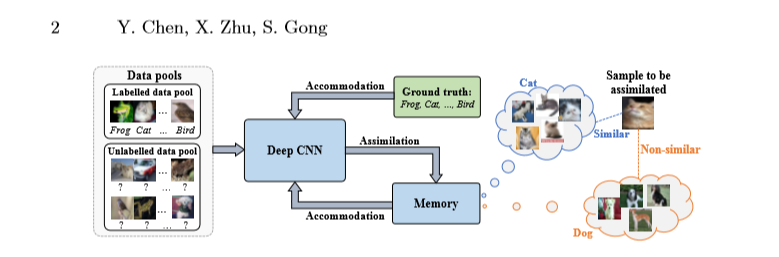
**Semi-Supervised Deep Learning with Memory**

1. **Introduction**

Học tập bán giám sát (SSL) nhằm mục đích tăng hiệu suất mô hình bằng cách sử dụng số lượng lớn dữ liệu không được gắn nhãn khi chỉ có một số lượng dữ liệu có nhãn hạn chế [4,37]. Đó là động lực mà dữ liệu không có nhãn hiệu có sẵn ở quy mô lớn nhưng dữ liệu được dán nhãn là khan hiếm do chi phí ghi nhãn cao. Lược đồ học tập này hữu ích và hữu ích đối với nhiều ứng dụng như tìm kiếm hình ảnh [6], phân loại trang web [2], truy xuất tài liệu [21], bộ gen [29], v.v. Trong tài liệu SSL, thuật toán SSL đơn giản nhất là tự đào tạo nơi mô hình đích được đào tạo từng bước bằng dữ liệu tự dán nhãn bổ sung được đưa ra bởi các dự đoán riêng của mô hình với con số cao [21,2,25]. Phương pháp này có khuynh hướng tuyên truyền lỗi trong học mô hình do dự đoán sai về sự kết hợp cao. Các phương pháp phổ biến khác bao gồm Transductive SVM [10,3] và phương pháp graphbased [39,1], tuy nhiên, có khả năng chịu thiệt hại từ khả năng mở rộng kém đến dữ liệu không được đánh dấu quy mô lớn do tối ưu hóa không hiệu quả.

Gần đây, các phương pháp SSL dựa trên mạng nơron [23,35,15,12,30,24,19,26,16, 9,32] bắt đầu thống trị tiến trình nhờ khả năng học tập biểu diễn mạnh mẽ của mạng thần kinh sâu. Hầu hết các phương pháp này thường sử dụng mạng đào tạo cập nhật để xây dựng hình phạt bổ sung không giám sát



để cho phép học tập bán giám sát. Chúng tôi xem xét rằng loại lược đồ SSL sâu này không tối ưu với điều kiện khả năng ghi nhớ của các mạng sâu thường không đầy đủ và không bị ngăn cản để thể hiện kiến ​​thức được tích luỹ trong quá trình học lặp lại [34]. Để thúc đẩy kiến ​​thức như vậy, chúng tôi giới thiệu một cơ chế bộ nhớ vào quá trình đào tạo mạng sâu để cho phép học tập bán giám sát từ dữ liệu đào tạo có nhãn có kích thước nhỏ và dữ liệu đào tạo không có nhãn có kích thước lớn. Theo tinh thần của lý thuyết của Piaget về khả năng học tập liên tục của con người [7], chúng tôi hướng tới thiết kế một sơ đồ SSL cho phép mô hình sâu tìm hiểu thêm từ bộ nhớ của nó (đồng hóa) và tự điều chỉnh để tối ưu dữ liệu đào tạo (chỗ ở) theo cách gia tăng. Để đạt được mục đích này, chúng tôi xây dựng một khung học tập sâu được giám sát bán bộ nhớ mới hỗ trợ bộ nhớ: Mạng nơ-ron sâu được hỗ trợ bộ nhớ (MA-DNN) như được minh họa trong hình 1. MA-DNN được đặc trưng bởi sự tương tác giữa đồng hóa và chỗ ở giữa mạng và một mô-đun bộ nhớ ngoài.

Chìa khóa để thiết kế khung công tác của chúng tôi là hai khía cạnh: (1) biểu diễn tính năng phân biệt đối xử cấp lớp và sự không chắc chắn suy luận mạng được tích lũy dần dần trong một mô-đun bộ nhớ ngoài; (2) thông tin ghi nhớ này được sử dụng để đồng hóa các mẫu hình ảnh mới đến và tạo ra một mất mát bộ nhớ không giám sát thông tin để hướng dẫn việc học mạng cùng với sự mất phân lớp giám sát.

Đóng góp của chúng tôi là hai lần: (1) Chúng tôi đề xuất khai thác bộ nhớ của học mô hình để cho phép học tập sâu bán giám sát từ một ít dữ liệu đào tạo có nhãn và nhiều dữ liệu đào tạo không có nhãn, đồng thời hoàn toàn áp dụng quy trình đào tạo từ đầu đến cuối. Điều này trái ngược với hầu hết các phương thức SSL sâu hiện có, thường bỏ qua bộ nhớ của việc học mô hình. (2) Chúng tôi xây dựng một mạng lưới thần kinh sâu bộ nhớ mới hỗ trợ bộ nhớ (MA-DNN) đặc trưng bởi một cơ chế bộ nhớ. Chúng tôi giới thiệu một mất mát bộ nhớ không giám sát tương thích với sự mất mát phân loại theo tiêu chuẩn giám sát để cho phép học tập bán giám sát. Các thí nghiệm so sánh mở rộng thể hiện những ưu điểm của mô hình MA-DNN được đề xuất của chúng tôi qua nhiều phương pháp học sâu rộng được giám sát hiện đại nhất.

