗi lần lặp.

Các ánh xạ mẫu của ma trận T được thể hiện trong hình 1. Như có thể thấy, phép biến đổi tuyến tính không âm tính xấp xỉ CNN GAB nhiều lớp của CycleGAN ở một mức độ nào đó. Kiểm tra các phần tử của T, chúng chia sẻ một số thuộc tính với hoán vị: tổng trung bình của các hàng là 1.06 (SD 0.08) và 99.5% các phần tử dưới 0,01.

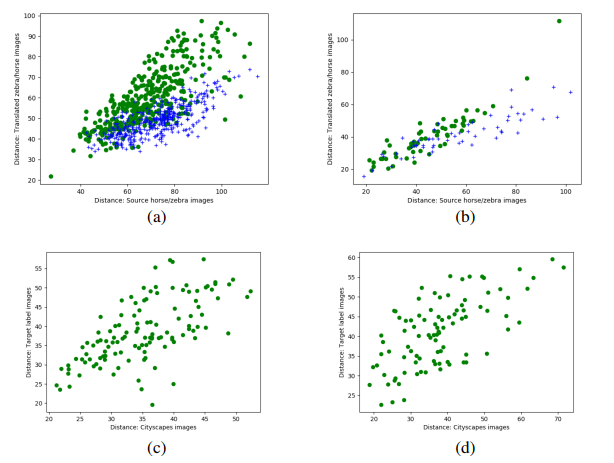


Trong trường hợp thêm kính hoặc thay đổi giới tính hoặc màu tóc (Hình 3 bên dưới), một sửa đổi hình ảnh tương đối nhỏ, không thay đổi đáng kể phần lớn thông tin hình ảnh, đủ để tạo hiệu ứng hình ảnh mong muốn. Thay đổi như vậy có khả năng duy trì phần lớn khoảng cách hình ảnh hai chiều trước và sau khi chuyển đổi.



Khi ánh xạ giữa ngựa với ngựa vằn, kết cấu của da ngựa được chuyển thành hình ngựa vằn. Trong trường hợp này, hầu hết các thông tin hình ảnh không bị ảnh hưởng và phần được thay đổi được sửa đổi bởi một kết cấu đồng nhất, một lần nữa duy trì khoảng cách hai chiều. Trong hình 2 (a), chúng tôi so sánh khoảng cách L1 trong không gian RGB của các cặp hình ảnh con ngựa với khoảng cách của các mẫu sau khi ánh xạ bởi mạng CycleGAN được thực hiện. Rõ ràng là mối tương quan giữa các miền chéo với các cặp khoảng cách là cao. Chúng tôi cũng xem xét hình ảnh Cityscapes và các cặp ground truth trong Hình 2 (c), và thấy rằng có sự tương quan cao giữa khoảng cách. Đây cũng là trường hợp trong nhiều ánh xạ dựa trên tài liệu khác giữa các tập dữ liệu mà chúng ta đã thử nghiệm và các cặp ground truth cơ bản. (ground truth chính là nhãn/label/đầu ra thực sự của các điểm trong test data)

Mặc dù có ít nhược điểm khi làm việc với các cặp hình ảnh đào tạo so với làm việc với các hình ảnh đơn lẻ, để nghiên cứu thêm lượng thông tin cần thiết để căn chỉnh thành công, chúng tôi cũng xem xét khoảng cách giữa hai nửa của cùng một hình ảnh. Chúng tôi so sánh khoảng cách L1 giữa nửa bên trái và bên phải như được tính trên hình ảnh đầu vào với hình ảnh thu được trên hình ảnh được tạo hoặc hình ảnh chân lý tương ứng. Hình 2(b) và hình 2(d) trình bày kết quả cho ngựa để dịch ngựa vằn và cho cặp hình ảnh và nhãn Cityscapes tương ứng. Như có thể thấy, sự tương quan cũng rất quan trọng trong trường hợp này.



1. Experiment (Thí nghiệm)

Chúng tôi so sánh nhiều phương pháp: DiscoGAN hoặc CycleGAN, ánh xạ một chiều sử dụng Ldistance (A → B hoặc B → A).

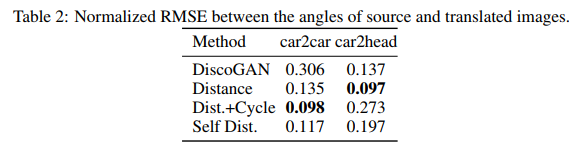
Đối với DiscoGAN, chúng tôi sử dụng hình dạng trọng lượng cố định cho tất cả thử nghiệm.

