Neural Image Colorization with Adversarial Training

Tushar Verma

22 Nov 2024

1 Introduction

This project addresses automatic colorization of grayscale images using deep learning. A custom Generative Adversarial Network (GAN) framework was developed, where the generator is a U-Net architecture enhanced with self-attention modules, and the discriminator distinguishes real from generated color images. The goal is to reconstruct realistic colorizations without manual annotations, maintaining structural consistency and plausible color distributions.

2 Methodology

2.1 Dataset

A dataset of natural images was used, primarily sourced from the ImageNet Mini 1000 dataset. This dataset is a smaller subset of the original ImageNet, containing 1000 classes with a manageable number of images per class, making it suitable for faster experimentation. Input images were converted to grayscale, and the target outputs were the corresponding RGB images. Random affine transformations and partial occlusion (random pixel removal) were applied as data augmentation techniques to improve model robustness.

2.2 Model Architecture

- **Generator (U-Net with Attention):**
  - **Encoder:** Based on VGG-16 (BatchNorm) pretrained on ImageNet, adapted for single-channel grayscale input.
  - **Bottleneck:** Additional convolutional layers to enrich feature representations.
  - **Decoder:** U-Net-style decoder with:
    * Skip connections linking encoder and decoder stages
    * Transposed convolutions for upsampling
    * Self-Attention modules at multiple decoder levels for enhanced feature modeling
- **Discriminator:** A convolutional neural network that classifies real vs. generated colored images, encouraging realistic outputs through adversarial training.

2.3 Training Strategy

The model was trained in a GAN setup:

- **Generator Loss:** Combination of Perceptual Loss, L1 Loss, and weighted Adversarial Loss:

  $$\text{Generator Loss} = (\text{Smooth L1 Loss}) + (\text{Weighted Perceptual Loss}) + 0.01 \times (\text{Adversarial Loss})$$

  where:

  - *Smooth L1 Loss* minimizes direct pixel-level differences between generated and target images.

- *Perceptual Loss* compares high-level feature representations extracted from intermediate layers of a pretrained VGG19 network (layers: `relu1_2`, `relu2_2`, `relu3_3`, `relu4_3`). Deeper layers are assigned higher weights to emphasize semantic content.
- *Adversarial Loss* promotes realism by encouraging the generator to fool the discriminator.

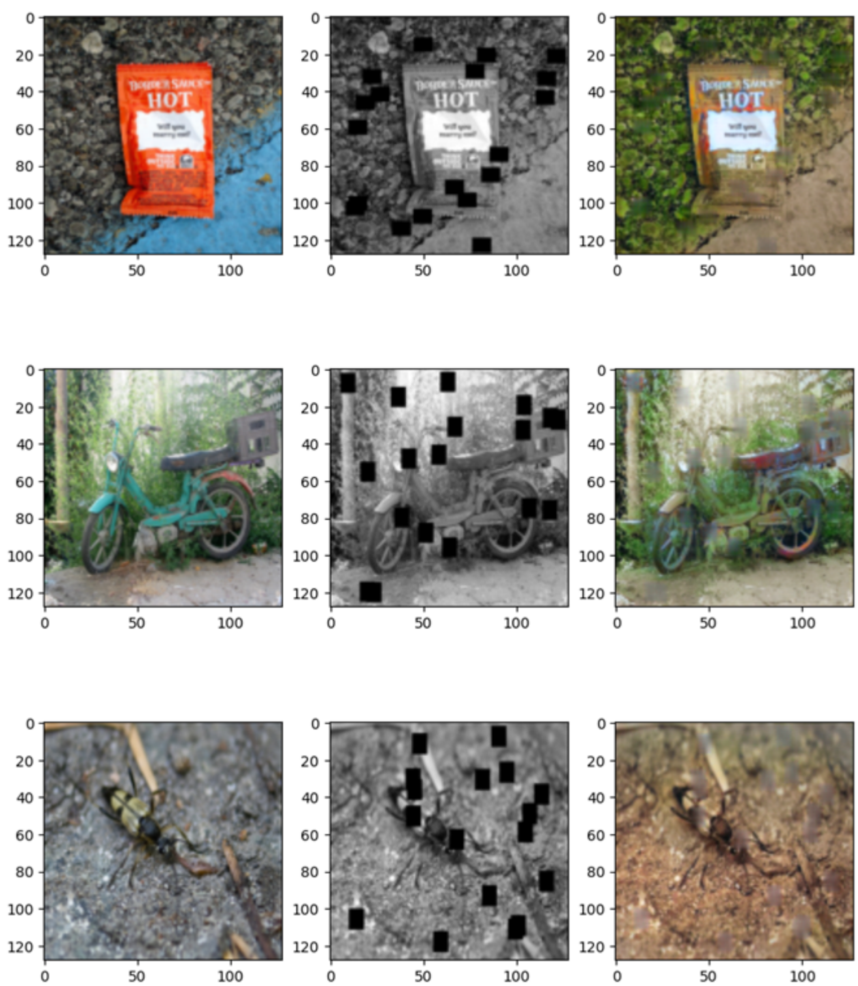
- **Discriminator Loss:** Binary cross-entropy loss for classifying real versus generated images.
- **Optimization:** Alternate optimization of generator and discriminator using Adam optimizer.