1. **Related Works**

Học sâu bán giám sát gần đây đã đạt được sự thu hút ngày càng tăng do sức mạnh khái quát hóa mạnh mẽ của các mạng lưới thần kinh sâu [35,15,12,30,24,19, 14]. Chiến lược chung là đào tạo các mạng lưới thần kinh sâu bằng cách tối ưu hóa đồng thời một phân loại mất mát được giám sát tiêu chuẩn trên các mẫu được dán nhãn cùng với một điều khoản mất mát không giám sát bổ sung được áp dụng cho dữ liệu không ghi nhãn [15,27,5] hoặc cả dữ liệu được dán nhãn và không dán nhãn [35,24 , 19,14]. Các điều khoản tổn thất bổ sung này được coi là tín hiệu giám sát không giám sát, vì nhãn sự thật không nhất thiết phải có để lấy các giá trị tổn thất. Ví dụ, Lee [15] sử dụng cross-entropy loss được tính toán trên nhãn giả (các lớp có xác suất dự đoán tối đa được cung cấp bởi mạng cập nhật) của các mẫu không ghi nhãn làm tín hiệu giám sát bổ sung. Rasmus et al. [24] chấp nhận tổn thất tái cấu trúc giữa một lần truyền thuận sạch và một lần truyền được chuyển tiếp ngẫu nhiên có nguồn gốc từ cùng một mẫu. Miyato et al. [19] định nghĩa sự thông suốt phân phối chống lại sự nhiễu loạn ngẫu nhiên cục bộ như một hình phạt không giám sát. Laine và cộng sự. [14] giới thiệu một mất L2 không được giám sát để xử phạt sự không thống nhất giữa các dự đoán mạng và các dự đoán mạng được sắp xếp theo thời gian. Nhìn chung, lý do căn bản của các thuật toán SSL này là thường xuyên hóa mạng bằng cách thực thi các ranh giới phân loại trơn tru và nhất quán, mạnh mẽ đến nhiễu loạn ngẫu nhiên [24,19]; hoặc để làm phong phú các tín hiệu giám sát bằng cách khai thác kiến ​​thức mà mạng học được, chẳng hạn như sử dụng nhãn giả [15] hoặc các dự đoán được sắp xếp theo thời gian

Đối với việc chia sẻ chung một mục đích của việc phạt không giám sát , phương pháp của chúng ta là duy nhất trong số các cách cơ bản:

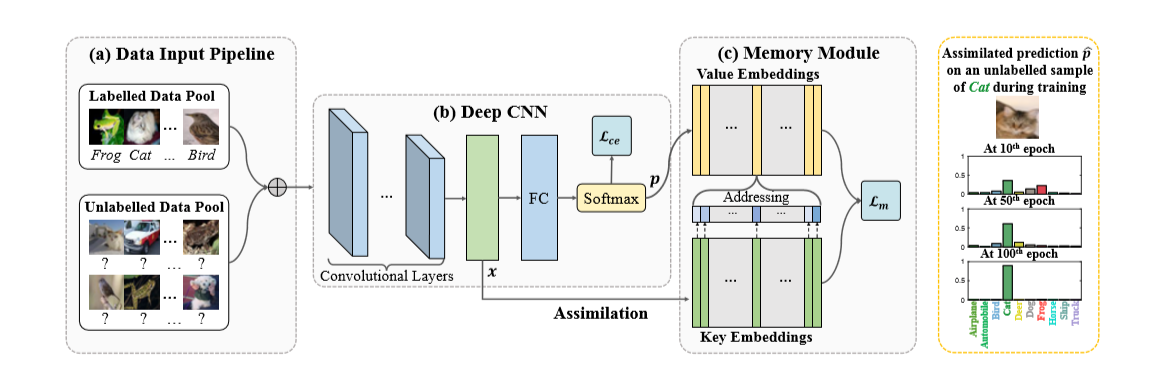
1. Khai thác bộ nhớ của học mô hình: Thay vì dựa vào kiến không đầy đủ của một mạng đơn cập nhật để suy ra mất mát bổ sung, chúng tôi sử dụng một mô-đun bộ nhớ để lấy được một sự mất mát bộ nhớ dựa trên biểu diễn đặc trưng mức tích lũy và mô hình không chắc chắn suy luận tổng hợp tất cả thông qua các lần lặp đào tạo trước đó
2. Chi phí tính toán thấp: Bằng cách sử dụng cơ chế bộ nhớ, chỉ cần một lan truyền về phía trước để tính toán thời gian mất thêm để đào tạo mạng, trái ngược với nhiều hơn một phép truyền tiếp theo yêu cầu của các mô hình khác
3. Mức tiêu thụ bộ nhớ thấp: Thay vì lưu trữ tất cả các dự đoán của tất cả các mẫu đào tạo trong một bản đồ lớn, mô-đun bộ nhớ được cập nhật trực tuyến của chúng ta tiêu thụ rất ít dung lượng bộ nhớ, do đó có khả năng mở rộng dữ liệu hơn.

Các mạng thần kinh với bộ nhớ được giới thiệu gần đây để cho phép khả năng học tập và lý luận mạnh mẽ hơn để giải quyết một số nhiệm vụ đầy thách thức, chẳng hạn như trả lời câu hỏi [34,31,18] và học một lần [28,11]. Tăng cường một mạng với một thành phần bộ nhớ ngoài rất hấp dẫn nhờ khả năng lưu trữ, trừu tượng hóa và sắp xếp các kiến ​​thức trong quá khứ thành một dạng cấu trúc và địa chỉ. Như các công trình trước đó, Weston et al. [34] đề xuất Mạng bộ nhớ, tích hợp các thành phần suy luận với một thành phần bộ nhớ có thể đọc và viết để ghi nhớ các sự kiện hỗ trợ từ quá khứ để trả lời câu hỏi. Kaiser và cộng sự. [11] đề xuất một mô-đun bộ nhớ suốt đời để ghi lại kích hoạt mạng các sự kiện hiếm gặp cho việc học một lần. Công việc của chúng tôi được lấy cảm hứng từ các tác phẩm này, nhưng đó là nỗ lực đầu tiên để khám phá cơ chế bộ nhớ trong học tập sâu bán giám sát. Bên cạnh chức năng lưu trữ cơ bản, mô-đun bộ nhớ của chúng tôi tạo ra sự tương tác đồng hóa-điều ứng (assimilation-accommodation) để khai thác bộ nhớ của việc học mô hình và tạo ra một sự mất mát bộ nhớ không giám sát nâng cao cho phép học bán giám sát.

1. **Memory-Assisted Deep Neural Network**

Chúng tôi xem xét việc học tập sâu bán giám sát trong bối cảnh phân loại hình ảnh đa lớp. Trong bối cảnh này, chúng tôi có quyền truy cập vào một số lượng giới hạn các mẫu hình ảnh được dán nhãn DL = {(Ii, l, yi, l)} nl i nhưng một lượng lớn các mẫu hình ảnh không gắn nhãn DU = {(Ii, u)} nu i, nơi nu ≫ nl. Mỗi hình ảnh không gắn nhãn là giả định thuộc về một trong cùng một loại đối tượng K (lớp) Y = {yi} K i là dữ liệu được dán nhãn, trong khi các dán nhãn dự đoán của họ không có sẵn để đào tạo. Mục tiêu chính của SSL là nâng cao hiệu suất mô hình bằng cách học từ dữ liệu hình ảnh có nhãn DL và dữ liệu hình ảnh không được gắn thêm DU cùng một lúc. Để đạt được mục tiêu đó, chúng tôi xây dựng một khuôn khổ học tập sâu được giám sát bán bộ nhớ có hỗ trợ bộ nhớ tích hợp mạng nơron sâu với mô-đun bộ nhớ, Chúng tôi gọi đây là mạng nơ-ron sâu hỗ trợ bộ nhớ (MA-DNN).

