**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

A logo with a fire and flames

Description automatically generated

**ĐỒ ÁN CUỐI KÌ**

**DEEP LEARNING**

*Người thực hiện*: **Ngô Hoàng Nhật Huy – 20079071**

**Đoàn Ánh Dương - 20082631**

Lớp : 420300411601

Khóa : 16

*Người hướng dẫn*: **TS Bùi Thanh Hùng**

## Thành phố Hồ Chí Minh – 2023

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

A logo with a fire and flames

Description automatically generated

**ĐỒ ÁN CUỐI KÌ**

**DEEP LEARNING**

*Người thực hiện*: **Ngô Hoàng Nhật Huy – 20079071**

**Đoàn Ánh Dương - 20082631**

Lớp : 420300411601

Khóa : 16

*Người hướng dẫn*: **TS Bùi Thanh Hùng**

## Thành phố Hồ Chí Minh – 2023

# LỜI CẢM ƠN

Chúng tôi xin chân thành cảm ơn đến quý Thầy TS. Bùi Thanh Hùng– Giảng viên hướng dẫn đã truyền đạt những kiến thức nền tảng, những bài học lập trình, qua đó đã giúp đỡ chúng tôi rất nhiều trong quá trình thực hiện Đồ án này.

# ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH

Chúng tôi xin cam đoan đây là sản phẩm đồ án của chúng tôi và được sự hướng dẫn của TS Bùi Thanh Hùng. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường Đại học Sư phạm TP. Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày tháng năm Tác giả*

*(ký tên và ghi rõ họ tên)*

*Ngô Hoàng Nhật Huy*

*Đoàn Ánh Dương*

# PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

TP. Hồ Chí Minh, ngày tháng năm (kí và ghi họ tên)

## TÓM TẮT

Bài toán nhận diện các vệt trong ảnh vệ tinh để giảm hiện tượng nóng lên toàn cầu (Identify Contrails to Reduce Global Warming) là bài toán cải thiện độ hiệu quả mô hình của các nhà nghiên cứu. Đây là một bài toán quan trọng trong việc xác định những vệt được ngưng tụ từ những đám mây tinh thể đá hình thành trong khí thải động cơ máy bay, chúng góp phần tạo nên sự nóng lên toàn cầu bằng cách giữ nhiệt trong bầu khí quyển. Bài báo này trình bày về việc áp dụng mô hình Unet, một mô hình mạng nơ-ron học sâu, để dự đoán các vệt trong ảnh vệ tinh nhằm cải thiện độ hiệu quả của việc đánh giá sự hiện diện của chúng. Bài toán được đánh giá dựa trên hệ số Dice toàn cầu, phương pháp này giúp so sánh sự trùng khớp pixel-wise giữa kết quả dự đoán và đối chiếu với thực tế, từ đó đo lường hiệu suất của mô hình.

# MỤC LỤC

[LỜI CẢM ƠN 1](#_Toc153314325)

[PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN 3](#_Toc153314327)

[TÓM TẮT 4](#_Toc153314328)

[MỤC LỤC 1](#_Toc153314329)

[DANH MỤC HÌNH VẼ 4](#_Toc153314330)

[DANH MỤC CÁC BẢNG 5](#_Toc153314331)

[1.1 Giới thiệu về bài toán 6](#_Toc153314332)

[1.2 Phân tích yêu cầu của bài toán 7](#_Toc153314333)

[1.2.1 Yêu cầu của bài toán 7](#_Toc153314334)

[1.2.2 Phương pháp giải quyết bài toán 7](#_Toc153314335)

[1.2.2.1 Các nghiên cứu trước đó 7](#_Toc153314336)

[1.2.2.2 Các phương pháp đề xuất 8](#_Toc153314337)

[1.3 Phương pháp giải quyết bài toán 9](#_Toc153314338)

[1.3.1 Mô hình tổng quát 9](#_Toc153314339)

[1.3.2 Đặc trưng của mô hình đề xuất 9](#_Toc153314340)

[1.4 Thực nghiệm 13](#_Toc153314341)

[1.4.1 Dữ liệu 13](#_Toc153314342)

[1.4.2 Xử lý dữ liệu 14](#_Toc153314343)

[1.4.3 Công nghệ sử dụng 15](#_Toc153314344)

[1.4.4 Cách đánh giá 15](#_Toc153314345)

[1.5 Kết quả đạt được 16](#_Toc153314346)

[1.5.1 Tham số thực nghiệm 16](#_Toc153314347)

[1.5.2 Kết quả đạt được 16](#_Toc153314348)

[1.6 Kết luận 17](#_Toc153314349)

[1.6.1 Kết quả đạt được 17](#_Toc153314350)

[1.6.2 Hạn chế 18](#_Toc153314351)

[1.6.3 Hướng phát triển 18](#_Toc153314352)

[TÀI LIỆU THAM KHẢO 19](#_Toc153314353)

[LÀM VIỆC NHÓM 21](#_Toc153314354)

[TỰ ĐÁNH GIÁ 22](#_Toc153314355)

