

**ĐẠI HỌC QUỐC GIA HÀ NỘI,
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



Phạm Thị Quỳnh Trang

**GIẢM THIỂU QUÊN NGHIÊM TRỌNG TRONG HỌC LIÊN TỤC SỬ DỤNG
CÁC PHƯƠNG PHÁP PHÁT LẠI BỘ NHỚ, CHUYỂN GIAO TRI THỨC
VÀ KIẾN TRÚC ĐỘNG**

TÓM TẮT LUẬN ÁN NGÀNH HỆ THỐNG THÔNG TIN

Ngành đào tạo: Hệ thống thông tin

Mã số: 9480104

Hà Nội - 2026

Công trình được hoàn thành tại: Trường Đại học Công Nghệ, Đại học Quốc gia Hà Nội

Người hướng dẫn	PGS.TS. Nguyễn Trí Thành	Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội
Đồng hướng dẫn	TS. Lê Đức Trọng	Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội
Phản biện	PGS.TS. Bùi Thu Lâm	Học viện Kỹ thuật Mật mã
Phản biện	PGS.TS. Trần Đăng Hưng	Trường Đại học Công nghiệp Hà Nội
Phản biện	PGS.TS. Nguyễn Long Giang	Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Luận án được bảo vệ trước Hội đồng cấp Đại học Quốc gia chấm luận án tiến sĩ họp tại phòng 212-E3 vào hồi 10 giờ 30 ngày 26 tháng 08 năm 2026

NGHIÊN CỨU SINH

CÁN BỘ HƯỚNG DẪN

XÁC NHẬN CỦA ĐƠN VỊ ĐÀO TẠO

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội

Tóm tắt

Học liên tục giải quyết vấn đề **quên nghiêm trọng (catastrophic forgetting)** khi mô hình học máy phải học các tác vụ mới liên tiếp mà không được huấn luyện lại từ đầu trên toàn bộ dữ liệu; các hướng tiếp cận hiện có là phát lại bộ nhớ, chuyển giao tri thức, và mở rộng kiến trúc, lần lượt gặp hạn chế về quyền riêng tư/bộ nhớ, dung lượng tham số, và chi phí tính toán tăng tuyến tính. Luận án giải quyết vấn đề này qua ba giai đoạn tương ứng với các mức độ ràng buộc về bộ nhớ và độ dài chuỗi tác vụ: ở giai đoạn không thể lưu dữ liệu thật, luận án đề xuất tăng cường bộ nhớ bằng nguyên mẫu kết hợp mô-đun căn chỉnh dựa trên hàm năng lượng để sửa hiện tượng trôi dạt biểu diễn; ở giai đoạn không có bất kỳ bộ nhớ nào, luận án phát triển CL-plugin (chuyển giao tri thức có chọn lọc qua hai hàm mất mát CED và CSC cho phân loại cảm xúc theo khía cạnh) và ZeroRFF (biến học tăng dần thành hồi quy giải bằng công thức đóng); và ở giai đoạn chuỗi tác vụ dài và đa dạng, luận án đề xuất bốn phương pháp với kiến trúc linh hoạt và tham số phân bố động, khung OPE và APF (cùng biến thể tích hợp căn chỉnh không gian ẩn cho phân loại ảnh y tế) tách biệt nhận dạng tác vụ khỏi phân loại nhãn, DAKTA kết hợp LoRA với bộ phân loại Kolmogorov-Arnold cùng chính quy hóa Fisher Information Matrix, và SO-LoRA xây dựng cơ sở trực giao chung giúp dung lượng adapter không đổi theo độ dài chuỗi tác vụ. Các thực nghiệm toàn diện trên nhiều bộ dữ liệu chuẩn về văn bản và ảnh, cùng các phân tích ablation và lý thuyết, xác nhận các phương pháp đề xuất đạt hiệu suất tốt nhất so với các phương pháp hiện có, cung cấp giải pháp có cơ sở lý luận vững chắc và tính ứng dụng cao cho các môi trường khắt khe về quyền riêng tư, bộ nhớ và tính toán.

MỞ ĐẦU

Động lực nghiên cứu

Học máy tĩnh huấn luyện mô hình trên tập dữ liệu cố định và triển khai mà không thích nghi về sau, dẫn đến bốn hạn chế liên quan chặt chẽ: thiếu khả năng thích nghi với thay đổi phân phối dữ liệu, tổng quát hóa kém trên dữ liệu mới, không có cơ chế cập nhật liên tục, và dễ tổn thương trước concept drift. **Học liên tục (Continual Learning - CL)** khắc phục các hạn chế này bằng cách cho phép mô hình tiếp thu tri thức mới một cách tăng dần trong khi vẫn duy trì tri thức cũ, giảm nhu cầu huấn luyện lại toàn bộ và hiện thực hóa các hệ thống AI thích nghi trong môi trường phi tĩnh. Hình 1 đối chiếu hai mô hình học, còn Hình 2 cho thấy sự tăng trưởng nhanh của nghiên cứu CL, từ 243 công bố năm 2018 lên hơn 1.800 công bố năm 2024.



Figure 1: (a) Học máy tĩnh huấn luyện trên toàn bộ dữ liệu một lần. (b) Học máy liên tục học trên dữ liệu mới khi chúng xuất hiện.

Luận án được thúc đẩy bởi ba khoảng trống nghiên cứu. **Thứ nhất, quyền riêng tư dữ liệu:** phát lại bộ nhớ đòi hỏi lưu trữ mẫu thô từ tác vụ cũ, không khả thi trong các lĩnh vực nhạy cảm như y tế, tài chính. **Thứ hai, sự phụ thuộc vào bộ nhớ:** ngay cả phương pháp xấp xỉ nguyên mẫu cũng kém tin cậy khi không gian đặc trưng phức tạp, và bộ đệm bộ nhớ còn lại có thể không khả thi dưới ngân sách tính toán hạn chế. **Thứ ba, dung**

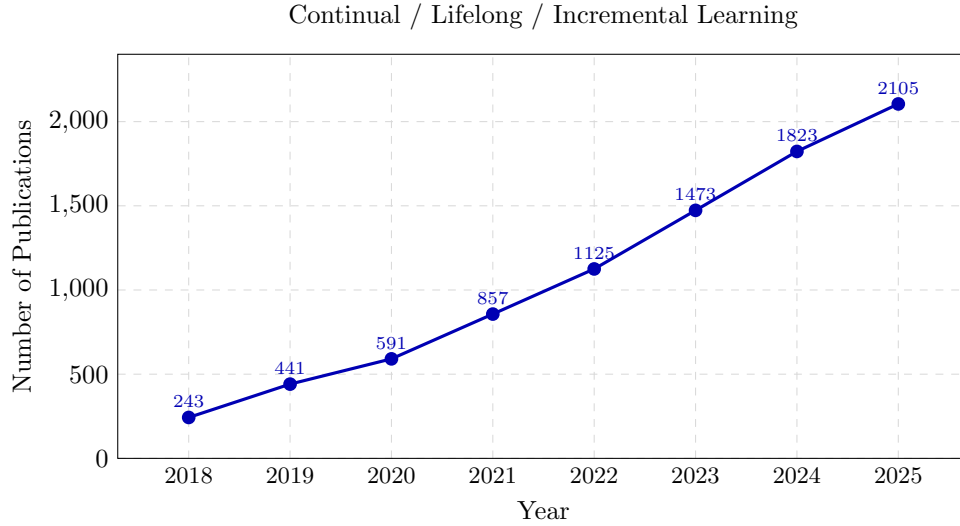


Figure 2: Xu hướng công bố hàng năm của các nghiên cứu về *continual learning*, *lifelong learning* và *incremental learning* trên DBLP (2018–2025).

lượng mô hình: với chuỗi tác vụ dài hoặc đa dạng, ngân sách tham số cố định buộc đánh đổi giữa ổn định và dẻo dai mà chính quy hóa hay chất lọc tri thức không thể giải quyết triệt để.

