Name: Umang Shikarvar

Roll no: 23110301

Problem Statement:

This task focuses on improving the generalisation of an object detection model across different geographic regions. Models trained on brick kiln images from one region (e.g., West Bengal) experience performance drops when deployed in another region (e.g., Bihar) due to domain shifts, such as variations in environmental conditions, image quality, and kiln appearances.

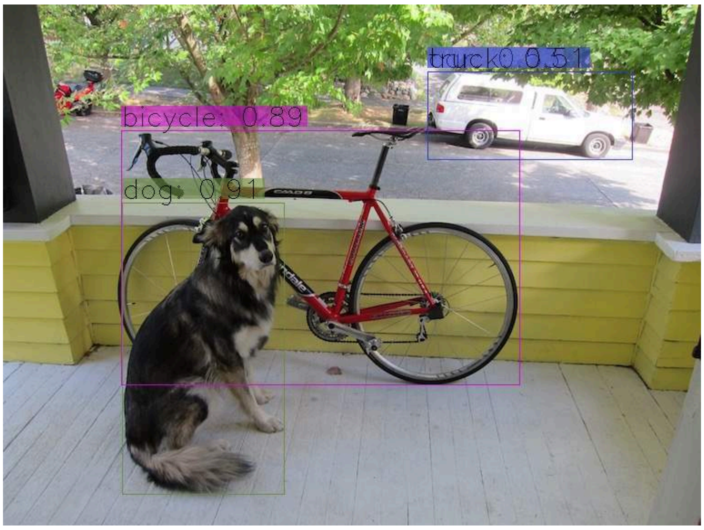
Proposed Solution:

Using unpaired image-to-image translation models, such as CycleGAN and CUT, to generate synthetic images that visually resemble the target domain while preserving the structural integrity of objects in the source domain. These adapted images, retaining the original annotations, will be used to train a YOLO-based object detection model, enabling it to generalize better across different geographic regions and mitigate domain shifts.

Related Work and Background Study:

1. YOLO (You Only Look Once)

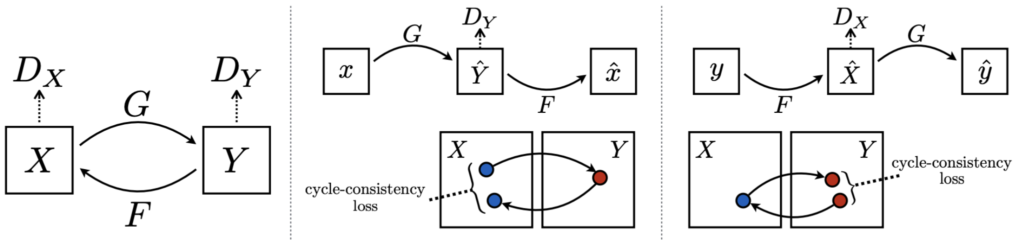
- YOLO is a real-time object detection model that provides fast and accurate predictions by treating detection as a single-stage problem.
- It predicts bounding boxes and class probabilities directly from an image using a convolutional neural network (CNN).



2. CycleGAN (Cycle-Consistent Generative Adversarial Networks)

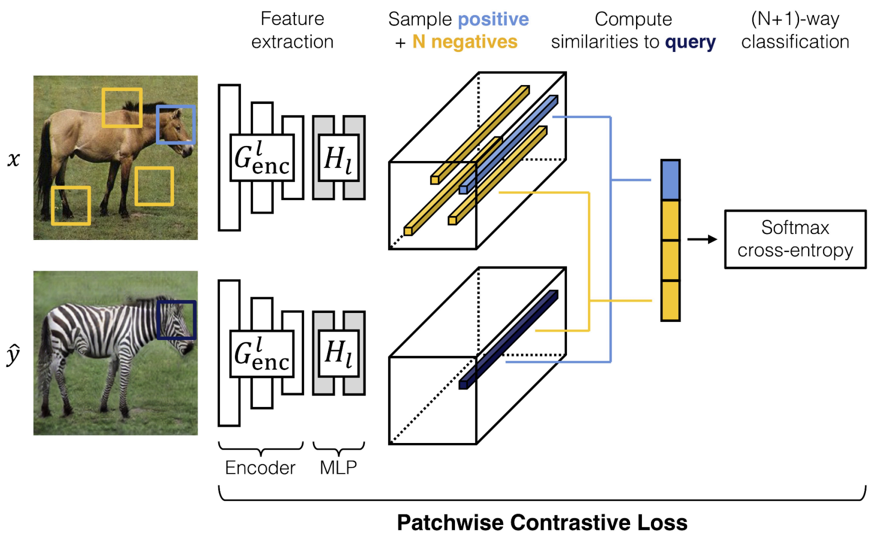
- CycleGAN is a deep learning-based generative model that can transform images from one domain to another without paired data.

- It uses two generators to convert images from source to target domain and back, along with two discriminators to distinguish real from generated images. A cycle consistency loss ensures that transforming an image X → Y → X and Y → X → Y results in minimal change, preserving structural details while adapting visual characteristics.



3. **CUT (Contrastive Learning for Unpaired Image-to-Image Translation)**

- CUT is an alternative to CycleGAN that performs unpaired image-to-image translation using contrastive learning.
- Unlike CycleGAN, CUT focuses on key regions in an image and learns feature correspondences in a more data-efficient manner. Instead of enforcing cycle consistency, CUT uses contrastive learning to maximize mutual information between corresponding patches in the source and target images. This method enables high-quality domain adaptation while using only a single generator and one-sided translation.



Preliminary Work:

https://github.com/Umang-Shikarvar/Domain-Adaption/blob/main/Results.ipynb