Đối với CycleGAN, có nhiều độ nhạy với các tham số, mẫu chung được giữ nguyên, chúng tôi sử dụng trọng số khác nhau cho giới hạn khoảng cách tùy thuộc vào thử nghiệm, chữ số hoặc ngựa + ngựa vằn.

DiscoGAN:

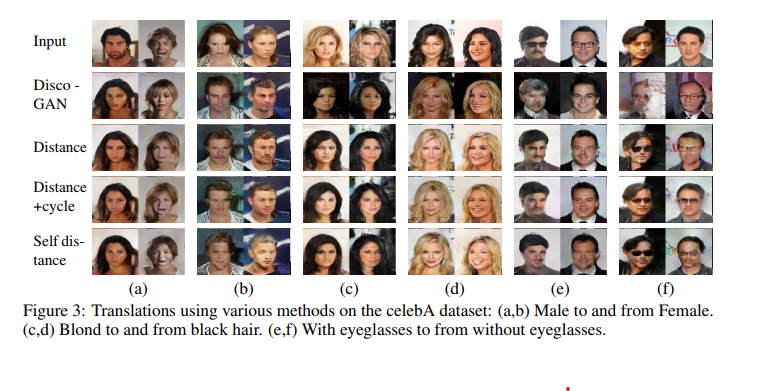
Đối với thử nghiệm car2car, bộ dữ liệu xe hơi được chia thành hai phần, một trong số đó được sử dụng cho A và một cho B (Nó được chia tách thành bộ huấn luyện và kiểm tra). Vì góc quay thể hiện nguồn biến thiên lớn nhất, và do phép quay được chia sẻ giữa các tập dữ liệu, chúng tôi mong đợi nó không thay đổi mà mạng biết được, tức là ánh xạ ngữ nghĩa sẽ bảo toàn các góc.

Một phép hồi quy đã được huấn luyện để tính toán góc của một hình ảnh chiếc xe dựa trên dữ liệu huấn luyện. Bảng 2 cho thấy Sai số trung bình gốc (RMSE) giữa góc của hình ảnh nguồn và hình ảnh đã dịch. Như có thể thấy, ánh xạ dựa trên cặp khoảng cách dẫn đến lỗi thấp hơn so với ánh xạ DiscoGAN, kết quả sẽ được cải thiện nếu kết hợp cả hai, khoảng cách vượt trội hơn cả DiscoGAN và khoảng cách hai chiều. Việc thực hiện DiscoGAN ban đầu đã được sử dụng, nhưng do sự khác biệt về đánh giá (các biến hồi quy khác nhau) những con số này không tương thích với biểu đồ được hiển thị trong DiscoGAN.

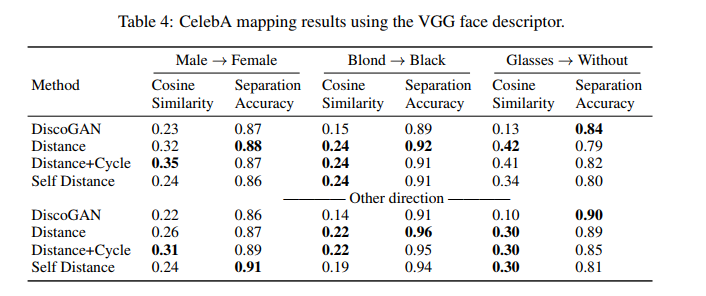


Đối với car2head, giải pháp của DiscoGAN tạo ra hình ảnh phản chiếu và kết hợp ràng buộc tròn của DiscoGAN với giới hạn khoảng cách tạo ra một giải pháp được xoay 90o. Chúng tôi coi những dự kiến này là sự mơ hồ trong ánh xạ chứ không phải là sai lầm và do đó loại bỏ lỗi trung bình trước tính toán RMSE. Trong thử nghiệm này, khoảng cách (Distance) hoạt động tốt hơn tất cả các phương pháp khác. Sự kết hợp của cả hai phương pháp là ít cạnh tranh hơn cả hai, có lẽ vì mỗi phương pháp kéo về phía một giải pháp khác nhau. Self-Distance tệ hơn trong tập dữ liệu này.