Ba khoảng trống này định hướng cấu trúc luận án. Chương 2 giải quyết khoảng trống thứ nhất bằng prototype lớp nhiều Gaussian thay cho mẫu thực, kết hợp mô-đun căn chỉnh ẩn dựa trên năng lượng. Chương 3 giải quyết khoảng trống thứ hai qua chuyển giao tri thức không cần exemplar: CL-plugin với chất lọc đối lập cho task-incremental, và ZeroRFF (random Fourier features trên backbone đóng băng, nghiệm phân tích dạng đóng) cho class-incremental. Chương 4 giải quyết khoảng trống thứ ba qua các chiến lược kiến trúc thích nghi: khung đa đầu phân loại với Online Prototype Equilibrium và Adaptive Prototypical Feedback; DAKTA kết hợp LoRA với bộ phân loại Kolmogorov-Arnold; và SO-LoRA tách biệt không gian con qua cơ sở trực giao cố định và độ thưa nhóm.

Ba câu hỏi nghiên cứu xuyên suốt luận án: (Q1) tăng cường prototype giảm thiểu quên nghiêm trọng trong khi bảo toàn quyền riêng tư như thế nào; (Q2) chuyển giao tri thức chọn lọc giảm dư thừa thông tin đến mức nào mà không cần exemplar; (Q3) phân bổ tham số động ngăn quên nghiêm trọng, hạn chế chi phí tính toán và chia sẻ tri thức giữa các tác vụ như thế nào. Các đóng góp tương ứng được công bố tại bảy công bố. Hình 3 minh họa cấu trúc tổng thể luận án.

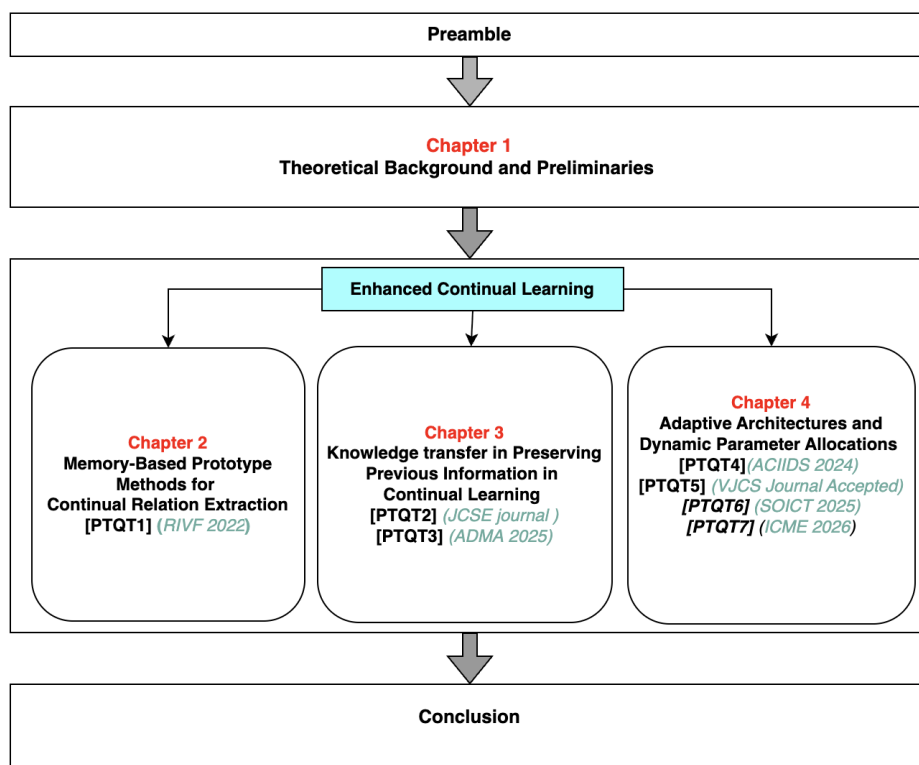


Figure 3: Sơ đồ cấu trúc luận án và các công bố liên quan trong từng chương.

Chapter 1

Nền tảng lý thuyết và các kiến thức tiên đề

Học liên tục (Continual Learning - CL) là mô hình học máy trong đó mô hình được huấn luyện tuần tự trên một chuỗi tác vụ, hướng tới hiệu suất tương đương khi dữ liệu được quan sát đồng thời. Trở ngại cốt lõi là **catastrophic forgetting**: cập nhật tham số cho tác vụ mới làm suy giảm hiệu suất tác vụ cũ, phản ánh **stability-plasticity dilemma**. Thrun (1996) và Chen và Liu (2018) hình thức hóa CL như hệ thống học tuần tự, duy trì Cơ sở Tri thức tích lũy và khai thác tri thức quá khứ cho tương lai, mà không lưu dữ liệu thô. CL còn đối mặt các thách thức về chuyển giao tri thức, xung đột giữa các tác vụ, dung lượng bộ nhớ hữu hạn và trôi dạt phân phối.

Các kịch bản chuẩn được phân theo cách nhận dạng tác vụ và cấu trúc nhân, gồm IIL, DIL, TIL, CIL (được nghiên cứu nhiều nhất), TFCL/OCL, BBCL và CPT. Về phương pháp, năm nhóm chính là: **dựa trên chính quy hóa** (EWC, SI, LwF) phạt thay đổi tham số quan trọng; **dựa trên replay** (iCaRL, DER, GEM) lưu/tổng hợp mẫu cũ để ôn luyện, đánh đổi bằng chi phí bộ nhớ và rủi ro riêng tư; **dựa trên tối ưu hóa** điều chỉnh huấn luyện qua chiều gradient hoặc meta-learning; **dựa trên biểu diễn** tận dụng tiền huấn luyện đối lập/tự giám sát kết hợp prototype; và **dựa trên kiến trúc** (Progressive NN, PackNet, adapter) phân bổ mô-đun riêng cho từng tác vụ. Đánh giá dựa trên ba chiều: hiệu quả tổng thể (AA, AIA), tính ổn định (FM, BWT) và tính mềm dẻo (IM, FWT).

Tổng kết

Ba khoảng trống nghiên cứu định hướng luận án: (1) các phương pháp replay lưu dữ liệu thực gây *lo ngại quyền riêng tư*, cần giải pháp sinh dữ liệu tổng hợp; (2) chất lọc tri

thực hiện tại *thiếu hiệu quả*, cần chiến lược chọn lọc tận dụng mô hình tiền huấn luyện;
(3) các phương pháp kiến trúc *thiếu cơ chế chia sẻ tri thức* giữa các tác vụ, cần thiết kế
động cân bằng giữa đặc thù hóa và tri thức chung.

Chapter 2

Phương pháp Replay dựa trên Bộ nhớ cho Trích xuất Quan hệ Liên tục

Trích xuất quan hệ (Relation Extraction - RE) là bài toán nhận diện và phân loại các quan hệ ngữ nghĩa giữa các cặp thực thể trong văn bản, đóng vai trò là bài toán con cốt lõi của Trích xuất thông tin trong các ứng dụng như hỏi đáp, phân tích cảm xúc và tìm kiếm có cấu trúc. Các mô hình RE truyền thống hoạt động trên một tập quan hệ cố định được định nghĩa trước, điều này hạn chế khả năng áp dụng trong thực tế khi các loại quan hệ mới liên tục xuất hiện. **Trích xuất quan hệ liên tục (Continual Relation Extraction - CRE)** giải quyết hạn chế này bằng cách huấn luyện mô hình tuần tự trên các tập quan hệ mới trong khi vẫn ngăn chặn hiện tượng quên thảm khốc (catastrophic forgetting) đối với các quan hệ đã học trước đó. Trong số các hướng giải pháp chính – dựa trên điều chuẩn (regularization), kiến trúc động (dynamic architecture) và dựa trên bộ nhớ (memory-based) – các phương pháp dựa trên bộ nhớ luôn đạt hiệu năng tốt nhất trong các tác vụ NLP nhờ việc duy trì một bộ nhớ episodic lưu trữ các mẫu dữ liệu cũ để replay lại.