**3.1. Approach Overview**

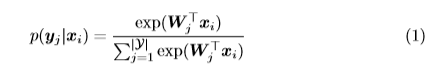


**Mô tả Hình** : Tổng quan về Mạng lưới thần kinh sâu được hỗ trợ bởi bộ nhớ (MA-DNN) cho việc học tập sâu bán giám sát. Trong quá trình huấn luyện, Cho ít dữ liệu huấn luyện có gắn nhãn và nhiều dữ liệu đào tạo không có gắn nhãn, các lô nhỏ dữ liệu được gắn nhãn/không gắn nhãn chuyển tiếp vào *(b)* Deep CNN để để có được đại diện đăc trưng cập nhật **x** và dự đoán xác suất **p** cho mỗi mẫu. Cho *(c)* mô đun bộ nhớ được cập nhật, Đồng hóa (assimilation) bộ nhớ tạo ra một dự đoán nhiều lớp khác (ký hiệu ˆp ) cho mỗi cặp Địa chỉ Khóa (Key addressing) và giá trị. Trong Điều ứng (accommodation), một bộ nhớ mất **Lm** được tính từ ˆp và được sử dụng như một tín hiệu giám sát bổ sung để hướng dẫn việc học mạng cùng với sự phân loại giám sát mất mát. Tại thời điểm thử nghiệm, mô-đun bộ nhớ không còn cần thiết, do đó nó không ảnh hưởng đến hiệu quả triển khai.

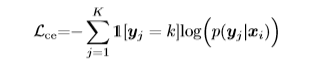
Thiết kế tổng thể của kiến trúc MA-DNN được mô tả trong hình . MA-DNN được đề xuất có ba phần: (1) Mạng nơron sâu; (2) Một mô-đun bộ nhớ được thiết kế để ghi lại bộ nhớ của học mô hình; và (3) Cơ chế tương tác đồng hóa-điều ứng (assimilation-accommodation) được giới thiệu để khai thác bộ nhớ một cách hiệu quả nhằm tạo thuận lợi cho việc tối ưu hóa mạng trong học tập bán giám sát.

* 1. **Conventional Deep Neural Network**

Khung đề xuất nhằm mục đích làm việc với các mạng nơ ron sâu tiêu chuẩn hiện có. Chúng tôi chọn Mạng thần kinh chuyển đổi (CNN) trong công việc này do khả năng học tập đại diện mạnh mẽ cho dữ liệu hình ảnh. Để huấn luyện một CNN cho phân loại hình ảnh, hàm cross-entropy loss có giám sát thường được áp dụng. Trong quá trình đào tạo, với bất kỳ mẫu đào tạo nào, chúng tôi chuyển tiếp nó qua mạng sâu cập nhật để có được một vectơ đặc trưng **x** và một vectơ dự đoán xác suất đa giác **p** trên tất cả các lớp. Đặc biệt, chúng tôi dự đoán xác suất hậu nghiệm của lớp thứ j của mẫu hình ảnh được dán nhãn là



trong đó xi đề cập đến biểu diễn đặc trưng sâu được nhúng của được trích xuất bởi CNN sâu và là tham số chức năng dự đoán lớp thứ j. **Cross-entropy** loss trên so với nhãn lớp chân thực **K** được tính là



Rõ ràng, hàm **Cross-entropy loss** bị hạn chế để học từ các mẫu được dán nhãn một mình. Để tận dụng các mẫu đào tạo không ghi nhãn, một cách đơn giản là sử dụng các nhãn dự đoán được đưa ra bởi mô hình cập nhật trong đào tạo. Tuy nhiên, điều này có thể dễ bị lỗi và không đáng tin cậy khi ước tính nhãn chưa trưởng thành, đặc biệt khi bắt đầu đào tạo mô hình. Điều này trình bày một vấn đề Catch-22. Chúng tôi khắc phục vấn đề này bằng cách đưa mô-đun bộ nhớ vào quy trình đào tạo mạng để ước tính dần các dự đoán đáng tin cậy hơn về dữ liệu không ghi nhãn.

* 1. **Memory Module**

Để tận dụng lợi thế của các thông tin ghi nhớ được tạo ra trong mô hình học tập, nó là cần thiết để chúng tôi giới thiệu một mô-đun bộ nhớ. Chúng tôi xem xét hai loại bộ nhớ có kinh nghiệm của mạng-trong-đào tạo: (1) đại diện đặc trưng cấp lớp, và (2) mô hình không chắc chắn suy luận.

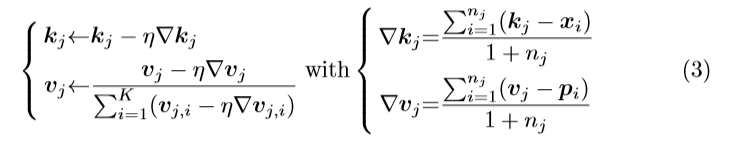
Để quản lý các thông tin ghi nhớ này, chúng tôi xây dựng mô-đun bộ nhớ trong cấu trúc khóa-giá trị(Key-value).

+ Key: sự mô tả đặc trưng được cập nhật của mỗi lớp trong không gian đặc trưng.

+ Value: ghi lại dự đoán xác suất đa lớp được cập nhật tương tự liên quan với mỗi lớp.

Để đại diện cho không gian đặc trưng tăng dần và tính không chắc chắn của mô hình tổng thể cập nhật, cập nhật bộ nhớ được thực hiện mỗi lần lặp để chứa các bản cập nhật mới nhất của mạng. Chúng tôi chỉ sử dụng dữ liệu được dán nhãn để cập nhật bộ nhớ, quy định rằng các mẫu không được gắn nhãn không có độ đảm bảo trong việc gán lớp và do đó có khả năng gây ra nguy cơ truyền bá lỗi.

Chính thức, giả sử tồn tại các mẫu hình ảnh được dán nhãn từ lớp thứ j (j ∈ {1, ···, K}) với các vectơ đặc trưng và dự đoán xác suất của chúng là , vị trí cắm bộ nhớ thứ j (kj, vj) được cập nhật tích lũy trên tất cả các lần lặp đào tạo như sau.



