**CT\_UAV\_Data\_AI\_Project: Bùi Đình Bảo – 0975 742 607 – dinhbao.itech@gmail.com**

***Tổng quan***

Dựa theo bài báo [Few-shot object detection on aerial imagery via deep metric learning and knowledge inheritance - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1569843223002212), ta có thể phân loại các phương pháp FSOD như sau:  
Few-shot learning Object Detection  
- Meta learning  
 - Single Branch  
 - Dual Branch  
- Transfer learning

Theo bài báo [A comprehensive review of few-shot object detection on aerial imagery - ScienceDirect](https://www.sciencedirect.com/science/article/abs/pii/S157401372500036X#preview-section-introduction), ta thấy có rất nhiều phương pháp few-shot để có thể phát hiện đối tượng trong ảnh vệ tinh. Cụ thể, tác giả đã liệt kê ra 55 phương pháp hiện đại trải dài từ meta learning đến transfer learning, đồng thời là 12 bộ dữ liệu tiêu chuẩn để đánh giá và 3 bộ dữ liệu viễn thám phổ biến nhất bao gồm: DIOR, NWPU VHR-10, and DOTA.

Trong giới hạn việc truy cập vào các paper, sau đây là 3 phương pháp few-shot khả thi được sử dụng để phát hiện đối tượng trong ảnh viễn thám:

1. [Few-shot object detection on aerial imagery via deep metric learning and knowledge inheritance - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1569843223002212)
2. [[PDF] Few-shot Oriented Object Detection with Memorable Contrastive Learning in Remote Sensing Images | Semantic Scholar](https://www.semanticscholar.org/paper/Few-shot-Oriented-Object-Detection-with-Memorable-Zhou-Li/02b2861e56a241408a1e067695f643a64498de24)
3. [Multi-Modal Prototypes for Few-Shot Object Detection in Remote Sensing Images](https://www.mdpi.com/2072-4292/16/24/4693)

Benchmark (dataset đánh giá) chung: NWPU VHR-10

Cách tải Dataset: [TorchGeo: How to Download the NWPU VHR-10 Dataset | by Byeong-Hyeok Yu | Medium](https://medium.com/@byeonghyeokyu/torchgeo-how-to-download-the-nwpu-vhr-10-dataset-0cfc16b2f568" \t "_blank)

**GitHub Link: [fabyanbui/CT\_UAV\_Data\_AI\_Project](https://github.com/fabyanbui/CT_UAV_Data_AI_Project" \t "_blank)**

**Phương pháp 1: Few-shot object detection on aerial imagery via deep metric learning and knowledge inheritance**

**Cơ chế hoạt động**

Phương pháp được trình bày trong bài báo là một giải pháp **phát hiện vật thể ít mẫu (Few-shot Object Detection - FSOD)** dành cho ảnh hàng không. Nó giải quyết hai thách thức chính của các phương pháp học chuyển tiếp (transfer-learning): phân loại sai đối tượng mới và hiện tượng quên thảm khốc (catastrophic forgetting).

Cơ chế hoạt động của phương pháp này dựa trên hai module chính:

1. **Module mã hóa đa tương tự (Multi-Similarity - MS) và MS loss:**
   * Mục tiêu là học các biểu diễn đặc trưng (embedding) mạnh mẽ và có tính phân biệt cho các đề xuất vật thể (object proposals).
   * Module này được tích hợp vào mạng Faster R-CNN.
   * **MS loss** được thiết kế để kéo các embedding của các đề xuất thuộc cùng một lớp lại gần nhau và đẩy các embedding của các lớp khác nhau ra xa nhau trong không gian embedding.
   * Quá trình này tạo ra một ranh giới quyết định rõ ràng hơn, giúp cải thiện hiệu suất phát hiện các vật thể mới chưa từng thấy.
2. **Module kế thừa tri thức (Knowledge Inheritance) và Coherence loss:**
   * Mục tiêu là chống lại hiện tượng "quên thảm khốc", tức là sự suy giảm hiệu suất trên các lớp cơ sở (base classes) khi mô hình được tinh chỉnh (fine-tuned) trên các lớp mới.
   * Module này giữ lại bộ trích xuất đặc trưng của các lớp cơ sở (base head) và song song đó huấn luyện một bộ trích xuất mới cho các lớp mới (novel head).
   * **Coherence loss** được sử dụng để điều chỉnh quá trình học, đảm bảo rằng phân phối xác suất dự đoán của mô hình mới cho các lớp cơ sở vẫn phù hợp với mô hình ban đầu đã được huấn luyện trên dữ liệu dồi dào.

