

VIETNAM NATIONAL UNIVERSITY
HANOI UNIVERSITY OF SCIENCE



Báo cáo kết thúc tiểu luận
Đề tài: Khai phá ảnh viễn thám

Vũ Minh Hiếu - 22001582

Nguyễn Xuân Hồng - 22001589

Ngô Quang Huy - 22001592

Môn học: Thị giác máy tính

Giảng viên: TS. Đỗ Thanh Hà

HÀ NỘI - 2025

TÓM TẮT

Việc xác định chính xác cường độ bão là một thách thức quan trọng trong công tác dự báo và phòng chống thiên tai. Các phương pháp truyền thống như kỹ thuật Dvorak thủ công còn phụ thuộc nhiều vào kinh nghiệm chuyên gia và có thể thiếu khách quan. Tiểu luận này đề xuất ứng dụng Trí tuệ Nhân tạo (AI) để xây dựng một hệ thống tự động phân loại cấp độ bão dựa trên ảnh vệ tinh hồng ngoại (IR) từ vệ tinh MTSAT và thông tin lịch sử về chỉ số Cường độ Hiện tại (CINumber). Chúng tôi đã nghiên cứu và phát triển một mô hình học sâu kết hợp kiến trúc Vision Transformer (ViT) hoặc ResNet làm bộ trích xuất đặc trưng (backbone), cùng với một nhánh chính để phân loại bốn cấp độ bão (TD, TS, STS, TY) và một nhánh phụ (Auxiliary Task) để dự đoán chỉ số CINumber nhằm tăng cường khả năng học của mô hình. Kết quả thực nghiệm trên bộ dữ liệu ảnh bão cho thấy mô hình ViT kết hợp nhánh phụ (ViT+Aux) đạt hiệu năng tốt nhất với độ chính xác 77.88% và F1-score 76.19% trên tập kiểm thử. Việc tích hợp nhánh phụ đã cải thiện đáng kể hiệu suất so với các mô hình gốc. Nghiên cứu này khẳng định tiềm năng của các mô hình học sâu, đặc biệt khi kết hợp với kiến thức chuyên môn thông qua nhiệm vụ phụ, trong việc nâng cao độ chính xác và tự động hóa quy trình phân loại cường độ bão.

LỜI CẢM ƠN

Nhóm sinh viên chúng em xin bày tỏ lòng biết ơn sâu sắc tới **TS. Đỗ Thanh Hà** đã định hướng và đóng góp những ý kiến chuyên môn quý báu trong quá trình nhóm thực hiện nghiên cứu về đề tài ảnh viễn thám.

Chúng em cũng xin cảm ơn sự hợp tác, nỗ lực và chia sẻ của các thành viên trong nhóm đã cùng nhau vượt qua khó khăn để hoàn thành tiểu luận này.

Nhóm sinh viên thực hiện

Mục lục

Tóm tắt	1
Lời cảm ơn	2
Bảng ký hiệu và chữ viết tắt	5
Danh sách Hình và Bảng Chú thích	6
Danh sách Hình	6
Danh sách Bảng	6
1 MỞ ĐẦU	7
1.1 Bối cảnh và tính cấp thiết của đề tài	7
1.2 Mục tiêu nghiên cứu	8
1.3 Đối tượng và phạm vi nghiên cứu	9
2 Cơ sở lý thuyết	10
2.1 Kỹ thuật Dvorak và chỉ số CINumber	10
2.2 Các kiến trúc nền tảng (ResNet, ViT)	11
2.2.1 ResNet (Residual Network)	11
2.2.2 Vision Transformer (ViT)	12
2.3 Học đa nhiệm và Nhiệm vụ phụ	13
2.4 Tổng kết và khoảng trống nghiên cứu	14
3 PHƯƠNG PHÁP ĐỀ XUẤT	15
3.1 Nguồn và đặc tính dữ liệu nghiên cứu	15
3.2 Kiến trúc tổng quan của mô hình	16
3.3 Tiền xử lý dữ liệu và Tăng cường dữ liệu	18
3.3.1 Tiền xử lý Dữ liệu Best-track	18
3.3.2 Tiền xử lý Dữ liệu Ảnh	18
3.3.3 Tăng cường và Cân bằng Dữ liệu	19
3.4 Các mô hình so sánh và Cấu hình huấn luyện	20
4 KẾT QUẢ THỰC NGHIỆM VÀ PHÂN TÍCH	21
4.1 Số lượng ảnh và phân chia dữ liệu	21
4.2 So sánh hiệu năng các mô hình	21

4.3	Phân tích quá trình huấn luyện	24
4.4	Thảo luận kết quả	26
5	KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	28
5.1	Kết luận	28
5.2	Hướng phát triển	28
	Phân công công việc và Đánh giá đóng góp	30
	Tài liệu tham khảo	33

BẢNG KÝ HIỆU VÀ CHỮ VIẾT TẮT

AI	Artificial Intelligence (Trí tuệ Nhân tạo)
IR	Infrared (Hồng ngoại)
MTSAT	Multi-functional Transport Satellite (Vệ tinh vận tải đa chức năng)
CI	Current Intensity (Cường độ Hiện tại - thường dùng trong Dvorak)
CINumber	Current Intensity Number
TD	Tropical Depression (Áp thấp nhiệt đới)
TS	Tropical Storm (Bão nhiệt đới)
STS	Severe Tropical Storm (Bão nhiệt đới dữ dội)
TY	Typhoon (Bão lớn/Siêu bão)
CNN	Convolutional Neural Network (Mạng Neural Tích chập)
ViT	Vision Transformer
ResNet	Residual Network
Aux	Auxiliary (Phụ / Hỗ trợ)
ROI	Region of Interest (Vùng quan tâm)
MSE	Mean Squared Error (Sai số toàn phương trung bình)

DANH SÁCH HÌNH VÀ BẢNG CHÚ THÍCH

Danh sách Hình

1	Minh họa các mẫu hình thái mây trong kỹ thuật Dvorak (Nguồn: NOAA).	10
2	Cấu trúc tổng quan của mạng ResNet-18, thể hiện các khối dư và kết nối tắt.	12
3	Minh họa kiến trúc tổng quan của Vision Transformer (ViT). Ảnh đầu vào được chia thành các patch, sau đó được xử lý bởi các khối Transformer Encoder dựa trên cơ chế self-attention để tạo ra vector đặc trưng.	13
4	Tổng quan kiến trúc mô hình đề xuất. Mô hình bao gồm một backbone (ViT hoặc ResNet), một nhánh phụ (Auxiliary Task) để dự đoán chỉ số Dvorak (CI), và một nhánh chính (Final Head) để dự đoán cường độ bão cuối cùng (Grade) dựa trên các đặc trưng kết hợp.	17
5	Biểu đồ Accuracy trên tập Train và Validation cho 4 mô hình .	25
6	Biểu đồ Loss trên tập Train và Validation cho 4 mô hình	25

Danh sách Bảng

1	Các cấu hình mô hình thử nghiệm và luồng xử lý đặc trưng. . .	20
2	Số lượng ảnh theo từng lớp (Grade) trước và sau tăng cường dữ liệu	21
3	Số lượng ảnh theo từng lớp (Grade) trước và sau tăng cường dữ liệu	21

1 MỞ ĐẦU

1.1 Bối cảnh và tính cấp thiết của đề tài

Bão nhiệt đới (Tropical Cyclones - TC) là một trong những hiện tượng thời tiết cực đoan nguy hiểm nhất, thường xuyên gây ra những thiệt hại nghiêm trọng về người và của, đồng thời tác động tiêu cực đến đời sống kinh tế - xã hội của các khu vực chịu ảnh hưởng. Ở khu vực Tây Bắc Thái Bình Dương, các cấp độ bão thường gặp bao gồm Áp thấp nhiệt đới (Tropical Depression - TD), Bão nhiệt đới (Tropical Storm - TS), Bão nhiệt đới dữ dội (Severe Tropical Storm - STS) và Bão lớn/Siêu bão (Typhoon - TY). Việc xác định chính xác và kịp thời cường độ của bão có vai trò vô cùng quan trọng trong công tác dự báo, cảnh báo sớm và triển khai các biện pháp phòng chống, từ đó góp phần giảm thiểu tối đa những tổn thất có thể xảy ra.

Trong nhiều thập kỷ qua, kỹ thuật Dvorak [Dvorak1984] đã trở thành một công cụ chủ yếu để ước tính cường độ bão dựa trên việc phân tích thủ công các đặc điểm hình thái của mây từ ảnh vệ tinh. Mặc dù đã chứng minh được tính hữu dụng, phương pháp này bộc lộ một số hạn chế đáng kể như phụ thuộc nhiều vào kinh nghiệm và đánh giá chủ quan của chuyên gia, tốn thời gian thực hiện và khó đảm bảo tính nhất quán, khách quan giữa các trung tâm dự báo khác nhau. Đặc biệt, trong bối cảnh các cơn bão có xu hướng ngày càng diễn biến phức tạp và khó lường, việc cập nhật liên tục và nhanh chóng thông tin về cường độ bão là một yêu cầu cấp thiết.

Sự phát triển mạnh mẽ của Trí tuệ Nhân tạo (AI), đặc biệt là các kỹ thuật Học sâu (Deep Learning), đã mở ra những hướng tiếp cận mới đầy tiềm năng cho nhiều lĩnh vực khoa học, bao gồm cả khí tượng thủy văn. Học sâu cho phép các mô hình máy tính tự động học và trích xuất những đặc trưng phức tạp, tinh vi từ lượng lớn dữ liệu đầu vào, chẳng hạn như ảnh vệ tinh. Điều này mang lại khả năng tự động hóa quy trình phân tích, giải quyết một phần những thách thức liên quan đến ranh giới phân loại mờ và sự biến thiên đa dạng về hình thái giữa các cấp độ bão. Do đó, việc nghiên cứu và ứng dụng AI để phát triển các hệ thống hỗ trợ phân loại cường độ bão tự động, nhanh chóng và chính xác hơn là một nhiệm vụ khoa học có tính cấp thiết cao, hứa hẹn mang lại những cải tiến đáng kể so với các phương pháp truyền thống.