**DANH MỤC CÁC CHỮ VIẾT TẮT**

## DANH MỤC HÌNH VẼ

[**Hình 1**: Các vệt từ khói thải động cơ máy bay 6](#_Toc153306787)

[**Hình 2**: Mô hình tổng quát 8](https://iuhedu-my.sharepoint.com/personal/20079071_huy_student_iuh_edu_vn/Documents/BaoCaoGroup13.docx#_Toc153306788)

[**Hình 3**: Kiến trúc U-Net 9](#_Toc153306789)

[**Hình 4**: Kiến trúc VGG19 10](#_Toc153306790)

[**Hình 5:** Kiến trúc ResNet50 11](#_Toc153306791)

[**Hình 6:** Tóm tắt bước tiền xử lý dữ liệu 13](#_Toc153306792)

[**Hình 7:** Kết quả dự đoán 15](#_Toc153306793)

## DANH MỤC CÁC BẢNG

[**Bảng 1**: Công nghệ sử dụng 13](#_Toc153306842)

[**Bảng 2:** So sánh kết quả thực nghiệm 14](#_Toc153306843)

**XÁC ĐỊNH CÁC VỆT – IDENTIFY CONTRAILS**

# 1.1 Giới thiệu về bài toán

Hiện tượng đám mây tạo ra từ hơi ngưng tụ “condensation trails”, hay còn được viết tắt là “contrails”, là những đám mây mảnh tạo ra từ hơi trong không khí, được tạo ra bởi các phản ứng hóa học và vật lý khi khí thải từ động cơ máy bay tương tác với môi trường xung quanh.

Xác định các vệt là bài toán nhận diện sự tồn tại của các vệt do khói thải động cơ máy bay tạo ra, đóng vai trò quan trọng việc nóng lên toàn cầu bằng cách giữ lại nhiệt trong khí quyển. Sáng kiến nghiên cứu của Google nhằm giải quyết thách thức môi trường này bằng cách sử dụng mô hình học máy để nhận diện vệt có trong hình ảnh vệ tinh. Mục tiêu cuối cùng là hỗ trợ các biện pháp chủ động để ngăn chặn sự hình thành của chúng, từ đó giảm nhẹ ảnh hưởng của chúng đối với biến đổi khí hậu.

Kết quả nghiên cứu của bài toán này chính là cơ sở để hiểu rõ hơn về phân bố không gian và thời gian của các vệt, từ đó ảnh hưởng tích cực cho ngành hàng không, cung cấp thông tin về khu vực dễ xuất hiện các vệt và hỗ trợ trong việc phát triển chiến lược để tránh hoặc giảm thiểu sự hình thành của chúng. Nghiên cứu này đồng bộ với các nỗ lực của các nhà nghiên cứu, các nhà khoa học trên toàn cầu để giải quyết vấn nạn biến đổi khí hậu bằng cách nhắm vào một nhân tố cụ thể và có ảnh hưởng lớn đến sự nóng lên toàn cầu.

A diagram of a engine exhaust

Description automatically generated

**Hình 1**: Các vệt từ khói thải động cơ máy bay

Trong hơn một thập kỷ qua, các nhà khoa học khí hậu Châu Âu đã xây dựng các mô hình mạnh mẽ, dựa trên dữ liệu thời tiết, để xác định khi nào các vệt sẽ hình thành và chúng sẽ tạo ra hiện tượng nóng lên đến mức nào. Nghiên cứu của họ đã được các phòng thí nghiệm khác xác nhận và hiện đã được chấp nhận rộng rãi rằng các vệt đóng góp khoảng 1% vào việc nóng lên toàn cầu do con người gây ra. Động cơ đằng sau việc sử dụng hình ảnh từ vệ tinh là để xác nhận thực nghiệm các dự đoán từ những mô hình này. Với sự xác minh đáng tin cậy, phi công có thể tin tưởng vào những mô hình này và ngành công nghiệp hàng không có thể có một cách đo lường đáng tin cậy để đánh giá việc tránh các vệt tương đối thành công

# 1.2 Phân tích yêu cầu của bài toán

## 1.2.1 Yêu cầu của bài toán

Mục tiêu của bài toán sử dụng các mô hình học sâu, học máy để xác định các vệt từ khói thải động cơ máy bay, cải thiện độ tin cậy trong việc dự đoán các vùng hình thành vệt và các kỹ thuật để tránh tạo ra chúng. Cụ thể, các phương pháp phải đáp ứng đủ yêu cầu như: xác định các vệt được cung cấp từ hình ảnh của bộ dữ liệu và phải đúng với những vệt đã được đánh dấu từ trước của các nhãn gia. Bài toán Identify Contrails là một bài toán tương đối khó nhưng mang lại tiềm năng vô cùng lớn cho con người tránh khỏi hiện tượng nóng lên toàn cầu cũng như phát triển mạnh mẽ cho ngành hàng không.

