Project Report - Comparative Analysis of CycleGANs and RevNets in Image-to-Image Translation

Yuval Uner

May 2024

1 Abstract

This report explores the performance of CycleGANs and RevNets in image-to-image translation. CycleGANs excel in unpaired image translation but involve complex architectures. RevNets, with their reversible architecture, may thus provide a viable alternative to CycleGAN's 2 generator architecture. We compare CycleGANs with various RevNet configurations. Our findings show that CycleGANs produce superior image translations, while RevNet-based models struggle with maintaining visual fidelity due to architectural constraints, indicating that while RevNets have theoretical advantages, they require further refinement to match CycleGAN performance.

2 Introduction

Image-to-image translation is crucial in computer vision, with applications like style transfer and medical imaging. CycleGANs enable translation without paired examples using adversarial learning and cycle consistency. Despite their effectiveness, they involve complex, resource-intensive architectures.
RevNets, as a bijective mapping, may offer an alternative by potentially simplifying the architecture needed for image translation tasks.

In this study, we perform a comparative analysis of CycleGANs and RevNets to evaluate their capabilities and limitations in image-to-image translation tasks. Specifically, we investigate whether a single reversible network can replace the dual-generator structure of CycleGANs,

3 Method

We extended the original CycleGAN code to include the following models:

3.1 Models

3.1.1 Pure RevNet GAN

The Pure RevNet GAN architecture consisted of a single generator and a single discriminator, and was trained only on the A to B direction. The generator was built using only RevNet layers. However, due to the constraint that RevNets must not perform downsampling to remain reversible, the number of channels remained at 3, matching the input image's channels. This restriction led to a very low parameter count and necessitated splitting the input along height or width dimensions instead of channels, as recommended by the original RevNet paper.

3.1.2 RevNet GAN

To address the limitations of the Pure RevNet GAN, we introduced an initial 1x1 convolution layer in the generator to increase the number of channels, followed by RevNet layers, and a final convolution layer to restore the channel count to 3. This modification allowed the input to be split by channels and to significantly increase the number of parameters (although to a number that is still significantly lower than that of a CycleGAN generator), while ideally keeping the network mostly reversible.

3.1.3 Bi-Directional RevNet GAN

This model extended the RevNetGAN by incorporating two discriminators, one for each translation direction (A to B and B to A). Additionally, each direction had separate initial and final convolution layers.

3.1.4 RevNet GAN and Bi-Directional RevNet GAN with Reconstruction Loss

Both the RevNetGAN and the Bi-Directional RevNetGAN were further augmented with reconstruction loss. In this setup, the generated output was fed back into the network in the reverse direction and compared with the original input.
This loss term was defined as $L_{\mathrm{rec}}(G^{-1}(G(A), A))$ and $L_{\mathrm{rec}}(G(G^{-1}(B), B))$.

3.2 Training

We used the horse2zebra dataset [3] for training and evaluation.
All training was done on a single P100 GPU over 70 epochs.
The learning rate was 0.0002 for the first 50 epochs and then linearly decreased over the final 20 epochs.
For comparison, a CycleGAN model was trained under the same conditions.
Each model required approximately 8 to 10 hours to train.

4 Results

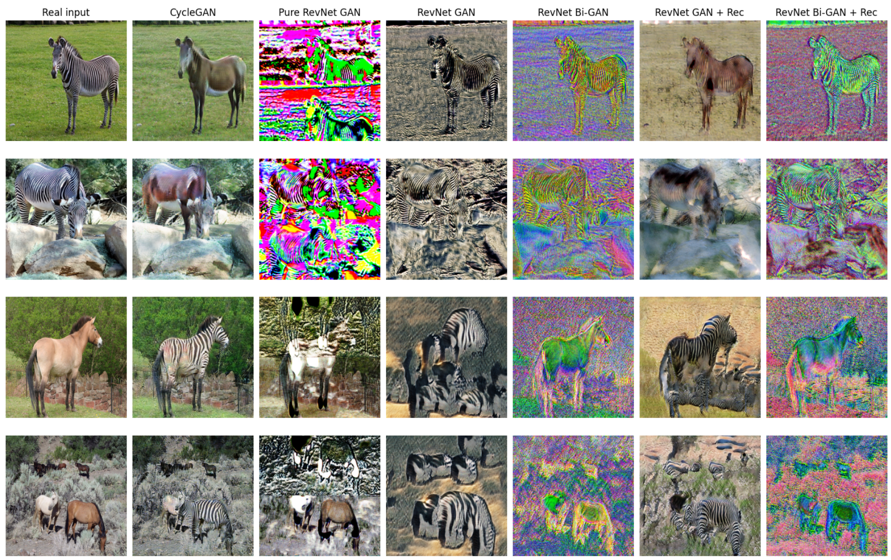

Figure 1: Side by side comparison of the various models' outputs

Figure 1 summarizes the experimental results, showing a side-by-side comparison of outputs from CycleGAN, Pure RevNet GAN, RevNetGAN, Bi-Directional RevNetGAN, RevNetGAN with reconstruction loss, and Bi-Directional RevNetGAN with reconstruction loss.
Compared to the baseline of CycleGANs, it is evident that the other models all performed poorly.

4.1 Pure RevNet GAN

Due to the architectural constraints discussed earlier, the Pure RevNet GAN significantly distorted the images. In the horse-to-zebra direction, it failed to achieve the desired translation, and in the reverse direction, it completely distorted the images. The limitations in parameter count and the inability to downsample or effectively manipulate channels are likely what led to these poor results.

4.2 RevNet GAN

With the addition of initial and final 1×1 convolution layers and an increased parameter count, the RevNetGAN model showed some improvement. It managed to capture the color scheme change in the horse-to-zebra direction but failed entirely in the reverse direction, only producing distorted images. This indicates that while increasing parameters helps, it is insufficient for achieving high-quality translations.

4.3 Bi-Directional RevNet GAN

The Bi-Directional RevNetGAN models, both with and without reconstruction loss, retained the outlines of the input images but only managed to distort the coloring. They failed to achieve the desired image-to-image translation. This could be due to the competing loss values from both directions, causing the model to be pulled in conflicting directions during training and preventing it from learning a coherent transformation.

4.4 RevNet GAN with Reconstruction Loss

The best performing model was the RevNetGAN with reconstruction loss. While it still caused distortions, it best managed to capture the shape and colors of zebras in the horse-to-zebra direction and approximated the color and shape of a horse in the reverse direction. The reconstruction loss helped maintain some fidelity to the original images, but the overall visual quality was still lacking compared to the CycleGAN.

4.5 Conclusions

The experiment's results demonstrate that RevNet-based models, constrained by the requirements for reversibility and low parameter counts, are not viable alternatives to the CycleGAN's dual-generator architecture for image-to-image translation tasks. The limitations inherent in the RevNet architecture, such as the inability to downsample and the need to maintain reversibility, severely restrict their performance.

Attempts to increase the number of RevNet blocks or the number of filters in the initial convolution would result in greatly increased training and inference times, far exceeding those of CycleGANs and beyond the scope of this experiment. Therefore, the conclusions from this experiment suggest that, given the current architectural constraints and performance outcomes, RevNets are not suitable replacements for the CycleGAN framework in bidirectional image-to-image translation tasks.

5 Code

All code can be found here: https://github.com/YuvalUner/pytorch-CycleGAN-and-pix2pix-fork-revnet.

References

[1] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

[2] Gomez, Aidan N., et al. "The reversible residual network: Backpropagation without storing activations." Advances in neural information processing systems 30 (2017).

[3] horse2zebra dataset