Nguyễn Huy Quang \_ 15110107

Trần Nguyên Quang \_ 15110108

GVHD : Nguyễn Thiên Bảo

**Đồ án 3**

**Learning Disentangled Representations with Semi-Supervised Deep Generative Models**

Tính trừu tượng

Các bộ mã tự động biến đổi (VAE) học các biểu diễn dữ liệu bằng cách cùng nhau đào tạo một bộ mã hóa và bộ giải mã xác suất. Thông thường, các mô hình này mã hóa tất cả các tính năng của dữ liệu thành một biến duy nhất. Ở đây, chúng tôi quan tâm đến việc học các biểu diễn bị xáo trộn mã hóa các khía cạnh riêng biệt của dữ liệu thành các biến riêng biệt. Chúng tôi đề xuất tìm hiểu các biểu diễn như vậy bằng cách sử dụng các kiến ​​trúc mô hình tổng quát từ các VAE chuẩn, sử dụng một cấu trúc mô hình đồ họa chung trong bộ mã hóa và bộ giải mã. Điều này cho phép chúng tôi đào tạo các mô hình một phần cụ thể đưa ra các giả định tương đối mạnh về một tập con các biến có thể diễn giải và dựa vào tính linh hoạt của mạng nơron để tìm hiểu các biểu diễn cho các biến còn lại. Chúng tôi tiếp tục xác định mục tiêu chung cho việc học bán giám sát trong lớp mô hình này, có thể xấp xỉ bằng cách sử dụng quy trình lấy mẫu quan trọng , có thể được ước tính gần bằng cách sử dụng quy trình lấy mẫu quan trọng. Chúng tôi đánh giá khả năng của chúng tôi trong việc tìm hiểu các biểu diễn bị xáo trộn, cả bằng cách thăm dò chất lượng khả năng sinh sản của nó và đánh giá định lượng khả năng phân biệt đối xử của nó trên nhiều mô hình và tập hợp dữ liệu khác nhau.

**1. Giới thiệu**

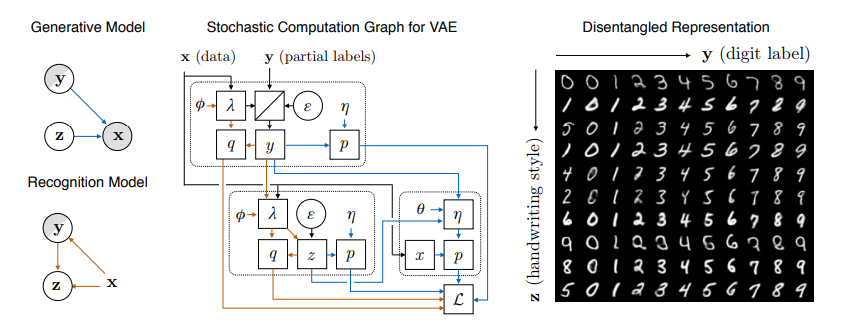
Việc biểu diễn học tập từ dữ liệu là một trong những thách thức cơ bản trong học máy và trí tuệ thông minh. Đặc điểm của đại diện đã học có thể phụ thuộc vào mục đích sử dụng của chúng. Với mục đích giải quyết một nhiệm vụ duy nhất, đặc điểm chính cần thiết là sự phù hợp cho nhiệm vụ đó. Tuy nhiên, nỗ lực lặp đi lặp lại lãng phí. Một biểu diễn có một số cấu trúc đáng tin cậy, và ngữ nghĩa nhất quán liên quan đến các phần khác nhau, có nhiều khả năng khái quát hóa một nhiệm vụ mới. Các mô hình sinh sản xác suất cung cấp một khuôn khổ chung cho các biểu diễn học tập: một mô hình được xác định bằng phân bố xác suất chung trên dữ liệu và các biến ngẫu nhiên tiềm ẩn, và biểu diễn có thể được tìm thấy bằng cách xem xét sau về biến tiềm ẩn cho dữ liệu đặc biệt. Biểu diễn được học - có nghĩa là, giá trị phỏng đoán của các biến tiềm ẩn - phụ thuộc không chỉ trên dữ liệu mà còn trên mô hình sinh sản trong việc lựa chọn biến tiềm ẩn và mối quan hệ giữa biến tiềm ẩn và dữ liệu. Có hai thái cực của phương pháp tiếp cận để xây dựng các mô hình sinh sản. Ở một đầu là các mô hình đồ họa xác suất đầy đủ cụ thể, quyết định tất cả các biến tiềm ẩn có trong phân phối chung, mối quan hệ giữa chúng, và dạng chức năng của các phân phối có điều kiện mà không xác định mô hình. Ở đầu kia là mô hình sinh trưởng sâu [8, 17, 20, 21], áp đặt rất ít giả định về cấu trúc của mô hình, thay vì sử dụng mạng nơron làm appro các bộ xấp xỉ hàm có thể được sử dụng để huấn luyện phân phối có điều kiện trên dữ liệu, thay vì chỉ định nó bằng tay. Sự cân bằng là rõ ràng.

Trong một mô hình đồ họa được xây dựng một cách rõ ràng, cấu trúc và hình thức của sự phân bố chung đảm bảo rằng các biến tiềm ẩn sẽ có ngữ nghĩa cụ thể, tạo ra một biểu diễn bị xáo trộn. Thật không may, việc xác định một mô hình xác suất tốt là khó khăn: trong các lĩnh vực cảm nhận phức tạp như thị giác, kỹ thuật tính năng mở rộng (ví dụ: Berantetal. [1], Siddharthetal. [31]) có thể cần thiết để xác định chức năng thích hợp. Mô hình sinh trưởng sâu hoàn toàn vượt qua những khó khăn của kỹ thuật tính năng. Mặc dù chúng giải quyết các biểu diễn học tập, sau đó cho phép chúng tái tạo lại dữ liệu tốt hơn, nhưng bản thân các bài trình bày không có ý nghĩa thống nhất theo các trục biến thể: chúng tạo ra các biểu diễn vướng víu. Trong khi các cách tiếp cận như vậy có giá trị đáng kể, đặc biệt khi phải đối mặt với sự vắng mặt của bất kỳ thông tin phụ nào về dữ liệu, thường có các tình huống khi các khía cạnh của biến thể trong dữ liệu có thể được, hoặc muốn được mô tả. Việc thu hẹp khoảng cách này là một thách thức.