Công trình liên quan gần nhất là **Consistent Representation Learning (CRL)**, kết hợp học tương phản có giám sát (supervised contrastive learning) với chưng cất tri thức (knowledge distillation) để duy trì các biểu diễn quan hệ ổn định qua các tác vụ tuần tự. Dù đạt hiệu năng tốt, CRL vẫn tồn tại ba hạn chế chính: (1) phụ thuộc hoàn toàn vào việc lưu trữ dữ liệu thật, khiến phương pháp dễ bị giới hạn bởi bộ nhớ và gặp vấn đề về quyền riêng tư; (2) ngân sách bộ nhớ cho mỗi quan hệ bị thu hẹp dần khi số tác vụ tăng lên, dẫn

đến nguy cơ overfitting trên các mẫu thưa; và (3) hiệu năng kém trên các tập dữ liệu mất cân bằng như TACRED, nơi các lớp chiếm ưu thế lấn át việc học các quan hệ thiểu số trong quá trình replay tương phản.

Để khắc phục các hạn chế trên, chương này đề xuất **ProtoAug**, tích hợp kỹ thuật tăng cường nguyên mẫu (prototype augmentation) vào khung CRL nhằm giảm sự phụ thuộc vào dữ liệu thật được lưu trữ. Phương pháp gồm bốn bước: (1) một bộ mã hóa BERT trích xuất biểu diễn quan hệ, tối ưu bằng hàm mất mát tương phản có giám sát; (2) phân cụm k-means chọn ra mẫu đại diện nhất cho mỗi quan hệ để lưu vào bộ nhớ; (3) sinh các mẫu ảo bằng cách nhiễu hóa (perturb) các nguyên mẫu quan hệ với nhiễu Gaussian, làm phong phú bộ nhớ mà không cần lưu thêm dữ liệu thật; và (4) sử dụng cả mẫu thật lẫn mẫu tổng hợp cho việc replay tương phản và chưng cất tri thức nhằm duy trì khả năng phân biệt của các quan hệ đã học trước đó. Hình 2.3 minh họa kiến trúc tổng thể.

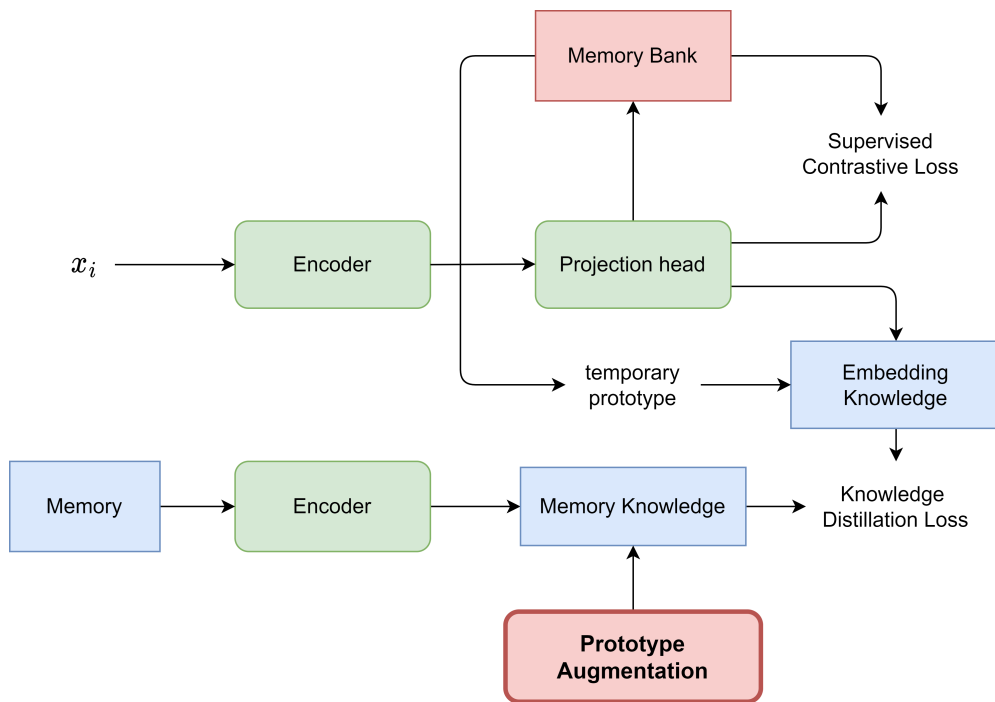


Figure 2.3: Kiến trúc mô-đun biểu diễn nhất quán kết hợp tăng cường nguyên mẫu.

Một mở rộng tiếp theo giới thiệu mô-đun **Căn chỉnh không gian tiềm ẩn dựa trên năng lượng (Energy-based Latent space Alignment - ELI)** nhằm giải quyết hiện tượng trôi biểu diễn (representational drift) giữa các tác vụ. Sau khi huấn luyện mỗi tác vụ, một mô hình dựa trên năng lượng (Energy-Based Model - EBM) được huấn luyện để gán năng lượng thấp cho các biểu diễn tiềm ẩn từ mô hình ở bước trước và năng lượng cao cho biểu diễn từ mô hình hiện tại, sử dụng động lực học Langevin để lấy mẫu xấp xỉ. Ở giai đoạn suy luận, thuật toán ELI di chuyển dần vector tiềm ẩn của mỗi mẫu kiểm tra theo hướng

gradient năng lượng về vùng năng lượng thấp, giúp căn chỉnh lại các biểu diễn bị trôi về gần với không gian đặc trưng của tác vụ trước mà không cần lưu thêm dữ liệu. Hình 2.4 thể hiện kiến trúc đầy đủ kết hợp cả tăng cường nguyên mẫu và ELI.