Một tập hợp các thí nghiệm khác phát sinh từ việc xem xét các hình ảnh khuôn mặt có và không có một thuộc tính nhất định. Dataset CelebA được chú thích cho nhiều thuộc tính bao gồm giới tính, màu tóc và sự tồn tại của mắt kính trong ảnh. Chúng tôi thực hiện ánh xạ giữa hai giá trị của từng thuộc tính trong số ba giá trị này. Các kết quả được hiển thị trong Hình 3. Rõ ràng là phương pháp DiscoGAN trình bày nhiều trường hợp lỗi hơn phương pháp dựa trên cặp của chúng ta. Phương pháp tự khoảng cách (Self – Distance) được thực hiện với nửa hình ảnh trên cùng và dưới cùng, thay vì từ trái sang phải, vì khuôn mặt là đối xứng. Phương pháp này cũng có vẻ tốt hơn DiscoGAN.



Để đánh giá mức độ dịch của khuôn mặt, chúng tôi sử dụng lớp biểu diễn của VGG trên hình ảnh A và đầu ra của nó trong B. Người ta có thể giả định rằng hai hình ảnh phù hợp sẽ có nhiều tính năng tương tự và do đó biểu diễn VGG sẽ tương tự. Các điểm tương đồng cosin, như được đánh giá giữa các hình ảnh đầu vào và các phiên bản được ánh xạ của chúng, hiển thị trong Tab. 4. Trong mọi trường hợp, cặp khoảng cách sẽ tạo ra nhiều khuôn mặt đầu vào - đầu ra tương tự hơn. Tự thực hiện khoảng cách (Self – distance) không tốt so với các khoảng cách, nhưng thường tốt hơn DiscoGAN. Áp dụng vòng tròn (cycle) cùng với khoảng cách, cung cấp kết quả tốt nhất nhưng yêu cầu, không giống như khoảng cách, học cả hai bên cùng một lúc.

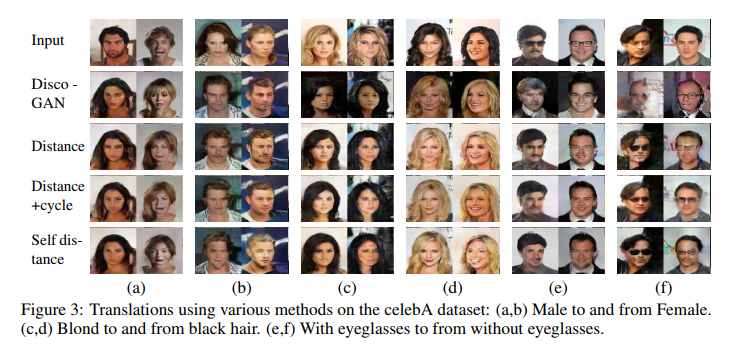


Mặc dù chúng tôi tạo ra hình ảnh phù hợp hơn với chỉ số mô tả khuôn mặt, khả năng tạo hình ảnh chân thật với phân phối thứ hai của chúng tôi không bị suy giảm. Điều này được chứng minh bằng cách học một phân loại tuyến tính giữa hai miền dựa trên các mẫu huấn luyện và sau đó áp dụng nó cho một tập hợp các hình ảnh thử nghiệm trước và sau khi ánh xạ. Độ chính xác phân tách giữa hình ảnh thử nghiệm đầu vào và phiên bản được ánh xạ được hiển thị trong Tab. 4. Có thể thấy, khả năng tách của phương thức của chúng tôi tương tự như của DiscoGAN

Chúng tôi thực hiện 1 nghiên cứu người dùng để đánh giá chất lượng kết quả của chúng tôi. Người dùng lần đầu tiên được xem một tập các hình ảnh thực từ tập dữ liệu. Sau đó, 50 cặp hình ảnh ngẫu nhiên được trình bày cho 1 người dùng trong một giây, một hình ảnh được huấn luyện bằng cách sử dụng DiscoGAN và một hình ảnh sử dụng phương pháp của chúng tôi. Người dùng được yêu cầu quyết định hình ảnh nào trông thực tế hơn. Bài kiểm tra được thực hiện trên 22 người dùng. Trong trường hợp giày để dịch túi xách, bản dịch của chúng tôi hoạt động tốt hơn trên 65% các trường hợp. Đối với túi xách dịch qua giày, điểm số là 87%. Từ nam sang nữ, cả hai phương pháp đều cho thấy điểm số thực tế tương tự (51% đến 49% của DiscoGAN). Do đó, chúng tôi đã hỏi một câu hỏi thứ hai: với khuôn mặt của một người đàn ông, cái mà trong số hai biến thể nữ được tạo ra là phù hợp hơn với khuôn mặt nguyên thủy. Phương pháp của chúng tôi thắng 88% của các lần thí nghiệm.