Một cách để thực thi một biểu diễn bị rối là giữ các trục khác nhau của biến thể trong khi huấn luyện [21]. Johnson et al. [14] kết hợp một khả năng mạng thần kinh với một mô hình gia đình hàm mũ liên hợp cho các biến tiềm ẩn. Trong lớp mô hình này, sự chênh lệch hiệu quả đối với các biến tiềm ẩn có thể được thực hiện bằng cách học một phép chiếu lên cùng một gia đình hàm mũ trong bộ mã hóa. Ở đây chúng tôi đề xuất một lớp tổng quát hơn về các mô hình đồ họa từng phần: các mô hình đồ họa xác suất trong đó người modeller chỉ cần xác định mối quan hệ chính xác cho một số tập con của các biến ngẫu nhiên trong mô hình. Các yếu tố còn lại không được xác định trong mô hình sau đó được học, được xác định bởi các mạng nơron có thể phát hiện được. Điều này cung cấp khả năng tự xác định vị trí tại một điểm cụ thể trên quang phổ, bằng cách chỉ định chính xác các trục biến thể đó (và phụ thuộc của chúng), chúng ta có thông tin về hoặc muốn trích xuất và học các biểu diễn không bị rối, được học một cách vướng víu.

Một lớp con của các mô hình một phần đặc biệt phổ biến là chúng ta có thể thu thập dữ liệu giám sát cho một số tập hợp con của các biến. Trong thực tế, thường có sự thay đổi trong dữ liệu (ít nhất là khái niệm) dễ giải thích, và do đó chú thích, trong khi các biến thể khác ít rõ ràng hơn. Ví dụ, hãy xem xét số liệu MNIST của các chữ số viết tay: các hình ảnh khác nhau cả về nội dung (chữ số), và kiểu (chữ số được viết), như được hiển thị ở phía bên tay phải của Hình 1. Có biến tiềm ẩn "chữ số" rõ ràng ghi lại một trục biến thể có ý nghĩa và nhất quán, độc lập với kiểu; sử dụng một mô hình đồ họa một phần đặc biệt có nghĩa là chúng ta có thể xác định biến “chữ số” ngay cả khi không loại bỏ các câu trả lời của các kiểu khác nhau và quá trình hiển thị chữ số cho một hình ảnh. Với việc học không giám sát, không có sự đảm bảo rằng suy luận về một mô hình có 10 lớp sẽ tạo ra các biểu diễn tiềm ẩn được thừa nhận với các thừa số tương ứng với 10 chữ số. Tuy nhiên, với một số lượng nhỏ các ví dụ được dán nhãn, nhiệm vụ này trở nên rõ ràng dễ dàng hơn. Về cơ bản, phương pháp tiếp cận của chúng tôi phù hợp với ý tưởng rằng các khái niệm rõ ràng về sự mất cân bằng đòi hỏi phải xác định một nhiệm vụ để đo lường nó [4]. Ví dụ: khi xem xét hình ảnh khuôn mặt của mọi người, chúng tôi có thể muốn nắm bắt danh tính của người đó trong một ngữ cảnh và điều kiện ánh sáng trên khuôn mặt trong khuôn mặt khác, các đặc điểm khuôn mặt của người khác hoặc các kết hợp này trong các bối cảnh khác. Các mô hình một phần đặc biệt và giám sát yếu có thể được xem như một cách để vận hành trực tiếp nhiệm vụ phụ thuộc vào mục tiêu học tập.

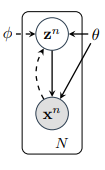
Trong bài báo này, chúng tôi giới thiệu một công thức cho việc học và suy luận trong các mô hình một phần, một khuôn khổ có thể hiểu được các biểu diễn dữ liệu bằng cách sử dụng các cấu trúc mô hình đồ họa để mã hóa các ràng buộc để giải thích dữ liệu. Chúng tôi trình bày khuôn khổ này trong bối cảnh các bộ tự động biến đổi (VAEs), phát triển một công thức tổng quát về học bán giám sát với DGM cho phép khuôn khổ của chúng tôi tự động sử dụng hệ số chính xác của mục tiêu cho bất kỳ sự lựa chọn nào của mô hình và tập hợp các điểm đến được quan sát. Về mặt này, công việc của chúng tôi mở rộng những nỗ lực trước đây để giới thiệu sự giám sát vào các bộ tự động biến thiên [18, 24, 32]. Chúng tôi giới thiệu một mục tiêu biến đổi có thể áp dụng cho một lớp mô hình tổng quát hơn, cho phép chúng ta xem xét các cấu trúc mô hình đồ họa với các phụ thuộc tùy ý giữa các vị trí tiềm ẩn, các miền ẩn liên tục và những người phụ thuộc thay đổi động. Chúng tôi cung cấp một đặc điểm về cách biên dịch các mô hình sinh sản được giám sát một phần thành các đồ thị tính toán ngẫu nhiên, thích hợp cho đào tạo từ đầu đến cuối. Cách tiếp cận này cho phép chúng ta suy luận suy luận [7, 23, 29, 34], đồng thời học một mạng thực hiện suy luận gần đúng về các biểu diễn cùng lúc chúng ta biết được các yếu tố không xác định của chính mô hình đó. Chúng tôi chứng minh tính hiệu quả của khuôn khổ của chúng tôi về một loạt các nhiệm vụ, liên quan đến phân loại, hồi quy và tổng hợp dự báo, bao gồm cả khả năng mã hóa các vĩ độ có chiều hướng thay đổi.

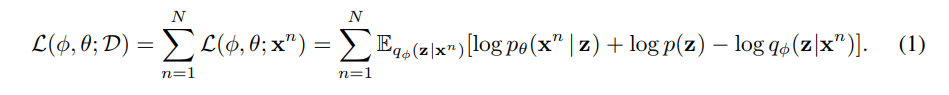


Hình 1: Nghiên cứu bán giám sát trong các bộ mã hóa biến đổi có cấu trúc, được minh họa trên các chữ số MNIST. Trên cùng bên trái: Mô hình sinh học. Dưới cùng bên trái: Mô hình nhận dạng. Giữa: Đồ thị tính toán ngẫu nhiên, cho thấy việc mở rộng từng nút đến biểu đồ phụ tương ứng của nó. Phụ thuộc mô hình sinh học được hiển thị bằng màu xanh và phụ thuộc vào mô hình nhận dạng được hiển thị bằng màu cam. Xem Phần 2.2 để có giải thích chi tiết. Phải: học đại diện.

**2. Framework and Formulation : Khuôn khổ và xây dựng**

VAE [17, 28] là một lớp mô hình sinh trưởng sâu đồng thời đào tạo cả bộ mã hóa và bộ giải mã xác suất cho một phần tử của tập dữ liệu D = {x1,. . . xN}. Tương tự trung tâm là một mã hóa z có thể được coi là một biến tiềm ẩn, đúc bộ giải mã dưới dạng mật độ xác suất có điều kiện pθ (x | z). Các tham số ηθ (z) của phân bố này là đầu ra của mạng nơron xác định với các tham số θ (phổ biến nhất là MLP hoặc CNN) lấy z làm đầu vào. Bằng cách đặt một điểm yếu trước trên z, bộ giải mã xác định một phần sau và phân phối chung pθ (z | x) ∝ pθ (x | z) p (z).