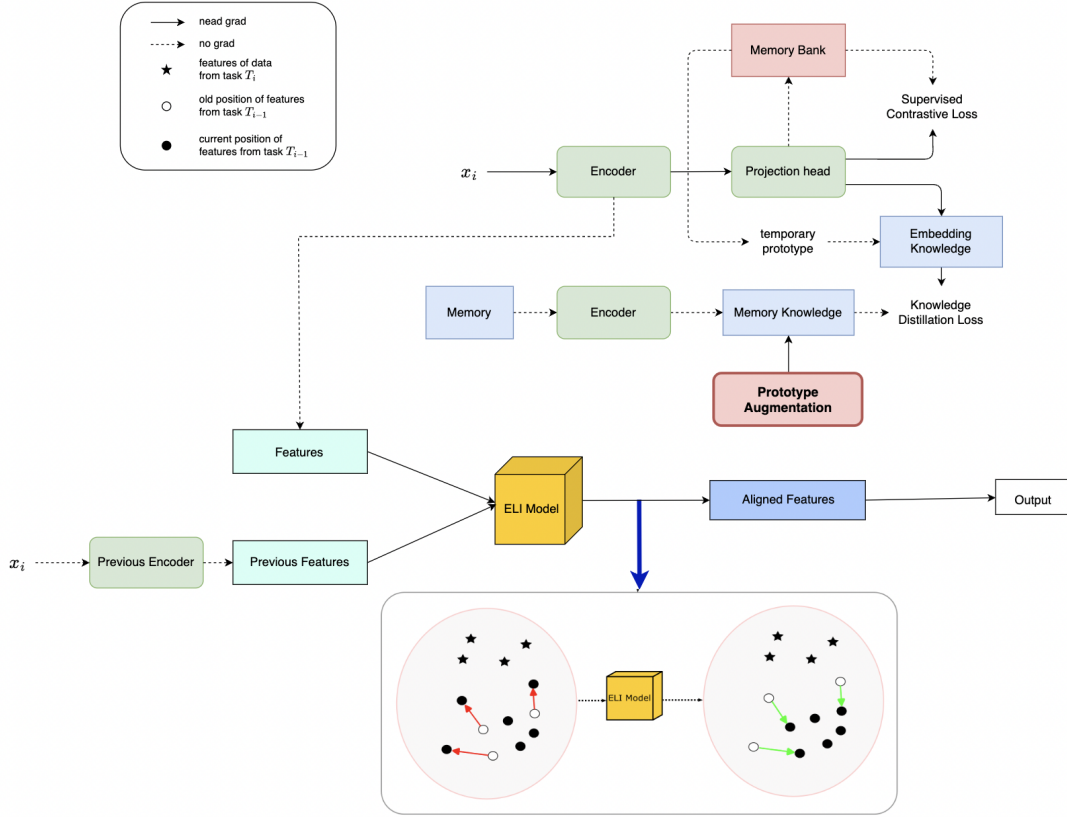


Figure 2.4: Kiến trúc tổng thể của mô hình trích xuất quan hệ liên tục được đề xuất, kết hợp tăng cường nguyên mẫu và căn chỉnh không gian tiềm ẩn dựa trên năng lượng (ELI).

Thí nghiệm trên hai bộ dữ liệu chuẩn FewRel và TACRED theo thiết lập task-incremental với mười tác vụ tuần tự đã khẳng định hiệu quả của phương pháp đề xuất. Bảng 2.1 và 2.2 báo cáo độ chính xác của từng phương pháp tại mỗi bước tác vụ. Trên FewRel, mô hình đề xuất vượt trội hơn CRL ở tám trên mười tác vụ với mức chênh lệch 1–2%, và vượt qua tất cả các baseline khác trên toàn bộ chuỗi tác vụ. Trên TACRED – một bộ dữ liệu khó hơn và mất cân bằng hơn – mức cải thiện còn rõ rệt hơn, vượt CRL tới 4% ở Task 4 và liên tục vượt EMAR+BERT với biên độ từ 1.1% đến 11.1%. Mức cải thiện càng rõ rệt khi số lượng tác vụ tăng lên, cho thấy khả năng ổn định lâu dài và hiệu quả vượt trội trong việc giảm thiểu hiện tượng quên thảm khốc.

Tóm tắt

Tóm lại, chương này chứng minh rằng việc thay thế các mẫu dữ liệu thật được lưu trữ bằng dữ liệu tổng hợp tăng cường từ nguyên mẫu không chỉ đủ mà còn mang lại lợi

Table 2.1: Độ chính xác (%) của từng mô hình qua các tác vụ trên bộ dữ liệu **FewRel**. In đậm = tốt nhất; gạch dưới = tốt nhì.

Mô hình	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EA-EMR	89.0	69.0	59.1	54.2	47.8	46.1	43.1	40.7	38.6	35.2
EMAR	88.5	73.2	66.6	63.8	55.8	54.3	52.9	50.9	48.8	46.3
CML	91.2	74.8	68.2	58.2	53.7	50.4	47.8	44.4	43.1	39.7
EMAR+BERT	98.8	89.1	89.5	85.7	83.6	84.8	79.3	80.0	77.1	73.8
RP-CRE	97.9	92.7	91.6	89.2	88.4	86.8	85.1	84.1	82.2	81.5
CRL	<u>98.2</u>	<u>94.6</u>	<u>92.5</u>	<u>90.5</u>	<u>89.4</u>	<u>87.9</u>	<u>86.9</u>	<u>85.6</u>	<u>84.5</u>	<u>83.1</u>
Đề xuất	<u>98.2</u>	94.7	93.9	92.2	90.1	89.6	88.6	87.1	86.1	84.6

ích cho bài toán trích xuất quan hệ liên tục. Sự kết hợp giữa lựa chọn mẫu bằng k-means, tăng cường nguyên mẫu bằng nhiễu Gaussian và căn chỉnh không gian tiềm ẩn dựa trên năng lượng tạo ra một phương án thay thế bảo vệ quyền riêng tư, tiết kiệm bộ nhớ so với các phương pháp replay hiện có, đặt nền tảng cho chương tiếp theo – nghiên cứu các phương pháp chưng cất tri thức loại bỏ hoàn toàn nhu cầu lưu trữ dữ liệu.

Table 2.2: Độ chính xác (%) của từng mô hình qua các tác vụ trên bộ dữ liệu **TACRED**.
In đậm = tốt nhất; gạch dưới = tốt nhì.

Mô hình	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
EA-EMR	47.5	40.1	38.3	29.9	24.0	27.3	26.9	25.8	22.9	19.8
EMAR	73.6	57.0	48.3	42.3	37.7	34.0	32.6	30.0	27.6	25.1
CML	57.2	51.4	41.3	39.3	35.9	28.9	27.3	26.9	24.8	23.4
EMAR+BERT	96.6	85.7	81.0	78.6	73.9	72.3	71.7	72.2	72.6	71.0
RP-CRE	97.6	90.6	86.1	82.4	79.8	77.2	75.1	73.7	72.4	72.4
CRL	<u>97.7</u>	<u>93.2</u>	<u>89.8</u>	<u>84.7</u>	<u>84.1</u>	<u>81.3</u>	<u>80.2</u>	<u>79.1</u>	<u>79.0</u>	<u>78.0</u>
Đề xuất	97.8	95.2	90.6	88.2	84.6	81.7	82.8	80.5	80.0	79.6

Chapter 3

Chuyển giao tri thức trong bảo tồn tri thức cũ trong học liên tục

Chương này khảo sát chuyển giao tri thức như một cơ chế giảm thiểu catastrophic forgetting mà không cần lưu trữ exemplar. Hai chiến lược bổ trợ nhau được đề xuất: (1) chất lọc tập hợp đối lập trong khung **Task-Incremental Learning (TIL)** cho bài toán phân loại cảm xúc theo khía cạnh (ASC); (2) học phân tích dạng đóng trên backbone đóng băng trong khung **Class-Incremental Learning (CIL/GCIL)**. Hai hướng này thể hiện hai bản chất khác nhau để đạt được sự ổn định, chất lọc dựa trên đạo hàm và tính toán phân tích không cần đạo hàm, nhưng cùng đạt mục tiêu học liên tục không exemplar.

Phần 1: Chất lọc tri thức kết hợp học đối lập (TIL).

Một **CL-plugin** được tích hợp vào BERT tại hai vị trí, thay thế adapter chuẩn, gồm hai mô-đun: **KSM** (lớp capsule với định tuyến động, chia sẻ tri thức chung giữa các tác vụ) và **TSM** (lớp kết nối đầy đủ với mặt nạ mềm mức nơ-ron, bảo vệ tham số then chốt khỏi cập nhật gradient ở tác vụ sau). Hình 3.3 và 3.4 minh họa kiến trúc.