**Tính khả thi**

Phương pháp này có tính khả thi cao đối với bài kiểm tra về phát hiện vật thể trên ảnh UAV, đặc biệt khi yêu cầu dữ liệu ít mẫu, bởi vì:

* **Xử lý tốt dữ liệu ít mẫu:** Các mô hình học sâu truyền thống yêu cầu một lượng lớn dữ liệu được gán nhãn, việc này rất tốn thời gian và chi phí. Phương pháp FSOD này giải quyết trực tiếp vấn đề đó bằng cách chỉ cần một lượng mẫu có giới hạn cho các lớp mới. Điều này cực kỳ phù hợp với bài toán trên UAV, nơi việc thu thập và gán nhãn dữ liệu có thể gặp nhiều khó khăn.
* **Giải quyết vấn đề phân loại nhầm:** Trong ảnh hàng không, các vật thể thường phức tạp và khó phát hiện do độ phân giải và điều kiện ánh sáng. Bài báo chỉ ra rằng các phương pháp trước đây dễ phân loại nhầm các vật thể tương tự nhau. Việc sử dụng học metric sâu (deep metric learning) giúp mô hình học được các biểu diễn đặc trưng có tính phân biệt cao hơn, từ đó giảm thiểu lỗi phân loại.
* **Duy trì hiệu suất trên các lớp đã học:** Module kế thừa tri thức cho phép mô hình thích nghi với các đối tượng mới mà không làm giảm hiệu suất đã đạt được trên các lớp đối tượng cơ sở (như máy bay, tàu, xe cộ,...) đã được học trước đó. Điều này đảm bảo rằng mô hình có thể phát hiện cả các đối tượng quen thuộc lẫn các đối tượng mới phát sinh một cách hiệu quả.
* **Hiệu suất vượt trội:** Bài báo đã thực hiện thử nghiệm trên các bộ dữ liệu ảnh hàng không tiêu chuẩn (NWPU VHR-10 và DIOR) và cho thấy phương pháp đề xuất đạt hiệu suất tốt, thậm chí vượt trội so với các phương pháp FSOD tiên tiến khác, đặc biệt trong các thiết lập ít mẫu (3-shot, 5-shot, 10-shot). Điều này chứng minh rằng phương pháp này là một lựa chọn mạnh mẽ và đáng tin cậy.

**Phương pháp 2: Few-shot Oriented Object Detection with Memorable Contrastive Learning in Remote Sensing Images**

**Cơ chế hoạt động**

Phương pháp **Few-shot Oriented Object Detection with Memorable Contrastive Learning (FOMC)** được đề xuất để giải quyết hai thách thức chính trong việc phát hiện vật thể ít mẫu trên ảnh viễn thám:

1. **Hộp bao ngang (HBB):** Các hộp bao truyền thống (HBB) có thể bao gồm các vật thể liền kề và khu vực nền phức tạp, đặc biệt là trong ảnh chụp từ trên cao, nơi các đối tượng có thể xuất hiện với nhiều hướng khác nhau và được đóng gói dày đặc.
2. **Phân loại sai các đối tượng mới:** Việc thiếu dữ liệu huấn luyện cho các lớp mới có thể dẫn đến phân loại sai.