Suy luận trong VAE có thể được thực hiện bằng phương pháp variational xấp xỉ phân phối hậu nghiệm pθ (z | x) sử dụng bộ mã hóa qφ (z | x), với tham số λφ (x) là đầu ra của mạng (với tham số φ) được gọi là "mạng suy luận" hoặc "mạng nhận diện". Mạng lưới sinh sản và suy luận, được biểu thị bằng đường nét liền và đứt nét tương ứng trong mô hình đồ họa, được đào tạo chung bằng cách thực hiện độ dốc ngẫu nhiên lên trên bằng chứng thấp hơn (ELBO) L (φ, θ; D) ≤ log pθ (D), 

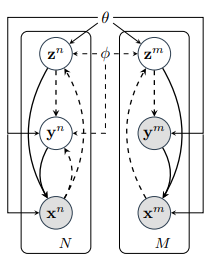


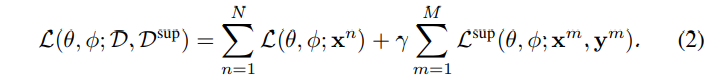
Thông thường, thuật ngữ đầu tiên Eqφ (z | xn) [log pθ (xn | z)] xấp xỉ bằng ước tính của Monte Carlo và hai số còn lại được biểu diễn dưới dạng phân kỳ −KL (qφ (z | xn) kp (z) ), có thể được tính toán phân tích khi mô hình bộ mã hóa và trước là Gaussian.

Trong bài báo này, chúng ta sẽ xem xét các mô hình trong đó cả hai mô hình sinh sản pθ (x, y, z) và xấp xỉ sau qφ (y, z | x) có thể có cấu trúc phụ thuộc tùy ý liên quan đến các biến ngẫu nhiên được xác định trên một số phân bố khác nhau loại. Chúng tôi quan tâm đến việc xác định các kiến trúc VAE trong đó một tập con của các biến y có thể diễn giải được. Đối với các biến này, chúng tôi giả định rằng nhãn giám sát có sẵn cho một phần dữ liệu. VAE sẽ bổ sung thêm một số biến số z mà suy luận được thực hiện theo cách hoàn toàn không được giám sát. Điều này phù hợp với mục tiêu trung tâm của chúng tôi là xác định và học tập trong các mô hình được chỉ định một phần. Trong ví dụ đang chạy cho MNIST, y tương ứng với nhãn phân loại, trong khi z chụp tất cả các tính năng tiềm ẩn khác, chẳng hạn như kiểu bút và kiểu chữ viết tay.

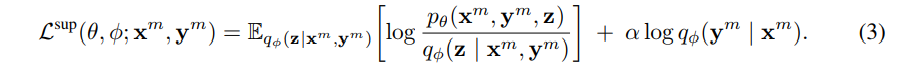
Lớp mô hình này tổng quát hơn các mô hình trong tác phẩm của Kingma et al. [18], người xem xét ba mẫu thiết kế với cấu trúc phụ thuộc cụ thể. Chúng tôi cũng không yêu cầu p (y, z) là một mô hình gia đình hàm mũ liên hợp, như trong tác phẩm của Johnson et al. [15]. Để thực hiện việc học bán giám sát trong lớp mô hình này, chúng ta cần i) xác định một mục tiêu phù hợp với đồ thị phụ thuộc chung và ii) xác định phương pháp để xây dựng một đồ thị tính toán ngẫu nhiên [30] kết hợp cả cấu trúc phụ thuộc điều kiện trong mô hình sinh sản và mô hình công nhận vào mục tiêu này.

**2.1 Objective Function : Hàm mục tiêu**

Công việc trước đây về học tập bán giám sát cho các mô hình sinh trưởng sâu [18] định nghĩa một mục tiêu trên N điểm dữ liệu không giám sát D = {x1,. . . , xN} và M điểm dữ liệu được giám sát Dsup = {(x1, y1),. . . , (xM,yM)}, 



Phân phối chung của mô hình của chúng tôi đưa vào bộ sưu tập các cụm từ không được giám sát và giám sát trên D và Dsup như được trình bày trong mô hình đồ họa. Sự biến đổi chuẩn biến đổi trên các bằng chứng chung của tất cả các dữ liệu quan sát (bao gồm cả giám sát) cũng được tính như trong phương trình. (2). Vì hệ số tương ứng với phần không giám sát của mô hình đồ họa chính xác là Eq. (1), chúng tôi tập trung vào thuật ngữ được giám sát trong phương trình. (2), được mở rộng bên dưới, kết hợp thêm một thành phần có trọng số như trong Kingma et al. [18].



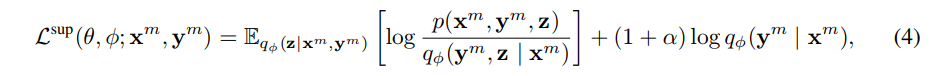
Lưu ý rằng công thức trong phương trình. (2) giới thiệu một hằng số γ điều khiển cường độ tương đối của thuật ngữ được giám sát. Mặc dù sự phân bố chung trong mô hình của chúng tôi có ý nghĩa thống kê hai thuật ngữ, trong trường hợp các kích thước tương đối của D và Dsup khác nhau rất nhiều, việc kiểm soát trọng số tương đối của các điều khoản có thể giúp cải thiện sự khác biệt đó.

Định nghĩa này trong phương trình. (3) ngầm giả định rằng chúng ta có thể đánh giá xác suất có điều kiện qφ (z | x, y) và q\_ biên (y | x) = R dz qφ (y, z | x). Đây thực sự là trường hợp cho các mô hình được Kingma và cộng sự xem xét. [18], có hệ số yếu tố q (y, z | x) = qφ (z | x, y) qφ (y | x).

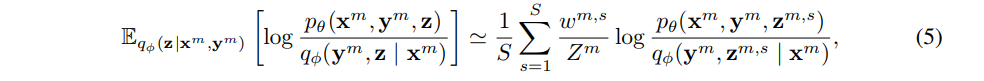
Ở đây chúng ta sẽ lấy được một bộ ước lượng cho L sup để tổng quát hóa các mô hình trong đó qφ (y, z | x) có thể có một cấu trúc phụ thuộc có điều kiện tùy ý. Với mục đích trình bày, chúng ta sẽ cho thời điểm xem xét trường hợp qφ (y, z | x) = qφ (y | x, z) qφ (z | x). Đối với yếu tố này, tạo mẫu z m, s ∼ qφ (z | x m, y

m) yêu cầu suy luận, có nghĩa là chúng ta không còn có thể tính toán một ước lượng đơn giản của Monte Carlo bằng cách lấy mẫu từ phân phối vô điều kiện qφ (z | x m). Hơn nữa, chúng ta cũng không thể đánh giá mật độ qφ (z | x m, y m).

