Where Pixels Meet Perception: Understanding CycleGAN Translations Across Aesthetics

Sathammai Sathappan
Independent Research Project
sathammai2005@gmail.com

***Abstract*—This study explores unpaired image-to-image translation using CycleGAN across aesthetic domains of varying complexity, ranging from minimalism (Fashion MNIST) to maximalism (Bridal Couture). It investigates the impact of dataset diversity and visual quality on translational stability, realism, and style fidelity using both quantitative and qualitative evaluations.**

I. Introduction

CycleGAN has proven its effectiveness in unpaired image translations, but its strength strongly depends on dataset richness and clarity. Understanding this relationship is crucial for designing reliable translation systems, especially for aesthetic domains like fashion and art, where data imbalance and visual variance are frequently encountered challenges.

II. Datasets

- Fashion-MNIST: A standard low-resolution grayscale dataset. Sandals were chosen as domain-A and sneakers as domain-B.
- Bridal Couture: A custom high resolution dataset curated with Indian bridal attire as domain-A and Western bridal attire as domain-B. Preprocessing involved manual and automated scraping, background removal using *rembg*, and face anonymization using *retinaface*.

III. Methodology

A. Architecture

- Generators: ResNet-based architectures with two downsampling layers, multiple residual blocks and 2 upsampling layers.
- Discriminators: PatchGAN discriminators.

B. Loss Functions

Adversarial loss, cycle-consistency loss, and identity loss were used for training balance.

C. Metrics

Four categories of metrics were used: Pixel-based, Structural, Perceptual and Colour-based metrics. This helped asses reconstruction accuracy and perceptual quality.

IV. Results

A. Fashion MNIST

Despite producing strong metrics for identity mapping and cycle reconstruction, visual inspection revealed limited translations. Sandals were often converted into alternate sandal designs instead of being converted to sneakers. Sneakers were sometimes converted into unconventional sandals. In several cases, the images remained untranslated in both the translation directions.

B. Bridal Couture

Cycle and identity metrics slightly favoured the Western domain due to its simpler structure. Edge-based metrics indicated better performance in Western-to-Indian translations, consistent with the minimalism to maximalism direction. Qualitatively, both the translations functioned as near inverses of each other. Western-to-Indian introduced ornaments, color richness and embroidery, whereas Indian-to-Western suppressed all those details. Overall, translations maintained stylistic coherence and realism.

V. Discussion

Fashion MNIST's low-resolution nature limited the discriminator's learning capacity. Consequently, the generator, already constrained by cycle-consistency, exploited this weakness and produced minimal stylistic changes. Thus, the cycle-consistency constraint, originally implemented to enforce uniqueness, paradoxically becomes a limitation in such cases.

Conversely, the Bridal Couture dataset, being inherently rich and detailed, combined with the effective preprocessing, allowed CycleGAN to express its full theoretical potential. Despite minor segmentation noise from *rembg*, the model produced robust and visually rich translations, effectively resisting model collapse. Dataset richness and preprocessing rigor jointly ensured stability and stylistic authenticity.

VI. Conclusion

- Dataset richness and diversity have a significant impact on the depth of the style transfer.
- Robust preprocessing enhances CycleGAN's resilience to noise and model collapse.
- Perfect reconstruction metrics does not always correlate with high perceptual realism. Often, qualitative fidelity requires deviation from numerical perfection.