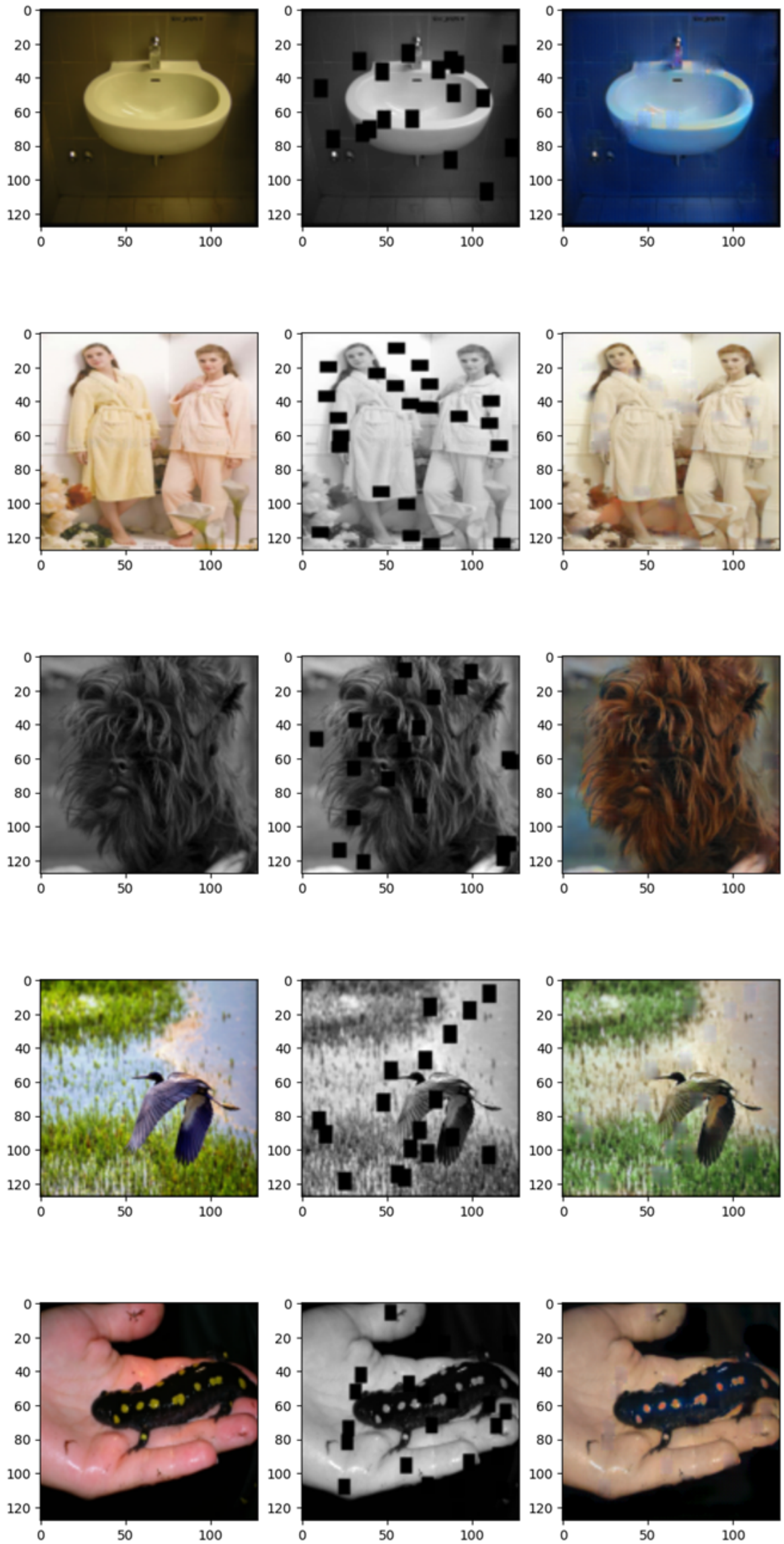
3 Results

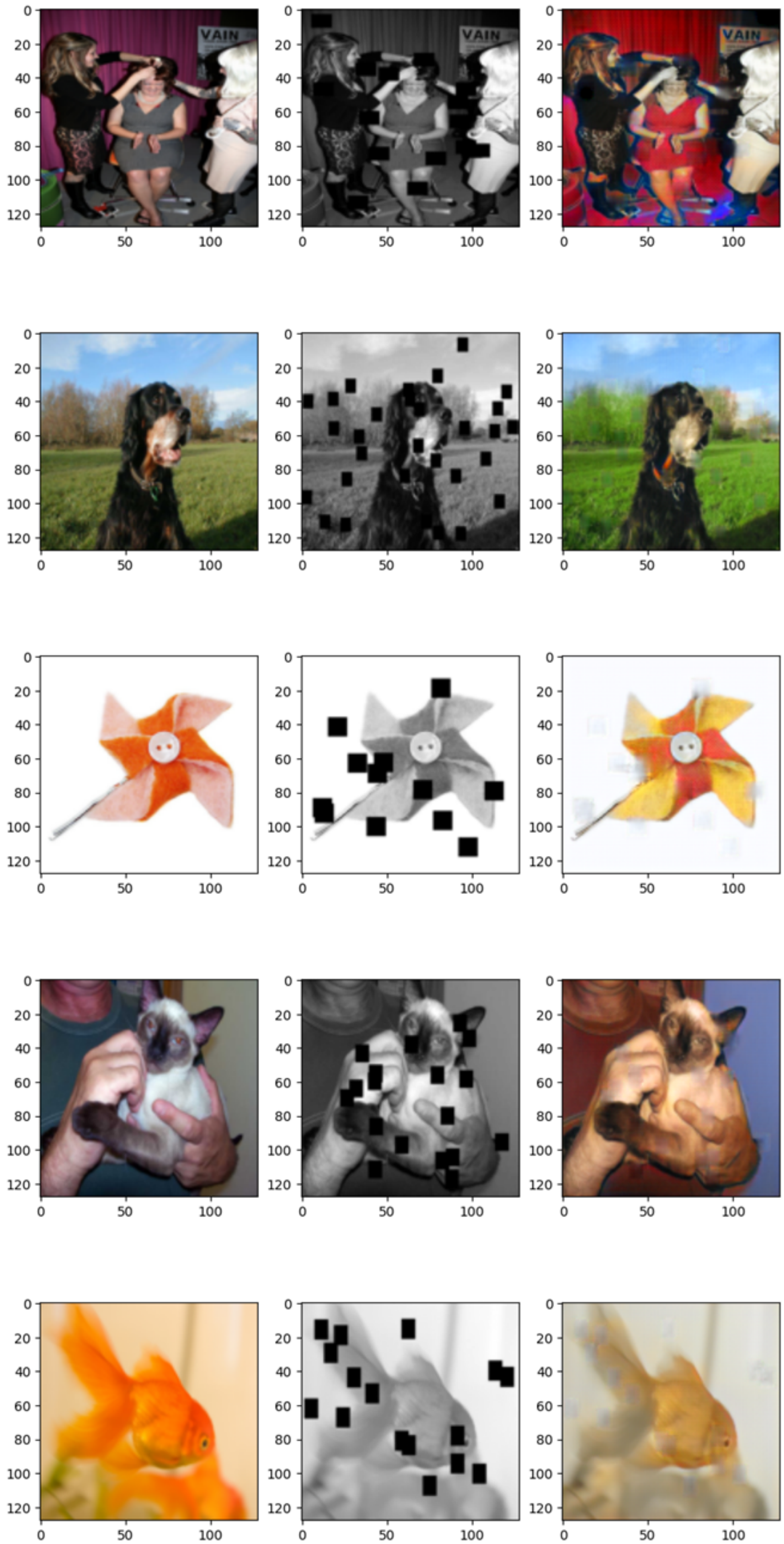
3.1 Visual Results

The generated colorized images maintained strong structural consistency and produced realistic colors for object-centric regions. Some minor artifacts appeared in ambiguous or heavily textured areas.

Some sample outputs of neural image colorization: **(Left)** Original, **(Middle)** Patchy Grayscale input, **(Right)** Colorized output.






VAIN
0
20
40
60
80
100
120
0
50
100

3.2 Quantitative Results

- Achieved an average **PSNR of approximately 25** on the test dataset.

3.3 Challenges

- Hallucination of unrealistic colors on ambiguous backgrounds.
- Loss of fine gradient transitions in highly detailed regions.

4 Conclusion

This project successfully demonstrated neural image colorization using a GAN framework with an attention-enhanced U-Net generator. The addition of self-attention improved spatial coherence, while adversarial training enhanced global realism. Future extensions could explore training in LAB color space, utilizing multi-scale discriminators, and improving fine detail reconstruction.

5 References

1. P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," CVPR 2017.
2. O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," MICCAI 2015.
3. A. Vaswani et al., "Attention Is All You Need," NeurIPS 2017.
4. J. Johnson, A. Alahi, L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," ECCV 2016.