Hai hàm mất mát đối lập bổ sung cho cross-entropy: **CED** (Contrastive Ensemble Distillation) chất lọc tri thức từ các mô hình tác vụ trước, coi logit cùng đầu vào từ mô hình cũ/mới là cặp dương; **CSC** (Contrastive Supervised Classification) tối đa hóa tương đồng nội lớp, tối thiểu hóa tương đồng liên lớp trên đầu ra TSM. Hàm mục tiêu tổng:

$$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{CED} + \mathcal{L}_{CSC}. \quad (3.1)$$

Thực nghiệm trên 19 tác vụ ASC từ bốn benchmark (HL5Domains, Liu3Domains, Ding9Domains, SemEval14). Trung bình trên năm thứ tự tác vụ ngẫu nhiên, mô hình

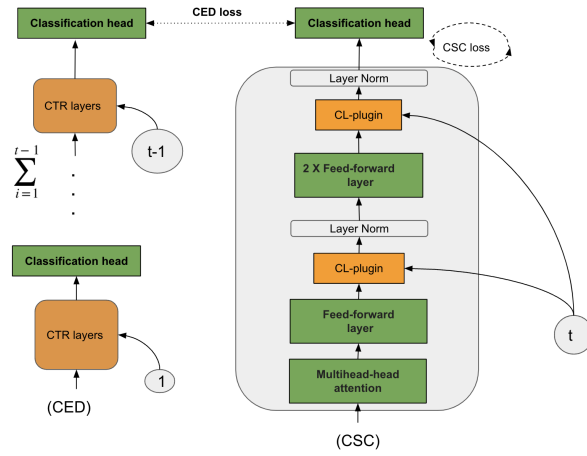


Figure 3.3: Kiến trúc tổng thể của mô hình CL-plugin đề xuất.

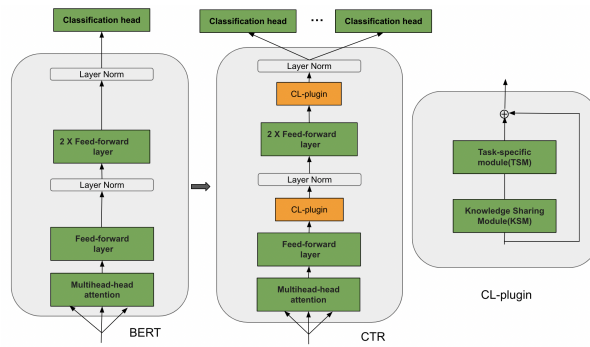


Figure 3.4: BERT (trái), CTR (giữa), CL-plugin tích hợp vào BERT (phải).

đề xuất đạt Acc. 88,50% và Macro-F1 tốt nhất 82,35% trong nhóm Adapter-BERT; CTR* (đối thủ gần nhất cùng điều kiện) đạt 88,21%/81,19%. LAMOL có Acc. cao hơn (88,91%) nhưng dựa vào backbone sinh GPT-2, chi phí tính toán lớn hơn đáng kể.

Ablation: loại bỏ CL-plugin làm giảm 2,67% Acc. và 2,60% MF1; loại bỏ cả CED và CSC làm giảm 2,02% Acc. và 3,03% MF1 — xác nhận mỗi thành phần đều đóng góp tích cực.

Phần 2: Lưu giữ tri thức không cần exemplar qua học phân tích (GCIL).

Nhằm khắc phục hạn chế của GACL (bộ phân loại tuyến tính trên đặc trưng thô, khó khăn với phân phối không phân tách tuyến tính), **ZeroRFF** tích hợp **Random Fourier Features (RFF)** vào khung học liên tục phân tích (Hình 3.8).

ZeroRFF gồm bốn bước cho mỗi tác vụ: (1) backbone DeiT-S/16 đóng băng trích xuất đặc trưng $X_k^{(E)}$; (2) mô-đun RFF ánh xạ sang không gian $D=5000$ chiều qua $z(x) = \sqrt{2/D} \cos(\omega^\top x + \beta)$ với $\omega \sim \mathcal{N}(0, \sigma^{-2}I)$, $\beta \sim \text{Uniform}(0, 2\pi)$, xấp xỉ nhân RBF để dữ liệu phân tách tuyến tính hơn; (3) học phân tích cập nhật đệ quy ma trận tương

Table 3.1: Kết quả trung bình (Acc., MF1) trên năm chuỗi ngẫu nhiên gồm 19 tác vụ ASC. In đậm = tốt nhất.

Kịch bản	Backbone	Mô hình	Acc.(%)	MF1(%)
Không CL	BERT	MTL	91.91	88.11
Học liên tục	BERT (Đóng băng)	OWM	87.02	79.31
		HAT	86.14	78.52
		BCL	88.29	81.40
	Adapter-BERT	LAMOL	88.91	80.59
		CTR	89.47	83.62
		CTR*	88.21	81.19
		Ours	88.50	82.35

Table 3.2: Kết quả ablation.

Biến thể	Acc.(%)	MF1(%)
Ours (đầy đủ)	88.50	82.35
Không CL-plugin	85.83	79.75
Không CSC + CED	86.48	79.32
Không CSC	87.52	79.93
Không CED	87.98	80.10

quan R_k (Sherman-Morrison) và suy ra trọng số $\hat{W}_{\text{FCN}}^{(k)}$ dạng đóng, không cần exemplar hay tối ưu lặp; (4) lặp qua K tác vụ, trả về trọng số cuối cùng. Thuật toán 1 tóm tắt quy trình.

Trên CIFAR-100, kịch bản Si-Blurry GCIL (50% lớp rời rạc, 10% mẫu mờ, 5 seed): ZeroRFF đạt $A_{\text{AUC}}=60,48$, $A_{\text{AVG}}=58,55$, $A_{\text{Last}}=72,44$, cải thiện $\sim 2\%$ so với GACL (baseline không bộ nhớ tốt nhất), vượt hầu hết phương pháp dùng buffer 500 mẫu, và tiệm cận MVP-R (bộ nhớ 2000) dù không lưu trữ dữ liệu.

Tổng kết. Hai chiến lược chuyển giao tri thức không exemplar chiếm giữ hai vị trí khác biệt trong không gian thiết kế học liên tục: CL-plugin phù hợp khi có sự chồng chéo ngữ nghĩa liên tác vụ phong phú; ZeroRFF được ưu tiên khi có backbone đóng băng mạnh và hiệu quả huấn luyện/quyền riêng tư dữ liệu là ưu tiên hàng đầu, cho bảo đảm

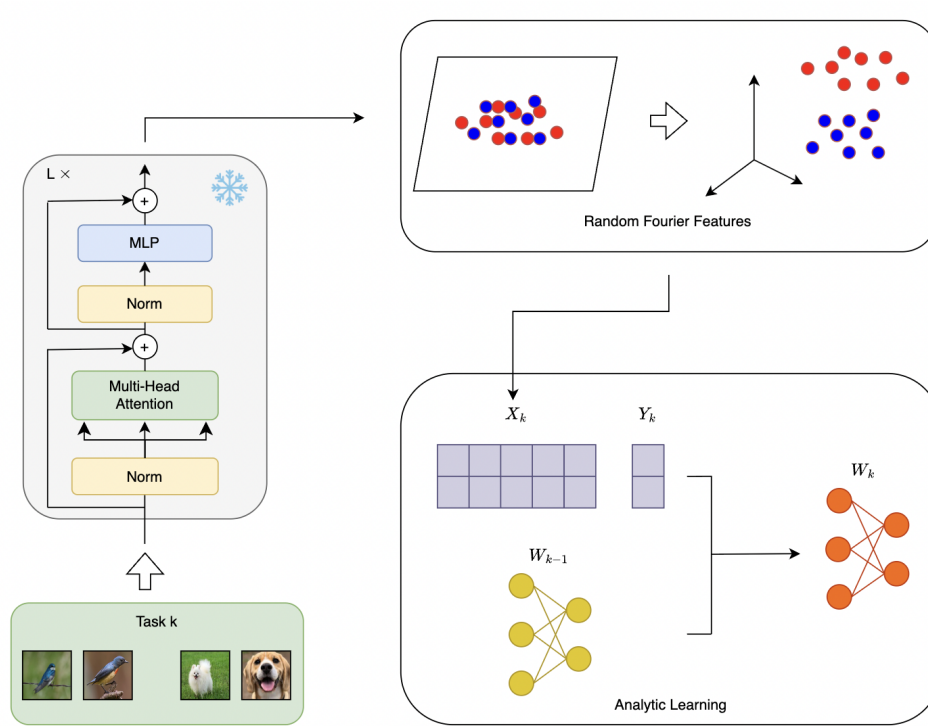


Figure 3.8: Kiến trúc tổng thể của ZeroRFF cho GCIL.