1.2 Mục tiêu nghiên cứu

Dựa trên những phân tích về bối cảnh và tính cấp thiết đã trình bày, nghiên cứu này đề ra các mục tiêu chính như sau:

1. **Nghiên cứu và đề xuất mô hình học sâu hiệu quả:** Xây dựng một mô hình học sâu tiên tiến, có khả năng tự động phân loại cấp độ bão. Mô hình sẽ dựa trên các kiến trúc mạng mạnh mẽ như Vision Transformer (ViT) hoặc Mạng Residual (ResNet) làm bộ trích xuất đặc trưng từ ảnh vệ tinh hồng ngoại (IR).
2. **Tích hợp kỹ thuật học với nhiệm vụ phụ (Auxiliary Task):** Nghiên cứu việc kết hợp một nhiệm vụ phụ là dự đoán chỉ số Cường độ Hiện tại (Current Intensity Number - CINumber) từ ảnh. Mục tiêu của việc này là nhằm tăng cường khả năng học các đặc trưng liên quan đến cường độ bão của mô hình, từ đó cải thiện hiệu năng của nhiệm vụ chính là phân loại cấp độ bão.
3. **Xây dựng và đánh giá hệ thống phân loại tự động 04 cấp độ bão:** Áp thấp nhiệt đới (TD), Bão nhiệt đới (TS), Bão nhiệt đới dữ dội (STS) và Bão lớn/Siêu bão (TY), dựa trên dữ liệu ảnh IR từ vệ tinh MTSAT.
4. **Phân tích và so sánh hiệu năng:** Đánh giá một cách có hệ thống hiệu quả của mô hình đề xuất, bao gồm việc so sánh giữa các kiến trúc backbone khác nhau (ViT so với ResNet) và đánh giá cụ thể vai trò, mức độ đóng góp của việc tích hợp nhiệm vụ phụ dự đoán CINumber vào kết quả cuối cùng.

Từ các mục tiêu trên, câu hỏi nghiên cứu chính mà tiểu luận này hướng đến giải quyết bao gồm:

- Liệu việc tích hợp một nhiệm vụ phụ, cụ thể là dự đoán CINumber, có thực sự cải thiện hiệu năng phân loại cấp độ bão của mô hình học sâu so với việc chỉ sử dụng nhiệm vụ chính không?
- Trong bối cảnh bài toán và bộ dữ liệu cụ thể được sử dụng, kiến trúc mạng nào (Vision Transformer hay ResNet) cho thấy sự phù hợp và mang lại kết quả tốt hơn?

1.3 Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của tiểu luận này bao gồm:

- **Dữ liệu đầu vào chính:** Các ảnh vệ tinh hồng ngoại (IR) đa kênh thu được từ vệ tinh địa tĩnh MTSAT (Multi-functional Transport Satellite). Các kênh cụ thể được sử dụng và đặc tính của chúng sẽ được mô tả chi tiết trong Chương 3.
- **Thông tin phụ trợ:** Dữ liệu lịch sử về chỉ số Dvorak, cụ thể là chỉ số CINumber, tương ứng với các ảnh bão được nghiên cứu. Thông tin này chủ yếu được sử dụng để thiết kế và đánh giá hiệu quả của nhiệm vụ phụ.
- **Các cấp độ bão mục tiêu:** Bốn cấp độ bão theo thang đo quốc tế thường dùng ở khu vực Tây Bắc Thái Bình Dương: TD, TS, STS, và TY.
- **Các mô hình và kỹ thuật học máy:** Tập trung vào các kiến trúc mạng học sâu hiện đại như Mạng CNN Residual (ví dụ: ResNet-18) và Vision Transformer (ViT), cùng với kỹ thuật học đa nhiệm thông qua việc tích hợp nhiệm vụ phụ (Auxiliary Task Learning).

Phạm vi nghiên cứu được giới hạn như sau:

- Nghiên cứu tập trung chủ yếu vào bài toán **phân loại cường độ bão** dựa trên các đặc trưng hình ảnh được trích xuất từ một khung ảnh vệ tinh hồng ngoại đơn lẻ (single-frame image analysis).
- Chỉ số CINumber được khai thác như một nguồn thông tin bổ sung, định hướng quá trình học của mô hình thông qua một nhánh nhiệm vụ phụ, thay vì là một đầu vào trực tiếp cho mô hình phân loại cấp độ bão chính (trừ khi có các thiết kế thử nghiệm cụ thể được nêu rõ).
- Nghiên cứu này *không* đi sâu vào các bài toán phức tạp hơn như: dự báo quỹ đạo di chuyển của bão, phân tích và dự báo cường độ bão dựa trên chuỗi thời gian của nhiều khung ảnh (multi-frame temporal analysis) hay tích hợp đa dạng các nguồn dữ liệu khí tượng khác (ví dụ: tốc độ gió, áp suất bề mặt biển từ các nguồn khác). Những khía cạnh này có

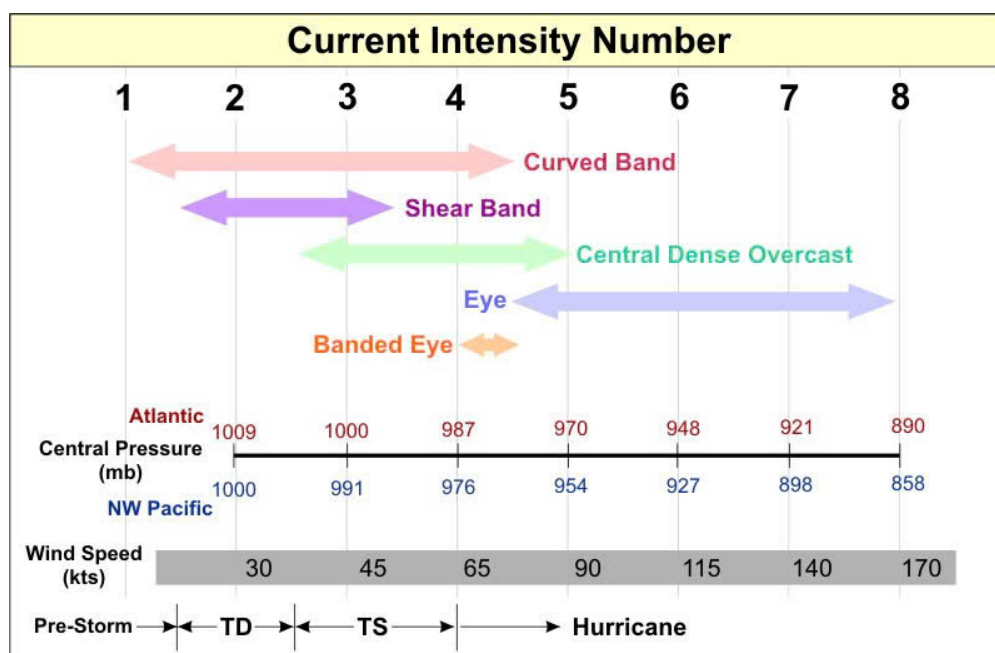
thể được xem xét như những hướng phát triển tiềm năng trong tương lai.

Cuối cùng là danh mục Tài liệu tham khảo và Phụ lục (nếu có).

2 Cơ sở lý thuyết

2.1 Kỹ thuật Dvorak và chỉ số CINumber

Kỹ thuật Dvorak, được phát triển bởi Vernon Dvorak vào đầu những năm 1970, là một phương pháp chủ quan được sử dụng rộng rãi để ước tính cường độ của xoáy thuận nhiệt đới (TC) dựa trên hình ảnh vệ tinh khả kiến và hồng ngoại. Kỹ thuật này dựa trên việc nhận dạng các mẫu hình thái mây đặc trưng liên quan đến các giai đoạn phát triển khác nhau của bão.



Hình 1: Minh họa các mẫu hình thái mây trong kỹ thuật Dvorak (Nguồn: NOAA).

Trọng tâm của kỹ thuật Dvorak là chỉ số **Current Intensity Number (CINumber)** hay còn gọi là số T (T-number). Các nhà dự báo phân tích các đặc điểm của hệ thống mây, chẳng hạn như độ cong của các dải mây, kích thước và hình dạng của mắt bão (nếu có), và nhiệt độ đỉnh mây để xác định một giá trị CINumber. Giá trị này thường dao động từ 1.0 đến 8.0, với các giá trị cao hơn biểu thị cơn bão mạnh hơn.

Sau khi CINumber được xác định, nó sẽ được quy đổi thành các ước tính về vận tốc gió duy trì tối đa (Maximum Sustained Wind - MSW) và áp suất

tối thiểu tại tâm bão (Minimum Sea Level Pressure - MSLP) thông qua các bảng chuyển đổi hoặc công thức đã được thiết lập. Mặc dù là một phương pháp chủ quan và đòi hỏi kinh nghiệm của người phân tích, kỹ thuật Dvorak vẫn là một công cụ vô giá, đặc biệt đối với các khu vực không có máy bay trinh sát bão.

2.2 Các kiến trúc nền tảng (ResNet, ViT)

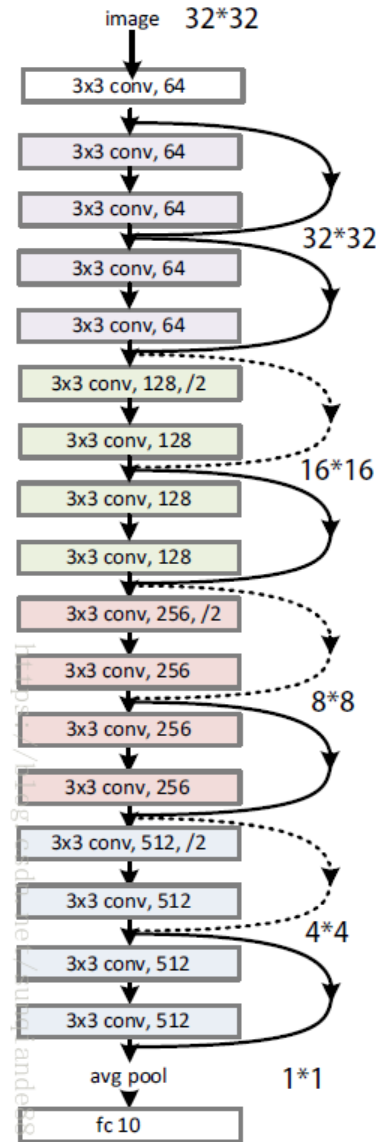
Các mô hình học sâu, đặc biệt là mạng nơ-ron tích chập (Convolutional Neural Networks - CNNs) và Transformer, đã cho thấy tiềm năng to lớn trong việc tự động hóa và cải thiện độ chính xác của việc ước tính cường độ bão từ hình ảnh vệ tinh. Trong nghiên cứu này, hai kiến trúc nổi bật được xem xét làm thành phần "backbone" để trích xuất đặc trưng từ ảnh bão là ResNet và Vision Transformer (ViT).