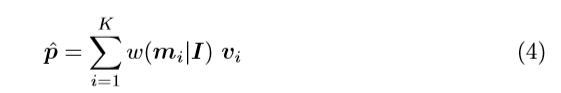
Trong đó **η** biểu thị tốc độ học tập (được đặt thành η = 0,5 trong các thí nghiệm của chúng tôi). Giá trị nhúng vj được chuẩn hóa để đảm bảo tính chất phân phối xác suất của nó. Trong quá trình đào tạo, khi các gradient (∇kj, ∇vj) dần dần nhỏ hơn, các khóa nhúng và giá trị sẽ trở nên đáng tin cậy hơn để tái tạo cấu trúc tính năng cơ bản và phân phối nhiều lớp. Để bắt đầu quá trình đào tạo mà không áp đặt kiến thức trước đó, chúng tôi khởi tạo tất cả các khóa nhúng và giá trị thành **0** và **. 1** (phân phối xác suất thống nhất trên các lớp K), tương ứng. Điều này cho thấy thông tin ghi nhớ được mạng phát hiện đầy đủ trong quá trình đào tạo, mà không có bất kỳ giả định cụ thể nào về cài đặt vấn đề, do đó có khả năng áp dụng cho các tác vụ phân loại hình ảnh bán giám sát.

* 1. **The Assimilation-Accomodation Interaction**

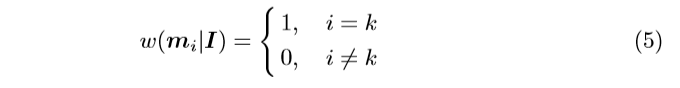
Với bộ nhớ cập nhật về học tập mô hình, chúng tôi tiếp tục sử dụng nó để cho phép học sâu bán giám sát. Điều này đạt được bằng cách giới thiệu một cơ chế tương tác **assimilation-accomodation** với hai thao tác được thực hiện mỗi lần lặp đào tạo:

* **Memory Assimilation**: Tính toán bộ nhớ dự đoán cho mỗi mẫu đào tạo bằng cách đọc địa chỉ và đọc giá trị;
* **Accommodation**: Tính toán bộ nhớ mất mát để xây dựng mục tiêu học tập bán giám sát cuối cùng.

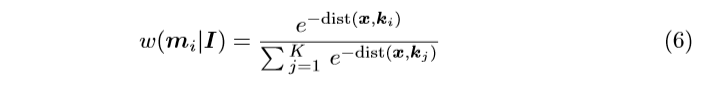
**(1) Memory Assimilation**. Với biểu diễn hình ảnh lan truyền **x** và dự đoán mạng p của hình ảnh **I**, sự đồng hóa bộ nhớ tạo ra một dự đoán xác suất đa lớp khác dựa trên bộ nhớ được cập nhật. Chúng tôi có được điều này bởi key addressing và value reading [18]. Đặc biệt, key addressing là để tính toán xác suất địa chỉ **w(mi|I)**, nghĩa là gán xác suất cho từng vị trí bộ nhớ **mi = (ki, vi), i ∈ {1, ···, K}**, dựa trên độ tương tự cặp viết mỗi key embedding. Về bản chất, **w(mi|I)** là phép gán cụm trong không gian đặc trưng. Với các xác suất địa chỉ trên tất cả các vị trí nhớ **K**, việc đọc giá trị sau đó được áp dụng để tính toán dự đoán bộ nhớ bằng cách lấy tổng của tất cả các giá trị nhúng như sau.



Theo tính sẵn có của nhãn, chúng tôi áp dụng hai chiến lược giải quyết. Đầu tiên là địa chỉ dựa trên vị trí được áp dụng cho các mẫu đào tạo được dán nhãn. Chính thức, giả sử mẫu đào tạo **I** được dán nhãn là lớp thứ **k**, xác suất địa chỉ được đạt được dựa trên vị trí **k** là

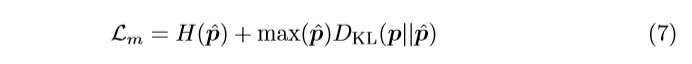


Thứ hai là địa chỉ dựa trên nội dung được áp dụng cho các mẫu hình ảnh không ghi nhãn. Chiến lược này tính toán xác suất đánh địa chỉ dựa trên độ tương tự cặp đôi giữa mẫu hình ảnh **I** và các khóa nhúng **ki** là



Trong đó **x** là vectơ đặc trưng được trích xuất của **I** và **dist ()** biểu thị khoảng cách **Euclide**. Phương trình (6) có thể được coi là một hình thức truyền nhãn [38] dựa trên giả định cụm [35,36], theo nghĩa là khối lượng xác suất được phân phối theo khoảng cách gần với từng tâm của cụm trong không gian đặc trưng. Đó là, phân công xác suất được tính dựa trên thành viên cụm.

**(2) Accommodation.** Hoạt động này cung cấp cho mạng sâu với một **memory loss** để hình thành mục tiêu học tập bán giám sát cuối cùng để mạng có thể học thêm từ dữ liệu không ghi nhãn. Đặc biệt, chúng tôi giới thiệu một **memory loss** trên mỗi mẫu đào tạo x như sau.



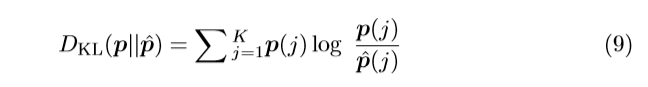
Trong đó **H()** dùng để chỉ biện pháp entropy; **max()** là hàm tối đa trả về giá trị tối đa của vectơ đầu vào; **DKL()** là phân kỳ Kullback-Leibler (KL). Cả **H() và DKL()** đều có thể được tính mà không có nhãn thực tế và do đó áp dụng cho việc học bán giám sát. Hai điều khoản loss trong phương trình. (7) được đặt tên là The Model Entropy (ME) loss and The Memory-Network Divergence (MND) (Phân kỳ mạng bộ nhớ) loss như được giải thích dưới đây.

1. The Model Entropy (ME) loss term **H(ˆ p)** được tính theo công thức:



mà định lượng lượng thông tin được mã hóa trong **ˆ p**. Từ quan điểm thông tin hóa, entropy phản ánh sự không chắc chắn của mô hình tổng thể. Một entropy cao trên một mẫu hình ảnh được dán nhãn chỉ ra rằng **ˆ p** là phân phối xác suất đa phương thức mơ hồ, tương ứng với việc **value embedding** được lấy của một lớp cụ thể. Điều này chỉ ra rằng mạng không thể phân biệt tốt giữa lớp này và các lớp khác, điều này xuất phát từ việc gán các dự đoán xác suất không nhất quán cho các mẫu hình ảnh trong cùng một lớp. Mặt khác, một entropy cao trên một mẫu không ghi nhãn cho thấy sự phân bổ lớp nghiêm trọng chồng chéo giữa các lớp khác nhau trong không gian đặc trưng. Điều này là do mẫu không ghi nhãn không thể được gán cho một lớp nhất định với xác suất cao. Do đó, giảm thiểu entropy mô hình H tương đương với việc giảm sự chồng chéo phân phối lớp trong không gian đặc trưng và xử phạt các dự đoán mạng không nhất quán ở cấp lớp, về cơ bản được thúc đẩy bởi nguyên tắc tối thiểu hóa entropy [8].