## 1.2.2 Phương pháp giải quyết bài toán

### 1.2.2.1 Các nghiên cứu trước đó

Để giải quyết bài toán này, một số nghiên cứu đã sử dụng phương pháp học sâu để xác định các vệt từ khói thải động cơ máy bay. Cụ thể, bài báo “Combining UPerNet and ConvNeXt for Contrails Identification to reduce Global Warming” của tác giả Zhenkuan Wang được công bố trên tạp chí Computer Vision and Pattern Recognition năm 2023 [1]. Bài báo này trình bày các nghiên cứu sử dụng các phương pháp học sâu như: UperNet [2] và ConvNext [3] để xác định các vệt khói thải này. Nghiên cứu này đã sử dụng tập dữ liệu GOES-16 [4] được giới thiệu trong cuộc thi này. Các mô hình đã được đánh giá trên tập kiểm tra và đạt được kết quả khá tốt. Kết quả của nghiên cứu này cho thấy rằng mô hình học sâu có tiềm năng lớn để cải thiện độ chính xác và hiệu quả của việc xác định vệt khói thải. Tuy nhiên, nghiên cứu này còn một số hạn chế nhất định của việc xây dựng mô hình học sâu đó là cần tập dữ liệu lớn và đa dạng để đào tạo. Ngoài ra, mô hình học sâu còn có thể bị ảnh hưởng bởi nhiễu và một số yếu tố khác.

### 1.2.2.2 Các phương pháp đề xuất

Convolution Neural Network (CNN) là một trong những mô hình thuộc lĩnh vực Học sâu. Cấu trúc CNN bao gồm 3 thành phần chính là lớp convolutional, lớp pooling, lớp fully connected:

Đối với lớp Convolutional sẽ tính “element-wise” của ma trận con từ ma trận đặc trưng có kích thước được điều chỉnh trước với ma trận kernel có cùng kích thước ma trận con. Tùy vào dạng ma trận kernel ta nhận được các đặc trưng khác nhau. Với mỗi phép tính ta sẽ được một đặc trưng được trích xuất. Sau khi đã tạo ra ma trận đặc trưng mới sẽ đi qua hàm kích hoạt đưa ra đầu ra của lớp Convolutional.

Lớp Pooling sẽ giúp CNN giảm số lượng tính toán nhưng vẫn có thể trích xuất đặc trưng và giảm số chiều ma trận đặc trưng. Điểm bất lợi mang lại mặc giảm được số lượng tính toán nhưng phải chấp nhận có thể mất đi nhiều đặc trưng quan trọng lúc trích xuất hơn lớp Convolutional.

Lớp Fully connected sẽ có dạng như một mạng lưới nơ-ron thông thường có đâu vào lúc này được nối lại tạo thành vector chứa các đặc trưng trích xuất từ câu và đầu ra là các lớp cần dự đoán. Thành phần bao gồm 3 phần: lớp đầu vào, lớp ẩn, lớp đầu ra.

CNN có nhiều ưu điểm trong bài toán phân đoạn ảnh bao gồm:

Khả năng học các đặc trưng cục bộ: CNN sử dụng các bộ lọc tích chập để học các đặc trưng cục bộ trong hình ảnh. Điều này giúp CNN có thể học các đặc trưng quan trọng cho bài toán phân đoạn, chẳng hạn như các cạnh, các đường, các hình dạng.

Khả năng học các đặc trưng không gian: CNN có thể học các đặc trưng không gian, tức là các đặc trưng liên quan đến vị trí của các đặc trưng trong hình ảnh. Điều này giúp CNN có thể phân đoạn các đối tượng có kích thước khác nhau hoặc có vị trí chồng chéo lên nhau.

Khả năng học các đặc trưng cấp độ cao: CNN có thể học các đặc trưng cấp độ cao, tức là các đặc trưng tổng hợp từ các đặc trưng cục bộ. Điều này giúp CNN có thể phân đoạn các đối tượng phức tạp, chẳng hạn như khuôn mặt người, các bộ phận cơ thể.

Trong bài toán này, chúng tôi đề xuất sử dụng các phương pháp học sâu để giải quyết là các biến thể của CNN. Một số phương pháp chúng tôi sử dụng trong bài toán này: Unet, RestNet50, VGG19. Vì các phương pháp trên đều được sử dụng hầu hết cho các bài toán phân đoạn ảnh, có thể phân lớp được nhiều đối tượng và có kiến trúc linh hoạt. Các kết quả cho thấy những mô hình này dều đạt được hiệu quả tốt trong bài toán xác định các vệt khói thải từ động cơ máy bay.