dạng đóng và dung lượng bộ nhớ không đổi theo số tác vụ. Cả hai đều bảo tồn tri thức tích lũy mà không gánh chịu chi phí lưu trữ và rủi ro quên

Algorithm 1 Huấn luyện ZeroRFF

Require: Các tác vụ $\mathcal{D}_1^{\text{train}}, \dots, \mathcal{D}_K^{\text{train}}$; trọng số backbone đóng băng Θ_{backbone}

Ensure: $\hat{W}_{\text{FCN}}^{(K)}$

- 1: **Khởi tạo:** $R_0 \leftarrow \gamma I, \quad \hat{W}_{\text{FCN}}^{(0)} \leftarrow 0$
 - 2: **for** $k = 1$ **đến** K **do**
 - 3: $X_k^{(E)} \leftarrow f_{\text{backbone}}(X_k^{\text{train}}, \Theta_{\text{backbone}})$
 - 4: $X_k^{(B)} \leftarrow \text{rff}(X_k^{(E)})$
 - 5: $R_k \leftarrow R_{k-1} - R_{k-1} X_k^{(B)\top} (I + X_k^{(B)} R_{k-1} X_k^{(B)\top})^{-1} X_k^{(B)} R_{k-1}$
 - 6: $\hat{W}^{(k)} \leftarrow \hat{W}^{(k-1)} - R_{k-1} X_k^{(B)\top} X_k^{(B)} \hat{W}^{(k-1)} + R_k X_k^{(B)\top} Y_k^{\text{train}}$
 - 7: **end for**
 - 8: **return** $\hat{W}_{\text{FCN}}^{(K)}$
-

Table 3.3: Kết quả trên CIFAR-100, kịch bản Si-Blurry (%). In đậm = tốt nhất trong nhóm không bộ nhớ.

Bộ nhớ	Phương pháp	A_{AUC}	A_{AVG}	A_{Last}
2000	MVP-R	60.62 ± 1.03	57.58 ± 0.56	64.30 ± 0.29
500	MVP-R	56.20 ± 1.44	53.61 ± 0.04	55.35 ± 0.43
0 (không bộ nhớ)	SLDA	53.00 ± 3.85	50.09 ± 2.77	61.79 ± 3.81
	GACL	57.99 ± 2.46	56.24 ± 3.12	70.31 ± 0.06
	ZeroRFF (Ours)	60.48 ± 4.33	58.55 ± 6.05	72.44 ± 0.13

Chapter 4

Kiến trúc động và phân bổ tham số động

Học liên tục dựa trên kiến trúc giải quyết catastrophic forgetting bằng cách phân bổ động dung lượng mô hình cho mỗi tác vụ mới, thay vì ràng buộc cập nhật (chính quy hóa) hay lưu trữ dữ liệu quá khứ (replay). Chương này trình bày ba đóng góp bổ trợ nhau trong kịch bản **Học tăng dần theo lớp (Class-Incremental Learning - CIL)**: hai phương pháp trên backbone ResNet-18 huấn luyện từ đầu, và một phương pháp trên backbone ViT-B/16 tiền huấn luyện theo hướng tinh chỉnh hiệu quả tham số (PEFT).

Phần 1: Multi-Head với OPE, APF và ELI.

Đóng góp thứ nhất mở rộng khung OnPro nhằm khắc phục hai hạn chế: đầu phân loại chung làm lẫn lộn dự đoán trong tác vụ (WP) với nhận dạng tác vụ (TP), và xử lý đồng đều mọi cặp lớp bất kể mức độ dễ nhầm lẫn. Phương pháp đề xuất (Hình 4.2) gồm ba cơ chế:

- **OPE (Online Prototype Equilibrium)**: cân bằng việc học lớp mới và bảo toàn lớp cũ bằng cách kéo gần prototype cùng lớp, đẩy xa prototype khác lớp, thông qua $\mathcal{L}_{\text{OPE}} = \mathcal{L}_{\text{pro}}^{\text{new}} + \mathcal{L}_{\text{pro}}^{\text{seen}}$.
- **APF (Adaptive Prototypical Feedback)**: nhận dạng cặp lớp dễ nhầm lẫn qua khoảng cách prototype, sau đó lấy mẫu quá mức các cặp này khi replay.
- **ELI (Energy-based Latent Alignment)**: dùng mô hình năng lượng (EBM) và Langevin dynamics để hiệu chỉnh trôi dạt biểu diễn về đa tạp đặc trưng của tác vụ trước, không cần lưu thêm dữ liệu.

Lớp phân loại tách thành hai đầu (OOD và WP), huấn luyện theo ba bước: (1) huấn luyện chung bộ trích xuất đặc trưng và đầu OOD; (2) tinh chỉnh đầu WP; (3) tinh chỉnh các đầu OOD trước đó trên dữ liệu replay.

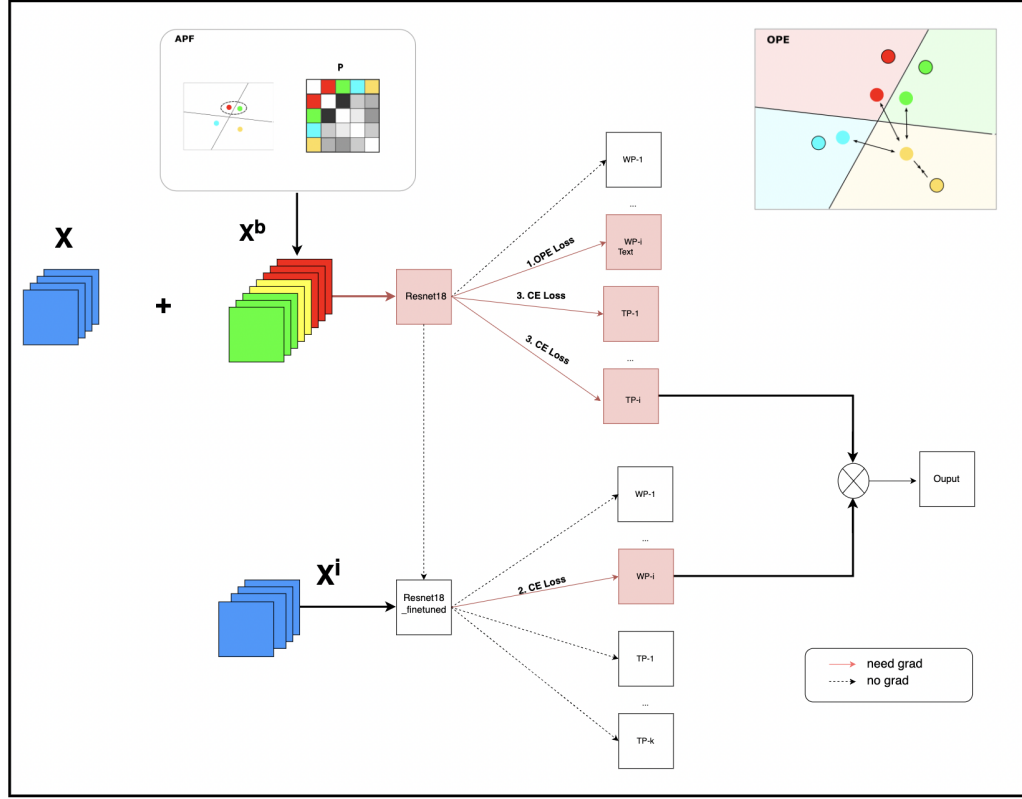


Figure 4.2: Kiến trúc multi-head với OPE, APF và hai đầu phân loại (WP và OOD).