Cơ chế hoạt động của FOMC bao gồm hai thành phần chính để giải quyết các thách thức này:

* **Hộp bao định hướng (OBB):** Thay vì sử dụng HBB, FOMC sử dụng OBB để tạo ra các hộp bao chặt chẽ hơn và căn chỉnh tốt hơn với các đối tượng có hướng tùy ý, giúp giảm nhiễu nền.
* **Module học tương phản đáng nhớ (MCL):** Module này sử dụng **học tương phản có giám sát (supervised contrastive learning)** để học các đặc trưng có tính phân biệt cao. Nó duy trì một "ngân hàng bộ nhớ" (memory bank) cập nhật động để lưu trữ các đặc trưng đề xuất, cho phép sử dụng một số lượng lớn các mẫu âm (negative samples) từ nhiều mini-batch khác nhau. Điều này giúp mô hình học được ranh giới quyết định tốt hơn giữa các lớp và cải thiện khả năng phân loại các lớp mới chưa từng thấy.

**Tính khả thi**

Phương pháp này có tính khả thi cao đối với bài kiểm tra về phát hiện vật thể trên ảnh UAV, đặc biệt khi yêu cầu phát hiện các đối tượng có hướng cụ thể (như máy bay, tàu, xe cộ trong bãi đỗ...) và với số lượng dữ liệu hạn chế.

* **Phát hiện đối tượng có hướng:** Ảnh chụp từ UAV thường chứa các đối tượng có hướng tùy ý. Việc sử dụng OBB thay vì HBB giúp mô hình phát hiện chính xác hơn, giảm nhầm lẫn với các đối tượng liền kề hoặc nhiễu nền, một lợi thế quan trọng trong các tình huống thực tế.
* **Hiệu quả với dữ liệu ít mẫu:** Phương pháp FSOD này được thiết kế để hoạt động hiệu quả khi chỉ có một vài mẫu được gán nhãn cho các lớp mới. Điều này làm giảm đáng kể chi phí và thời gian thu thập, gán nhãn dữ liệu, vốn là một rào cản lớn đối với các mô hình học sâu truyền thống.
* **Hiệu suất vượt trội:** Bài báo đã thực hiện các thử nghiệm trên các bộ dữ liệu ảnh viễn thám tiêu chuẩn như DOTA và HRSC2016, và mô hình FOMC đã đạt được hiệu suất tốt nhất (state-of-the-art) trong nhiệm vụ phát hiện đối tượng định hướng ít mẫu. Các nghiên cứu chứng minh rằng phương pháp này không chỉ cải thiện hiệu suất trên các lớp mới mà còn duy trì hiệu suất tốt trên các lớp cơ sở đã được học trước đó.

**Phương pháp 3: Multi-Modal Prototypes for Few-Shot Object Detection in Remote Sensing Images**

**Cơ chế hoạt động**

Phương pháp được trình bày trong bài báo có tên là **Multi-Modal Prototypes for Few-Shot Object Detection (MP-FSDet)** ,. Mục tiêu của phương pháp này là tăng cường hiệu suất phát hiện đối tượng ít mẫu (Few-shot Object Detection - FSOD) bằng cách kết hợp ưu điểm của các nguyên mẫu (prototypes) hình ảnh và văn bản ,.

Cơ chế hoạt động của MP-FSDet dựa trên ba thành phần chính:

* **Nguyên mẫu đa phương thức (Multi-Modal Prototypes):** Thay vì sử dụng nguyên mẫu đơn lẻ từ hình ảnh hoặc văn bản, phương pháp này kết hợp cả hai ,.
  + **Nguyên mẫu văn bản (Textual Prototypes):** Được tạo ra từ tên lớp (class names) bằng cách sử dụng một bộ mã hóa văn bản đã được huấn luyện trước (pre-trained text encoder) như BERT , ,. Các nguyên mẫu này mang lại kiến thức ngữ nghĩa tổng quát và khả năng khái quát hóa mạnh mẽ ,.
  + **Nguyên mẫu hình ảnh (Visual Prototypes):** Được trích xuất từ các hình ảnh hỗ trợ (support images) với số lượng hạn chế. Các nguyên mẫu này cung cấp thông tin chi tiết về không gian của đối tượng trong ảnh viễn thám ,.
* **Module tổng hợp nguyên mẫu (Prototype Aggregating Module - PAM):** Module này có chức năng tích hợp kiến thức tổng quát từ nguyên mẫu văn bản với các chi tiết không gian từ nguyên mẫu hình ảnh để tạo ra một nguyên mẫu đa phương thức ,. Nguyên mẫu đa phương thức này sau đó được tích hợp vào bộ mã hóa và giải mã của mô hình Grounding DINO để phát hiện đối tượng.
* **Chiến lược huấn luyện hai giai đoạn hiệu quả (Efficient Two-Stage Training Strategy - ETS):** Phương pháp này đề xuất một chiến lược huấn luyện mới, có tính đến đặc điểm của bộ mã hóa văn bản đã được huấn luyện trước (pre-trained text encoder), để tránh làm giảm khả năng khái quát hóa của nó ,. Điều này giúp giảm rủi ro overfitting và rút ngắn thời gian huấn luyện tổng thể.

**Tính khả thi**

Phương pháp này có tính khả thi cao đối với đề bài kiểm tra AI UAV, nơi cần phát hiện các đối tượng trên ảnh chụp từ trên cao với dữ liệu hạn chế.

* **Tăng cường hiệu suất phát hiện với dữ liệu ít mẫu:** Bằng cách kết hợp các nguyên mẫu từ cả văn bản và hình ảnh, MP-FSDet tận dụng ưu điểm của cả hai để tạo ra các nguyên mẫu có khả năng phân biệt tốt hơn. Điều này giúp cải thiện hiệu suất phát hiện các lớp đối tượng mới chỉ với một lượng nhỏ dữ liệu được gán nhãn, một thách thức lớn trong ảnh UAV.
* **Hiệu quả cao trên ảnh viễn thám:** Bài báo đã chứng minh rằng phương pháp này hoạt động hiệu quả trên các bộ dữ liệu viễn thám tiêu chuẩn như DIOR và NWPU VHR-10.v2 ,. Đặc biệt, nó đã vượt trội hơn các phương pháp tiên tiến khác với mức tăng hiệu suất lên tới 8.7% trong nhiệm vụ FSOD trên ảnh viễn thám.
* **Khả năng thích ứng:** Phương pháp này được xây dựng trên một bộ phát hiện tiên tiến (Grounding DINO), cho phép nó thích ứng tốt với các nhiệm vụ phát hiện đối tượng. Bằng cách học cách kết hợp các đặc trưng tổng quát (từ văn bản) và chi tiết (từ hình ảnh), mô hình có thể phát hiện các đối tượng một cách mạnh mẽ hơn.

***Cách chạy code***

* Chạy file setup.ipynb để cài đặt các thư viện cần thiết
* Chạy lần lượt các file demo ở mỗi thư mục Method
  + Method1/demo.ipynb
  + Method2/demo.ipynb
  + Method3/demo.ipynb

ENGLISH VERSION:

**Overview**

Based on the article **"Few-shot object detection on aerial imagery via deep metric learning and knowledge inheritance"** from ScienceDirect, few-shot object detection (FSOD) methods can be classified as follows:

**Few-shot learning Object Detection**

* Meta learning
  + Single Branch
  + Dual Branch
* Transfer learning

According to the paper **"A comprehensive review of few-shot object detection on aerial imagery"** also from ScienceDirect, there are numerous few-shot methods for detecting objects in satellite imagery. Specifically, the authors listed 55 state-of-the-art methods spanning from meta-learning to transfer learning, along with 12 standard evaluation datasets and the three most popular remote sensing datasets: **DIOR**, **NWPU VHR-10**, and **DOTA**.