# 1.3 Phương pháp giải quyết bài toán

## 1.3.1 Mô hình tổng quát

Data

Data Pre-processing

Compare and Evaluate

Find Optimal Thershold

Deep Learning Models

VGG19

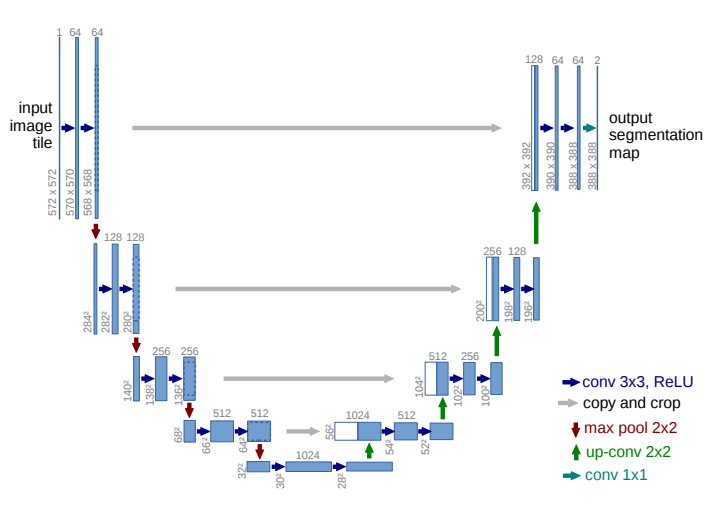
Unet

ResNet50

**Hình 2**: Mô hình tổng quát

## 1.3.2 Đặc trưng của mô hình đề xuất

**U-Net [5]** là một kiến trúc mạng nơ-ron sâu được sử dụng cho các tác vụ phân đoạn hình ảnh (image segmentation). Kiến trúc này được giới thiệu lần đầu tiên bởi Olaf Ronneberger và cộng sự trong bài báo "U-Net: Convolutional Networks for Biomedical Image Segmentation" năm 2015. U-net là một mạng nơ-ron tích chập (CNN) được phát triển để giải quyết các vấn đề liên quan đến xử lý ảnh, đặc biệt là trong tác vụ phân đoạn ảnh. Kiến trúc này đã được chứng minh là hiệu quả trong việc phân đoạn các đối tượng trong hình ảnh y tế, tự nhiên và vệ tinh.



**Hình 3**: Kiến trúc U-Net

Kiến trúc U-Net có thể được chia thành hai phần chính:

**Phần downsampling**:

Phần này chịu trách nhiệm thu nhỏ kích thước của biểu diễn ảnh và tăng số kênh. Điều này giúp mô hình học các đặc trưng toàn cục của ảnh.

**Phần upsampling:**

Phần này chịu trách nhiệm phóng to kích thước của biểu diễn ảnh và giảm số kênh. Điều này giúp mô hình khôi phục chi tiết và dự đoán phân đoạn cho từng pixel.

Kiến trúc U-Net có một số ưu điểm sau:

Khả năng học các đặc trưng ở các mức độ khác nhau: Phần downsampling của U-Net giúp mô hình học các đặc trưng toàn cục của ảnh, trong khi phần upsampling của U-Net giúp mô hình khôi phục chi tiết và dự đoán phân đoạn cho từng pixel. Điều này giúp mô hình có thể dự đoán phân đoạn chính xác hơn.

Kiến trúc đơn giản: Kiến trúc U-Net có cấu trúc đơn giản và dễ hiểu, giúp việc huấn luyện và triển khai mô hình trở nên dễ dàng hơn.

**Quá trình huấn luyện**

Trong đồ án này, mô hình U-Net của chúng tôi sử dụng các lớp

DoubleConv, lớp Down, lớp UP, lớp outc.

**Lớp DoubleConv:**

Là khối xây dựng cơ bản của mô hình.Gồm hai lớp cuộn tích chập, hai lớp chuẩn hóa BatchNorm2d và hai lớp ReLU.Giúp trích xuất đặc trưng và áp dụng các hàm kích hoạt không tuyến tính.

**Lớp Down:**

Giảm kích thước của biểu diễn ảnh và tăng số kênh. Gồm một lớp MaxPooling2d và một lớp DoubleConv. Giúp tổng hợp thông tin và tập trung vào các đặc trưng quan trọng hơn.

**Lớp UP:**

Tăng kích thước của biểu diễn ảnh và giảm số kênh.Gồm một lớp Upsample hoặc ConvTranspose2d và một lớp DoubleConv.Giúp khôi phục chi tiết và dự đoán phân đoạn cho từng pixel.

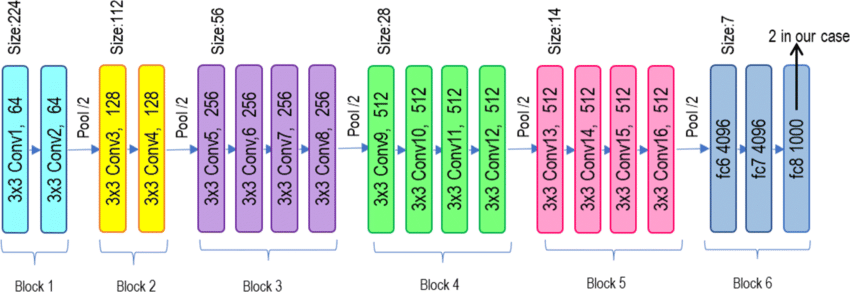
**Lớp outc:**

Lớp cuộn tích chập cuối cùng. Dự đoán phân đoạn với một kênh đầu ra.

Mô hình sử dụng thuật toán tối ưu Adam với tốc độ học là 0.01. Quá trình huấn luyện diễn ra qua nhiều epochs, sau mỗi epochs tốc độ học được cập nhật lại. Sau khi kết thúc quá trình huấn luyện chúng tôi đánh giá hiệu xuất mô hình trên tập validation với hệ số đánh giá Dice.

