


THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
[Tại đây!](#)
- Link slides (dạng .pdf đặt trên Github của nhóm):
[Tại đây!](#)
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

<ul style="list-style-type: none">● Họ và Tên: Nguyễn Viết Đức● MSSV: 22520273 	<ul style="list-style-type: none">● Lớp: CS519.P11● Tự đánh giá (điểm tổng kết môn): 9/10● Số buổi vắng: 0● Link Github:● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:<ul style="list-style-type: none">○ Lên ý tưởng.○ Viết tóm tắt, giới thiệu, nội dung & phương pháp, kết quả mong đợi.○ Làm poster.
--	---

<ul style="list-style-type: none">● Họ và Tên: Đoàn Văn Hoàng● MSSV: 22520459	<ul style="list-style-type: none">● Lớp: CS519.P11● Tự đánh giá (điểm tổng kết môn): 8.5/10● Số buổi vắng: 1● Link Github: https://github.com/dangchoigame/CS519.P11● Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
--	---



- Viết giới thiệu, nội dung & phương pháp.
- Làm slide.
- Làm video.

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

NÂNG CAO HIỆU SUẤT KNOWLEDGE DISTILLATION THÔNG QUA ATTENTION TRANSFER CHO BÀI TOÁN PHÂN LOẠI.

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

ENHANCING KNOWLEDGE DISTILLATION PERFORMANCE THROUGH ATTENTION TRANSFER FOR CLASSIFICATION TASKS.

TÓM TẮT (Tối đa 400 từ)

Knowledge Distillation [1] là kỹ thuật chuyển giao kiến thức từ mô hình lớn và phức tạp (teacher) sang mô hình nhỏ hơn (student), giúp giảm thiểu chi phí tính toán và bộ nhớ mà vẫn duy trì hiệu suất. KD đặc biệt hữu ích trong các ứng dụng Internet of Things (IoT)[1], nơi các thiết bị nhỏ không có khả năng xử lý phân cứng mạnh mẽ.

Tuy nhiên, phương pháp KD truyền thống chỉ tập trung vào việc khớp các logits (đầu ra cuối cùng của mô hình), điều này không phản ánh đầy đủ các đặc trưng không gian và ngữ cảnh quan trọng mà mô hình teacher đã học được.

Để giải quyết vấn đề này, chúng tôi đề xuất tích hợp Attention Transfer (AT)[2] vào Knowledge Distillation, đặc biệt hướng đến các bài toán phân loại. Phương pháp này cho phép mô hình student học không chỉ các logits mà còn các attention maps từ teacher, giúp mô hình tập trung vào các vùng quan trọng của dữ liệu. Điều này sẽ cải thiện khả năng tổng quát hóa và độ chính xác, đặc biệt trên các lớp khó phân loại[2], đồng thời vẫn đảm bảo tính hiệu quả về tài nguyên khi triển khai trên các thiết bị IoT.

GIỚI THIỆU (Tối đa 1 trang A4)

Knowledge Distillation (KD) đã trở thành một kỹ thuật quan trọng trong việc nén mô hình và học chuyển giao. Ý tưởng chính của KD là giảm thiểu khoảng cách giữa hiệu suất của một mô hình lớn và một mô hình nhỏ hơn bằng cách chuyển giao kiến thức đã học từ mô hình teacher. Trong khi **phương pháp Knowledge Distillation (KD)** truyền thống chủ yếu tập trung vào việc khớp các logits giữa teacher và student, nhưng chúng lại bỏ qua những **đặc trưng không gian và ngữ cảnh** mà mô hình teacher chú ý đến trong không gian đặc trưng. Điều này dẫn đến việc thiếu sự chuyển giao kiến thức chi tiết, vì student không được dạy cách chú ý vào những vùng quan trọng trong dữ liệu. Vậy nên câu hỏi đặt ra chúng ta có

thể cải thiện Knowledge Distillation bằng cách hướng dẫn mô hình student chú ý vào các đặc trưng quan trọng, từ đó nâng cao hiệu suất và hiệu quả không?

Attention mechanisms, đặc biệt là **attention maps**, đã chứng minh khả năng nổi bật trong việc nhận diện và làm nổi bật các đặc trưng quan trọng từ dữ liệu đầu vào trong các mô hình học sâu[2][3]. Việc tận dụng attention maps trong KD có thể mang lại một cách tiếp cận có cấu trúc hơn để hướng dẫn mô hình student, giúp mô hình này chú ý đến những vùng quan trọng trong dữ liệu giống như teacher. Điều này sẽ giúp cải thiện việc đồng nhất giữa các đại diện bên trong của teacher và student.

Bài toán phân loại là một ứng dụng phổ biến và thách thức, đặc biệt khi xử lý các lớp khó phân loại hoặc dữ liệu không cân bằng[2]. Do đó, nghiên cứu này tập trung vào việc nâng cao hiệu suất của mô hình phân loại thông qua việc tích hợp Attention Transfer (AT) vào quá trình Knowledge Distillation.

MỤC TIÊU *(Viết trong vòng 3 mục tiêu)*

Khám phá vai trò của attention maps trong việc chuyển giao kiến thức: Phân tích cách các attention maps có thể được tạo ra từ mô hình teacher và cách chúng hướng dẫn quá trình học của mô hình student.

Tích hợp Attention Transfer (AT) vào Knowledge Distillation (KD): Phát triển một framework nơi mô hình student được huấn luyện để mô phỏng không chỉ các logits của teacher mà còn các attention maps mà teacher đã học được.

Cải thiện hiệu suất phân loại của mô hình student: Đánh giá tác động của AT-KD trong việc tăng độ chính xác và khả năng tổng quát trên các lớp khó phân loại.