2.2.1 ResNet (Residual Network)

ResNet, được giới thiệu bởi He và cộng sự (2016) [He2016], là một kiến trúc CNN mang tính cách mạng, giải quyết hiệu quả vấn đề suy giảm độ dốc (vanishing gradient) trong quá trình huấn luyện các mạng rất sâu. Ý tưởng cốt lõi của ResNet là sử dụng các **kết nối tắt (skip connections)** hoặc **khối dư (residual blocks)**.

Thay vì cố gắng học một ánh xạ trực tiếp $H(\mathbf{x})$, khối dư học một hàm dư $\mathcal{F}(\mathbf{x}) = H(\mathbf{x}) - \mathbf{x}$. Khi đó, ánh xạ gốc được tính lại thành $H(\mathbf{x}) = \mathcal{F}(\mathbf{x}) + \mathbf{x}$. Việc học hàm dư thường dễ dàng hơn, đặc biệt khi hàm đồng nhất (identity mapping) là tối ưu. Công thức này cho phép tín hiệu gradient được lan truyền trực tiếp qua kết nối tắt, giúp huấn luyện các mạng sâu hơn mà không làm giảm hiệu suất.

Trong nghiên cứu này, **ResNet-18** (một phiên bản ResNet với 18 lớp) được sử dụng làm một trong các bộ trích xuất đặc trưng ảnh. Trọng số của ResNet-18 có thể được **khởi tạo trước (pretrained)** trên một tập dữ liệu lớn như ImageNet, sau đó được **tinh chỉnh (fine-tuned)** trên dữ liệu ảnh bão cụ thể. Điều này giúp mô hình tận dụng các đặc trưng tổng quát đã học từ ImageNet và thích ứng chúng với nhiệm vụ chuyên biệt là phân tích ảnh bão. ResNet nổi trội trong việc trích xuất các đặc trưng không gian cục bộ



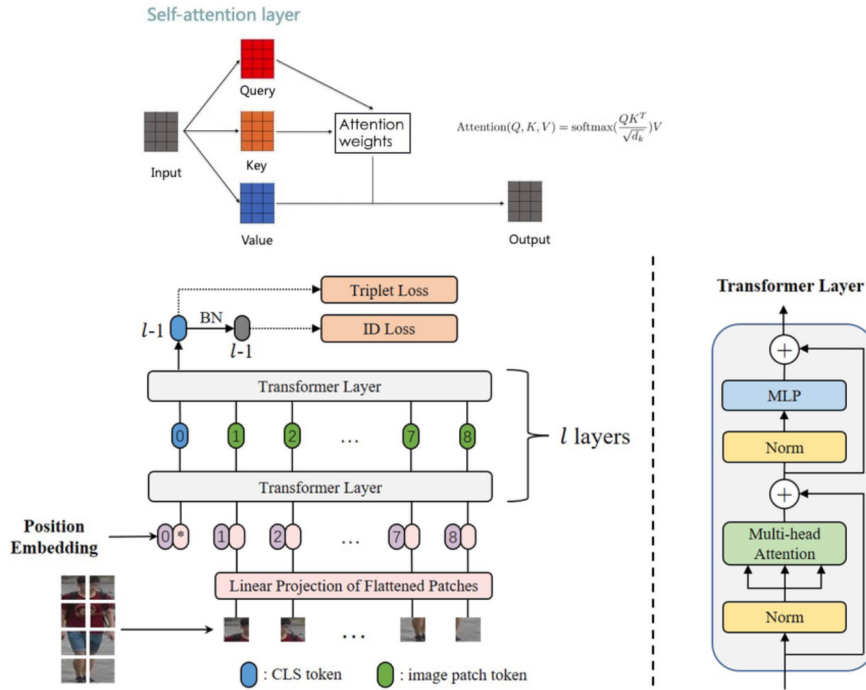
Hình 2: Cấu trúc tổng quan của mạng ResNet-18, thể hiện các khối dư và kết nối tắt.

và phân cấp từ ảnh.

2.2.2 Vision Transformer (ViT)

Vision Transformer (ViT), được giới thiệu bởi Dosovitskiy và cộng sự (2020) [Dosovitskiy2020], đã đánh dấu một bước đột phá bằng cách áp dụng thành công kiến trúc Transformer, vốn thống trị trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP), vào các bài toán thị giác máy tính.

Thay vì xử lý ảnh bằng các tích chập cục bộ như CNN, ViT chia ảnh đầu vào thành một chuỗi các "patch"(mảnh vá) không chồng chéo có kích thước cố định. Mỗi patch được làm phẳng và chiếu tuyến tính thành một vector embedding. Một embedding vị trí (positional embedding) được thêm



Hình 3: Minh họa kiến trúc tổng quan của Vision Transformer (ViT). Ảnh đầu vào được chia thành các patch, sau đó được xử lý bởi các khối Transformer Encoder dựa trên cơ chế self-attention để tạo ra vector đặc trưng.

vào để giữ lại thông tin không gian. Chuỗi các vector này, cùng với một "class token" có thể học được, sau đó được đưa vào một chuỗi các khối Transformer Encoder tiêu chuẩn. Mỗi khối Transformer Encoder bao gồm một lớp Multi-Head Self-Attention (MHSA) và một lớp Multi-Layer Perceptron (MLP). Các kết nối dư và chuẩn hóa lớp (Layer Normalization) cũng được sử dụng.

Cơ chế self-attention cho phép ViT mô hình hóa các mối quan hệ **toàn cục** giữa các patch khác nhau trong ảnh, giúp nó nắm bắt các phụ thuộc ở phạm vi xa hiệu quả hơn so với các receptive field cục bộ của CNN truyền thống. Điều này có thể đặc biệt hữu ích cho việc phân tích các cấu trúc mây quy mô lớn trong ảnh bão.

2.3 Học đa nhiệm và Nhiệm vụ phụ

Học đa nhiệm (Multi-Task Learning - MTL) là một phương pháp trong học máy mà một mô hình duy nhất được huấn luyện để thực hiện đồng thời nhiều nhiệm vụ chính liên quan. Ý tưởng cơ bản là các nhiệm vụ này có thể chia sẻ một biểu diễn chung, giúp cải thiện hiệu suất tổng thể của mô hình và khả năng khái quát hóa. Trong bối cảnh dự đoán cường độ bão, các **nhiệm vụ chính** thường là ước tính trực tiếp Vận tốc gió duy trì Tối đa (MSW) và

Áp suất Tối thiểu tại Tâm bão (MSLP). Học đa nhiệm có thể được áp dụng bằng cách huấn luyện một mô hình để đồng thời dự đoán các đại lượng này, tận dụng mối tương quan giữa chúng để học các đặc trưng mạnh mẽ hơn.

Nhiệm vụ phụ (Auxiliary Tasks) là các nhiệm vụ bổ sung được đưa vào quá trình huấn luyện, không nhất thiết là mục tiêu cuối cùng nhưng giúp mô hình chính học được các biểu diễn tốt hơn để phục vụ các nhiệm vụ chính. Trong nghiên cứu này, một **nhiệm vụ phụ then chốt** được xem xét là **dự đoán chỉ số CINumber** (Current Intensity Number) của kỹ thuật Dvorak.

- **Mục đích:** Việc huấn luyện mô hình dự đoán CINumber cho phép tận dụng một nguồn dữ liệu lịch sử quý giá là "best track". Dữ liệu "best track" thường chứa các bản ghi về CINumber (hoặc T-number tương đương) được các chuyên gia phân tích và ghi nhận qua nhiều thập kỷ. Bằng cách học dự đoán CINumber, mô hình được khuyến khích học các đặc trưng hình thái mây mà các chuyên gia đã dựa vào, qua đó gián tiếp tích hợp kiến thức chuyên môn này vào quá trình huấn luyện cho các nhiệm vụ chính (dự đoán MSW, MSLP).

Bằng cách giải quyết các nhiệm vụ phụ này, đặc biệt là dự đoán CINumber từ dữ liệu best track, mạng buộc phải học các đặc trưng chi tiết, có ý nghĩa lịch sử và phù hợp với phân tích chuyên môn từ dữ liệu đầu vào, từ đó có thể cải thiện đáng kể hiệu suất của các nhiệm vụ chính là ước tính cường độ bão (MSW, MSLP).

Sự kết hợp giữa backbone mạnh mẽ (như ResNet hoặc ViT) với chiến lược học đa nhiệm cho các mục tiêu cường độ chính, cùng với việc sử dụng nhiệm vụ phụ dự đoán CINumber để khai thác dữ liệu lịch sử, có thể dẫn đến các mô hình dự đoán cường độ bão chính xác, đáng tin cậy và có cơ sở diễn giải tốt hơn.

2.4 Tổng kết và khoảng trống nghiên cứu

Kỹ thuật Dvorak đã cung cấp một phương pháp nền tảng và có giá trị lịch sử trong việc ước tính cường độ bão, để lại một di sản dữ liệu "best track" phong phú chứa thông tin CINumber. Sự phát triển của học sâu đã mở ra những hướng tiếp cận mới, tự động hóa và có khả năng vượt qua những hạn chế của phương pháp thủ công. Việc sử dụng các kiến trúc tiên tiến như ResNet và

Vision Transformer làm backbone cho phép trích xuất các đặc trưng phong phú từ ảnh vệ tinh. Đặc biệt, việc áp dụng **nhiệm vụ phụ là dự đoán chỉ số CINumber** nổi lên như một chiến lược quan trọng để **tận dụng hiệu quả kho tàng dữ liệu "best track" trong quá khứ**. Bằng cách này, mô hình không chỉ học từ các nhãn cường độ hiện tại (MSW, MSLP) mà còn được "dạy" để nhận biết các mẫu hình thái mây tương ứng với các đánh giá CINumber lịch sử, giúp tăng cường sự hiểu biết của mô hình về các đặc điểm cốt lõi của bão. Kết hợp với các khái niệm học đa nhiệm cho các đầu ra cường độ chính, cách tiếp cận này hứa hẹn nâng cao hơn nữa hiệu suất mô hình bằng cách khai thác các thông tin liên quan và biểu diễn tiềm ẩn trong dữ liệu một cách toàn diện.