**VGG19 [6]** là một trong những mô hình nổi tiếng trong lĩnh vực thị giác máy tính và được phát triển bởi nhóm nghiên cứu Visual Geometry Group tại Đại học Oxford. Mô hình này đã đạt được thành công lớn trong cuộc thi ImageNet Large Scale Visual Recognition Challenge (ILSVRC) năm 2014. Mô hình VGG19 giống với VGG16 ngoại trừ việc nó hỗ trợ nhiều lớp chập hơn VGG16 ba lớp. Điểm đặc biệt VGG19 có thể phân loại tới 1000 đối tượng.

Kiến trúc mạng của VGG19 có kích thước đầu vào là 224x224x3, gồm 16 lớp convolutional đặt liên tiếp, mỗi lớp có bộ lọc 3x3 và hàm kích hoạt ReLU, Lớp pooling kiểu max pooling với cửa sổ 2x2 và bước nhảy 2, ba lớp fully connected với 4096 đơn vị ẩn mỗi lớp, hàm softmax thường được áp dụng để tạo ra phân phối xác suất trên các lớp phân loại. Bằng cách này, mục tiêu của VGG19 là đạt được hiệu suất tốt trong việc phân loại hình ảnh và các nhiệm vụ thị giác máy tính khác, thông qua việc sử dụng một kiến trúc mạng sâu với nhiều lớp tích chập và lớp kết nối đầy đủ.



**Hình 4**: Kiến trúc VGG19

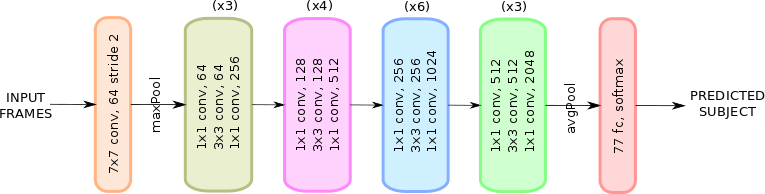
**Quá trình huấn luyện**

Đầu vào của mô hình là một ảnh với 24 kênh và kích thước 256x256. Quá trình huấn luyện mô hình mô hình sử dụng thuật toán tối ưu Adam với tốc độ học là 0.01. Mô hình được huấn luyện trên tập dữ liệu huấn luyện bằng cách gọi phương thức fit. Quá trình huấn luyện diễn ra qua nhiều epochs, sau mỗi epochs tốc độ học được cập nhật lại. Thời gian hoàn thành quá trình huấn luyện và các thống kê hiệu suất khác như loss và chỉ số dice được ghi nhận và đánh giá.

**ResNet50 [7]** (Residual Network với 50 lớp) giới thiệu khái niệm của residual learning, trong đó các residual block được sử dụng để tránh vấn đề vanishing gradient trong quá trình đào tạo mô hình. Các residual block cho phép thông tin chuyển dẫn không biến đổi thông qua một đường truyền thẳng (identity shortcut).

Với hiệu suất cao và có khả năng mở rộng dễ dàng, ResNet50 là một tiến bộ quan trọng trong học sâu, mang lại giải pháp cho những thách thức trong việc xây dựng và đào tạo các mô hình sâu. Với kiến trúc linh hoạt và hiệu suất ấn tượng, ResNet50 không chỉ là một công cụ quan trọng trong nghiên cứu mà còn là một công cụ hữu ích trong nhiều ứng dụng thực tế. Nó mở ra một kỷ nguyên mới trong phát triển và triển khai mô hình học sâu.

ResNet50 có một kiến trúc sâu và linh hoạt với 50 lớp, bao gồm cả các lớp tích chập và các lớp kết nối đầy đủ. Các residual block có số lượng và kích thước lớp biến đổi được điều chỉnh để tối ưu cho hiệu suất và khả năng mở rộng.



**Hình 5**: Kiến trúc ResNet50

**Quá trình huấn luyện**

Đầu vào của mô hình là một ảnh với 24 kênh và kích thước 256x256. Quá trình huấn luyện mô hình sử dụng thuật toán tối ưu Adam với tốc độ học là 0.01. Mô hình được huấn luyện trên tập dữ liệu huấn luyện bằng cách gọi phương thức fit. Quá trình huấn luyện diễn ra qua nhiều epochs, sau mỗi epochs tốc độ học được cập nhật lại. Thời gian hoàn thành quá trình huấn luyện và các thống kê hiệu suất khác như loss và chỉ số Dice được ghi nhận và đánh giá.

# 1.4 Thực nghiệm

## 1.4.1 Dữ liệu

Đồ án tiến hành thực nghiệm trên bộ dữ liệu sử dụng hình ảnh vệ tinh đồng trạm từ Thiết Bị Hình ảnh Cơ Bản Nâng Cao (ABI) của GOES-16 để nhận diện dải vệt máy bay. Hình ảnh được cung cấp theo chuỗi 10 phút để cung cấp bối cảnh thời gian. Mỗi ví dụ chứa một khung hình được gắn nhãn, và dữ liệu đào tạo có chứa thông tin cá nhân và chân thực đánh giá từ nhiều nhãn gia.