Để giải quyết những khó khăn này, chúng tôi thể hiện lại các thuật ngữ được giám sát trong mục tiêu



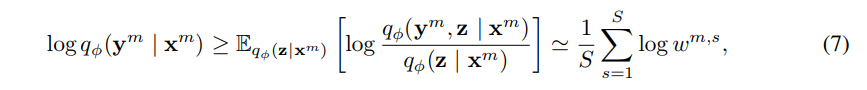
loại bỏ nhu cầu đánh giá qφ (z | x m, y m). Sau đó, chúng tôi có thể sử dụng lấy mẫu tầm quan trọng (tự chuẩn hóa) để ước tính mức độ mong đợi. Để làm như vậy, chúng tôi lấy mẫu các đề xuất z m, s ∼ qφ (z | x m) từ phân phối bộ mã hóa không điều kiện và xác định bộ ước lượng



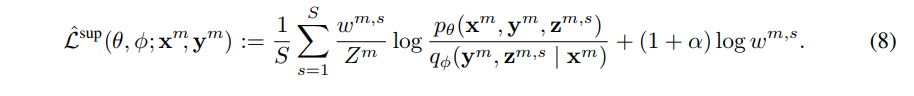
trong đó trọng số quan trọng không bình thường w m, s và bình thường Z m được định nghĩa là



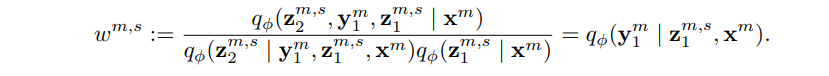
Để ước tính nhật ký qφ (y m | x m), chúng tôi sử dụng một ước lượng Monte Carlo của giới hạn dưới thường được sử dụng trong ước tính khả năng tối đa,



sử dụng cùng mẫu z m, s và trọng lượng w m, s như trong phương trình. (5). Khi chúng ta kết hợp các từ trong Eqs. (5) và (7), chúng tôi có được ước tính

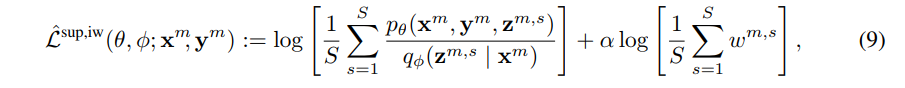


Chúng tôi lưu ý rằng ước tính này áp dụng cho bất kỳ cấu trúc phụ thuộc có điều kiện nào. Giả sử chúng ta định nghĩa một bộ mã hóa qφ (z2, y1, z1 | x) với hệ số factorφ (z2 | y1, z1, x) qφ (y1 | z1, x) qφ (z1 | x). Nếu chúng ta đề xuất z2 ∼ qφ (z2 | y1, z1, x) và z1 ∼ qφ (z1 | x), thì trọng số quan trọng w m, s đối với bộ ước lượng trong phương trình. (8) được định nghĩa là



Nói chung, trọng số quan trọng chỉ đơn giản là sản phẩm của xác suất có điều kiện của các biến được giám sát y trong mô hình. Lưu ý rằng điều này cũng áp dụng cho các mô hình trong Kingma et al. [18], mục tiêu chúng ta có thể phục hồi bằng cách lấy trọng số là hằng số w m, s = qφ (y m | x m).

Chúng ta cũng có thể xác định một mục tiêu tương tự với một mục tiêu được sử dụng trong các autoencoders có trọng số quan trọng [2], trong đó chúng ta tính toán logarit của một ước tính của Monte Carlo, thay vì ước tính của Monte Carlo về logarit. Mục tiêu này có dạng



có thể được bắt nguồn bằng cách di chuyển tổng trong Eq. (8) vào logarit và áp dụng sự thay thế w m, s / qφ (y m, z m, s | x m) = 1 / qφ (z m, s | x m).

**2.2 Construction of the Stochastic Computation Graph : Xây dựng biểu đồ tính toán ngẫu nhiên**

Để thực hiện độ dốc lên trên mục tiêu trong phương trình. (8), chúng tôi lập bản đồ các mô hình đồ họa cho pθ (x, y, z) và qφ (y, z | x) vào một đồ thị tính toán ngẫu nhiên trong đó mỗi nút ngẫu nhiên tạo thành một biểu đồ con.

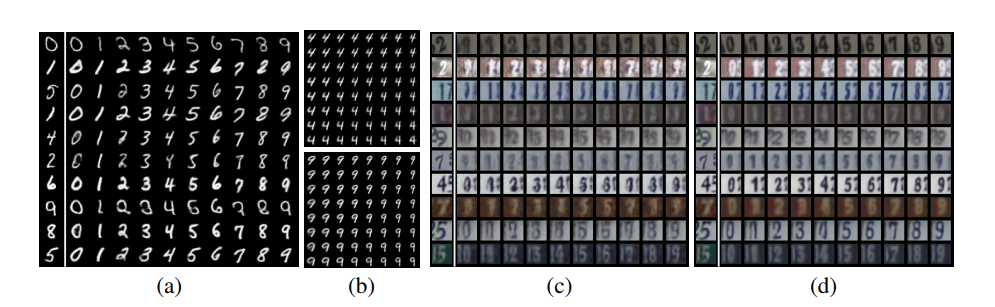
Hình 1 cho thấy sự mở rộng này cho VAE đơn giản cho các chữ số MNIST từ [17]. Trong mô hình này, y là một biến rời rạc đại diện cho chữ số cơ bản, biến quan tâm tiềm ẩn của chúng ta, mà chúng ta có dữ liệu giám sát một phần. Biến không được phân phối Gaussian không được phân phối z nắm bắt phần còn lại của thông tin tiềm ẩn. Điều này bao gồm các tính năng như kiểu viết tay và độ dày nét vẽ. Trong mô hình sinh sản (Hình 1 trên cùng bên trái), chúng ta giả định một hệ số pθ (x, y, z) = pθ (x | y, z) p (y) p (z) trong đó y và z độc lập theo trước đó. Trong mô hình nhận dạng (Hình 1 dưới cùng bên trái), chúng tôi sử dụng cấu trúc phụ thuộc có điều kiện qφ (y, z | x) = qφz (z | y, x) qφy (y | x) để loại bỏ nhãn chữ số y khỏi kiểu chữ viết tay z (Hình 1 bên phải).