Phương pháp đề xuất (ModifiedOnPro) vượt OnPro tới 11,2 điểm trên CIFAR-10 ($\mathcal{M}=100$: 52,4% $\rightarrow$ 63,6%) và 6,4 điểm trên CIFAR-100 ($\mathcal{M}=500$: 20,7% $\rightarrow$ 27,1%), đồng thời đạt điểm quên tốt nhất (2,9) trên CCH5000 ở kịch bản hai lớp/tác vụ.

Table 4.1: Độ chính xác trung bình (%). In đậm = tốt nhất.

Phương pháp	CIFAR-10	CIFAR-10	CIFAR-100	CIFAR-100
	$\mathcal{M}=100$	$\mathcal{M}=500$	$\mathcal{M}=500$	$\mathcal{M}=2000$
OCM	47.5	70.1	19.7	34.4
OnPro	52.4	67.7	20.7	34.9
Ours	63.6	72.8	27.1	36.3

Phần 2: DAKTA: Bộ phân loại Kolmogorov-Arnold cho Số học tác vụ.

DAKTA (Hình 4.6) xây dựng trên khung ITA-LoRA, thay đầu tuyến tính bằng **Bộ phân loại Kolmogorov-Arnold (KAC)** dựa trên các hàm cơ sở RBF học được, cho ranh giới quyết định nhạy cảm cục bộ. Cơ chế **Hợp nhất vector có định hướng** kết hợp logit RBF với độ tương đồng cosine: $l_c(x) = l_c^{\text{init}}(x) + \gamma_c \cdot \cos(x, W_c)$. **Chính quy hóa Fisher (FIM) bậc hai** được áp dụng cho cả backbone và đầu KAC để hạn chế lệch tham số quan trọng. Các vector tác vụ được tổng hợp bằng lấy trung bình tại thời điểm suy luận (số học tác vụ).

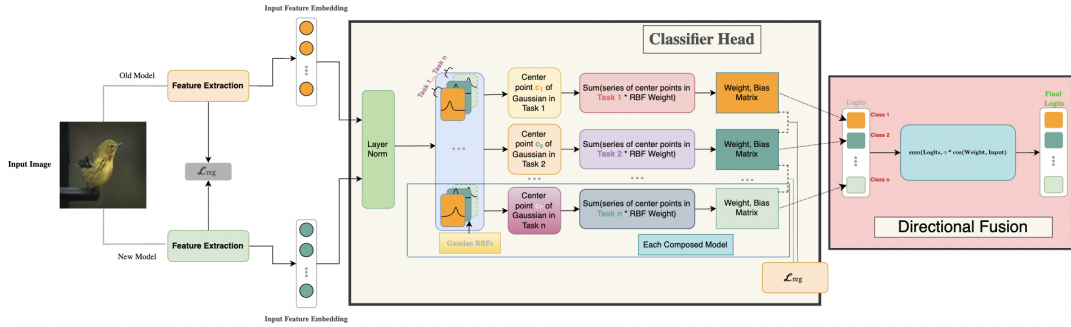


Figure 4.6: Pipeline DAKTA: ViT+LoRA, KAC, số học tác vụ tăng dần và chính quy hóa Fisher.

DAKTA đạt tốt nhất trên 5/6 benchmark, vượt ITA-LoRA tới 2,04 điểm trên CIFAR-100 và CUB-200 (benchmark nhận dạng chi tiết), đồng thời có mức quên thấp nhất trên cả sáu benchmark. Ablation xác nhận cả KAC lẫn chính quy hóa FIM đều thiết yếu.

Table 4.2: Độ chính xác cuối cùng (%) — tóm tắt. In đậm = tốt nhất.

Mô hình	C-100	CUB	RESISC
ITA-LoRA	89.96	85.55	82.00
ITA-(IA) ³	90.66	85.67	83.73
DAKTA (Ours)	91.21	87.59	85.82

Phần 3: SO-LoRA — LoRA Trực giao Thừa thốt.

SO-LoRA (Hình 4.8) giải quyết mâu thuẫn giữa cô lập adapter theo tác vụ (tổn $\mathcal{O}(T)$ bộ nhớ) và adapter dùng chung (dễ nhiễu loạn). Phương pháp duy trì cơ sở trực chuẩn cố định $A \in \mathbb{R}^{d \times r}$ (khởi tạo qua QR, đóng băng) và mỗi tác vụ chỉ học ma trận hệ số thừa thốt $B^{(t)}$: $W^{(t)} = W_0 + AB^{(t)}$. Nhiễu loạn giữa hai tác vụ bị chặn theo chuẩn

$\ell_{2,1}$ của $B^{(t)}$ (group lasso), và một **mặt nạ mềm nhận biết gradient** M bảo vệ các chiều rank quan trọng dựa trên trung bình động mũ của chuẩn gradient.

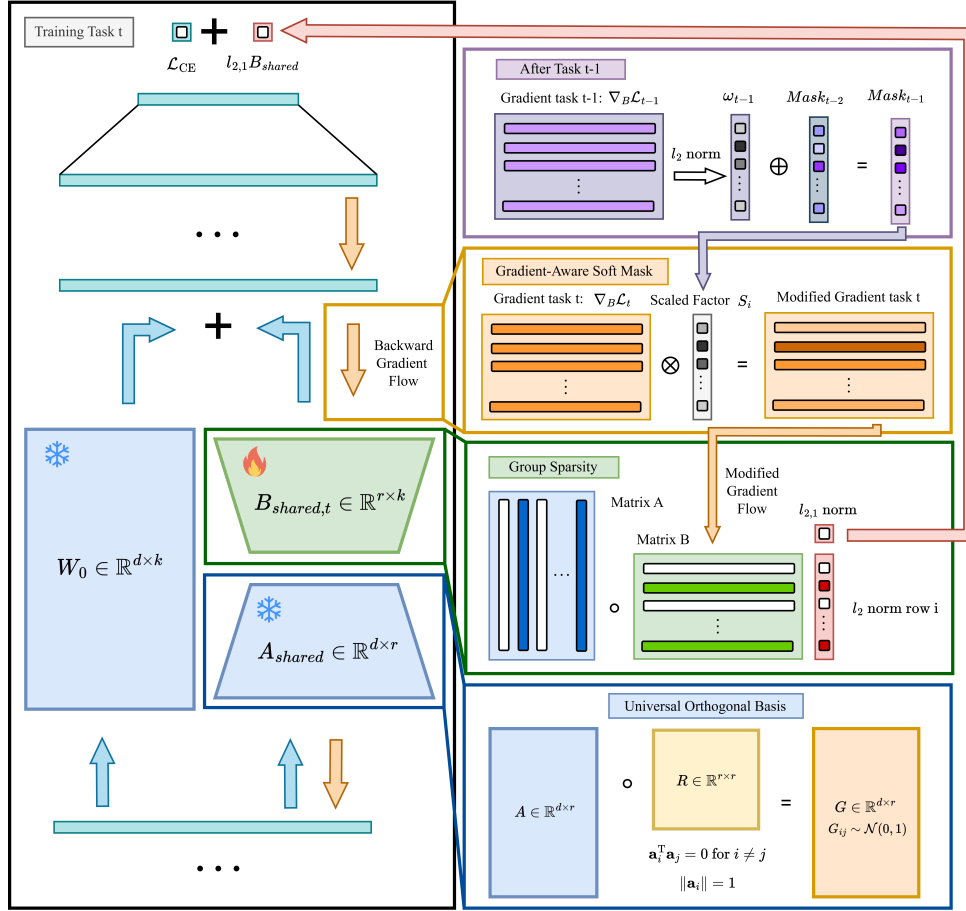


Figure 4.8: SO-LoRA: hệ số thừa thớt $B^{(t)}$ trên cơ sở trực chuẩn cố định A , với mặt nạ nhận biết gradient M .