3 PHƯƠNG PHÁP ĐỀ XUẤT

3.1 Nguồn và đặc tính dữ liệu nghiên cứu

Nghiên cứu này tập trung vào việc phân tích hình ảnh vệ tinh để ước tính cường độ bão. Dữ liệu đầu vào chính bao gồm:

- **Ảnh hồng ngoại (IR) đa kênh từ vệ tinh địa tĩnh MTSAT (Multi-functional Transport Satellite):** Vệ tinh MTSAT cung cấp hình ảnh hồng ngoại thường xuyên, cho phép theo dõi liên tục sự phát triển của các hệ thống mây bão. Các kênh hồng ngoại khác nhau có thể cung cấp thông tin về nhiệt độ đỉnh mây, cấu trúc mây, và độ ẩm, là những yếu tố quan trọng trong việc đánh giá cường độ bão.
- **Thông tin Dvorak trong quá khứ (dữ liệu "best track"):** Dữ liệu lịch sử về chỉ số cường độ hiện tại (Current Intensity Number - CINumber hoặc T-number) được xác định bởi các chuyên gia theo kỹ thuật Dvorak. Dữ liệu này đóng vai trò quan trọng trong việc huấn luyện nhiệm vụ phụ, giúp mô hình học cách nhận biết các mẫu hình thái mây tương ứng với các đánh giá cường độ đã được kiểm chứng.

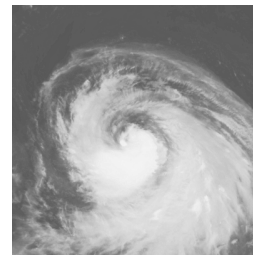
Một đặc điểm quan trọng của dữ liệu bão là **tính mất cân bằng (data imbalance)** đáng kể. Sự mất cân bằng này có thể khiến mô hình học máy bị thiên lệch về phía các lớp đa số và gặp khó khăn trong việc nhận diện chính

xác các trường hợp bão mạnh hiếm gặp nhưng lại có tác động nghiêm trọng nhất. Do đó, việc áp dụng các kỹ thuật xử lý mất cân bằng dữ liệu (như đã đề cập trong mục Tiền xử lý và Tăng cường dữ liệu) là rất cần thiết.

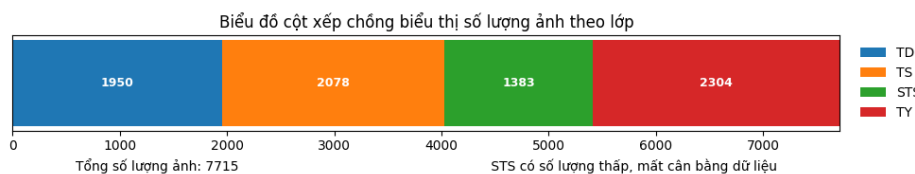
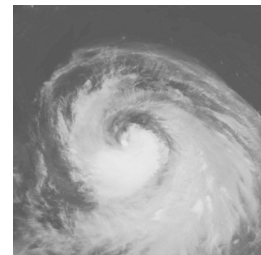
Thêm vào đó, **bản chất phức tạp và đa dạng của hình ảnh bão** cũng đặt ra những thách thức không nhỏ. Các cơn bão có thể khác nhau đáng kể về kích thước, hình dạng, cấu trúc mắt bão (rõ ràng, bị che phủ, hoặc không có), sự đối xứng của các dải mây, và sự tương tác với các yếu tố môi trường xung quanh (ví dụ: đất liền, các hệ thống thời tiết khác). Ngoài ra, các yếu tố như góc nhìn của vệ tinh, điều kiện chiếu sáng (đối với kênh khả kiến, mặc dù nghiên cứu này tập trung vào IR), và nhiễu trong ảnh cũng góp phần làm tăng độ khó của bài toán. Mô hình cần phải đủ mạnh mẽ để học được các đặc trưng bất biến và khái quát hóa tốt trước sự đa dạng này để đưa ra các ước tính cường độ đáng tin cậy.

First images from MTSAT-2:	
Channel	Wavelength (micrometer)
Visible Channel	0.55 - 0.90
Infrared Channel 1	10.3 - 11.3
Infrared Channel 2	11.5 - 12.5
Infrared Channel 3	6.5 - 7.0
Infrared Channel 4	3.5 - 4.0

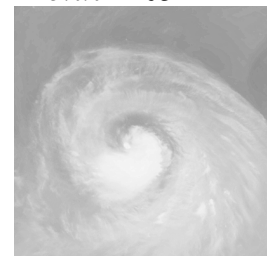
Kênh IR1



Kênh IR2

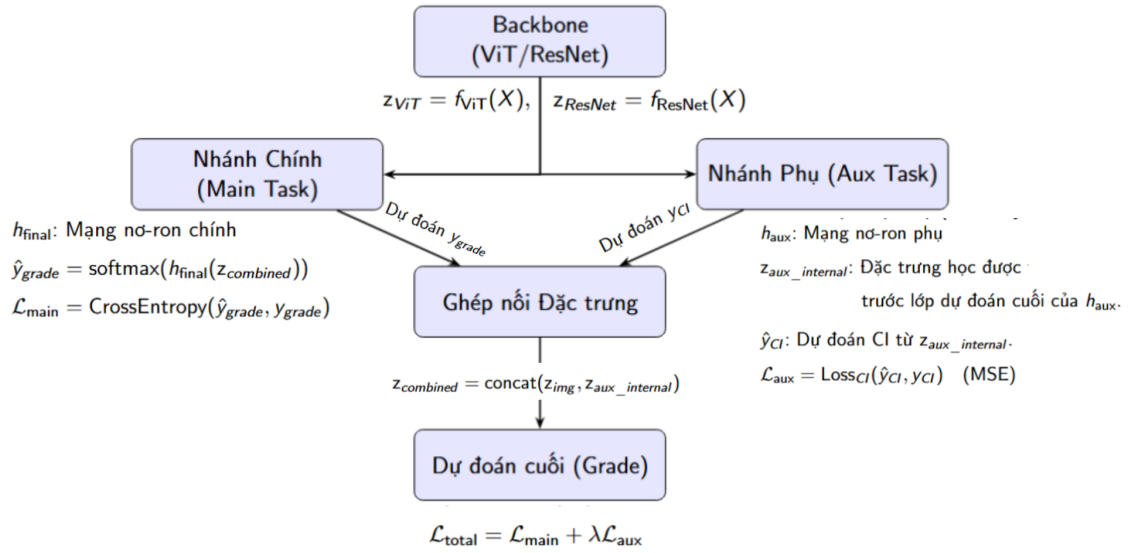


Kênh IR3



3.2 Kiến trúc tổng quan của mô hình

Mô hình đề xuất được thiết kế để tận dụng sức mạnh của các kiến trúc học sâu hiện đại và định hướng quá trình học thông qua một nhiệm vụ phụ. Luồng xử lý chính của mô hình được mô tả trong Hình 4.



Hình 4: Tổng quan kiến trúc mô hình đề xuất. Mô hình bao gồm một backbone (ViT hoặc ResNet), một nhánh phụ (Auxiliary Task) để dự đoán chỉ số Dvorak (CI), và một nhánh chính (Final Head) để dự đoán cường độ bão cuối cùng (Grade) dựa trên các đặc trưng kết hợp.

Các thành phần chính của kiến trúc bao gồm:

1. **Backbone (Thành phần trích xuất đặc trưng - f_{bb}):** Sử dụng một trong hai kiến trúc nền tảng: **ResNet-18** hoặc **Vision Transformer (ViT)**. Cả hai kiến trúc đều được khởi tạo với trọng số đã được **huấn luyện trước (pretrained)** trên tập dữ liệu ImageNet, sau đó được **tinh chỉnh (fine-tuned)** trên dữ liệu ảnh bão cụ thể của nghiên cứu. Backbone này nhận ảnh đầu vào đã qua tiền xử lý và tạo ra một vector đặc trưng z .
2. **Nhánh Phụ (Auxiliary Task Head - h_{aux}):** Nhánh này được thiết kế để thực hiện **nhiệm vụ phụ** là **dự đoán chỉ số Dvorak (CI)** từ vector đặc trưng F do backbone tạo ra. Mục tiêu của nhánh này là giúp backbone học các đặc trưng hình thái mây quan trọng mà các chuyên gia sử dụng trong kỹ thuật Dvorak, bằng cách tận dụng dữ liệu "best track" lịch sử. Hàm mất mát cho nhánh này được ký hiệu là \mathcal{L}_{aux} . Đầu ra của nhánh phụ (có thể là các đặc trưng nội bộ trước lớp dự đoán CI cuối cùng) được ký hiệu là z_{aux} .
3. **Ghép nối Đặc trưng (Feature Combination):** Các đặc trưng từ backbone (F) và các đặc trưng từ nhánh phụ (z_{aux}) được **ghép nối**

(**concatenate**) lại với nhau để tạo thành một vector đặc trưng tổng hợp $\mathbf{z}_{\text{final}} = \text{concat}(F, \mathbf{z}_{\text{aux}})$. Trong trường hợp mô hình không sử dụng nhánh phụ, thì $\mathbf{z}_{\text{final}} = F$.

4. **Nhánh Chính (Final Head - h_{final}):** Nhánh này nhận vector đặc trưng tổng hợp $\mathbf{z}_{\text{final}}$ và thực hiện nhiệm vụ chính là **dự đoán cường độ bão cuối cùng (Grade)**. "Grade" ở đây có thể là một lớp phân loại cường độ bão (ví dụ: theo thang Saffir-Simpson) hoặc một giá trị hồi quy (ví dụ: vận tốc gió). Hàm mất mát cho nhánh này được ký hiệu là $\mathcal{L}_{\text{main}}$.

Tổng hàm mất mát của mô hình khi có nhánh phụ sẽ là sự kết hợp của $\mathcal{L}_{\text{main}}$ và \mathcal{L}_{aux} , thường là một tổng có trọng số: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{main}} + \lambda \mathcal{L}_{\text{aux}}$, với λ là một siêu tham số điều chỉnh tầm quan trọng của nhiệm vụ phụ.

3.3 Tiền xử lý dữ liệu và Tăng cường dữ liệu

3.3.1 Tiền xử lý Dữ liệu Best-track

1. **Loại bỏ những hàng bị lỗi** Các hàng có giá trị Null cần được xử lý.
2. **Chuẩn hóa lại cột kinh độ, vĩ độ** Các giá trị tại cột kinh độ, Vĩ độ đưa về các đoạn hợp lý lần lượt là 0 đến 360 và -90 đến 90.