Các mô hình sinh sản và công nhận được cùng hình thành một đồ thị tính toán ngẫu nhiên (Hình 1 trung tâm) có chứa một biểu đồ phụ cho mỗi biến ngẫu nhiên. Đây có thể tương ứng với các biến được giám sát hoàn toàn, được giám sát và không giám sát. Biểu đồ ví dụ này chứa ba loại biểu đồ con, tương ứng với ba khả năng để giám sát và ước tính độ dốc :

* Đối với biến được giám sát đầy đủ x, chúng tôi tính toán khả năng p trong mô hình sinh sản, đó là pθ (x | y, z) = N (x; ηθ (y, z)). Ở đây ηθ (y, z) là một mạng nơron với các tham số θ trả về các tham số của phân bố chuẩn (tức là vectơ trung bình và hiệp phương sai đường chéo).
* Đối với biến không được quan sát z, chúng tôi tính cả xác suất trước p (z) = N (z; Î · z) và xác suất có điều kiện qÏ † (z | x, y) = N (z; Î »Ï † z ( x, y)). Ở đây reparametrisation thông thường được sử dụng để lấy mẫu z từ qÏ † (z | x, y) bằng cách lấy mẫu đầu tiên âˆ¼ N (0, I) bằng cách sử dụng thủ thuật reparametrisation thông thường z = g (, Î »Ï † (x, y)).
* Đối với biến y được quan sát một phần, chúng tôi cũng tính toán xác suất p (y) = Discrete (y; ηy) và qφy (y | x) = Discrete (y; λφz (x)). Giá trị y được xử lý như được quan sát khi có, và lấy mẫu khác. Trong ví dụ cụ thể này, chúng ta lấy mẫu y từ một qφy (y | x) bằng cách sử dụng thư giãn Gumbel-softmax [13, 25] của sự phân bố rời rạc.

Ví dụ trong Hình 1 minh họa một khuôn khổ chung để xác định VAE với các cấu trúc phụ thuộc tùy ý. Chúng ta bắt đầu bằng cách định nghĩa một nút cho mỗi biến ngẫu nhiên. Đối với mỗi nút, chúng tôi sau đó chỉ định loại phân phối và chức năng tham số η, xác định cách xác suất trong mô hình sinh sản phụ thuộc vào các biến khác trong mạng. Hàm này có thể là một mạng không đổi, hoàn toàn xác định hoặc một mạng nơron có tham số được học từ dữ liệu. Đối với mỗi biến không giám sát và bán giám sát, chúng ta phải bổ sung thêm một hàm λ trả về các giá trị tham số trong mô hình nhận dạng, cùng với một thủ tục lấy mẫu (reparametrised).

Với đặc điểm kỹ thuật này của biểu đồ tính toán, giờ đây chúng tôi có thể tính toán ước tính lấy mẫu quan trọng theo phương trình. (8) chỉ đơn giản là chạy mạng về phía trước nhiều lần để lấy mẫu từ qφ (· | λ) cho tất cả các biến không được quan sát. Sau đó chúng tôi tính toán pθ (x, y, z), qφ (y | x), qφ (y, z | x) và mức độ quan trọng



Hình 2: (a) Các tương tự trực quan cho dữ liệu MNIST, được giám sát một phần chỉ với 100 nhãn (trong số 50000). Chúng tôi suy ra biến kiểu z và sau đó thay đổi nhãn y. (b) Thăm dò trong không gian kiểu với nhãn y được giữ cố định và (2D) kiểu z khác nhau. Hình ảnh tương tự cho dữ liệu SVHN khi (c) được giám sát một phần chỉ với 1000 nhãn và (d) được giám sát hoàn toàn.

Trọng số w, là xác suất chung của tất cả các biến được bán giám sát mà nhãn có sẵn. Ước tính này sau đó có thể được tối ưu hóa đối với các biến θ và φ để đào tạo bộ mã hóa tự động.

**3. Experiments (Thử nghiệm)**

Chúng tôi đánh giá khuôn khổ của chúng tôi cùng một số trục khác nhau liên quan đến khả năng tìm hiểu các biểu diễn bị gỡ rối thông qua việc cung cấp các cấu trúc mô hình đồ họa một phần cho các điểm ẩn và giám sát yếu. Đặc biệt, chúng tôi đánh giá khả năng (i) của nó như một bộ phân loại / bộ hồi quy cho các vị trí cụ thể trong tập dữ liệu đã cho, (ii) tìm hiểu mô hinh sinh mẫu theo cách bảo tồn ngữ nghĩa của các vị trí liên quan đến dữ liệu được tạo ra, và (iii) thực hiện các nhiệm vụ này một cách linh hoạt, cho nhiều mô hình và dữ liệu khác nhau.

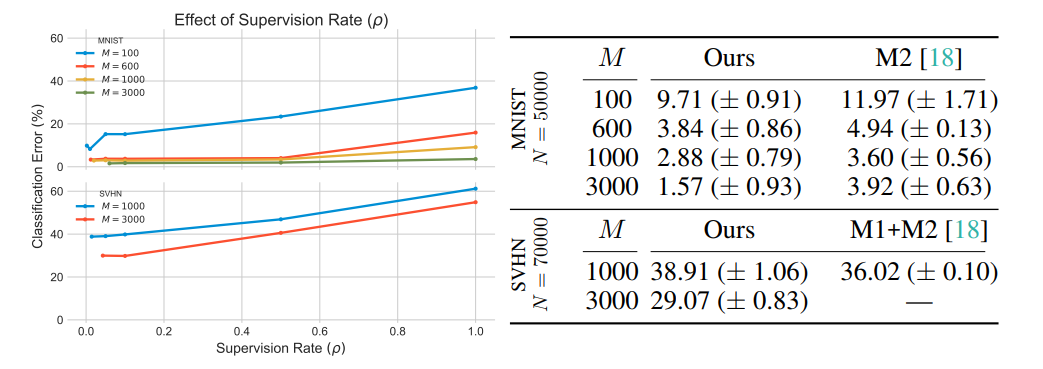
Đối với tất cả các thử nghiệm chạy, chúng tôi chọn kiến trúc và thông số được coi là tiêu chuẩn cho loại và kích thước của các tập dữ liệu tương ứng. Trường hợp hình ảnh có liên quan (ngoại trừ MNIST), chúng tôi sử dụng (de) các kiến trúc convolutional và sử dụng phép truy toán chuẩn GRU trong trường hợp Multi-MNIST. Đối với việc học, chúng tôi sử dụng AdaM [16] với một tỷ lệ học tập và các thuật ngữ động lượng được đặt thành các giá trị mặc định của chúng. Đối với các kích cỡ lô nhỏ, chúng thay đổi từ 100-700 tùy thuộc vào tập dữ liệu đang được sử dụng và kích thước của tập hợp con có nhãn Dsup. Tất cả những điều trên, bao gồm thêm chi tiết về các giá trị tham số chính xác và mã nguồn, bao gồm thư viện PyTorchbased của chúng tôi để chỉ định các mô hình đồ họa tùy ý trong khung công tác VAE, có sẵn tại <https://github.com/probtorch/probtorch>.