1. The Memory-Network Divergence (MND) loss term **DKL(p||ˆp)** được tính giữa dự đoán mạng **p** và dự đoán bộ nhớ **ˆp** như sau.



**DKL(p||ˆp)** là một hình phạt không âm, đo lường sự khác biệt giữa hai bản phân phối: **p** và **^p**. Nó đại diện cho thông tin bổ sung được mã hóa bằng **p** so với **ˆp** trong lý thuyết thông tin. Giảm thiểu sự phân kỳ KL này ngăn dự đoán mạng không bị sai lệch quá nhiều so với phân phối xác suất xuất phát từ mô-đun bộ nhớ. Khi **DKL (p || p) → 0**, nó cho biết dự đoán mạng phù hợp với dự đoán bộ nhớ của nó. Ngoài ra, chúng tôi cũng áp dụng trọng số động: **max (p)**, giá trị xác suất tối đa là **ˆp**, để giảm mức độ quan trọng của DKL () khi đưa ra dự đoán bộ nhớ mơ hồ, tức là, phân phối xác suất đa phương thức. Do đó, p được khuyến khích khớp với **ˆp** đặc biệt khi **ˆp** tương ứng với dự đoán bộ nhớ con, tức là, phân phối xác suất cực đại, trong đó đỉnh tương ứng với gán cho một lớp nhất định có xác suất cao.

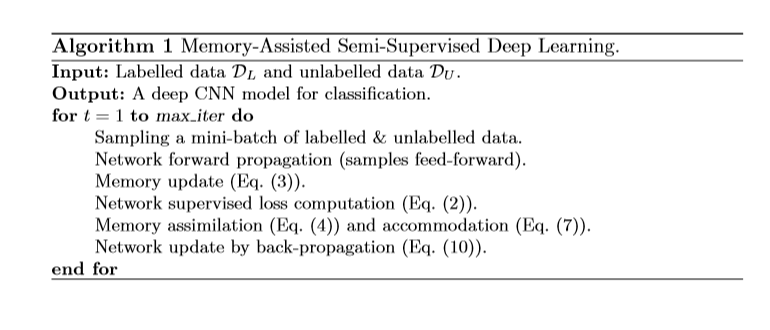
Hàm mục tiêu học tập bán giám sát cuối cùng được hình thành bằng cách hợp nhất phương trình. (7) và phương trình. (2) như sau.



trong đó **λ** là một siêu tham số được đặt thành 1 để đảm bảo tầm quan trọng tương đương của hai thuật ngữ mất trong quá trình đào tạo.

* 1. **Model Training**

**MA-DNN** được đề xuất được đào tạo bằng thuật toán Stochastic Gradient Descent tiêu chuẩn theo cách từ đầu đến cuối. Tổng quan về thuật toán của đào tạo mô hình được tóm tắt trong Thuật toán 1.



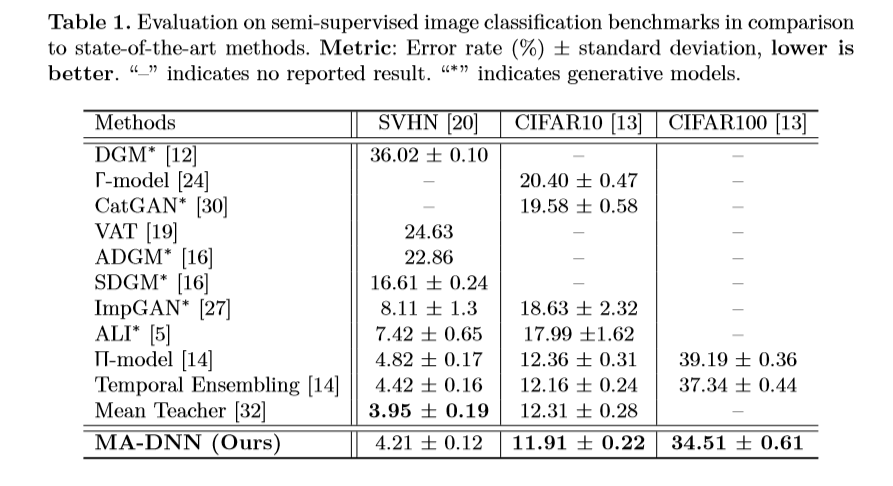
1. **Experiments**

Chúng tôi xác nhận tính hiệu quả của MA-DNN trên ba bộ dữ liệu điểm chuẩn phân loại hình ảnh được áp dụng rộng rãi, so với các phương pháp tiên tiến khác trong Phần 4.2 và nghiên cứu cắt bỏ trong Phần 4.2.

Dữ liệu. Để đánh giá MA-DNN được đề xuất của chúng tôi, chúng tôi chọn ba bộ dữ liệu điểm chuẩn phân loại hình ảnh được áp dụng rộng rãi như chi tiết sau đây.

1. **SVHN [20]:** Bộ dữ liệu Số nhà của Street View bao gồm 10 lớp (0∼9) hình ảnh chữ số màu từ Google Street View. Nhiệm vụ phân loại là nhận dạng chữ số trung tâm của mỗi hình ảnh. Chúng tôi sử dụng phiên bản định dạng-2 cung cấp hình ảnh được cắt có kích thước 32 × 32 và phân chia dữ liệu thử nghiệm / huấn luyện tiêu chuẩn 73.257 / 26.032.
2. **CIFAR10 [13]:** Một bộ dữ liệu hình ảnh tự nhiên chứa 50.000 / 10.000 mẫu hình ảnh đào tạo / thử nghiệm từ 10 lớp đối tượng. Mỗi lớp có 6.000 hình ảnh với kích thước 32 × 32.
3. **CIFAR100 [13]:** bộ dữ liệu (có cùng kích thước hình ảnh với CIFAR10) chứa 50.000 / 10.000 hình ảnh đào tạo / kiểm tra từ hơn 100 lớp học mới với sự khác biệt về hình ảnh giữa các lớp tinh tế.

**Giao thức thí nghiệm**. Theo giao thức phân loại bán giám sát tiêu chuẩn [12,24,30,19], chúng tôi chia ngẫu nhiên dữ liệu đào tạo thành một bộ có nhãn nhỏ và một bộ không ghi nhãn lớn. Số lượng hình ảnh đào tạo được dán nhãn là 1.000 / 4.000 / 10.000 trên SVHN / CIFAR10 / CIFAR100 tương ứng, với 72.257 / 46.000 / 40.000 hình ảnh còn lại là dữ liệu đào tạo không ghi nhãn. Chúng tôi áp dụng tỷ lệ lỗi phân loại phổ biến làm thước đo hiệu suất mô hình và báo cáo tỷ lệ lỗi trung bình trên 10 lần phân tách dữ liệu ngẫu nhiên.