Given the limited access to papers, here are three viable few-shot methods for object detection in remote sensing images:

* **Few-shot object detection on aerial imagery via deep metric learning and knowledge inheritance** - ScienceDirect
* [PDF] **Few-shot Oriented Object Detection with Memorable Contrastive Learning in Remote Sensing Images** | Semantic Scholar
* **Multi-Modal Prototypes for Few-Shot Object Detection in Remote Sensing Images**

**Common Benchmark (Evaluation Dataset):** NWPU VHR-10 **How to Download the Dataset:** TorchGeo: How to Download the NWPU VHR-10 Dataset | by Byeong-Hyeok Yu | Medium **GitHub Link:** fabyanbui/CT\_UAV\_Data\_AI\_Project

**Method 1: Few-shot object detection on aerial imagery via deep metric learning and knowledge inheritance**

**How It Works**

The method presented in this paper is a **Few-shot Object Detection (FSOD)** solution specifically designed for aerial imagery. It addresses two main challenges of transfer-learning methods: the misclassification of new objects and **catastrophic forgetting**.

The mechanism of this method is based on two core modules:

* **Multi-Similarity (MS) Module and MS Loss:** The goal is to learn robust and discriminative feature embeddings for object proposals. This module is integrated into a Faster R-CNN network. **MS loss** is designed to pull embeddings of proposals from the same class closer together while pushing embeddings of different classes farther apart in the embedding space. This process creates a clearer decision boundary, which helps improve the detection performance of new, unseen objects.
* **Knowledge Inheritance Module and Coherence Loss:** The objective is to combat "catastrophic forgetting," which is the performance degradation on **base classes** when the model is fine-tuned on **novel classes**. This module retains the feature extractor for the base classes (**base head**) and, in parallel, trains a new feature extractor for the novel classes (**novel head**). **Coherence loss** is used to regularize the learning process, ensuring that the new model's predicted probability distribution for the base classes remains consistent with the original model trained on abundant data.

**Feasibility**

This method is highly feasible for the UAV object detection task, especially when limited data is a requirement, because:

* **It handles few-shot data effectively:** Traditional deep learning models require a large amount of labeled data, which is time-consuming and expensive. This FSOD method directly solves this issue by only needing a limited number of samples for new classes. This is extremely suitable for UAV tasks where data collection and labeling can be challenging.
* **It addresses misclassification issues:** In aerial imagery, objects are often complex and difficult to detect due to resolution and lighting conditions. The paper points out that previous methods are prone to misclassifying similar objects. The use of deep metric learning helps the model learn more discriminative feature representations, thereby minimizing classification errors.
* **It maintains performance on learned classes:** The knowledge inheritance module allows the model to adapt to new objects without a drop in performance on the base object classes (like planes, ships, vehicles, etc.) that were learned beforehand. This ensures the model can effectively detect both familiar and newly emerging objects.
* **It provides superior performance:** The paper conducted experiments on standard aerial imagery datasets (NWPU VHR-10 and DIOR) and showed that the proposed method achieved excellent performance, even outperforming other advanced FSOD methods, especially in few-shot settings (3-shot, 5-shot, 10-shot). This proves the method is a strong and reliable choice.

**Method 2: Few-shot Oriented Object Detection with Memorable Contrastive Learning in Remote Sensing Images**

**How It Works**

The **Few-shot Oriented Object Detection with Memorable Contrastive Learning (FOMC)** method was proposed to address two key challenges in few-shot object detection in remote sensing images:

* **Horizontal Bounding Boxes (HBB):** Traditional HBBs can enclose adjacent objects and complex background areas, especially in aerial images where objects can appear in various orientations and are densely packed.
* **Misclassification of novel objects:** The lack of training data for new classes can lead to misclassification.