**3.1 MNIST and SVHN**

Chúng tôi bắt đầu với một thử nghiệm liên quan đến một cấu trúc phụ thuộc đơn giản, trên thực tế rất giống như trong Kingma et al. [18], để xác nhận hiệu suất của mục tiêu tầm quan trọng của chúng tôi trong trường hợp đặc biệt, nơi mạng lưới nhận dạng và các mô hình sinh sản được xác định như được chỉ ra trong Hình 1 (bên trái), cho chúng ta trọng số quan trọng là hằng số w^(m,s)= q\_∅ (y^m |x^n). Mô hình được kiểm tra về khả năng phân loại các chữ số và thực hiện tạo điều kiện trên bộ dữ liệu MNIST và Số nhà Phố của Google (SVHN). Như hình 1 (bên trái) cho thấy, các mô hình sinh sản và nhận dạng có nhãn "chữ số", biểu thị y, được chỉ định một phần (và được giám sát một phần).

Hình 2 (a) và (c) minh họa khả năng tạo điều kiện của mô hình đã học, ở đây chúng ta cho thấy hiệu ứng đầu tiên chuyển đổi đầu vào cho trước (cột ngoài cùng bên trái) sang không gian tiềm ẩn bị ẩn, và với biến cố định kiểu dáng cố định, thao tác chữ số thông qua mô hình sinh mẫu để tạo dữ liệu với các đặc tính hình ảnh được mong đợi. Lưu ý rằng cả hai kết quả thu được với sự giám sát từng phần - 100 (trong số 50000) điểm dữ liệu có nhãn trong trường hợp MNIST và 1000 (trong số 70000) điểm dữ liệu có nhãn trong trường hợp SVHN. Biến tiềm ẩn kiểu z được lấy là một Gaussian hiệp phương sai của các kích thước 10 và 15 tương ứng. Hình 2 (d) cho thấy tương tự cho SVHN với sự giám sát đầy đủ. Hình 2 (b) minh họa chế độ thay thế của thế hệ có điều kiện, trong đó kiểu dáng tiềm ẩn, ở đây được coi là một Gauss 2D, được thay đổi với số được giữ cố định.

Tiếp theo, chúng tôi đánh giá khả năng của mô hình để tìm hiểu hiệu quả một trình phân loại từ giám sát từng phần. Chúng tôi tính toán lỗi phân loại trên tác vụ dự đoán nhãn trên cả hai tập dữ liệu và kết quả là



Hình 3: Bên Phải: Tỷ lệ lỗi phân loại cho các kích thước được gán nhãn M khác nhau trên nhiều lần chạy, với tốc độ giám sát ρ= γ^M/(N+γ^M ),γ=1. Đối với SVHN, chúng tôi so sánh với quy trình nhiều giai đoạn (M1 + M2) [ 18], nơi mô hình của chúng tôi chỉ sử dụng một giai đoạn duy nhất. Còn lại: Lỗi phân loại so với các kích thước được đặt tên khác nhau và tỷ lệ giám sát cho MNIST (trên cùng) và SVHN (dưới cùng). Ở đây, tỉ lệ của mục tiêu phân loại được giữ cố định tại α = 50 (MNIST) và α = 70 (SVHN). Lưu ý rằng đối với dữ liệu được dán nhãn thưa thớt (M≪N), một phần ít của over-representation (γ>1) giúp cải thiện sự suy rộng với hiệu suất tốt hơn trên các thiết lập thử nghiệm. Ngược lại, quá nhiều over-representation dẫn đến overfitting.

Báo cáo trong bảng trong hình 3. Lưu ý rằng có một vài điểm khác biệt nhỏ trong thiết lập giữa phương pháp của chúng tôi và những điểm chúng tôi so sánh với [18]. Chúng tôi luôn chạy các mô hình của chúng tôi trực tiếp trên dữ liệu mà không cần xử lý trước hoặc học trước dữ liệu. Vì vậy, đối với MNIST, chúng tôi so sánh với mô hình M2 từ đường cơ sở chỉ giống nhau. Tuy nhiên, đối với SVHN, phương pháp cơ sở không báo cáo lỗi cho mô hình M2; chỉ mô hình M1 + M2 hai giai đoạn liên quan đến một bước trích xuất tính năng riêng biệt trên dữ liệu trước khi tìm hiểu một trình phân loại bán giám sát (a semi-supervised classifier).

Như các kết quả cho thấy, mô hình và mục tiêu của chúng tôi thực sự thực hiện ngang bằng với các thiết lập được xem xét trong Kingma et al. [18], phục vụ như là xác nhận cơ bản của khuôn khổ của chúng tôi. Tuy nhiên, chúng tôi lưu ý rằng từ quan điểm đạt được lỗi phân loại thấp nhất có thể, người ta có thể áp dụng bất kỳ số lượng thay thế nào [24] và các cải tiến trong kiến trúc mạng Neural [27, 33].

