

Horse-to-Zebra Image Translation Using CycleGAN

Md Naim Hassan Saykat

Abstract

This project implements Cycle-Consistent Adversarial Networks (CycleGAN) in PyTorch for unpaired image-to-image translation between horses and zebras. The notebook shows qualitative results, configures training with adversarial and cycle-consistency losses, and specifies generator and discriminator architectures. The Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) are used to assess the outputs of learning both horse-to-zebra and zebra-to-horse mappings.

Keywords: CycleGAN, unpaired image translation, GANs, horse2zebra, computer vision.

1 Introduction

Image-to-image translation learns mappings between visual domains. Unlike supervised approaches such as pix2pix [2], CycleGAN [5] enables training without paired samples by enforcing cycle-consistency. In this project, we apply CycleGAN to the horse2zebra dataset to demonstrate unpaired domain translation.

2 Methodology

Generator: U-Net-like encoder-decoder with residual blocks.

Discriminator: PatchGAN classifier that distinguishes real vs. fake local patches.

Losses:

- Adversarial loss (MSE)
- Cycle-consistency loss (L1)

Optimization: Adam optimizer with learning rate 2×10^{-4} , $\beta_1 = 0.5$, $\beta_2 = 0.999$.

3 Dataset

The project uses the horse2zebra dataset provided by the CycleGAN authors [5]. It consists of unpaired training images from both domains. Images are resized to 256×256 , normalized to $[-1, 1]$, and randomly shuffled for training.

4 Experiments

- **Training:** 100 epochs, batch size 1.
- **Checkpoints:** Generators saved every 10 epochs.
- **Inference:** After training, the models are set to evaluation mode for generation.

4.1 Qualitative Results: Horse-to-Zebra

Figure 1 shows a grid of real horses and their translated zebras produced by the trained model.

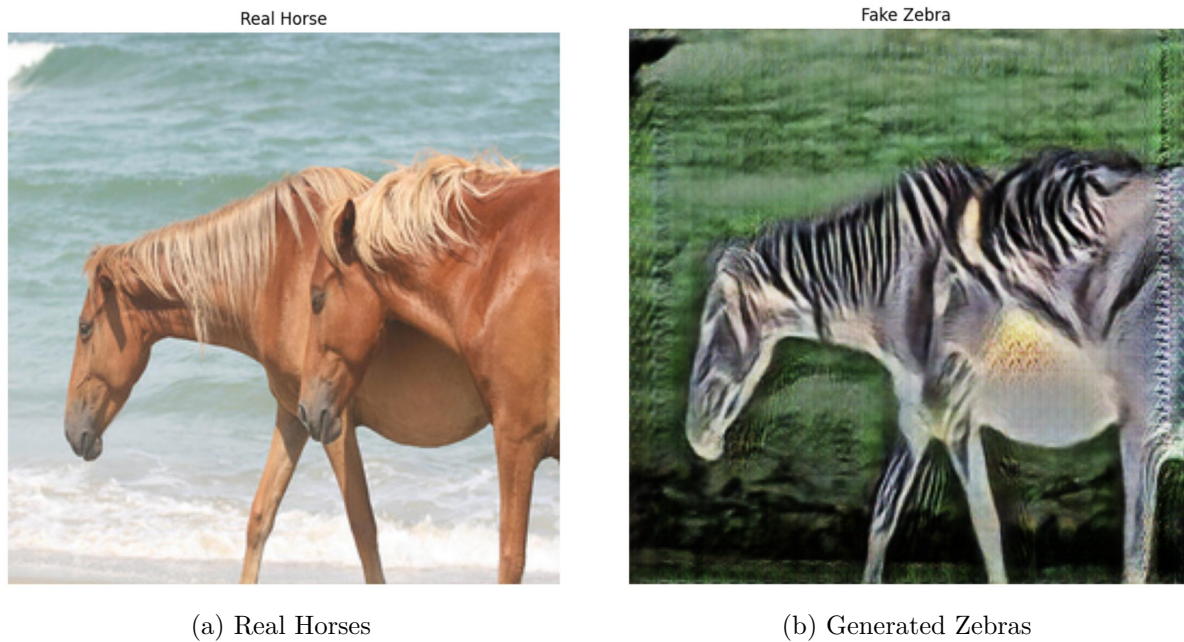


Figure 1: Qualitative example of horse-to-zebra translation using CycleGAN.