Dự liệu bao gồm:

Train/ - tập huấn luyện; mỗi thư mục đại diện cho một record\_id và chứa dữ liệu sau:

* band\_{08-16}.npy: mảng có kích thước H x W x T, trong đó T = n\_times\_Before + n\_times\_after + 1, biểu thị số lượng hình ảnh trong chuỗi. Có n\_times\_trước và n\_times\_after hình ảnh trước và sau khung được gắn nhãn tương ứng. Mỗi dải đại diện cho một kênh hồng ngoại ở các bước sóng khác nhau và được chuyển đổi thành nhiệt độ sáng dựa trên các thông số hiệu chuẩn. Số trong tên tệp tương ứng với số băng tần GOES-16 ABI.
* human\_individual\_mask.npy: mảng có kích thước H x W x 1 x R. Mỗi ví dụ được gắn nhãn bởi R người gắn nhãn riêng lẻ. R không giống nhau đối với tất cả các mẫu. Mặt nạ được gắn nhãn có giá trị 0 hoặc 1 và tương ứng với hình ảnh thứ (n\_times\_Before+1) trong band\_{08-16}.npy. Chúng chỉ có sẵn trong tập huấn luyện.
* human\_pixel\_masks.npy: mảng có kích thước H x W x 1 chứa giá trị cơ bản nhị phân. Một pixel được coi là pixel tương phản trong đánh giá nếu nó được hơn một nửa số người gắn nhãn dán nhãn là tương phản.

Validation/ - Giống như tập huấn luyện, không có chú thích nhãn riêng lẻ.

Test/ - Tập kiểm tra. Mục tiêu của tập này là xác định những dấu vết trái ngược được tìm thấy trong những tập này.

## 1.4.2 Xử lý dữ liệu

Dữ liệu hình ảnh được cung cấp theo chuỗi thời gian với khoảng thời gian 10 phút. Việc này mang lại lợi ích về bối cảnh thời gian cho việc nhận diện các vệt, nhưng đồng thời đòi hỏi xử lý đặc biệt để kết hợp thông tin từ các hình ảnh trước và sau khung hình được gắn nhãn. Mỗi hình ảnh được biểu diễn dưới dạng mảng với các kích thước và định dạng khác nhau. Đối với mỗi record\_id, có nhiều băng sóng (band\_{08-16}) và các mặt nạ chú thích từ nhiều nhãn gia. Điều này đòi hỏi quá trình xử lý để đảm bảo sự tương thích và đồng nhất giữa các mảng này. Dữ liệu chú thích từ nhiều nhãn gia (human\_individual\_masks.npy, human\_pixel\_masks.npy) cung cấp thông tin về vị trí và mức độ chắc chắn của các vệt. Để đạt được kết quả chính xác, quá trình xử lý cần đảm bảo sự nhất quán giữa các chú thích và chân thực. Bài toán được đánh giá bằng cách sử dụng hệ số Dice, yêu cầu phải so sánh sự trùng khớp pixel-wise giữa dự đoán và chân thực. Điều này đặt ra yêu cầu cụ thể về việc xử lý đầu ra của mô hình để tạo ra đầu ra có định dạng phù hợp với đánh giá này.

Vì vậy, quá trình xử lý dữ liệu là quan trọng để chuẩn bị dữ liệu cho việc đào tạo mô hình và đánh giá hiệu suất của nó trên bài toán nhận diện contrails từ hình ảnh vệ tinh.

A diagram of a diagram

Description automatically generated with medium confidence

**Hình 6:** Tóm tắt bước tiền xử lý dữ liệu

Thu nhận dữ liệu ảnh: Tạo ra 3 DataFrame chứa danh sách các tệp trong các thư mục dữ liệu tương ứng.

Tạo ranh giới: Nhận dữ liệu và ranh giới trong phạm vi [0,1].

Ánh xạ dữ liệu: Ánh xạ dữ liệu dựa trên ranh giới đã cung cấp.

Chuẩn hóa: Chuẩn hóa các dải bằng cách sử dụng ranh giới và dữ liệu đã được ánh xạ trước đó

Hình ảnh màu giả (RGB): Kết hợp các dải đã chuẩn hóa thành một hình ảnh màu giả được định dạng theo hệ màu RGB. Nếu đúng với mặt nạ đã được gắn nhãn bởi các nhãn gia trích xuất một kênh duy nhất từ mỗi dải vệt và trả về mảng 2D, ngược lại trả về hình ảnh màu giả.

## 1.4.3 Công nghệ sử dụng

|  |  |
| --- | --- |
| Ngôn ngữ | Python [8] |
| Thư viện | Pytorch [9] |
| Môi trường | Kaggle |

**Bảng 1**: Công nghệ sử dụng

## 1.4.4 Cách đánh giá

**Dice Coefficient (Dice) [10]**

Đồ án sẽ sử dụng hệ số Dice làm độ đo đánh giá hiệu suất của mô hình trong việc nhận diện các vệt từ hình ảnh vệ tinh. Hệ số Dice thường được sử dụng trong bài toán phân đoạn (segmentation) để đo lường độ tương đồng giữa kết quả dự đoán và mẫu thực tế.