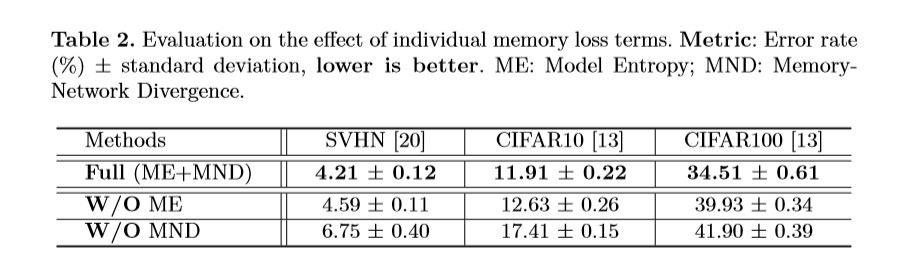


**Bảng 1**. Đánh giá về điểm chuẩn phân loại hình ảnh bán giám sát so với các phương pháp hiện đại. Số liệu: Tỷ lệ lỗi (%) ± độ lệch chuẩn, thấp hơn là tốt hơn. “–” chỉ ra không có kết quả báo cáo. “∗” để chỉ là các mô hình thế hệ.

**Chi tiết thực hiện**. Chúng tôi áp dụng kiến ​​trúc CNN 10 lớp giống như [14]. Nhiều chi tiết thực hiện được đưa ra trong các tài liệu bổ sung.

**So sánh với các Phương pháp hiện đại**. Trong Bảng 1, chúng tôi so sánh mô hình của chúng tôi với 11 phương pháp cạnh tranh tiên tiến nhất với kết quả được báo cáo của chúng trên SVHN, CIFAR10 và CIFAR100. Trong số tất cả các phương pháp này, **Mean Teacher** là người duy nhất vượt trội hơn **MA-DNN** của chúng tôi về nhiệm vụ phân loại chữ số. Trên các tác vụ phân loại hình ảnh tự nhiên, **MA-DNN** của chúng tôi vượt qua giải pháp thay thế tốt nhất (Tạm thời tạm thời) với biên độ 0,25% (12,16-11,91) và 2,83% (37,34-34,51) trên **CIFAR10** và **CIFAR100**. Điều này cho thấy sự vượt trội về hiệu suất của **MA-DNN** được đề xuất trong học sâu bán giám sát giữa các thuật toán học bán giám sát cạnh tranh khác nhau. Ngoài ra, cũng có thể thấy rằng **MA-DNN** vượt trội hơn đáng kể trên bộ dữ liệu **CIFAR100** đầy thách thức hơn với các cấu trúc ngữ nghĩa chi tiết hơn trong số nhiều lớp hơn. Điều này cho thấy rằng **memory loss** xuất phát từ bộ nhớ của việc học tập mô hình có thể tăng cường phân biệt và phân biệt lớp học nhiều hơn để tạo điều kiện cho việc học bán giám sát tốt hơn. Do đó, **MA-DNN** có khả năng mở rộng hơn so với các đối thủ khác trong các nhiệm vụ phân loại hình ảnh có liên quan đến số lượng lớp lớn hơn.

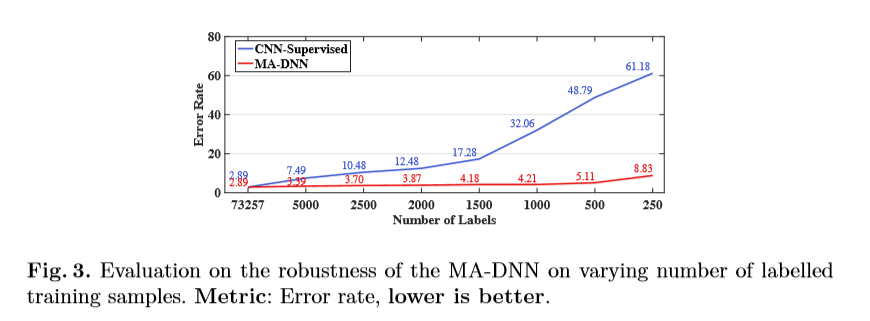
**Chi phí tính toán**. Độ phức tạp tính toán khoảng cách mỗi đợt được tạo ra bởi sự đồng hóa bộ nhớ và cập nhật bộ nhớ lần lượt là và , trong đó K là số lượng các vị trí bộ nhớ (**memory slots**), **Nl, Nu** là số lượng mẫu được dán nhãn và không ghi nhãn trong mỗi lô nhỏ. Đối với tính hiệu quả tính toán, tất cả các hoạt động của bộ nhớ được triển khai dưới dạng thao tác ma trận đơn giản trên GPU với độ chính xác điểm đơn lẻ. Nhìn chung, **MA-DNN** được tính toán một cách tính toán theo một số cách: **(i)** Chỉ cần một lan truyền về phía trước để tính toán tín hiệu giám sát bổ sung, trái với nhiều hơn một lần truyền theo yêu cầu của- Γ-model, VAT, Π-model and Mean-Teacher. **(ii)** Việc tiêu thụ dấu chân bộ nhớ bị hạn chế. Kích thước bộ nhớ của mô-đun bộ nhớ trong MA-DNN chỉ tỷ lệ thuận với số lượng lớp; trong khi Tạm thời yêu cầu lưu trữ dự đoán của tất cả các mẫu trong một bản đồ lớn với kích thước bộ nhớ tỷ lệ thuận với số lượng mẫu đào tạo. **(iii)** Không giống như các mô hình thế hệ bao gồm DGM, CatGAN, ADGM, SDGM, ImpGAN và ALI, MA-DNN của chúng tôi không cần tạo thêm hình ảnh tổng hợp trong quá trình đào tạo, do đó dẫn đến đào tạo mô hình hiệu quả hơn.