4.2 Qualitative Results: Zebra-to-Horse

Figure 2 presents real zebras and their translated horses.

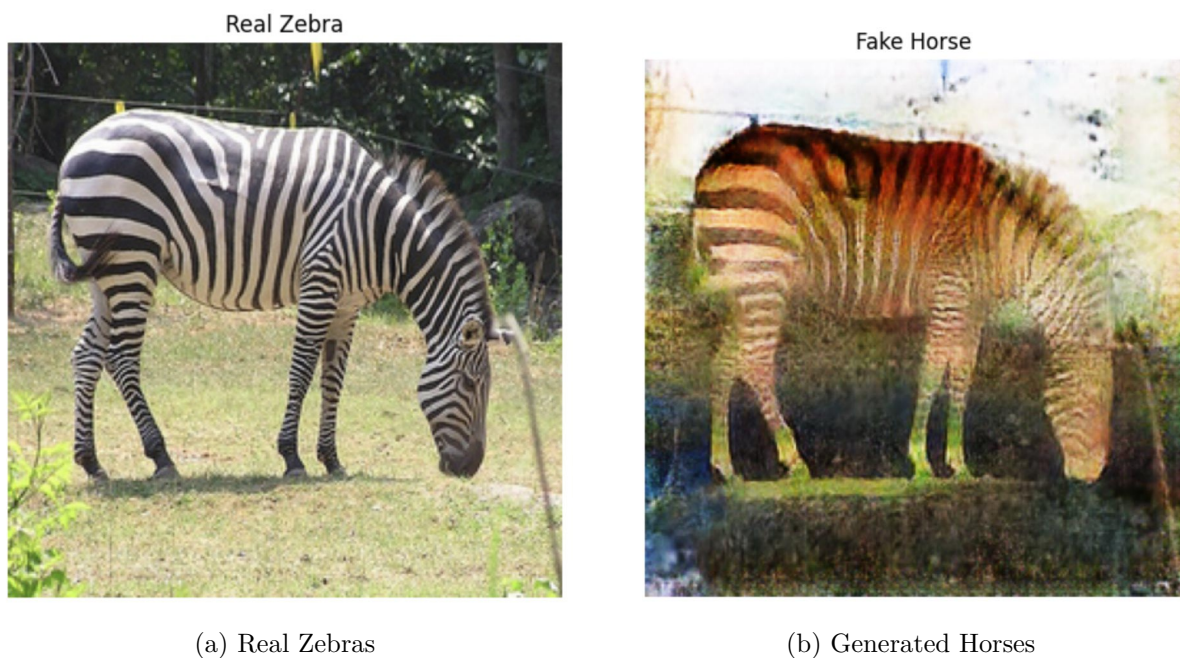


Figure 2: Qualitative example of zebra-to-horse translation using CycleGAN.

4.3 Quantitative Results

Average SSIM and PSNR were computed over 10 samples to provide an objective comparison of real vs. translated images. These pixel-level metrics offer a basic sanity check but are known to be limited for unpaired translation [3].

5 Results and Discussion

In both directions, the CycleGAN model exhibits believable style transfer. Although qualitative results demonstrate compelling stripe transfer, there are sporadic structural distortions and background artifacts. SSIM and PSNR scores confirm moderate similarity, consistent with known limitations of pixel-based evaluation in unpaired translation tasks.

6 Limitations and Future Work

- Background regions are sometimes incorrectly stylized.
- Fine details (e.g., legs, tails) can appear distorted.
- Future extensions may include perceptual metrics such as FID [1], LPIPS [4], and attention-based generators.

7 Conclusion

A working PyTorch implementation of CycleGAN for horse-zebra translation is presented in this notebook. Adversarial and cycle-consistency losses are used to learn both directions, and straightforward quantitative indicators and qualitative visualization are used to validate the findings. The research highlights opportunities for improvement while showcasing the efficacy of unpaired picture translation.

References

- [1] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 6626–6637, 2017.
- [2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.
- [3] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [4] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.
- [5] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.