Hệ số Dice được tính bằng công thức :

Nếu mô hình dự đoán chính xác giống với thực tế, X Y sẽ lớn và |X| +|Y| sẽ nhỏ, dẫn đến một hệ số Dice gần bằng 1.

Nếu mô hình dự đoán không đúng với thực tế, X Y giảm và |X| + |Y| tăng, dẫn đến một hệ số Dice giảm về 0.

# 1.5 Kết quả đạt được

## 1.5.1 Tham số thực nghiệm

Đồ án sẽ tiến hành thực nghiệm trên tập dữ liệu như đã trình bày ở phần trước.

Hàm Loss của đồ án sẽ sử dụng hàm Binary Cross Entropy [11] with Logits Loss. Thuật toán tối ưu Đồ án sẽ sử dụng thuật toán Adam [12].

Đồ án sẽ sử dụng số Epochs là 10

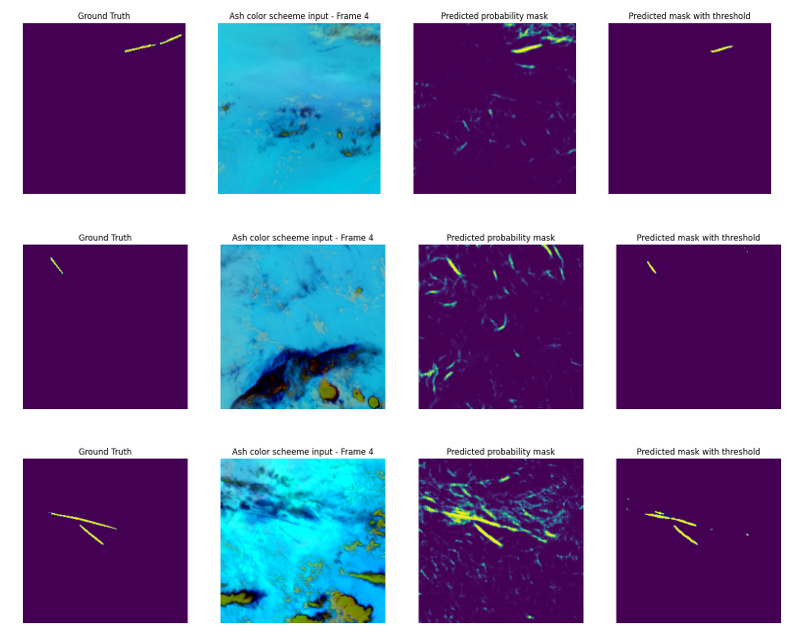
## 1.5.2 Kết quả đạt được

Bảng 2 so sánh kết quả thực nghiệm mô hình Đồ án đề xuất với một số phương pháp học sâu.

|  |  |
| --- | --- |
| **Phương pháp** | **Dice** |
| UNet | 0.405 |
| RestNet 50 | 0.101 |
| VGG19 | 0.003 |

**Bảng 2:** So sánh kết quả thực nghiệm

Như đã thấy, mô hình Unet có độ chính xác tốt hơn nhiều so với các phương pháp học sâu khác. Do đó, chúng tôi tiến hành dự đoan kết quả một số mẫu trên mô hình này.



**Hình 7:** Kết quả dự đoán

U-Net, ResNet50 và VGG19 đều là các mô hình mạng nơ-ron tích chập (CNN) phổ biến được sử dụng trong nhiều tác vụ liên quan đến hình ảnh. Tuy nhiên, U-Net thường cho kết quả tốt hơn so với ResNet50 và VGG19 trong các tác vụ phân đoạn hình ảnh, như phân đoạn đường hơi máy bay.

* **Kiến trúc đặc biệt của U-Net**: U-Net có một kiến trúc đặc biệt được thiết kế cho việc phân đoạn hình ảnh. Nó bao gồm hai phần: bộ mã hóa và bộ giải mã. Bộ mã hóa giúp nắm bắt ngữ cảnh trong hình ảnh, trong khi bộ giải mã giúp định vị chính xác đối tượng trong hình ảnh[1](https://medium.com/@resham.sundar/vgg16-vs-resnet50-when-used-as-encoder-in-unet-e93c95ee14a2).
* **Kết nối trực tiếp**: U-Net sử dụng các kết nối trực tiếp từ các lớp của bộ mã hóa đến các lớp tương ứng của bộ giải mã. Điều này giúp giữ lại thông tin chi tiết về vị trí, điều mà các mô hình như ResNet50 và VGG19 có thể mất trong quá trình gộp tối đa (max pooling) và chập (convolution).

# 1.6 Kết luận

## 1.6.1 Kết quả đạt được

Về mặt lý thuyết, Đồ án đã thực hiện phương pháp với 3 mô hình cho ra kết quả có chỉ số không mấy vượt trội nhưng cũng đủ để cho ra dự đoán tương đối chính xác.