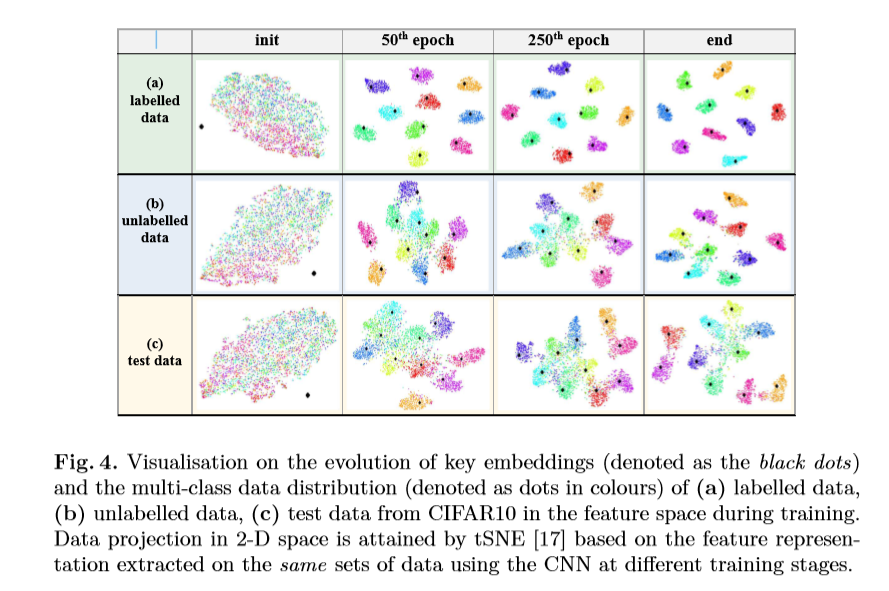
**Bảng 2.** Đánh giá về hiệu quả của các điều khoản mất trí nhớ cá nhân. Số liệu: Tỷ lệ lỗi (%) ± độ lệch chuẩn, thấp hơn là tốt hơn. ME: Model Entropy; MND: MemoryNetwork Divergence.

**4.2 Nghiên cứu đánh giá và phân tích sâu hơn**

**Ảnh hưởng của bộ nhớ mất (memory loss)**. Chúng tôi đánh giá sự đóng góp riêng lẻ của hai thuật ngữ mất trong công thức mất bộ nhớ (Phương trình (7)): (1) Entropy mô hình (ME) (Phương trình (8)) và (2) Phân kỳ mạng bộ nhớ (MND ) (Phương trình (9)). Chúng tôi đo lường tác động của từng thời hạn mất bằng cách giảm hiệu suất khi loại bỏ nó khỏi công thức mất bộ nhớ. Bảng 2 cho thấy kết quả đánh giá so với công thức mất bộ nhớ đầy đủ. Chúng tôi có các quan sát sau: (i) Cả hai thuật ngữ mất đều mang lại hiệu quả tích cực để tăng hiệu suất mô hình. Tỷ lệ lỗi phân loại tăng khi một trong hai điều khoản mất được loại bỏ. (ii) Thuật ngữ MND tăng cường hiệu suất mô hình. Loại bỏ thuật ngữ MND khiến hiệu suất giảm 2,54% (6,75-4,21), 5,50% (17,41-11,91), 7,39% (41,90-34,51) trên SVHN, CIFAR10 và CIFAR100. Điều này cho thấy hiệu quả của việc khuyến khích các dự đoán mạng phù hợp với các dự đoán bộ nhớ đáng tin cậy xuất phát từ bộ nhớ học tập mô hình. (iii) Thuật ngữ ME cũng hiệu quả. Loại bỏ thuật ngữ ME gây ra hiệu suất giảm 0,38% (4,59-4,21), 0,72% (12,63-11,91), 5,42% (39,93-34,51) trên SVHN, CIFAR10 và CIFAR100. Điều này cho thấy lợi ích của việc xử phạt chồng chéo phân phối lớp và tăng cường phân tách lớp, đặc biệt là khi số lượng lớp tăng lên - nhiều lớp khó phân tách hơn. Nhìn chung, việc đánh giá trong Bảng 2 cho thấy lợi ích chung của hai điều khoản tổn thất để cải thiện hiệu suất mô hình trong học tập sâu bán giám sát.



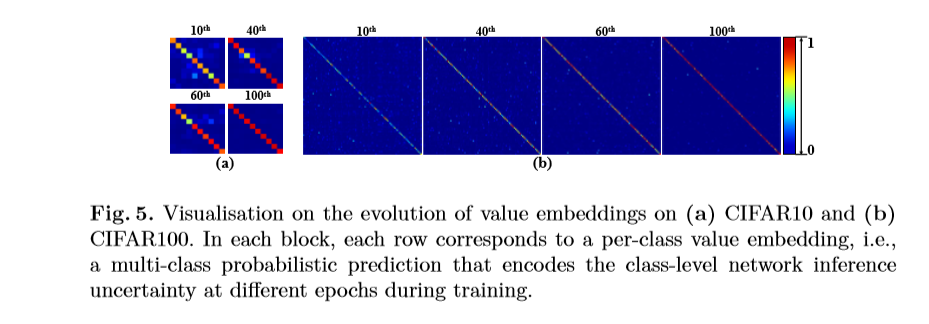
**Hình 3.** Đánh giá về độ mạnh của MA-DNN về số lượng mẫu đào tạo được dán nhãn khác nhau. Số liệu: Tỷ lệ lỗi, thấp hơn là tốt hơn.



**Hình.4**. Hình dung về sự phát triển của các key embedding (khóa nhúng) (được biểu thị là các chấm đen) và phân phối dữ liệu nhiều lớp (được biểu thị là các chấm màu) của (a) dữ liệu được gắn nhãn, (b) dữ liệu không ghi nhãn, (c) dữ liệu thử nghiệm từ CIFAR10 trong tính năng không gian trong quá trình đào tạo. Việc chiếu dữ liệu trong không gian 2 chiều được tSNE đạt được [17] dựa trên biểu diễn đặc trưng được trích xuất trên cùng một bộ dữ liệu bằng cách sử dụng CNN ở các giai đoạn đào tạo khác nhau.

**Kích cỡ mẫu đào tạo có dán nhãn.** Chúng tôi đánh giá sự mạnh mẽ của MA-DNN qua số lượng mẫu đào tạo được dán nhãn khác nhau. Chúng tôi tiến hành đánh giá này trên SVHN bằng cách thay đổi số lượng mẫu được dán nhãn từ 73.257 (tất cả các mẫu đào tạo được dán nhãn) đến 250. Để so sánh, chúng tôi áp dụng đối tác được giám sát CNN-Supervised chỉ sử dụng cùng một dữ liệu được dán nhãn mà không có mô-đun bộ nhớ. Hình 3 cho thấy khi kích thước của dữ liệu được dán nhãn giảm, hiệu suất mô hình của CNN-Supervised giảm từ 61,18% (cho 73.257 mẫu được dán nhãn) xuống 2,89% (cho 250 mẫu được dán nhãn), với tổng hiệu suất giảm 58,29% tỷ lệ. Ngược lại, hiệu suất của MA-DNN chỉ giảm 5,94% (8,83-2,89). Điều này cho thấy MA-DNN được đề xuất có thể tận dụng một cách hiệu quả dữ liệu không ghi nhãn bổ sung để tăng hiệu suất mô hình khi cả dữ liệu đào tạo không nhãn có kích thước nhỏ và kích thước lớn được cung cấp.