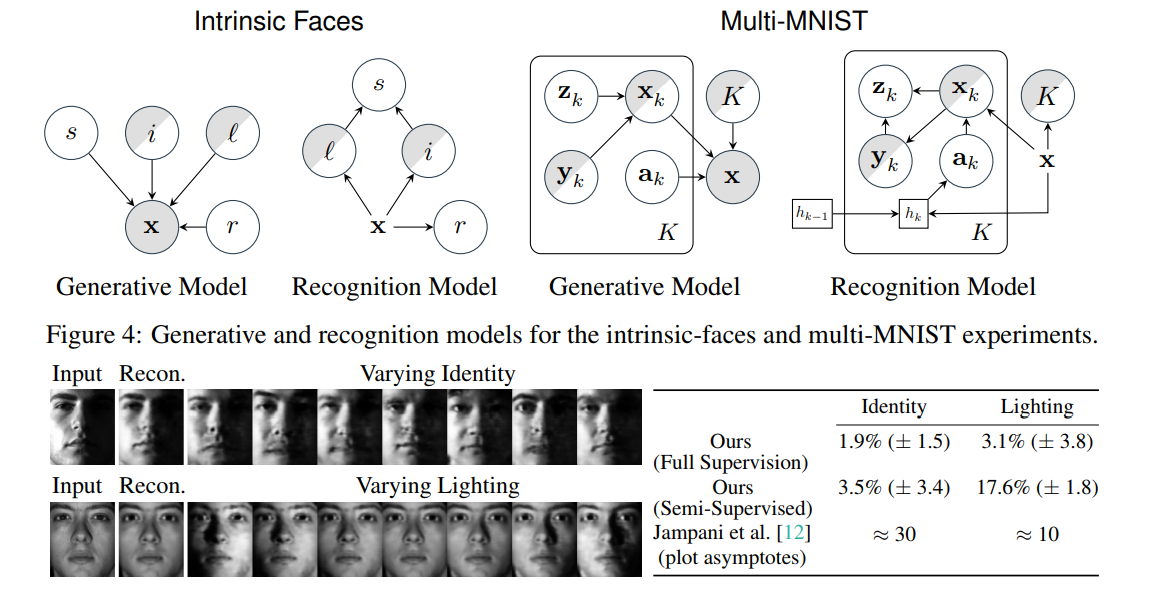
Tỷ lệ giám sát: Như đã thảo luận trong Phần 2.1, chúng tôi xây dựng mục tiêu của chúng tôi để cung cấp một xử lý về trọng lượng tương đối giữa các điều khoản được giám sát và không giám sát. Đối với một kích thước thiết lập chưa được giám sát N, kích thước thiết lập được giám sát M và thuật ngữ mở rộng γ, trọng số tương đối là ρ = γM / (N + γM). Hình 3 cho thấy sự khảo sát trọng số tương đối này trên các bộ dữ liệu MNIST và SVHN và trên các kích thước được giám sát khác nhau M. Mỗi dòng trong đồ thị đo lỗi phân loại cho một M, trên ρ, bắt đầu từ γ = 1, tức là ρ = M / (N + M). Trong dòng với Kingma et al.[18], chúng tôi sử dụng α = 0,1 / ρ. Khi dữ liệu được dán nhãn rất thưa thớt (M? N), việc biểu thị quá mức các ví dụ được dán nhãn trong quá trình đào tạo có thể giúp tổng quát hóa bằng cách cải thiện hiệu suất trên dữ liệu thử nghiệm. Trong các thí nghiệm của chúng tôi, phần lớn, chọn yếu tố này là ρ = M / (N + M) cho kết quả tốt. Tuy nhiên, như được mong đợi, over-fitting xảy ra khi ρ được tăng lên vượt ra ngoài một điểm nhất định.

**3.2 Intrinsic Faces**

Tiếp theo chúng ta chuyển sang một miền phức tạp hơn liên quan đến các mô hình sinh mẫu của các khuôn mặt. Như có thể thấy trong các mô hình đồ họa cho thí nghiệm này trong Hình 4, các cấu trúc phụ thuộc được sử dụng ở đây phức tạp hơn so với các mô hình từ thí nghiệm trước đó. Ở đây, chúng tôi sử dụng tập dữ liệu “Yale B” [6] được xử lý bởi Jampani et al. [12] cho kết quả trong Hình 5. Chúng tôi quan tâm đến việc mô hình của chúng tôi có thể tìm hiểu các biểu thị về nhận dạng và ánh sáng bị phân tán và đánh giá hiệu suất của nó trong các nhiệm vụ của (i) phân loại nhân dạng, và (ii) phương hướng.

Lưu ý rằng mô hình sinh mẫu của chúng tôi giả định không có cấu trúc đặc biệt – chúng tôi chỉ cần chỉ định một mô hình mà tất cả các biến tiềm ẩn độc lập theo trước. Công việc trước đây [12] giả định một mô hình sinh học với các biến tiềm ẩn nhận dạng i, ánh sáng l, bóng r, và phản xạ r, theo quan hệ (n.l)×r+ ∈ cho dữ liệu pixel. Ở đây, chúng tôi muốn chứng minh rằng mô hình sinh học của chúng tôi vẫn học được mối quan hệ chính xác với các biến tiềm ẩn này, nhờ cấu trúc trong mô hình nhận dạng và được giám sát (từng phần).

Lưu ý rằng trong mô hình nhận dạng (Hình 4), ánh sáng l là một biến tiềm ẩn với miền liên tục và một biến mà chúng ta giám sát một phần. Hơn nữa, chúng tôi mã hóa danh tính i dưới dạng biến ngẫu nhiên phân loại,



Hình 5: Trái: Khám phá khả năng sinh sản của mô hình được giám sát bằng cách điều khiển nhận dạng và ánh sáng cho một giá trị cố định (suy ra) của các biến tiềm ẩn khác. Phải: Phân loại và tỷ lệ lỗi hồi quy cho nhận dạng và biến tiềm ẩn ánh sáng, được giám sát hoàn toàn và được giám sát (với 6 hình ảnh ví dụ được dán nhãn cho mỗi cá nhân trong số 38 cá nhân, tỷ lệ giám sát ρ = 0.5 và α = 10). Phân loại là một lựa chọn trực tiếp 1 trong 38, trong khi đó so sánh, lỗi là sự mất mát gần hàng xóm nhất dựa trên phản xạ suy luận. Mất hồi quy là khoảng cách góc cạnh.

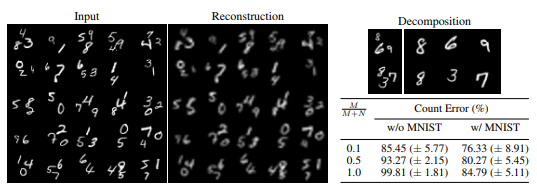
Thay vì xây dựng môt bản đồ pixel-wise surface-normal (mỗi giả định là độc lập Gaussian) như là phong tục. Công thức này cho phép chúng tôi giải quyết trực tiếp nhiệm vụ dự đoán danh tính, thay vì áp dụng các phương pháp đánh giá thay thế (ví dụ: phân loại hàng xóm gần nhất dựa trên phản xạ suy luận). Hình 5 trình bày cả đánh giá định tính và định lượng của khuôn khổ để cùng tìm hiểu cả mô hình nhận dạng có cấu trúc và các tham số mô hình sinh học.

**3.3 Multi-MNIST**

Cuối cùng, chúng tôi tiến hành một thử nghiệm mở rộng sự phức tạp từ các mô hình trước đó hơn nữa. Đặc biệt, chúng tôi khám phá khả năng của khung công tác của chúng tôi để xử lý các mô hình với chiều hướng ngẫu nhiên - có số lượng biến tiềm ẩn được xác định bởi một biến ngẫu nhiên và các mô hình có thể bao gồm các mô hình nhỏ hơn (phụ) khác. Chúng tôi tiến hành thử nghiệm này trong miền đa MNIST. Đây là một lựa chọn phù hợp vì nó đáp ứng cả hai yêu cầu trên - mỗi hình ảnh có thể có số lượng chữ số cá nhân khác nhau, về cơ bản chỉ ra rằng mô hình phải học đếm, và vì mỗi hình ảnh chính nó bao gồm các ví dụ (được chia tỷ lệ và được dịch) từ MNIST dữ liệu, chúng ta có thể sử dụng chính mô hình MNIST trong mô hình đa MNIST.