Về mặt thực nghiệm, Đồ án đã sử dụng bộ dữ liệu GOES-16 cho 3 mô hình khác để so sánh. Kết quả thực nghiệm cho thấy mô hình Unet đã mang lại kết quả tương đối tốt, gần giống với các vệt mà các nhà nghiên cứu đã vẽ.

## 1.6.2 Hạn chế

Nhìn chung, Unet vẫn còn vài nhược điểm dẫn đến độ chính xác không cao. Một vài lý do trong đó có thể kể đến như kích thước đầu vào cố định và không chú ý đến mối quan hệ không gian. Unet thường không giữ thông tin về mối quan hệ không gian giữa các pixel, điều này có thể làm mất thông tin về cấu trúc không gian của đối tượng trong hình ảnh.

## 1.6.3 Hướng phát triển

Hầu hết mô hình UNet được thiết kế đặc biệt cho các tác vụ phân đoạn hình ảnh, nơi mục tiêu là phân loại từng pixel trong hình ảnh và gán chúng vào các lớp tương ứng. Trong tương lai, Đồ án dự kiến sẽ tiếp tục nghiên cứu hướng phát triển chính như sau:

* Nghiên cứu thêm về các biến thể của UNet như UNet++, Nested UNet, và UNet 3D, với mục tiêu cải thiện hiệu suất và khả năng tổng hợp thông tin.
* Sử dụng cơ chế chú ý để tăng cường khả năng mô hình chú ý vào các vùng quan trọng trong hình ảnh, giúp cải thiện hiệu suất đặc biệt khi có sự biến động trong dữ liệu.

# TÀI LIỆU THAM KHẢO

1. Wang, Z. (2023). Combining UPerNet and ConvNeXt for Contrails Identification to reduce Global Warming. *arXiv preprint arXiv:2310.04808*.
2. Qiu, Y. (2023, April). Upernet-Based Deep Learning Method For The Segmentation Of Gastrointestinal Tract Images. In *Proceedings of the 2023 8th International Conference on Multimedia and Image Processing* (pp. 34-39).
3. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16133-16142).
4. Jure, P. R., Masuelli, S., & Cabral, J. (2022, September). A labeled dataset of cloud types using data from GOES-16 and CloudSat. In *2022 IEEE Biennial Congress of Argentina (ARGENCON)* (pp. 1-6). IEEE.
5. Soulami, K. B., Kaabouch, N., Saidi, M. N., & Tamtaoui, A. (2021). Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. *Biomedical Signal Processing and Control*, *66*, 102481.
6. Wen, L., Li, X., Li, X., & Gao, L. (2019, May). A new transfer learning based on VGG-19 network for fault diagnosis. In *2019 IEEE 23rd international conference on computer supported cooperative work in design (CSCWD)* (pp. 205-209). IEEE.
7. Mukti, I. Z., & Biswas, D. (2019, December). Transfer learning based plant diseases detection using ResNet50. In *2019 4th International conference on electrical information and communication technology (EICT)* (pp. 1-6). IEEE.
8. Python: <https://python.org/>
9. Pytorch: <https://pytorch.org/>
10. Shamir, R. R., Duchin, Y., Kim, J., Sapiro, G., & Harel, N. (2019). Continuous dice coefficient: a method for evaluating probabilistic segmentations. *arXiv preprint arXiv:1906.11031*.
11. Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng*, *9*(10).
12. Llugsi, R., El Yacoubi, S., Fontaine, A., & Lupera, P. (2021, October). Comparison between Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito. In *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM)* (pp. 1-6). IEEE.

# LÀM VIỆC NHÓM

Cách thức làm việc nhóm: gặp nhau trực tiếp để trao đổi, tìm kiếm tài liệu hoặc họp online qua discord, các buổi thực hành trên lớp.

Phân chia công việc của các thành viên trong nhóm:

|  |  |  |
| --- | --- | --- |
| Thành viên | Nhiệm vụ | Mức độ hoàn thành |
| Ngô Hoàng Nhật Huy | 1.1 Giới thiệu về bài toán  1.2 Phân tích yêu cầu của bài toán  1.4 Thực nghiệm  1.5 Kết quả đạt được  1.6 Kết luận | 100% |
| Đoàn Ánh Dương | 1.3 Phương pháp giải quyết bài toán  1.4 Thực nghiệm  1.5 Kết quả đạt được | 100% |

Tổng số lần gặp nhau:

Trực tiếp: 6 lần

Online: 12 lần

Tổng thời gian gặp nhau: Khoảng 52 giờ

# TỰ ĐÁNH GIÁ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Chương | Nội dung | Điểm chuẩn | Tự chấm | Ghi chú |
| **1** | **Giới thiệu về bài toán** | 0.5 | 0.5 |  |
| **2** | **Phân tích yêu cầu của bài toán** | 1.5 | 1.5 |  |
| **3** | **Phương pháp giải quyết bài toán** | 2.5 | 2 |  |
| **4** | **Thực nghiệm** | 4.5 | 4 |  |
| **5** | **Kết luận** | 0.5 | 0.5 |  |
| **Nhóm** | **Điểm nhóm** | 0.5 | 0.5 |  |
| **Tổng điểm** | 9 | | |  |