Trên ImageNet-R, SO-LoRA xếp hạng nhất ở mọi độ chi tiết tác vụ (5/10/20), suy giảm rất nhẹ từ 79,78% xuống 78,15% khi tăng từ 5 lên 20 tác vụ (so với O-LoRA: 72,76% $\rightarrow$ 60,98%). Trên ImageNet-A, SO-LoRA đạt 60,27%, cao hơn gần 5 điểm so với SD-LoRA. Trên CUB-200, SO-LoRA đạt độ chính xác cao nhất (81,58%). Ablation cho thấy mặt nạ nhận biết gradient đóng góp cải thiện nhất quán nhất.

Table 4.3: Hiệu suất tóm tắt (Acc, %). In đậm = tốt nhất.

Phương pháp	IN-R (5)	IN-R (20)	IN-A	CUB
O-LoRA	72.76	60.98	44.46	70.12
SD-LoRA	79.02	75.43	55.35	77.38
SO-LoRA	79.78	78.15	60.27	81.58

Tổng kết

Ba chiến lược dựa trên kiến trúc bổ trợ nhau, giải quyết các hạn chế then chốt của học liên tục: (1) multi-head OPE/APF/ELI tách biệt dự đoán trong tác vụ khởi nhận dạng tác vụ và hiệu chỉnh trôi dạt ẩn trong CIL trực tuyến; (2) DAKTA thay đầu tuyến tính bằng bộ phân loại Kolmogorov-Arnold, mở rộng chính quy hóa Fisher, đạt độ chính xác cao và mức quên thấp trên các benchmark nhận dạng chi tiết; (3) SO-LoRA cô lập hình học các biểu diễn tác vụ trong một cơ sở trực giao cố định, bảo vệ động các chiều quan trọng, cho khả năng mở rộng mạnh trên chuỗi tác vụ dài và điều kiện dịch chuyển phân phối nghiêm trọng.

Kết luận và Hướng phát triển

Luận án nghiên cứu các phương pháp giảm thiểu quên nghiêm trọng trong học liên tục thông qua một tiến trình ba giai đoạn tương ứng với ba mức độ ràng buộc thực tế. Giai đoạn thứ nhất (Chương 2) đề xuất replay prototype bảo toàn quyền riêng tư, kết hợp tăng cường nhiễu Gaussian và mô-đun Căn chỉnh Ấn dựa trên Năng lượng (ELI) để hiệu chỉnh trôi dạt biểu diễn, đạt hiệu suất ngang bằng hoặc vượt trội so với các phương pháp lưu dữ liệu thực trên FewRel và TACRED. Giai đoạn thứ hai (Chương 3) loại bỏ hoàn toàn việc lưu trữ dữ liệu với **CL-plugin** (đạt Macro-F1 tiên tiến nhất trên benchmark 19 tác vụ) và **ZeroRFF** (mở rộng nguyên lý không-exemplar sang thị giác class-incremental), khẳng định tính khả thi của học liên tục không-exemplar trên cả NLP và thị giác máy tính. Giai đoạn thứ ba (Chương 4) giải quyết ràng buộc dung lượng tham số bằng bốn phương pháp: **multi-head OPE/APF/ELI**, **DAKTA**, và **SO-LoRA**, lần lượt đạt kết quả tốt nhất trên CCH5000, sáu benchmark so với ITA-LoRA, và ImageNet-A/CUB-200/ImageNet-R. Các giai đoạn có tính tương hỗ, thể hiện qua việc ELI được tái sử dụng ở Chương 4 và ZeroRFF chia sẻ nền tảng khái niệm với SO-LoRA.

Tuy nhiên, luận án còn một số hạn chế: các phương pháp giả định ranh giới tác vụ đã biết trước, chưa được đánh giá hệ thống trên chuỗi hàng trăm tác vụ, phạm vi thực nghiệm được giới hạn ở văn bản và ảnh, và chưa nhất quán về backbone (ResNet-18 so với ViT-B/16) ở Chương 4. Hướng nghiên cứu tương lai bao gồm xây dựng một meta-khung thích nghi theo ngữ cảnh để tự động lựa chọn giữa replay, chất lọc và mở rộng kiến trúc; mở rộng số học tác vụ dựa trên LoRA sang các mô hình ngôn ngữ lớn; học liên tục không ranh giới tác vụ thông qua phát hiện ranh giới điều khiển bởi OPE; và các kịch bản đa phương thức.

Danh sách công bố

- [P1] **Quynh-Trang Pham Thi**, Anh-Cuong Pham, Ngoc-Huyen Ngo, and Duc-Trong Le. "Memory-Based Method using Prototype Augmentation for Continual Relation Extraction." In 2022 RIVF International Conference on Computing and Communication Technologies (RIVF), pp. 1-6. IEEE, 2022. Scopus.
- [P2] Thanh-Hai Dang, **Quynh-Trang Pham Thi**, Duc-Trong Le, and Tri-Thanh Nguyen. Contrastive Learning for Boosting Knowledge Transfer in Task-Incremental Continual Learning of Aspect Sentiment Classification Tasks. VNU Journal of Science: Computer Science and Communication Engineering 40, no. 1 (2024).
- [P3] Duc-Hung Nguyen, Tri-Thanh Nguyen, Thanh-Hai Dang and **Quynh-Trang Pham Thi**. ZeroRFF: A Random Fourier Features and Analytic Learning Method for Generalized Class Incremental Learning. The 21st International Conference on Advanced Data Mining and Applications 2025. Scopus Rank B.
- [P4] **Quynh-Trang Pham Thi**, Duc-Hung Nguyen, Thanh Hai Dang, Duc-Trong Le, Tri-Thanh Nguyen, and Quang-Thuy Ha. "Towards Robust Continual Learning: A Multi-Head Approach with Online Prototype Equilibrium and Adaptive Prototypical Feedback." In Asian Conference on Intelligent Information and Database Systems, pp. 275-286. Singapore: Springer Nature Singapore, 2024. Scopus Rank B.
- [P5] **Quynh-Trang Pham Thi**, Duc-Hung Nguyen, Anh-Duc Nguyen-Tran, Hung-Dung Nguyen, Tri-Thanh Nguyen, Quang-Thuy Ha and Thanh Hai Dang. "Addressing Catastrophic Forgetting in Continual Learning using Multihead Approach and Latent Alignment via Energy-Based Model." Vietnam journal of computer science. (Accepted). WoS Q3.
- [P6] **Quynh-Trang Pham Thi**, Van-Truong Le, Thanh-Huyen Pham, Thanh-Nga Hoang Thi and Duc-Trong Le. DAKTA: Directional Kolmogorov-Arnold Classifier for Task Arithmetic in Continual Learning. SOICT 2025. Scopus.

[P7] **Quynh-Trang Pham Thi**, Quang-Dat Mac, Thanh-Ha Nguyen and Duc-Trong Le.
SO-LoRA: Sparse Orthogonal LoRA for Parameter-Efficient Continual Learning.
In 2026 IEEE International Conference on Multimedia and Expo (ICME 2026).
Scopus Rank A.

Danh sách có 7 công bố.