Ngoài ra, chúng tôi so sánh các tổn thất phân biệt của GAN đối với các phương pháp khác nhau và cho thấy rằng các giá trị này gần giống nhau. Chúng tôi cũng đo lường các tổn thất của các phương pháp khác nhau trong quá trình thử nghiệm, ngay cả khi chúng không được tối ưu hóa trực tiếp. Ví dụ, mặc dù các ràng buộc này không được thực thi, các phương pháp dựa trên khoảng cách dường như trình bày một sự mất mát theo chu kỳ thấp, trong khi DiscoGAN trình bày một tổn thất khoảng cách tương đối cao hơn.

Nhìn chung, kết quả của cách tiếp cận dựa trên khoảng cách dường như tốt hơn DiscoGAN, trong khi kết quả của tự khoảng cách (Self-distance) không tốt lắm. Sự kết hợp giữa DiscoGAN và khoảng cách (Distance) thường hoạt động tốt nhất.

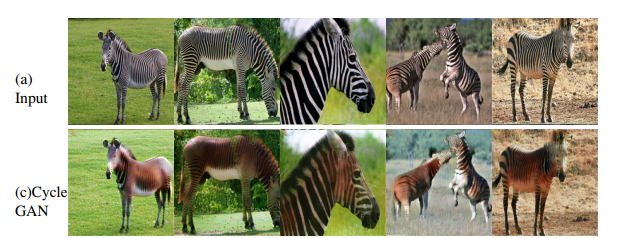


CycleGAN:

Sử dụng kiến trúc CycleGAN chúng tôi lập ánh xạ ngựa cho ngựa vằn, (xem Hình 4 bên dưới). Lưu ý rằng lập ánh xạ ngựa vằn và ngựa, tất cả các phương pháp không thành công mặc dù theo những cách khác nhau. Chủ quan, có vẻ như phương pháp khoảng cách (Distance) + chu kỳ (Cycle) cho thấy được lời giải tích cực nhất trong bản dịch này.

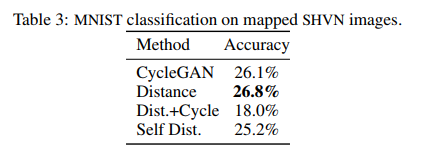


Hình xem thêm:



*Translation from zebra to horse based on the CycleGAN architecture.*

Để có được kết quả bằng số, chúng tôi sử dụng phương pháp CycleGAN cơ bản cũng như các phương pháp của chúng tôi để dịch từ Số nhà (SVHN – Street View House Numbers) sang MNIST. Độ chính xác sau đó được đo trong không gian MNIST bằng cách sử dụng mạng neural được huấn luyện cho nhiệm vụ này. Kết quả được hiển thị trong Tab. 3. Trong khi phương pháp dựa trên cặp khoảng cách được cải thiện theo phương pháp cơ sở, vẫn còn một khoảng cách lớn giữa thiết lập không giám sát và bán giám sát. Điều này có thể được giải thích bởi số lượng lớn thông tin không liên quan trong các hình ảnh SVHN. Kết hợp khoảng cách dựa trên khoảng cách với vòng tròn không hoạt động tốt trên tập dữ liệu này.





*Translating SVHN to MNIST with a CycleGAN architecture*

Chúng tôi cũng thực hiện đánh giá định tính bằng cách sử dụng điểm FCN. Chỉ số FCN đánh giá khả năng tương tác hình ảnh bằng cách chụp ảnh cảnh quan thành phố được tạo và tạo nhãn bằng thuật toán phân đoạn ngữ nghĩa. Nhãn được tạo sau đó có thể được so sánh với nhãn thật (ground truth label). Kết quả FCN được đưa ra: độ chính xác cho mỗi pixel, độ chính xác của lớp và IOU lớp. Phương pháp khoảng cách GAN của chúng tôi là thích hợp hơn trên cả ba điểm (0,53 so với 0,52, 0,19 so với 0,17 và 0,11 so với 0,11, tương ứng). Cặp giá trị t-test và p-values lần lượt là 0,29, 0,002 và 0,42. Trong một nghiên cứu sử dụng tương tự như DiscoGAN trên, bản dịch thành phố của chúng tôi đạt 71% khi so sánh với CycleGAN. Khi nhìn vào sự tương đồng với ground truth image, chúng tôi đạt 68%.