3.3.2 Tiền xử lý Dữ liệu Ảnh

Để chuẩn bị dữ liệu ảnh cho mô hình, các bước tiền xử lý sau được thực hiện:

1. **Tải ảnh IR từ vệ tinh MTSAT:** Thu thập ảnh gốc từ <ftp://hmwr829gr.cr.chiba-u.ac.jp/>.
2. **Chuyển ảnh từ định dạng NC sang PNG:** Thực hiện tìm các ngưỡng nhiệt để chuyển ảnh từ định dạng NC sang PNG giúp giảm thiểu khối lượng tính toán.
3. **Cắt vùng quan tâm (Region of Interest - ROI):** Một cửa sổ có kích thước 303×303 pixel được cắt ra quanh tâm bão đã được xác định trước (ví dụ, từ dữ liệu best track hoặc các phương pháp phát hiện tâm bão khác). Việc này giúp tập trung vào khu vực chứa các thông tin quan trọng nhất của cơn bão.

4. **Thay đổi kích thước ảnh (Resize):** Ảnh ROI sau đó được thay đổi kích thước thành 224×224 pixel. Kích thước này thường tương thích với đầu vào tiêu chuẩn của các mô hình pretrained như ResNet-18 và ViT. Quá trình resize sử dụng phương pháp **nội suy song tuyến tính (bilinear interpolation)**.
5. **Chuẩn hóa (Normalization):** Giá trị pixel của ảnh được chuẩn hóa. Cụ thể, giá trị pixel được điều chỉnh bằng cách trừ đi giá trị trung bình (mean) và chia cho độ lệch chuẩn (std) của tập dữ liệu ImageNet. Bước này giúp quá trình huấn luyện hội tụ nhanh hơn và ổn định hơn, đặc biệt khi sử dụng trọng số pretrained từ ImageNet.

3.3.3 Tăng cường và Cân bằng Dữ liệu

Để cải thiện khả năng khái quát hóa của mô hình và giải quyết vấn đề mất cân bằng lớp thường gặp trong dữ liệu bão (ví dụ: số lượng bão yếu nhiều hơn bão rất mạnh), các kỹ thuật tăng cường dữ liệu được áp dụng cho tập huấn luyện (train set):

- **Mục tiêu:** Giảm thiểu hiện tượng học vẹt (overfitting) và tăng cường sự đa dạng của dữ liệu huấn luyện, đặc biệt là đối với các lớp cường độ ít mẫu.
- **Các kỹ thuật được sử dụng:**
 - **Xoay ngẫu nhiên (Random Rotation):** Xoay ảnh một góc ngẫu nhiên trong khoảng $\pm 10^\circ$.
 - **Lật ảnh (Random Flipping):** Lật ảnh ngẫu nhiên theo chiều ngang hoặc chiều dọc.
 - **Dịch chuyển ảnh (Random Shift):** Dịch chuyển ảnh ngẫu nhiên theo cả chiều ngang và chiều dọc một vài pixel (ví dụ: ± 5 pixels).

Các phép biến đổi này được áp dụng ngẫu nhiên trong quá trình huấn luyện, giúp mô hình học được các đặc trưng bất biến với những thay đổi nhỏ về góc nhìn, vị trí.

3.4 Các mô hình so sánh và Cấu hình huấn luyện

Để đánh giá hiệu quả của các thành phần khác nhau trong kiến trúc đề xuất, bốn trường hợp (TH) mô hình được thiết lập và thử nghiệm, như trình bày trong Bảng 1. Các trường hợp này cho phép so sánh hiệu suất giữa việc sử dụng ResNet-18 và ViT làm backbone, cũng như đánh giá lợi ích của việc thêm nhánh phụ (Auxiliary Task) dự đoán chỉ số CI.

Bảng 1: Các cấu hình mô hình thử nghiệm và luồng xử lý đặc trưng.

Thành phần / Bước	TH 1: ResNet18	TH 2: ViT	TH 3: ResNet18+Aux	TH 4: ViT+Aux
Backbone (f_{bb})	ResNet18 $\rightarrow F$	ViT $\rightarrow F$	ResNet18 $\rightarrow F$	ViT $\rightarrow F$
Nhánh Phụ (h_{aux}) Mất mát \mathcal{L}_{aux}	<i>Không sử dụng</i>	<i>Không sử dụng</i>	Đầu vào: F Dự đoán CI Đầu ra (cho ghép nối):	Đầu vào: F Dự đoán CI Đầu ra (cho ghép nối):
Ghép nối Đặc trưng (\rightarrow)	$= F$	$= F$	$\text{concat}(F, \cdot) \rightarrow$	$\text{concat}(F, \cdot) \rightarrow$
Nhánh Chính (h_{final}) Mất mát \mathcal{L}_{main}	Đầu vào: Dự đoán Grade	Đầu vào: Dự đoán Grade	Đầu vào: Dự đoán Grade	Đầu vào: Dự đoán Grade

Ghi chú: F là vector đặc trưng từ Backbone. \cdot là vector đặc trưng nội bộ từ đầu ra của Nhánh Phụ trước lớp dự đoán CI cuối cùng, được sử dụng để ghép nối. CI là viết tắt của Current Intensity Number. Grade là cấp cường độ bão cuối cùng cần dự đoán.

Trong đó:

- **TH 1 (ResNet18):** Sử dụng ResNet-18 làm backbone, không có nhánh phụ. Đặc trưng F từ ResNet-18 được đưa thẳng vào nhánh chính h_{final} để dự đoán Grade.
- **TH 2 (ViT):** Sử dụng ViT làm backbone, không có nhánh phụ. Tương tự TH 1, đặc trưng F từ ViT được đưa vào nhánh chính.
- **TH 3 (ResNet18+Aux):** Sử dụng ResNet-18 làm backbone. Đặc trưng F được dùng cho cả nhánh phụ h_{aux} (dự đoán CI, tạo ra \mathcal{L}_{aux} và đặc trưng \cdot) và nhánh chính. Đặc trưng F và \cdot được ghép nối thành trước khi vào h_{final} để dự đoán Grade.
- **TH 4 (ViT+Aux):** Tương tự TH 3, nhưng sử dụng ViT làm backbone.

Việc huấn luyện và so sánh các trường hợp này sẽ giúp xác định kiến trúc backbone nào phù hợp hơn và liệu việc tích hợp nhiệm vụ phụ dự đoán CI có mang lại cải thiện đáng kể cho nhiệm vụ chính là dự đoán cường độ bão (Grade) hay không. Các thông số huấn luyện chi tiết như tốc độ học, hàm tối ưu, số epoch sẽ được xác định trong quá trình thực nghiệm.

4 KẾT QUẢ THỰC NGHIỆM VÀ PHÂN TÍCH

4.1 Số lượng ảnh và phân chia dữ liệu

Tập dữ liệu sử dụng trong thực nghiệm bao gồm ảnh vệ tinh địa tĩnh MTSAT thuộc 4 cấp độ bão (Grade): TD, TS, STS, và TY. Sau khi tiền xử lý và tăng cường dữ liệu (data augmentation), số lượng ảnh ở các tập dữ liệu được phân chia như Bảng 3.

Bảng 2: Số lượng ảnh theo từng lớp (Grade) trước và sau tăng cường dữ liệu

Lớp (Grade)	Origin Train	Augment Train	Validation	Test
TD	1532	3620	209	209
TS	1616	3680	231	231
STS	1058	3608	162	163
TY	1718	3618	293	293
Tổng	5924	14526	895	896

Tổng cộng, tập huấn luyện ban đầu gồm 5.924 ảnh, sau khi tăng cường đạt 14.526 ảnh. Các tập validation và test được chia đều theo từng lớp, với tổng số lần lượt là 895 và 896 ảnh. Việc tăng cường dữ liệu giúp cân bằng giữa các lớp và cải thiện khả năng tổng quát của mô hình khi huấn luyện.

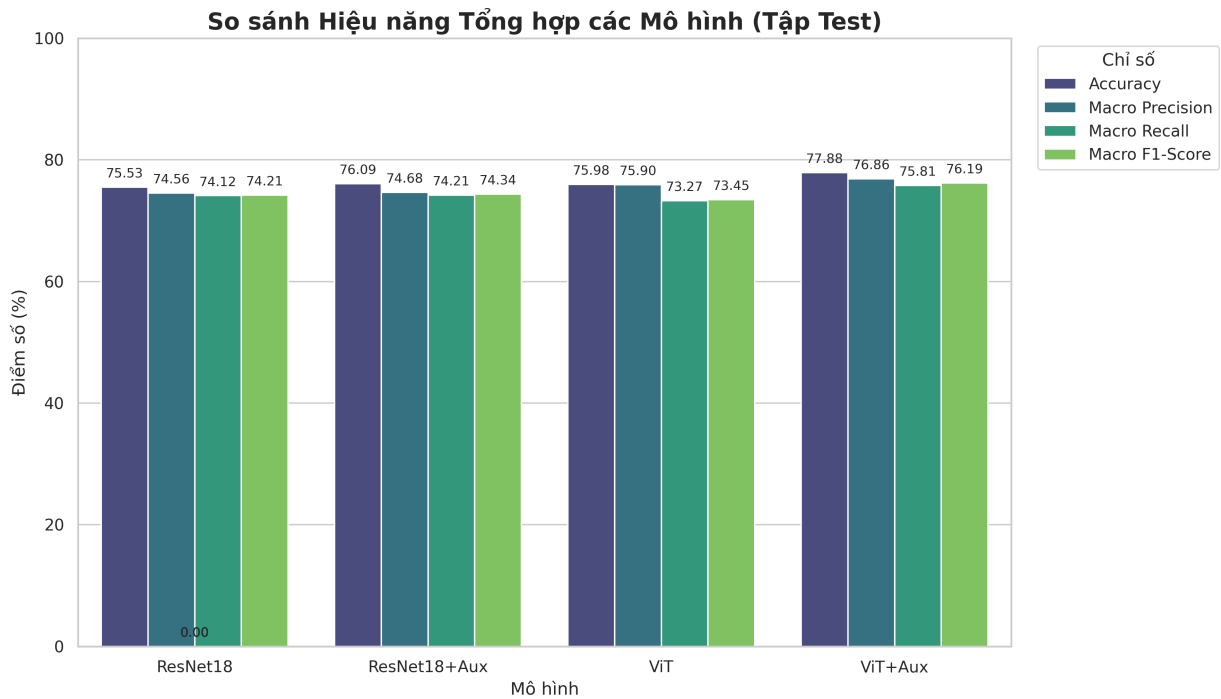
4.2 So sánh hiệu năng các mô hình

Bảng 3: Số lượng ảnh theo từng lớp (Grade) trước và sau tăng cường dữ liệu

Mô hình	TD	TS	STS	TY	Accuracy
ResNet18	81.19%	63.64%	62.73%	90.67%	75.53%
ResNet18+Aux	74.66%	66.97%	71.53%	85.58%	76.09%
ViT	70.96%	68.75%	80.0%	83.86%	75.98%
ViT+Aux	79.9%	67.9%	73.19%	86.45%	77.88%