NỘI DUNG VÀ PHƯƠNG PHÁP

Trong đề tài này, chúng tôi sẽ thực hiện nghiên cứu các nội dung chính sau:

1. Tìm hiểu tổng quan đề tài:

- Tìm hiểu tổng quan về Knowledge Distillation và Attention Transfer.
- Nghiên cứu các phương pháp nâng cao hiệu suất Knowledge Distillation thông qua Attention Transfer.

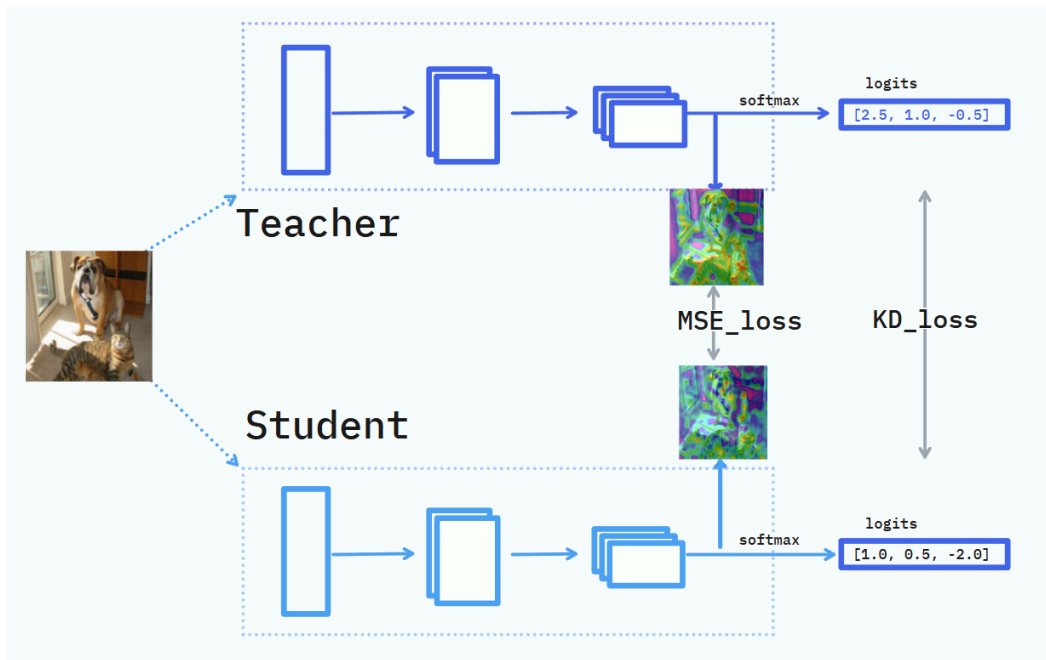
2. Tìm kiếm và xây dựng bộ dữ liệu huấn luyện và đánh giá

- **Bộ dữ liệu sử dụng:**
 - **CIFAR-10:** Bao gồm 60.000 hình ảnh thuộc 10 lớp với kích thước 32x32.
 - **ImageNet:** Bộ dữ liệu lớn hơn với hơn 1 triệu hình ảnh thuộc 1.000 lớp, phù hợp để kiểm tra khả năng tổng quát hóa của mô hình.

- **Tiền xử lý dữ liệu:**

- Tăng cường dữ liệu (Data Augmentation): Sử dụng các kỹ thuật như xoay, cắt ngẫu nhiên, thay đổi độ sáng để tăng sự đa dạng của dữ liệu.
- Chuẩn hóa dữ liệu: Sử dụng mean và standard deviation của từng kênh RGB để chuẩn hóa.

3. Xây dựng kiến trúc mô hình:



- **Teacher:** ResNet-50 (mô hình lớn, phức tạp) làm nguồn chuyển giao kiến thức, với các attention maps được trích xuất từ các lớp trung gian sử dụng cơ chế **self-attention** để làm nổi bật các vùng quan trọng trong không gian đặc trưng.
- **Student:** MobileNet (mô hình nhỏ hơn) được huấn luyện để mô phỏng:
 - **Logits:** Đầu ra cuối cùng của mô hình teacher, biểu diễn xác suất của các lớp.
 - **Attention maps:** Các vùng đặc trưng quan trọng mà teacher chú ý.
- **Hàm mất mát:**
 - **KL Divergence Loss:** Tính khoảng cách giữa logits của teacher (Z_t) và student (Z_s):
 - MSE Loss để giảm sự khác biệt giữa attention maps.

4. Huấn luyện và đánh giá:

- Huấn luyện student với sự kết hợp giữa logits matching và attention map matching.
- So sánh hiệu suất của phương pháp AT-KD với các phương pháp KD truyền thống và các kỹ thuật khác.

KẾT QUẢ MONG ĐỢI

- Độ chính xác (Accuracy) của student khi sử dụng AT-KD kì vọng cao hơn ít nhất 2-5% so với KD truyền thống trên cả CIFAR-10 và ImageNet.

- Mô hình student sử dụng AT-KD kì vọng đạt hiệu suất cao hơn, khả năng tổng quát hóa tốt trên cả các tập dữ liệu chưa được huấn luyện .
- Duy trì kích thước mô hình student nhỏ gọn, phù hợp với các thiết bị IoT, ổn định của mô hình khi triển khai trên dữ liệu thực tế.

TÀI LIỆU THAM KHẢO (*Định dạng DBLP*)

[1] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the Knowledge in a Neural Network. In *NIPS Workshop on Deep Learning and Representation Learning*.

[2] Sergey Zagoruyko, Nikos Komodakis (2016) Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer.

[3] Zhu Baozhou, Peter Hofstee1, Jinho Lee, Zaid Al-Ars (2021) An Attention Module for Convolutional Neural Networks.