**Sự phát triển của mô-đun bộ nhớ**. Như đã nói ở trên, hai loại thông tin có thể ghi nhớ cấp độ lớp được ghi trong mô-đun bộ nhớ là (1) biểu diễn tính năng cấp lớp (nhúng chính) và (2) độ không đảm bảo suy luận mô hình (nhúng giá trị). Để hiểu cách mô-đun bộ nhớ được cập nhật trong quá trình đào tạo, chúng tôi hình dung sự phát triển của các nhúng chính và các giá trị nhúng trong Hình 4, 5 và phân tích định tính các tác động của chúng như dưới đây.



**Hình 5:** Hình dung về sự phát triển của các giá trị nhúng trên (a) CIFAR10 và (b) CIFAR100. Trong mỗi khối, mỗi hàng tương ứng với việc nhúng giá trị mỗi lớp, tức là, một dự đoán xác suất nhiều lớp mã hóa sự không chắc chắn suy luận mạng cấp lớp ở các kỷ nguyên khác nhau trong quá trình đào tạo.

**Ảnh hưởng của các Key Embedding**. Như Hình 4 cho thấy, các nhúng chính (được biểu thị là các chấm đen) về cơ bản được cập nhật dưới dạng các cụm sao để nắm bắt cấu trúc đa tạp toàn cầu trong không gian đặc trưng. Cụ thể, chúng tôi có các quan sát sau: (i) Hình 4 (a) cho thấy mặc dù các nhúng chính được khởi tạo là 0 mà không áp dụng kiến ​​thức trước, chúng luôn được cập nhật để nắm bắt cấu trúc đa dạng cơ bản của dữ liệu được gắn nhãn trong chiếu không gian tính năng 2 chiều, như đã thấy ở các kỷ nguyên 50/250. (ii) Hình 4 (b) cho thấy có sự chồng chéo phân loại lớp nghiêm trọng của dữ liệu không ghi nhãn ban đầu; tuy nhiên, sự chồng chéo phân phối lớp như vậy có xu hướng giảm dần khi mô hình được đào tạo. (iii) Hình 4 (c) cho thấy các phần nhúng chính cũng nắm bắt được cấu trúc đa dạng toàn cầu của dữ liệu thử nghiệm không nhìn thấy, mặc dù mạng không được tối ưu hóa để phân phối dữ liệu thử nghiệm. Nhìn chung, những quan sát này phù hợp với động lực của chúng tôi trong việc ghi lại các trọng tâm cụm được cập nhật tích lũy như là các nhúng chính để có được sự phân công xác suất trên các mẫu không ghi nhãn dựa trên giả định cụm. Ngoài ra, sự phát triển của phân phối dữ liệu không ghi nhãn trong Hình 4 (b) cũng cho thấy một cách định tính rằng việc mất trí nhớ của chúng tôi sẽ xử phạt sự phân bổ lớp chồng chéo và khiến ranh giới quyết định của lớp nằm trong vùng mật độ thấp. Lưu ý rằng trực quan hóa dữ liệu chiều cao 2 chiều của dữ liệu chiều cao có thể không hoàn toàn phản ánh cấu trúc cơ bản về cách các lớp được phân tách trong không gian tính năng. Hiệu ứng của các Giá trị nhúng. Như Hình 5 cho thấy, các giá trị nhúng về cơ bản ghi lại sự không chắc chắn suy luận mô hình ở cấp độ lớp. Ở các giai đoạn đào tạo ban đầu, các giá trị nhúng sẽ tạo ra độ không đảm bảo suy luận cao hơn nhiều (phân phối đa phương thức với entropy cao hơn), nhưng dần dần tái hiện độ không chắc chắn suy luận thấp hơn (phân phối cực đại với entropy thấp hơn) khi mô hình được đào tạo dần dần. Trong thực tế, khi loại bỏ các giá trị nhúng, việc gán xác suất cho các mẫu không ghi nhãn có thể trở nên đặc biệt không đáng tin cậy ở các giai đoạn đào tạo trước đó, điều này thậm chí dẫn đến hiệu suất giảm 0,69 / 1,94 / 2,78% trên SVHN / CIFAR10 / CIFAR100 trong các thử nghiệm của chúng tôi. Do đó, các giá trị nhúng có thể phục vụ cho việc phân tách lớp trong không gian nhãn và được sử dụng để làm trơn tru việc gán xác suất với độ không đảm bảo suy luận mô hình để đưa ra dự đoán bộ nhớ đáng tin cậy hơn.

**Sự phát triển của dự đoán bộ nhớ**. Chúng tôi hình dung sự phát triển của các dự đoán bộ nhớ trên các mẫu không ghi nhãn từ CIFAR10 ở các giai đoạn đào tạo khác nhau trong Hình 6. Có thể thấy rằng các dự đoán bộ nhớ đang dần cải thiện từ không chắc chắn hơn (mơ hồ) sang các mẫu đào tạo không rõ ràng hơn. Điều này không chỉ thể hiện tính chất hội tụ tốt của MA-DNN, mà còn cho thấy mức độ mất trí nhớ trong học tập mô hình - (1) phạt phân phối lớp chồng chéo khi đưa ra dự đoán bộ nhớ không chắc chắn ở giai đoạn đào tạo trước đó trong khi (2) khuyến khích dự đoán mạng phù hợp với dự đoán bộ nhớ, sao cho dữ liệu không ghi nhãn được tối ưu hóa theo cấu trúc đa tạp bên dưới.

1. **Conclusions**

Trong tác phẩm này, chúng tôi trình bày một Mạng lưới thần kinh sâu hỗ trợ bộ nhớ (MADNN) mới để cho phép học sâu bán giám sát trên dữ liệu đào tạo không có nhãn và thưa thớt. MA-DNN được thành lập dựa trên ý tưởng khai thác bộ nhớ học tập mô hình để học hỏi một cách đáng tin cậy và hiệu quả hơn từ dữ liệu đào tạo không ghi nhãn. Cụ thể, chúng tôi xây dựng một tương tác đồng hóa mới giữa mạng và mô-đun bộ nhớ ngoài có khả năng tạo điều kiện cho việc học sâu bán giám sát hiệu quả hơn bằng cách áp dụng mất bộ nhớ xuất phát từ mô-đun bộ nhớ được cập nhật tăng dần. Đánh giá so sánh mở rộng trên ba bộ dữ liệu điểm chuẩn phân loại hình ảnh được giám sát bán xác nhận các ưu điểm của MA-DNN được đề xuất trên một loạt các phương pháp tiên tiến. Chúng tôi cũng cung cấp các nghiên cứu cắt bỏ chi tiết và phân tích sâu hơn để cung cấp cái nhìn sâu sắc về thiết kế mô hình và tăng hiệu suất.