|  |  |
| --- | --- |
| ĐẠI HỌC PHENIKAA  **TRƯỜNG CNTT PHENIKAA** | **CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập - Tự do - Hạnh phúc** |

**BÁO CÁO TIẾN ĐỘ**

**Quá trình thực tập tốt nghiệp**

**(Thời gian thực tập: 09/06-20/07/2025)**

**Nội dung báo cáo: Tình hình, tiến độ thực tập tuần 1 (09/06 – 15/06/2025)**

**1. THÔNG TIN SINH VIÊN THỰC TẬP**

Họ tên sinh viên: Nguyễn Minh Quân

Mã SV: 21010617

Lớp: CNTT1 Ngành: Công nghệ thông tin

Hệ: Chính quy Khóa: 15

Tên cơ sở thực tập: Lab Nghiên cứu liên ngành về Khoa học dữ liệu và trí tuệ nhân tạo ( IDSAI Lab)

**2. NỘI DUNG CÔNG VIỆC**

**2.1. Nội dung công việc được giao**

|  |  |  |  |
| --- | --- | --- | --- |
| **Thời gian** | **Nội dung/Kế hoạch thực tập** | **Kết quả thực hiện** | **Ghi chú** |
| Tuần 1  (từ 09/06/2025 đến 15/06/2025) | Nghiên cứu lý thuyết về Incremental Learning (IL) | Hiểu rõ nền tảng IL và vai trò trong thị giác máy tính và phân đoạn ảnh y tế | So sánh IL với Transfer Learning |

**2.2. Vấn đề đang giải quyết**

**? Bài toán**

Trong nhiều ứng dụng của thị giác máy tính (computer vision), đặc biệt là **phân đoạn ảnh y tế,** dữ liệu thường đến theo từng giai đoạn, ví dụ như:

* Ảnh từ các bệnh viện mới (domain khác, …)
* Dữ liệu về cơ quan/ bệnh lý mới (class mới)
* Nguồn ảnh khác nhau (modality: CT, MRI, …)

Tuy nhiên, các mô hình học sâu truyền thống:

* Cần huấn luyện lại toàn bộ khi có dữ liệu mới
* Dễ bị quên kiến thức cũ (catastrophic forgetting)
* Khó triển khai trong thực tế với chi phí cao và thời gian nhiều.

**?? Vấn đề cốt lõi**

Làm sao để mô hình học dữ liệu mới mà không quên kiến thức cũ, trong khi dữ liệu có thể mất cân bằng giữa các lớp hoặc phase.

**2.3. Phương pháp giải quyết**

**? Incremental Learning** (IL) là hướng tiếp cận cho phép mô hình học liên tục từ dữ liệu mới theo từng phase (gọi là task), và vẫn dữ lại hiệu quả từ các lớp cũ đã học.

**?? Các thành phần được nghiên cứu trong tuần:**

1. **Thách thức của IL:**

* **Catastrophic Forgetting:** mô hình quên lớp cũ sau khi học lớp mới.
* **Dual Imbalance:**
  + Inter-phase imbalance: dữ liệu lớp cũ ít hơn lớp mới.
  + Intra-phase imbalance: lớp phổ biến > lớp hiếm.
* **Stability – Plasticity Dilemma:**
  + Stability: giữ kiến thức cũ.
  + Plasticity: tiếp thu cái mới.

1. **Tổng hợp phương pháp từ bài báo** “Training Like a Medical Resident”**:**

* Đề xuất cách cân bằng gradient để giảm thiên lệch về lớp mới.
* Áp dụng DAKD (Distributeion-Aware Knowledge Distillation): điều chỉnh trọng số loss dựa theo số lượng dữ liệu đã mất -> giúp giữ lại tốt hơn các lớp cũ.
* Hiệu quả được kiểm chứng trên CIFAR100, ImageNetSubsset, Food101.

1. **So sánh với Transfer Learning:**

|  |  |  |
| --- | --- | --- |
| **Tiêu chí** | **IL** | **Transfer Learning** |
| **Học mới** | Tuần tự, từng lớp | Một lần, toàn bộ |
| **Dữ liệu** | Theo giai đoạn | Cần đầy đủ ngay từ đầu |
| **Thách thức** | Forgetting, imbalance | Domain mismatch |
| **Tính thực tiễn** | |  | | --- | |  |  |  | | --- | | Cao với dữ liệu cập nhật | | Tốt cho pretrain+finetune |

**2.4. Kết quả đạt được**

* Đọc hiểu khái niệm IL và phân loại các loại học liên tục (class-incremental, task-incremental, domain-incremental)
* Phân tích sâu các thách thức: catastrophic forgetting, dual imbalance
* Tổng hợp cơ chế Gradient Reweighting & DAKD từ bài báo CVPR 2024
* So sánh IL với Transfer Learning cả về lý thuyết và ứng dụng

**3. NHẬN XÉT CỦA GVHD**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**

**……………………………………………………………………………………**