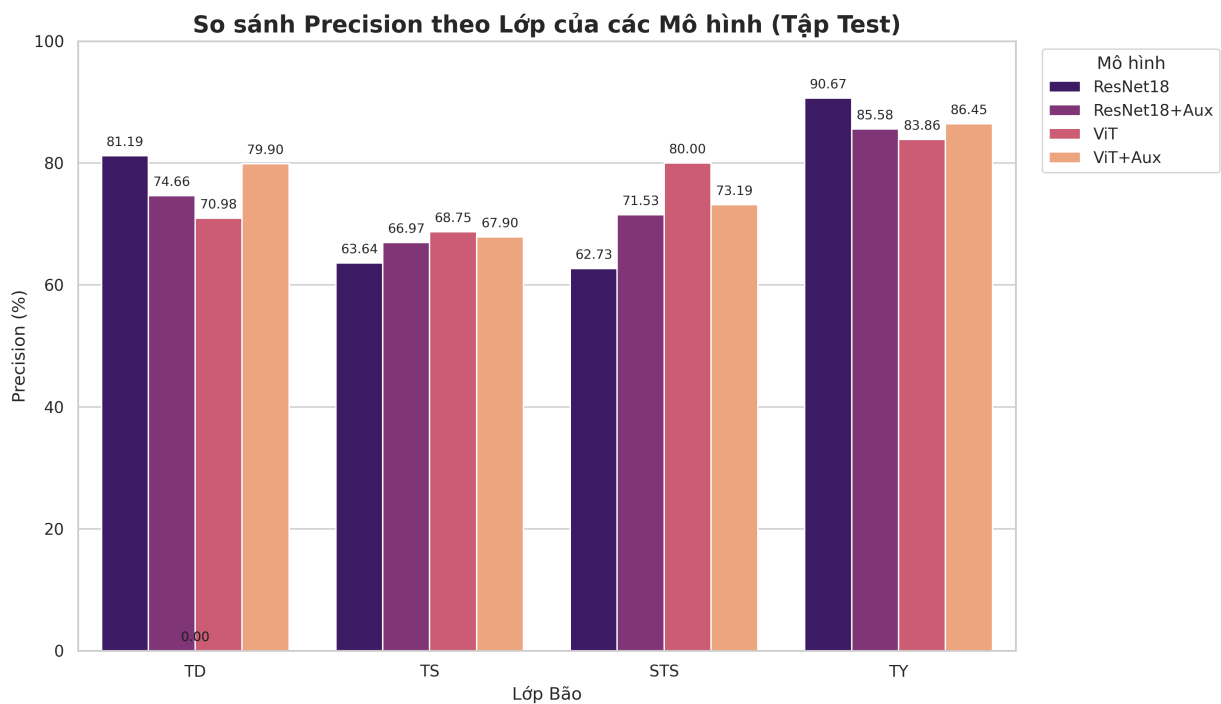
Độ chính xác tổng thể

- Mô hình **ViT+Aux** đạt kết quả cao nhất với **Accuracy 77.88%** và **F1-score 76.19%**.
- Việc bổ sung nhánh phụ (*+Aux*) giúp cải thiện hiệu năng rõ rệt cho cả ResNet và ViT.



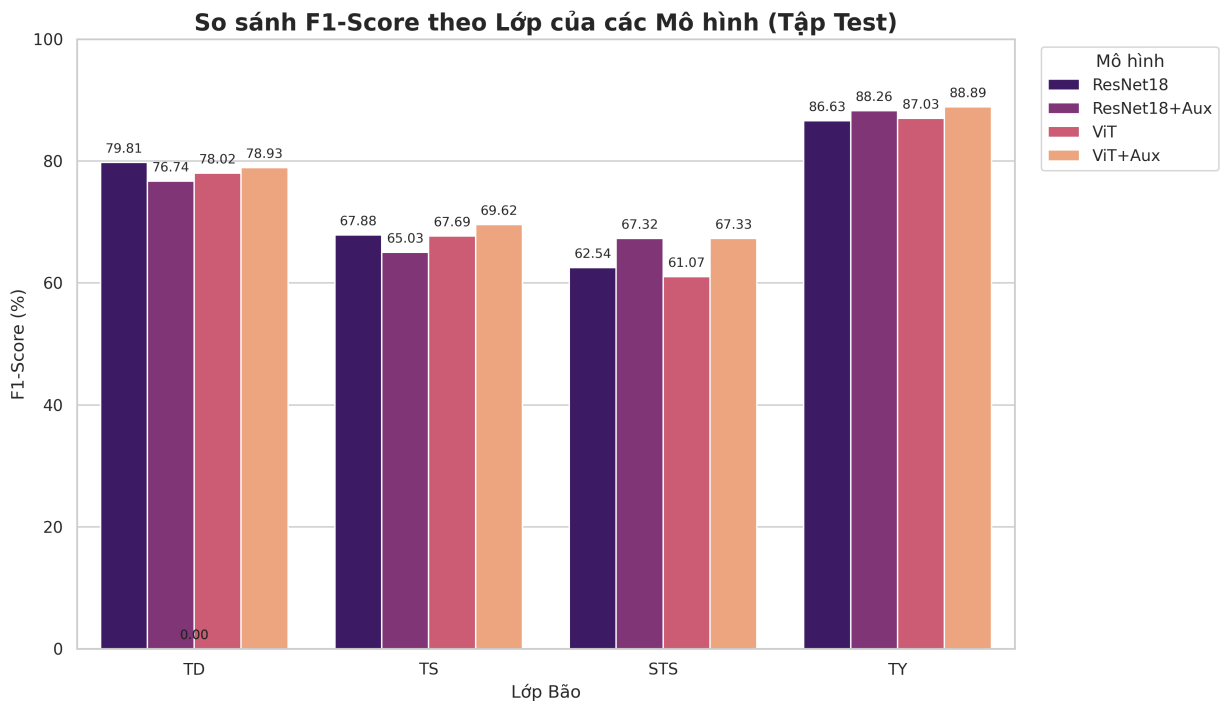
- Các chỉ số tổng hợp khác như **Macro Precision**, **Macro Recall**, và **Macro F1-score** cũng cho thấy ViT+Aux vượt trội so với các mô hình còn lại.

So sánh Precision theo lớp

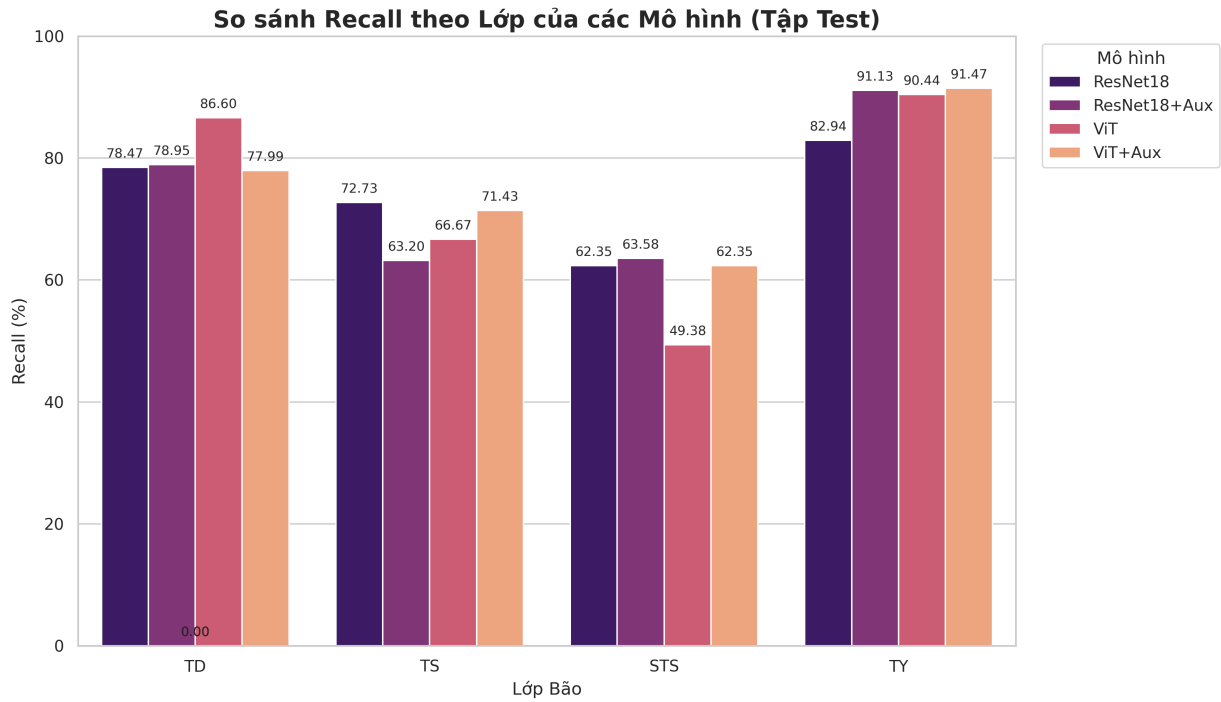


- **TY** và **TD** là hai lớp có độ chính xác cao nhất, cho thấy mô hình dễ phân biệt các lớp có đặc trưng rõ ràng.
- **TS** và **STS** có Precision thấp hơn, thường bị nhầm lẫn với nhau do đặc điểm hình thái mờ nhạt, đặc biệt là lớp STS.
- Việc tích hợp Auxiliary Task giúp **tăng Precision cho lớp STS**, cải thiện nhầm lẫn.

So sánh F1-Score theo lớp



- **Lớp TY** đạt **F1-score cao nhất** (87–89%), thể hiện khả năng nhận diện mạnh.
- **Lớp STS** có F1 thấp và biến động do khó phân biệt với các lớp trung gian như TS.
- Mô hình **ViT+Aux** cải thiện rõ **F1 của lớp STS** so với ViT gốc.



So sánh Recall theo lớp

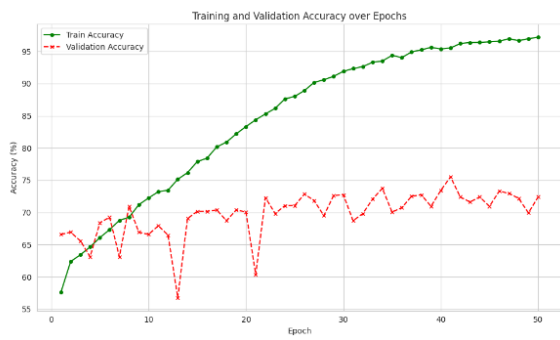
- **TY và TD** có Recall cao, nghĩa là mô hình ít bỏ sót mẫu thực.
- **Lớp STS** có Recall thấp, dễ bị dự đoán sai thành TS hoặc TY.
- Mô hình **ViT+Aux** cải thiện đáng kể Recall của STS, góp phần tăng hiệu năng tổng thể.