Cấu trúc mô hình mà chúng tôi giả định cho mạng lưới sinh sản và nhận dạng được thể hiện trong hình 4. Chúng tôi mở rộng các mô hình từ thí nghiệm MNIST bằng cách tạo ra nó bằng một chuỗi ngẫu nhiên máy phát điện, trong đó độ dài vòng lặp K là một biến ngẫu nhiên. Đối với mỗi vòng lặp lặp k = 1,. . . , K, mô hình sinh sản lặp đi lặp lại lấy mẫu một chữ số yk , kiểu zk và sử dụng các ký tự này để tạo hình ảnh chữ số xk theo cách tương tự như trong ví dụ MNIST trước đó. Ngoài ra, một sự biến dạng khác cũng lấy mẫu cho mỗi chữ số trong mỗi lần lặp để biến đổi hình ảnh chữ số xk thành phổ biến, kết hợp canvas đại diện cho hình ảnh được tạo cuối cùng x, sử dụng mạng biến áp không gian [11].

Trong mô hình nhận dạng, chúng tôi dự đoán số chữ số K từ các điểm ảnh trong hình ảnh. Cho mỗi lặp vòng lặp k = 1,. . . , K, chúng tôi định nghĩa một hình ảnh chữ số phân phối Bernoulli xk. Khi giám sát là có sẵn, chúng tôi tính toán xác suất của xk từ entropy chéo nhị phân theo cách tương tự như trong thời hạn khả năng cho mô hình MNIST. Khi không có giám sát, chúng tôi xác định đặt xk thành giá trị trung bình của phân phối. Điều này có thể được nhìn thấy giống như việc cung cấp các hộp giới hạn xung quanh các chữ số cấu thành như giám sát, mà phải được tính đến khi học affine các phép biến đổi phân tách một hình ảnh đa MNIST thành các hình ảnh giống như MNIST thành phần của nó. Điều này thiết kế mô hình tương tự như thiết kế được sử dụng trong DRAW [10], các VAE tái phát [3] và AIR [5].



Hình 6: Trái: Ví dụ nhập hình ảnh đa MNIST và tái tạo. Trên cùng bên phải : Phân hủy các hình ảnh Multi-MNIST thành các chữ số MNIST thành phần. Dưới cùng bên phải : Đếm độ chính xác qua các kích thước thiết lập được giám sát khác nhau M cho kích thước tập dữ liệu đã cho M + N = 82000.

Trong trường hợp không có bộ dữ liệu đa MNIST, ta cần tạo ra tập dữ liệu của riêng mình từ bộ dữ liệu MNIST theo điều chỉnh tỷ lệ và vị trí của các chữ số chuẩn thành một canvas kết hợp, cân bằng đồng đều trên số đếm (1-3) và chữ số. Sau đó, chúng tôi đã tiến hành hai thử nghiệm trong miền này, bên trong thí nghiệm đầu tiên, chúng tôi tìm cách đo mức độ phát sinh trình tự ngẫu nhiên học đếm một mình, không chú ý trả tiền để tháo gỡ các biểu diễn tiềm ẩn cho các chữ số cơ bản. Ở đây, mô hình sinh sản giả định tính khả dụng của các hình ảnh chữ số MNIST riêng lẻ, tạo ra kết hợp dưới biến đổi affine lấy mẫu. Trong thử nghiệm thứ hai, chúng tôi mở rộng phần trên mô hình cho đến nay cũng kết hợp cùng một mô hình MNIST được đào tạo trước đó từ phần trước, cho phép mô hình sinh sản lấy mẫu hình ảnh số MNIST, đồng thời cũng có thể dự đoán được chữ số cơ bản. Điều này cũng cho thấy cách chúng ta có thể tận dụng tính sáng tác của các mô hình: khi nào một mô hình phức tạp có một mô hình đơn giản hơn được gọi là một cấu trúc con, mô hình đơn giản hơn và được học trọng lượng có thể được giảm trực tiếp.

Lỗi chính xác về đếm trên các kích thước tập hợp được giám sát khác nhau, tái tạo cho một tập hợp ngẫu nhiên các đầu vào và sự phân hủy của một tập hợp các đầu vào đã cho thành các chữ số riêng lẻ của chúng, thể hiện trong hình 6. Tất cả các sự tái tạo và phân tách hình ảnh được hiển thị tương ứng với mô hình lồng nhau cấu hình. Chúng tôi quan sát thấy rằng không chỉ chúng tôi có thể suy ra một cách đáng tin cậy về số lượng chữ số trong cho hình ảnh, chúng tôi có thể tái tạo đồng thời các đầu vào cũng như các bộ phận cấu thành của nó.

**4. Thảo luận và kết luận**

Trong bài báo này, chúng tôi giới thiệu một khuôn khổ cho việc học đại diện cho dữ liệu sử dụng disentangled các cấu trúc mô hình đồ họa được chỉ định một phần và các lược đồ học tập bán giám sát trong miền các bộ tự động biến đổi (VAEs). Điều này được thực hiện bằng cách xác định các mô hình sinh sản lai kết hợp cả hai mô hình đồ họa có cấu trúc và các biến ngẫu nhiên không có cấu trúc trong cùng một ẩn không gian. Chúng tôi chứng minh tính linh hoạt của phương pháp này bằng cách áp dụng nó cho nhiều nhiệm vụ khác nhau trong lĩnh vực trực quan và đánh giá hiệu quả của nó khi tìm hiểu các biểu diễn bị gỡ rối trong một semisupervised theo cách, thể hiện hiệu suất mạnh mẽ. Các mô hình được chỉ định một phần này cho phép nhận dạng các mạng đưa ra các dự đoán trong một không gian có thể giải thích và bị rối loạn, bị ràng buộc bởi cấu trúc được cung cấp bởi mô hình đồ họa và sự giám sát yếu.

Khung được thực hiện như một thư viện PyTorch [26], cho phép xây dựng stochastic biểu đồ tính toán mã hóa cấu trúc và tính toán cần thiết. Điều này cung cấp một hướng để khám phá trong tương lai - sự mở rộng của khung đồ thị tính toán ngẫu nhiên đến lập trình xác suất [9, 35, 36]. Các chương trình xác suất vượt ra khỏi khung trình bày để cho phép nhiều mô hình biểu cảm hơn, kết hợp các cấu trúc đệ quy và các hàm bậc cao hơn. Các sự kết hợp của các khuôn khổ như vậy với các mạng thần kinh gần đây đã được nghiên cứu trong Le et al. [23] và Ritchie et al. [29], cho thấy một con đường đầy hứa hẹn để thăm dò thêm.