The FOMC mechanism includes two main components to tackle these challenges:

* **Oriented Bounding Boxes (OBB):** Instead of using HBBs, FOMC uses OBBs to create tighter bounding boxes that are better aligned with arbitrarily oriented objects, which helps reduce background noise.
* **Memorable Contrastive Learning (MCL) Module:** This module uses **supervised contrastive learning** to learn highly discriminative features. It maintains a dynamically updated **memory bank** to store proposal features, allowing a large number of negative samples from various mini-batches to be used. This helps the model learn a better decision boundary between classes and improves the classification capability for new, unseen classes.

**Feasibility**

This method is highly feasible for the UAV object detection task, especially when specific-oriented objects (such as planes, ships, or vehicles in a parking lot) need to be detected with limited data.

* **Oriented Object Detection:** Images captured from UAVs often contain objects with arbitrary orientations. Using OBBs instead of HBBs allows for more accurate detection, reducing confusion with adjacent objects or background noise. This is a crucial advantage in real-world scenarios.
* **Effectiveness with Few-Shot Data:** This FSOD method is designed to work efficiently with only a few labeled samples for new classes. This significantly reduces the cost and time of data collection and labeling, which is a major bottleneck for traditional deep learning models.
* **Superior Performance:** The paper conducted experiments on standard remote sensing datasets like DOTA and HRSC2016, and the FOMC model achieved **state-of-the-art** performance in the few-shot oriented object detection task. The studies show that this method not only improves performance on novel classes but also maintains good performance on the pre-learned base classes.

**Method 3: Multi-Modal Prototypes for Few-Shot Object Detection in Remote Sensing Images**

**How It Works**

The method presented in this paper is called **Multi-Modal Prototypes for Few-Shot Object Detection (MP-FSDet)**. Its goal is to enhance the performance of **Few-shot Object Detection (FSOD)** by combining the advantages of both visual and textual prototypes.

The mechanism of MP-FSDet is based on three main components:

* **Multi-Modal Prototypes:** Instead of using a single prototype from either images or text, this method combines both.
  + **Textual Prototypes:** These are generated from **class names** using a pre-trained text encoder like BERT. These prototypes provide general semantic knowledge and strong generalization capabilities.
  + **Visual Prototypes:** These are extracted from a limited number of **support images**. These prototypes provide detailed spatial information about objects in the remote sensing images.
* **Prototype Aggregating Module (PAM):** This module's function is to integrate the general knowledge from the textual prototypes with the spatial details from the visual prototypes to create a multi-modal prototype. This multi-modal prototype is then integrated into the encoder-decoder of the Grounding DINO model for object detection.
* **Efficient Two-Stage Training Strategy (ETS):** This method proposes a new training strategy that considers the characteristics of the pre-trained text encoder to avoid degrading its generalization ability. This helps reduce the risk of overfitting and shortens the overall training time.

**Feasibility**

This method is highly feasible for the UAV AI test, where detecting objects in aerial images with limited data is required.

* **Enhanced Performance with Few-Shot Data:** By combining prototypes from both text and images, MP-FSDet leverages the strengths of both to create more discriminative prototypes. This helps improve the detection performance for new object classes with only a small amount of labeled data, a major challenge in UAV imagery.
* **High Effectiveness on Remote Sensing Imagery:** The paper demonstrates that this method works effectively on standard remote sensing datasets like DIOR and NWPU VHR-10.v2. Specifically, it has outperformed other state-of-the-art methods with a performance increase of up to 8.7% in the FSOD task on remote sensing images.
* **Adaptability:** This method is built on an advanced detector (Grounding DINO), allowing it to adapt well to object detection tasks. By learning to combine general features (from text) and detailed features (from images), the model can detect objects more robustly.

**How to run the code:**

1. Run the **setup.ipynb** file to install the necessary libraries.
2. Run the **demo.ipynb** file sequentially in each method's directory:

Method1/demo.ipynb  
Method2/demo.ipynb  
Method3/demo.ipynb