Tổng kết

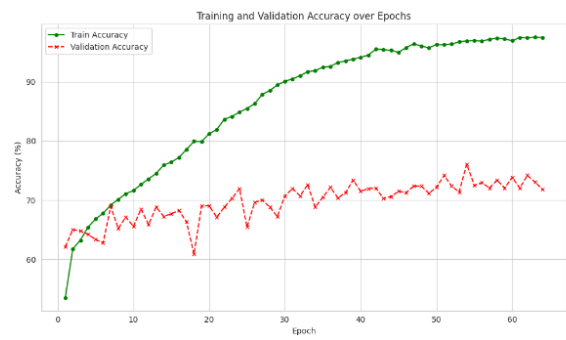
Các kết quả cho thấy mô hình **ViT+Aux** là lựa chọn tốt nhất trong bốn mô hình thử nghiệm. Kiến trúc transformer kết hợp với nhiệm vụ phụ hỗ trợ học biểu diễn tốt hơn, từ đó cải thiện độ chính xác và khả năng phân biệt giữa các lớp khó.

4.3 Phân tích quá trình huấn luyện

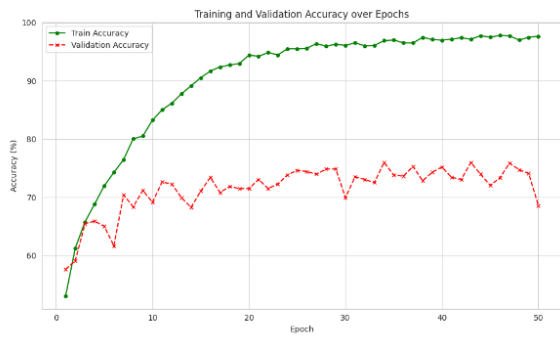
Quá trình huấn luyện của 4 mô hình được theo dõi thông qua hai chỉ số chính: **Accuracy** và **Loss** trên tập huấn luyện và validation. Kết quả được thể hiện trong Hình 5 và Hình 6.



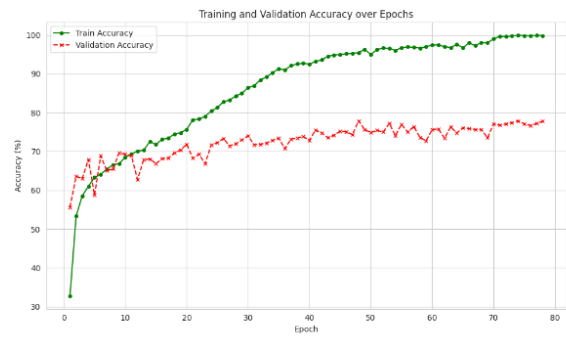
ResNet-18



ResNet18+Aux



ViT

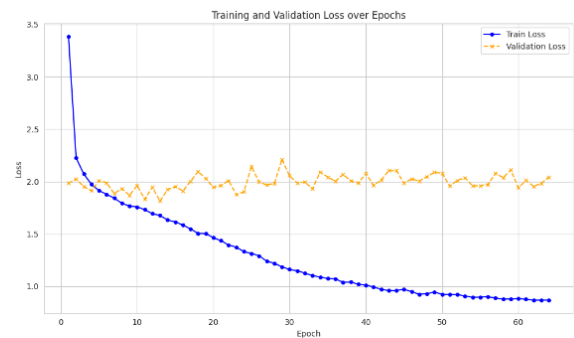


ViT+Aux

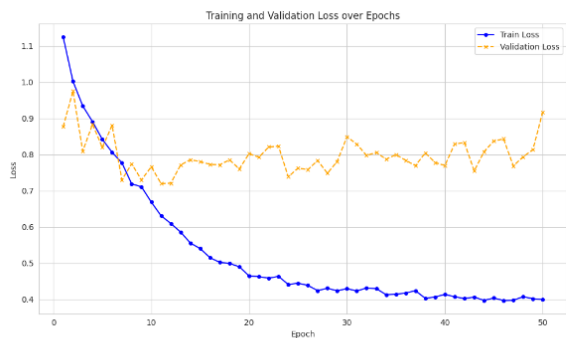
Hình 5: Biểu đồ Accuracy trên tập Train và Validation cho 4 mô hình



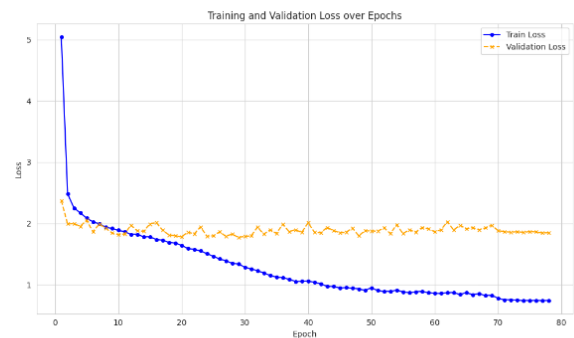
ResNet-18



ResNet18+Aux



ViT



ViT+Aux

Hình 6: Biểu đồ Loss trên tập Train và Validation cho 4 mô hình

Nhận xét

- Mô hình **ViT+Aux** mặc dù hội tụ chậm hơn trong giai đoạn đầu nhưng đạt Accuracy cao nhất trên validation và có độ dao động nhỏ nhất.
- **ResNet-18** huấn luyện nhanh, nhưng dễ gặp overfitting do mô hình đơn giản và không có cơ chế regularization bổ sung.
- **ViT (không có Aux)** có dấu hiệu **overfitting nhẹ** khi Accuracy trên validation thấp hơn rõ so với train và Loss không giảm ổn định ở cuối giai đoạn huấn luyện.
- Việc tích hợp **Auxiliary Task** đóng vai trò như một hình thức regularization, giúp mô hình học ổn định hơn và **giảm overfitting**.

Tổng kết

Qua các biểu đồ và phân tích, có thể thấy rằng:

- Các mô hình có tích hợp nhiệm vụ phụ (Auxiliary Task) không chỉ cải thiện hiệu năng tổng thể mà còn giúp **tăng khả năng tổng quát (generalization)** của mô hình trên dữ liệu chưa thấy.
- **ViT+Aux** thể hiện sự ổn định nhất trong quá trình huấn luyện, là mô hình ưu việt nhất trong thí nghiệm.

4.4 Thảo luận kết quả

Dựa trên các kết quả thực nghiệm, có thể rút ra một số nhận định và phân tích quan trọng như sau:

Tóm tắt các điểm nổi bật

- Việc tích hợp **nhiệm vụ phụ dự đoán chỉ số CINumber (Auxiliary Task)** giúp cải thiện đáng kể độ chính xác của mô hình, đóng vai trò như một hình thức regularization và cung cấp thêm tín hiệu huấn luyện.
- Mô hình **ViT+Aux Head** đạt kết quả tốt nhất trong tất cả các chỉ số đánh giá trên tập kiểm tra, cho thấy hiệu quả vượt trội khi kết hợp Transformer với hướng học phụ trợ.

- Việc tích hợp **kiến thức chuyên môn (Dvorak/CINumber)** thông qua nhánh phụ đã góp phần tăng khả năng học biểu diễn có ý nghĩa, từ đó nâng cao hiệu quả mô hình.

Hạn chế còn tồn tại

- Độ chính xác tổng thể vẫn còn hạn chế, chưa vượt quá 80%.
- Mô hình thường bị **nhầm lẫn giữa các lớp liền kề**, đặc biệt là giữa **TS và STS**, do ranh giới giữa hai cấp độ này không rõ ràng cả về mặt hình thái mây lẫn trong thực tế chuyên môn.
- **Dấu hiệu overfitting** vẫn tồn tại trong một số mô hình (đặc biệt là ViT không có Aux), thể hiện qua khoảng cách giữa train và validation accuracy.

Phân tích sâu hơn

- **ViT+Aux hoạt động tốt hơn** do khả năng học các mối liên kết toàn cục trong ảnh (global dependencies) của ViT và được dẫn dắt bởi nhiệm vụ phụ liên quan đến CINumber – một chỉ số chuyên ngành có ý nghĩa vật lý.
- Việc phân biệt **TS và STS khó khăn** là do các lớp này không có khác biệt rõ ràng trong cấu trúc mây, và sự phân định trong thực tế cũng dựa vào nhiều yếu tố ngoại vi không hiện rõ qua ảnh IR.
- **Mất cân bằng dữ liệu** vẫn là một yếu tố ảnh hưởng lớn đến hiệu năng, dù đã thực hiện tăng cường dữ liệu. Số lượng mẫu STS vẫn ít hơn đáng kể, khiến mô hình dễ thiên lệch khi dự đoán.

Kết luận rút ra

Kết hợp giữa mô hình học sâu hiện đại (ViT) và tri thức chuyên ngành (CINumber thông qua Auxiliary Task) là một hướng tiếp cận đầy hứa hẹn cho các bài toán nhận dạng hiện tượng thời tiết phức tạp. Tuy nhiên, vẫn cần cải tiến về kiến trúc, dữ liệu và kỹ thuật huấn luyện để khắc phục hạn chế hiện tại.

5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 Kết luận

Đề tài nghiên cứu nhằm ước tính cường độ bão (dưới dạng cấp độ) dựa trên việc phân tích ảnh vệ tinh hồng ngoại đa kênh. Phương pháp chính của đề tài sử dụng mô hình học sâu với các kiến trúc ResNet hoặc Vision Transformer làm backbone. Mô hình được huấn luyện theo hướng học đa nhiệm, kết hợp nhiệm vụ chính (dự đoán cường độ bão cuối cùng) và nhiệm vụ phụ (dự đoán chỉ số Dvorak CI) để tận dụng thông tin chuyên gia. Quá trình này bao gồm tiền xử lý ảnh tiêu chuẩn và các kỹ thuật tăng cường dữ liệu để cải thiện hiệu quả và đối phó với thách thức dữ liệu.

