

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

**Nguyễn Viết Đức - 22520273**

**Đoàn Văn Hoàng - 22520459**

# Tóm tắt

- Lớp: CS519.P11
- Link Github của nhóm:
- Link YouTube video:
- Nguyễn Viết Đức

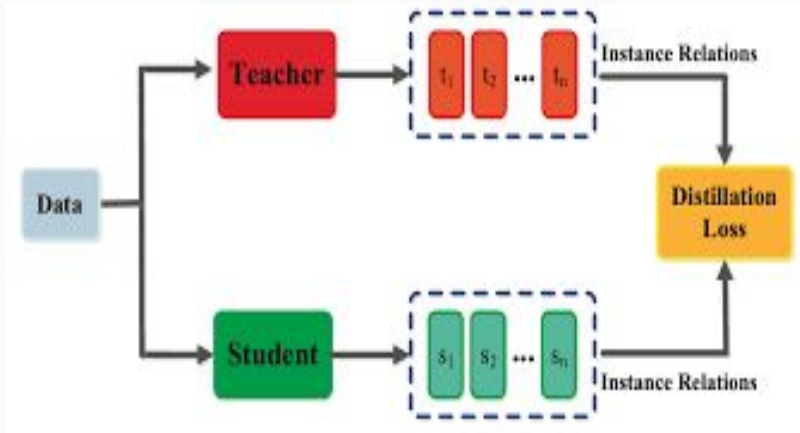


Đoàn Văn Hoàng

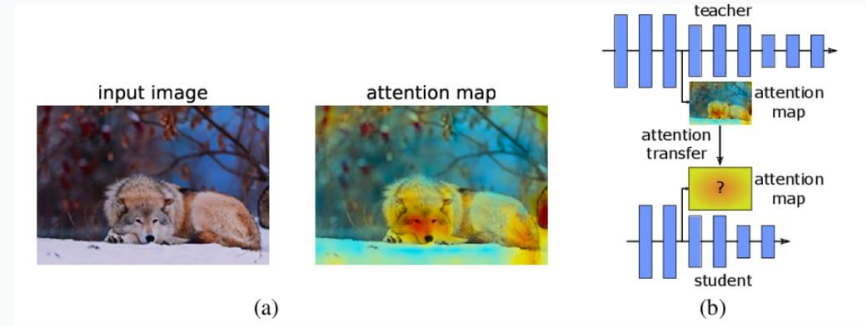


# Giới thiệu

## Knowledge Distillation:



## Attention Map:



# 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

## 1. Tìm hiểu tổng quan về 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:

- Bộ dữ liệu:
  - + CIFAR-10: gồm 60.000 ảnh thuộc 10 lớp, kích thước 32x32.
  - + ImageNet: hơn 1 triệu ảnh thuộc 1.000 lớp.
- Tiền xử lý dữ liệu: tăng cường (xoay, cắt ngẫu nhiên, đổi độ sáng), 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 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 không gian đặc trưng.

# Nội dung và Phương pháp

- Student: MobileNet (mô hình nhỏ hơn) huấn luyện để mô phỏng:
- + Logits: Đầu ra cuối của teacher, biểu diễn xác suất các lớp.
- + Attention maps: 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 của student ( $Z_s$ ); MSE Loss (giảm sự khác biệt giữa attention maps).

## 4. Huấn luyện, 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 dự kiến

1. Độ chính xác (Accuracy) của student khi sử dụng AT-KD kỳ vọng cao hơn ít nhất 2-5% so với KD truyền thống trên cả CIFAR-10 và ImageNet.
2. Mô hình student sử dụng AT-KD kỳ vọng đạt hiệu suất cao hơn, khả năng tổng quát hóa tốt trên cả các tập dữ liệu chưa được huấn luyện .
3. 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

- [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 Hofstee 1, Jinho Lee, Zaid Al-Ars (2021) An Attention Module for Convolutional Neural Networks.