ITI107 Assignment Report

# Data Collection and Annotation

The data collection process involved utilizing a combination of manual efforts and publicly available datasets. Specifically:

1. **Public Dataset:** Images of Pokémon were sourced from the Kaggle dataset titled [Pokémon Image Dataset](https://www.kaggle.com/datasets/hlrhegemony/pokemon-image-dataset).
2. **Manual Collection:** Images of Pokéballs were manually gathered from various online resources.

After collecting the images, annotation and labelling were completed entirely using the online tool **Roboflow**. Each image was meticulously labelled to indicate whether it contained a Pokémon or a Pokéball. A total of **2564 images** were annotated for this project, ensuring the dataset was well-prepared for training purposes.

# Model Training

The model training was conducted using the **YOLOv5** repository. However, several challenges and limitations influenced the training process:

1. **Training Configuration:**
   * **Batch Size:** The training utilized a batch size of 8.
   * **Epochs:** The model was trained for 5 epochs to balance between computational constraints and performance.
2. **Time Constraints:**
   * Training one epoch with a batch size of 16 took approximately 1.5 hours due to hardware limitations.
   * Consequently, the batch size was reduced, and the total epochs were capped at 5 to manage time effectively.
3. **Weights & Biases (wandb):**
   * While the initial training sequence included wandb for logging, errors with the wandb API caused retry loops at the end of every epoch, significantly increasing the training time. As a result, wandb was excluded from subsequent training runs.



1. **Fine-Tuning:**
   * Due to time constraints, fine-tuning the model was not performed. Efforts were focused on completing initial training within the available timeframe.

The training artifacts, including code, runs, weights, and validation results, have been included in the submission. They can be found under the code/YOLO/yolov5/runs folder.

# Results and Evaluation

The trained model’s performance was evaluated based on **mean average precision (mAP)** and qualitative observations:

* **Detection Capabilities:** The model successfully detected Pokémon and Pokéballs in test images and videos with reasonable accuracy, given the limited training epochs and lack of fine-tuning.
* **Inference Speed:** The exported model demonstrated efficient inference on a selected test image and a 1-minute video.

# Deployment

The final model was deployed on **Hugging Face Space** to demonstrate its functionality. The deployment involved:

1. Creating a Gradio-based demo application in the Hugging Face repository.
2. Exporting the trained model into a format suitable for real-time inference.
3. Hosting the demo app at the following link: [Hugging Face Space Deployment](https://huggingface.co/spaces/ITI107-2024S2/9442307k).

The application allows users to upload test images or videos and visualize the detection results.

# Recommendations for Improvement

Given the challenges faced during this assignment, several recommendations for future work include:

1. **Fine-Tuning:** Allocate more time for fine-tuning hyperparameters and conducting additional training epochs to improve model performance.
2. **Automated Logging:** Resolve wandb issues to enable seamless logging of experiments and facilitate better performance tracking.
3. **Augmentation:** Employ advanced data augmentation techniques to enhance model generalization capabilities.
4. **Hardware Utilization:** Leverage more powerful hardware or cloud-based GPUs to accelerate training and enable experimentation with larger batch sizes and epochs.

By addressing these areas, the model’s accuracy and robustness can be further enhanced.

This concludes the report for the ITI107 assignment. All required files and outputs have been included in the submission zip file for review.