Các kết quả trực quan cho thấy mô hình sử dụng Vision Transformer (ViT) làm backbone cho hiệu suất tốt hơn trên tập validation so với ResNet-18. Việc kết hợp nhánh phụ (Auxiliary Task) dường như có lợi ích đáng kể hơn khi sử dụng với ViT so với ResNet-18, giúp cải thiện độ chính xác và sự ổn định của quá trình huấn luyện. Tuy nhiên, hiện tượng quá khớp vẫn là một thách thức cần được giải quyết để mô hình có thể khái quát hóa tốt hơn trên dữ liệu bão thực tế.

Tóm lại

- Việc thêm nhiệm vụ phụ dự đoán CINumber không mang lại cải thiện rõ rệt cho mô hình dựa trên ResNet-18. Tuy nhiên, đối với mô hình dựa trên Vision Transformer (ViT), việc kết hợp nhiệm vụ phụ này dường như có đóng góp tích cực.
- Kiến trúc Vision Transformer (ViT) cho thấy hiệu suất phù hợp hơn cho bài toán này so với ResNet-18. Cả khi không sử dụng nhiệm vụ phụ (ViT so với ResNet-18) và khi có kết hợp nhiệm vụ phụ (ViT+Aux so với ResNet18+Aux), các mô hình sử dụng ViT đều đạt được mức độ chính xác cao hơn và ổn định hơn.

5.2 Hướng phát triển

Dựa trên những kết quả đạt được và các thách thức còn tồn tại, đặc biệt là hiện tượng quá khớp và sự khác biệt về hiệu năng giữa các kiến trúc, các

hướng phát triển tiềm năng cho đề tài trong tương lai bao gồm:

- **Nâng cấp Kiến trúc Mô hình:** Tiếp tục khám phá tiềm năng của các kiến trúc học sâu hiện đại. Kết quả sơ bộ cho thấy Vision Transformer (ViT) có hiệu quả triển vọng hơn so với ResNet-18. Do đó, các nghiên cứu tiếp theo có thể tập trung vào việc thử nghiệm các biến thể tiên tiến hơn của Transformer hoặc các kiến trúc mạng nơ-ron mới nổi khác phù hợp với dữ liệu ảnh viễn thám phức tạp [4, 3]. Bên cạnh đó, việc tối ưu hóa phương pháp ghép nối (feature fusion) giữa các đặc trưng từ backbone và nhánh phụ, hoặc giữa các nguồn dữ liệu tiềm năng khác, cũng là một hướng quan trọng để cải thiện hiệu năng tổng thể của mô hình [9, 6].
- **Khai thác và Xử lý Dữ liệu Hiệu quả hơn:** Một trong những thách thức lớn nhất vẫn là vấn đề về dữ liệu, bao gồm sự mất cân bằng lớp và yêu cầu về số lượng/đa dạng [2, 1, 7]. Hướng phát triển cần tập trung vào:
 - Tích hợp dữ liệu đa nguồn: Mở rộng dữ liệu đầu vào bằng cách tích hợp thông tin từ các kênh vệ tinh khác (ví dụ: kênh khả kiến - Visible), dữ liệu từ các loại cảm biến viễn thám khác (như SAR để ước tính gió bề mặt) hoặc dữ liệu tái phân tích (Reanalysis) để cung cấp cái nhìn đầy đủ hơn về cơn bão [9, 6]. Việc kết hợp các nguồn dữ liệu không gian địa lý khác cũng là một hướng đi tiềm năng [6].
 - Khai thác thông tin chuỗi thời gian: Cơn bão là một hiện tượng động lực học. Thay vì xử lý từng ảnh độc lập, cần phát triển các mô hình có khả năng phân tích chuỗi thời gian của ảnh vệ tinh để nắm bắt sự phát triển và biến đổi cường độ bão theo thời gian [5].
 - Xử lý mất cân bằng dữ liệu: Áp dụng các kỹ thuật xử lý mất cân bằng dữ liệu tiên tiến hơn (ngoài các phương pháp tăng cường cơ bản), chẳng hạn như lấy mẫu phức tạp hơn, các hàm mất mát có trọng số (weighted loss) được điều chỉnh, hoặc sử dụng các mô hình sinh (generative models) để tạo thêm mẫu cho các lớp bão mạnh hiếm gặp [2, 1, 7].

- **Cải tiến Quy trình Huấn luyện và Tối ưu hóa:** Hiện tượng quá khớp cho thấy cần có các chiến lược huấn luyện tốt hơn.
 - Điều chuẩn và Hàm mất mát: Nghiên cứu các kỹ thuật điều chuẩn (regularization) hiệu quả hơn để giảm overfitting, đặc biệt khi sử dụng các mô hình phức tạp như ViT [1, 7]. Khám phá các hàm mất mát mới hoặc điều chỉnh hàm mất mát hiện tại để tập trung giảm thiểu sai sót trong việc dự đoán các cấp độ bão quan trọng hoặc xử lý tốt hơn sự nhầm lẫn giữa các lớp cường độ gần nhau [3].
 - Học đa nhiệm: Tối ưu hóa chiến lược học đa nhiệm, bao gồm việc tìm ra hệ số trọng số λ tối ưu cho hàm mất mát phụ và nghiên cứu sâu hơn về cách nhiệm vụ phụ dự đoán CI tác động và cải thiện việc học đặc trưng cho nhiệm vụ chính dự đoán Grade, đặc biệt với kiến trúc ViT.
- **Tăng cường Khả năng Diễn giải (Explainability):** Để mô hình AI được tin cậy và ứng dụng hiệu quả trong thực tế bởi các nhà khí tượng thủy văn, khả năng hiểu được lý do mô hình đưa ra một dự đoán cụ thể là rất quan trọng.
 - Áp dụng các kỹ thuật AI giải thích được (Explainable AI - xAI) để làm sáng tỏ các đặc trưng ảnh hưởng nào mà mô hình học sâu (đặc biệt là ViT) dựa vào để đưa ra dự đoán cường độ [8, 7].
 - Liên kết các đặc trưng quan trọng mà mô hình học được với các khái niệm hình thái mây truyền thống trong kỹ thuật Dvorak hoặc các đặc trưng vật lý khác của bão để tăng tính minh bạch và tin cậy [8].

PHÂN CÔNG CÔNG VIỆC VÀ ĐÁNH GIÁ ĐÓNG GÓP

Phân công công việc

Thành viên	Công việc được giao	Đóng góp
Vũ Minh Hiếu (22001582)	Triển khai, hoàn thiện mô hình ViT, hoàn thiện cơ sở lý thuyết	34%
Ngô Quang Huy (22001592)	Triển khai, hoàn thiện mô hình ResNet, hoàn thiện cơ sở lý thuyết	33%
Nguyễn Xuân Hồng (22001589)	Tải và xử lý dữ liệu, tìm hiểu lý thuyết, hoàn thiện báo cáo và trực quan hóa kết quả, dữ liệu	33%

Tự đánh giá đóng góp của thành viên

Vũ Minh Hiếu - 22001582

Em đã tham gia vào các công việc:

- Tìm hiểu sâu về kiến trúc ResNet50, Vision Transformer (ViT) và ứng dụng cho bài toán phân loại ảnh.
- Thu thập dữ liệu ảnh bão, thực hiện các bước tiền xử lý như cắt ROI, resize, chuẩn hóa.
- Cài đặt và tinh chỉnh mô hình ViT, bao gồm cả việc tích hợp nhánh phụ (Auxiliary Task) để dự đoán CInumber.
- Tiến hành huấn luyện và đánh giá mô hình ViT (TH2) và ViT+Aux (TH4) trên bộ dữ liệu bão.
- Tham gia viết nội dung cho phần phương pháp liên quan đến ViT và phần kết quả của các mô hình này trong báo cáo.

Ngô Quang Huy - 22001592

Em đã tham gia vào các công việc:

- Tìm hiểu về nguồn dữ liệu ảnh vệ tinh MTSAT và kỹ thuật Dvorak, chỉ số CINumber.
- Tìm hiểu về kiến trúc ResNet-18 và áp dụng transfer learning.
- Cài đặt, huấn luyện và đánh giá mô hình ResNet-18 (TH1) và ResNet18+Aux (TH3).
- Phân tích và so sánh hiệu năng giữa ResNet và ViT, cũng như ảnh hưởng của nhánh phụ.
- Tham gia viết nội dung liên quan đến ResNet, cấu hình huấn luyện chung và hỗ trợ tổng hợp báo cáo.

Nguyễn Xuân Hồng - 22001589

Em đã tham gia vào các công việc:

- Tìm kiếm thêm một số tài liệu liên quan
- Xử lý dữ liệu, nghiên cứu và áp dụng các kỹ thuật tăng cường dữ liệu để xử lý mất cân bằng.
- Tổng hợp các kết quả thực nghiệm từ các mô hình.
- Trực quan hóa dữ liệu: đánh giá và vẽ các biểu đồ so sánh hiệu năng (Accuracy, Precision, F1-Score, Recall), đồ thị quá trình huấn luyện.
- Tham gia phân tích, thảo luận các kết quả thu được, đóng góp ý kiến cho phần Thảo luận và Kết luận.
- Hỗ trợ chuẩn bị slide thuyết trình và rà soát lỗi chính tả, định dạng báo cáo.

TÀI LIỆU THAM KHẢO

Tài liệu

- [1] Dvorak, V. F. (1975). Tropical cyclone intensity analysis and forecasting from satellite imagery. *Monthly Weather Review*, 103(5), 420-430.
- [2] Dvorak, V. F. (1984). Tropical cyclone intensity analysis using satellite data. *NOAA Technical Report NESDIS 11*, U.S. Department of Commerce, Washington, D.C., 47 pp.
- [3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- [5] Recent Trends Challenges and Limitations of Explainable AI in Remote Sensing, CVF Open Access
- [6] Remote Sensing Data Quality in the Era of AI, ResearchGate / ISPRS Annals, 2024
- [7] Artificial Intelligence for Remote Sensing Data Analysis: A Review of Challenges and Opportunities, ResearchGate
- [8] AI Foundation Models in Remote Sensing: A Survey, arXiv preprint arXiv:2408.03464, 2024
- [9] Research Trend of the Remote Sensing Image Analysis Using Deep Learning, Korean Journal of Remote Sensing
- [10] A Review of Practical AI for Remote Sensing in Earth Sciences, MDPI, 2023
- [11] Artificial Intelligence for Remote Sensing applications: recent developments and main challenges, YouTube video, 2025

- [12] Potential Applications and Limitations of Artificial Intelligence in Remote Sensing Data Interpretation: A Case Study, ejournal.csol.or.id
- [13] Opportunities and Challenges of Artificial Intelligence and Big Data for Remote Sensing, *Frontiers*