Unsupervised Domain Adaptation for Wildlife Image Enhancement

K. Hassan Qazi, Aidan Tsang

Affiliations: University of Toronto

Motivation

- **The Problem:** Automated wildlife monitoring relies on deep learning object detectors. However, performance degrades significantly on images captured at night (Near-Infrared/NIR) or in adverse weather.
- **The Constraint:** Supervised learning is impossible because collecting paired training data (the exact same animal, in the exact same pose, in both day and night conditions) is infeasible in the wild.
- **The Goal:** To develop an unsupervised Image-to-Image translation pipeline that converts low-quality NIR images into clear, RGB daytime equivalents.
- **Significance:** Successful translation makes "unreadable" images compatible with existing pre-trained wildlife detectors without the cost of re-labeling data.

Related Work

- **Cycle-Consistent GANs (CycleGAN):** The foundational baseline for unpaired translation. Uses a cycle-consistency loss to preserve structure without paired data.
- **Latent Space Models (MUNIT):** Disentangles "content" and "style" features to allow for **multimodal** outputs, modeling diverse lighting conditions (e.g., sunny vs. cloudy).
- **Diffusion Models (Instruct-Pix2Pix):** State-of-the-art text-guided editing. Offers high texture fidelity but is computationally expensive and prone to hallucinations.
- **The Gap:** Most methods optimize for **human perceptual quality**. Our approach uses a **Task-Aware Loss** to optimize specifically for **machine readability** (detection accuracy).

References

- [1] Zhu, J. Y., et al. (2017). *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. ICCV.
- [2] Huang, X., et al. (2018). *Multimodal Unsupervised Image-to-Image Translation (MUNIT)*. ECCV.
- [3] Li, B., et al. (2022). *ITTR: Unpaired Image-to-Image Translation with Transformers.* ArXiv.
- [4] Brooks, T., et al. (2023). *Instruct-Pix2Pix: Learning to Follow Image Editing Instructions*. CVPR.
- [5] MegaDetector (2023). Microsoft AI for Earth. *GitHub*.

New Technique

We propose a refined CycleGAN architecture that integrates a Task-Aware Perceptual Loss to preserve features specifically required by wildlife object detectors.

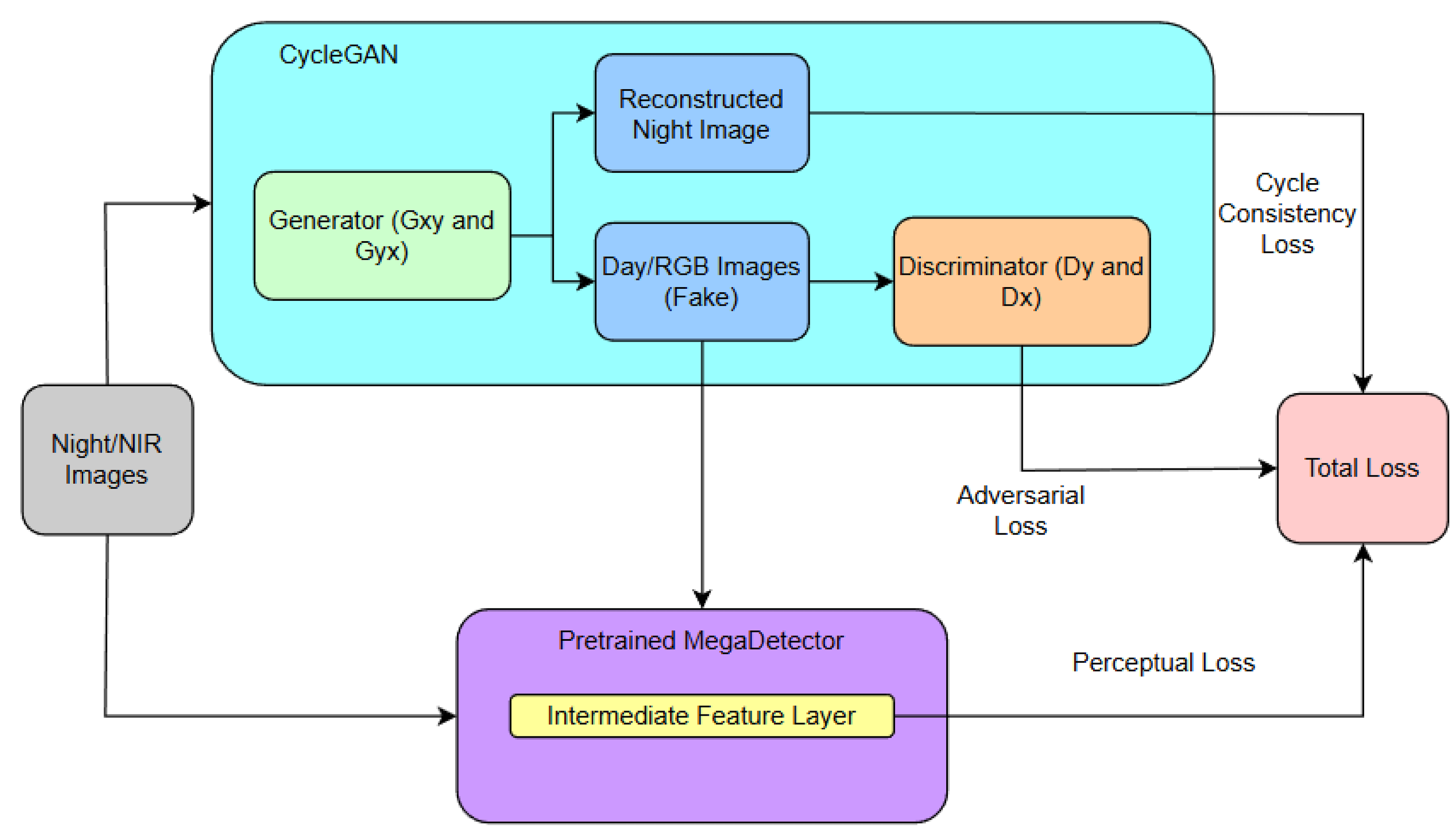
Architecture & Implementation:

- Generator: UNetGenerator with skip connections to preserve spatial information during the Night -> Day translation.
- Discriminator: PatchGAN operating on 70x70 patches to ensure high-frequency texture realism.
- Feature Extractor: We utilize MegaDetector v5a (YOLOv5) as a frozen feature extractor. We hook into the intermediate backbone (Layer 10) to extract semantic features relevant to animal detection.

We minimize a composite loss function comprising Adversarial, Cycle-Consistency, and our novel Perceptual Loss

Future Steps:

Implement Perceptual Loss in SOTA diffusion models



Experimental Results

Dataset:

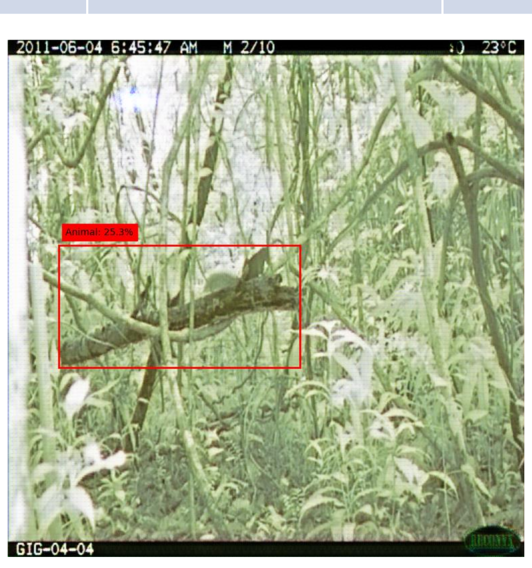
- **Source:** Caltech Camera Traps (CCT) - Missouri Subset.
- **Splits:** Unpaired training sets for Domain A (Night/Greyscale) and Domain B (Day/Color).

Baselines Compared:

- **Original NIR:** Raw nighttime images (Lower Bound).
- **Standard CycleGAN:** Vanilla implementation without perceptual loss.
- **Instruct-Pix2Pix:** A state-of-the-art diffusion model prompted with "turn this into a daytime photo".

Image Type	Accuracy	Precision	Recall	F1 Score
Base NIR	0.922	0.978	0.917	0.946
CycleGAN Generated Daytime	0.891	1.000	0.854	0.921
CycleGAN Reconstructed NIR	0.875	0.976	0.854	0.911
Perceptual Generated Daytime (lambda = 5)	0.828	1.000	0.771	0.871
Perceptual Reconstructed NIR (lambda = 10	0.906	1.000	0.875	0.933

Base NIR



CycleGAN generated daytime with perceptual

CycleGAN reconstructed NIR with perceptual