CartoonGAN: Generative Adversarial Networks for Photo Cartoonization

Brown Data Science Institute

Shiyu Liu, Sibo Zhou, Junhui Huang

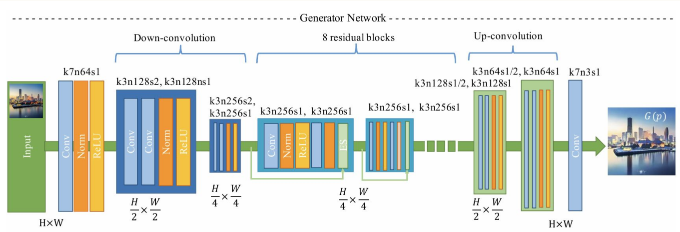
Introduction

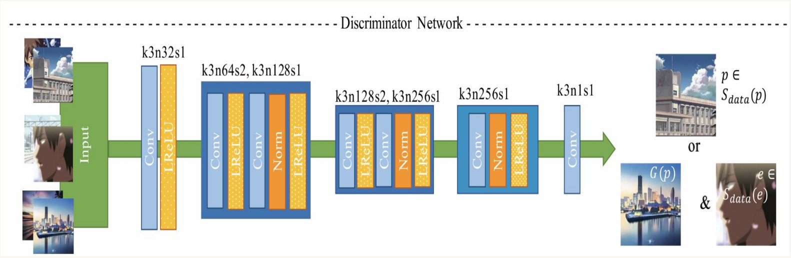
- **Goal**: Transform real-world photographs into stylized cartoon images using deep learning techniques.
- **Motivation**: Inspired by the rising popularity of cartoon-style outputs from generative models like ChatGPT.
- **Approach**:
  - Implement a custom image stylization model based on CartoonGAN, a Generative Adversarial Network architecture.
  - Learn the photo-to-cartoon transformation without requiring paired training data (unsupervised learning).
- **Challenge**: Cartoonization requires abstracting complex visual details into simplified shapes, colors, and edges — a task difficult to automate without compromising image content.
- **Contribution**:
  - Demonstrate a pipeline for unsupervised cartoon stylization.
  - Explore the trade-off between artistic abstraction and content preservation in deep generative models.

Dataset

- **Photo Dataset (COCO):** 4,000 real-world images from the COCO dataset, featuring diverse natural scenes, people, and objects.
  Used as the input domain to teach the generator realistic image structure and content.
- **Cartoon Dataset (Safebooru Anime):** 4,000 anime-style images from the Safebooru dataset on Kaggle, showcasing bright colors, sharp edges, and stylized character designs. Used as the target domain to guide cartoon stylization in an unsupervised setting.

Methods





Generator and Discriminator Networks

- **Content Loss**: Preserves structural features of the input photo using VGG-based feature maps.

$$\mathcal{L}_{\text{content}} = \|\phi(G(x)) - \phi(x)\|_1$$

- **Adversarial Loss:** Encourages the generator to produce images with sharp edges and stylized outlines.

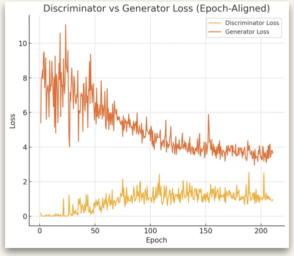
$$\mathcal{L}_{\text{adv}} = \log D(y) + \log(1 - D(G(x))) + \log(1 - D(y_s))$$

- **Total Loss:**

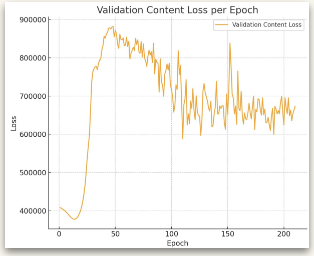
$$\mathcal{L}_G = \mathcal{L}_{\text{content}} + w\mathcal{L}_{\text{adv}}$$

Results

Sample after 210 training epochs (left:original photo, right: cartoonized images)



Discriminator vs Generator Loss



Validation Content Loss per Epoch

Discussion

- **Challenges Encountered:**
  - GAN instability and vanishing gradients during training.
  - Difficulty achieving sharp yet smooth edges in the cartoonized output.
  - Limited computational resources restricted the scope of hyperparameter tuning and experimentation.
- **Key Strategies:**
  - Careful tuning of loss weights to balance content preservation and stylization.
  - Iterative adjustment of hyperparameters to stabilize training and improve visual quality.
- **Outcomes:**
  - The final model successfully generates cartoonized images with clear stylistic abstraction.
  - Visual results validate the effectiveness of our training approach and architectural choices.
- **Future Work:**
  - Expand training datasets to improve generalization across photo styles.
  - Explore style control losses (e.g., stroke thickness, color palette) for finer control.
  - Investigate possibilities for real-time deployment on mobile or web platforms.

References

- Chen, Yang, Yu-Kun Lai, and Yong-Jin Liu. "CartoonGAN: Generative Adversarial Networks for Photo Cartoonization." CVPR 2018.
- https://www.kaggle.com/alamson/safebooru

Acknowledgements

We sincerely thank our TA, Dhruv Raghavan, for his mentorship throughout the semester, and Professor Eric Ewing for his invaluable guidance and support.