

Báo cáo Ariel Data Challenge 2025

*Note: Sub-titles are not captured in Xplore and should not be used

Đỗ Đức Thắng

VNU-UNET

FIT

HaNoi, VietNam

23020158@vnu.edu.vn

Tóm tắt nội dung—Việc đặc trưng hóa bầu khí quyển của các hành tinh ngoài hệ Mặt Trời thông qua phân tích phổ là một bài toán phức tạp. Cuộc thi Ariel Data Challenge 2025 do NeurIPS tổ chức phối hợp với Cơ quan Vũ trụ châu Âu (ESA) đã tạo điều kiện cho cộng đồng nghiên cứu áp dụng các kỹ thuật học máy nhằm trích xuất thành phần khí quyển từ dữ liệu phổ mô phỏng. Trong báo cáo này, tôi tập trung vào việc xây dựng và tối ưu hóa mô hình dự đoán phổ quá cảnh từ các tín hiệu đo đạc đã qua tiền xử lý nhằm loại bỏ nhiễu kết hợp với các đặc trưng của hành tinh và ngôi sao chủ.

Index Terms—học máy, học sâu

I. GIỚI THIỆU

Đặc trưng hóa bầu khí quyển các hành tinh ngoài hệ Mặt Trời là một trong những bài toán trọng yếu của thiên văn học hiện đại. Việc trích xuất chính xác các thành phần hóa học từ phổ quá cảnh (transit spectrum) quan sát được đóng vai trò then chốt để hiểu về tiến hóa hành tinh, thành phần khí quyển cũng như khả năng tồn tại sự sống. Cuộc thi **Ariel Data Challenge 2025**, thuộc chuỗi đấu trường NeurIPS và được tổ chức trên nền tảng Kaggle dưới sự hợp tác với ESA, cung cấp một sân chơi độc đáo để thử nghiệm các phương pháp học máy xử lý tín hiệu phổ mô phỏng với độ nhiễu cao và nhiều hiệu ứng vật lý giống thiết bị thực tế.

Trong cuộc thi, tôi phát triển một mô hình dự đoán phổ dựa trên các trụ cột chính:

- **Tiền xử lý dữ liệu (Signal pre-processing)**: Làm sạch dữ liệu, loại bỏ nhiễu do thiết bị đo đạc và tính toán độ sâu phổ trắng (White light transit depth)
- **Dự đoán phổ quá cảnh bằng Residual Network (ResNet)**: Mô hình sử dụng dữ liệu đầu vào gồm độ sâu phổ trắng và các đặc trưng của hành tinh để dự đoán phổ quá cảnh.
- **Ước lượng độ không chắc chắn của dự đoán**

II. XỬ LÝ DỮ LIỆU

2.1 Dữ liệu

- Đối với mỗi hành tinh, dữ liệu được cho gồm có:
 - **Tín hiệu FGS1**:(FGS1_signal_0.parquet) với kích thước [135000, 32 × 32] (135000 bước thời gian, mỗi bước tương ứng với 0,1 giây và 32 × 32 là dữ liệu từ cảm biến)

– Tín hiệu AIRS-CH0:

(AIRS-CH0_signal_0.parquet) với kích thước [11250, 32 × 356] (11250 bước thời gian, mỗi bước tương ứng với 1,2 giây và 32 × 356 là dữ liệu từ cảm biến, gồm 32 chiều không gian và 356 bước sóng khác nhau)

- Bộ dữ liệu hiệu chuẩn gồm các thành phần: ảnh tối dùng để loại bỏ nhiễu nhiệt và độ lệch cảm biến; ảnh xác định điểm ảnh chết hoặc nóng để loại bỏ các điểm không hoạt động chính xác; ảnh phẳng dùng để hiệu chỉnh sự không đồng đều về độ nhạy giữa các điểm ảnh; thông tin hiệu chỉnh phi tuyến giúp khôi phục tín hiệu chính xác khi cảm biến gần bão hòa; và ảnh nhiễu đọc ghi lại nhiễu điện tử phát sinh trong quá trình đọc dữ liệu từ cảm biến
- Các đặc trưng của hành tinh gồm mã định danh, bán kính ngôi sao, khối lượng ngôi sao, nhiệt độ ngôi sao, khối lượng hành tinh, độ lệch tâm quỹ đạo, chu kỳ quỹ đạo, bán kính trục lớn, độ nghiêng quỹ đạo.

2.2 Tiền xử lý

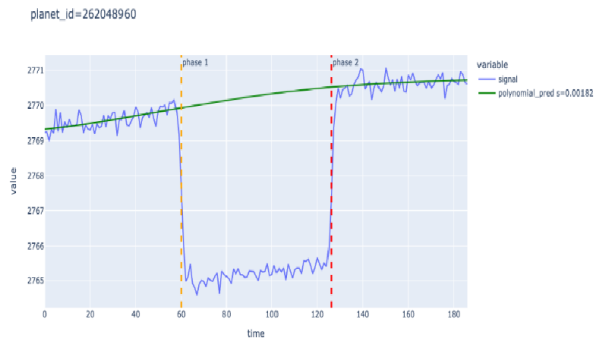
- Mô hình áp dụng quy trình tiền xử lý dữ liệu theo các bước:
 - **ADC (Analog-to-Digital Conversion)**: chuyển đổi tín hiệu thô sang giá trị số (đơn vị electrons) bằng hệ số gain và offset
 - **Loại bỏ dòng tối (Dark current subtraction)**: xác định và loại bỏ pixel lỗi (hot/dead) để loại bỏ tín hiệu nền không mong muốn từ cảm biến.
 - **Hiệu chỉnh phi tuyến (Linearity correction)**: hiệu chỉnh phi tuyến sử dụng polynomial model từ dữ liệu
 - **Loại bỏ dark-current (Dark-Frame Subtraction)**: Trừ ảnh tối đã được hiệu chỉnh (loại bỏ pixel chết) khỏi ảnh gốc theo thời gian phơi sáng để loại bỏ tín hiệu nền không mong muốn từ cảm biến.
 - **Lấy ảnh hiệu chỉnh đôi (Correlated Double Sampling - CDS)**: Tính hiệu giữa ảnh cuối và ảnh đầu mỗi chu kỳ phơi sáng để loại bỏ nhiễu và ổn định tín hiệu.
 - **Gộp theo thời gian (Time Binning)**: Gộp nhiều khung hình theo thời gian để giảm dung lượng và làm mượt chuỗi thời gian quan sát.
 - **Hiệu chỉnh trường phẳng (Flat Field Correction)**: Chia mỗi pixel cho giá trị tương ứng trong ảnh trường

Identify applicable funding agency here. If none, delete this.

phẳng để hiệu chỉnh sự không đồng đều trong độ
nhảy của từng pixel

- Loại bỏ các bước sóng ở ngoài rìa và tại các thời
điểm đo đầu và cuối để tránh nhiễu, chỉ giữ lại các
bước sóng ở trung tâm

2.3 Dự đoán *white light transit depth*



Hình 1. Mô tả hình ảnh.

—