Image-to-Image Translation using Pix2pix

19EAC381 Machine Learning Lab with Python

Department of Electronics and Communication Engineering

November 26, 2023

Introduction

- To generate images through image-to-image translation.
- Generative adversarial networks (GANs) can be employed to synthesize realistic images from input semantic layouts.

Applications

- Translate images from one domain to another.
- Fill in missing or corrupted parts of an image restoring old paintings or images.
- Enhance the quality of low-resolution images.
- Generate realistic images from random noise, such as creating faces, animals or landscapes.

Problem Statement

To generate images from a semantic input map using a Generative Adverserial Network (GAN).

Literature Survey

- Use of Generative Adversarial Networks (GANs) to enhance image quality, stability, and diversity.[1][2][3]
- Conditional image synthesis, artistic style transfer, and unsupervised/self-supervised learning show versatility.[1][3]
- Evaluating the generated images remains challenging, with metrics like Inception Score and Freshet Inception Distance commonly used but acknowledged for their limitations.[2][3]

Literature Survey

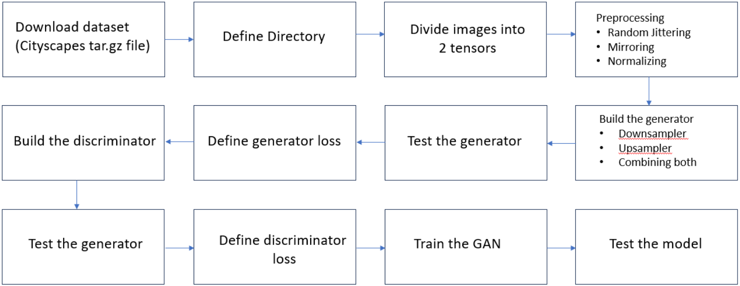
- Chong Bao et al. introduced NeRF coding for 2D and 3D image reconstruction, effectively addressing spatial ambiguity and limited multi-view supervision. Their approach achieved an accuracy of 89.5 percent for 2D and 83.3 percent for 3D, showcasing its effectiveness
- Yu Zeng et al. developed SceneComposer, a semantic image synthesis framework that offers a precision-driven synthesis of semantic layouts, allowing users to adjust precision levels and create diverse, customizable images.

Base Paper

Image-to-Image Translation with Conditional Adversarial Networks

Investigate conditional adversarial networks as a general-purpose solution to image-to-image translation problems.

Basic Flow Diagram



Implementation and Analysis

Generative Adversarial Networks (GANs)

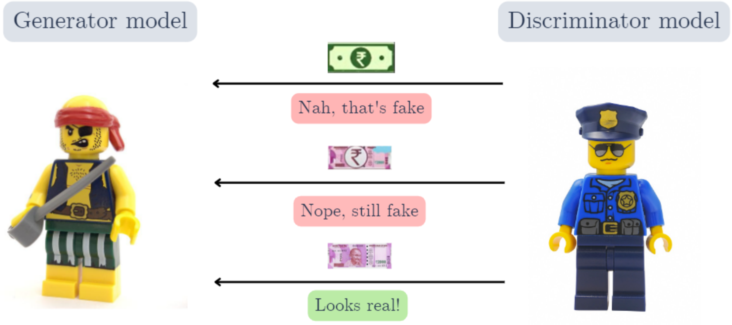
GANs generate a target image, conditional on a given input image.

pix2pix GAN

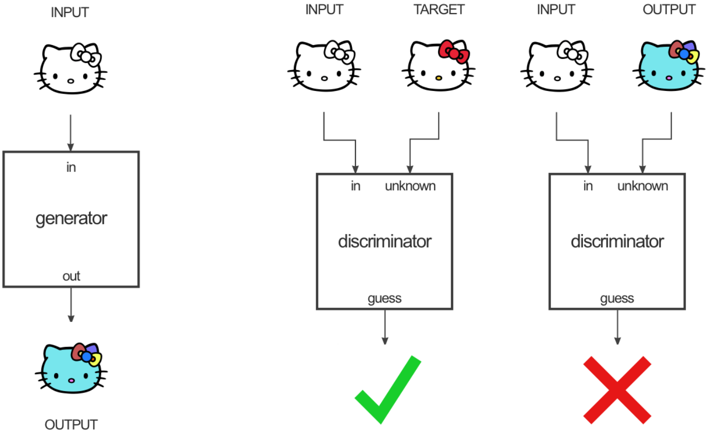
Generate images using Generator-Discriminator architecture.

Implementation and Analysis

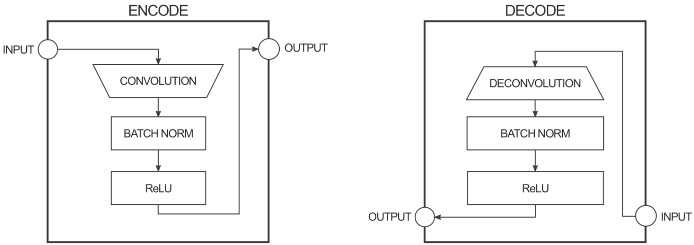
Analogy



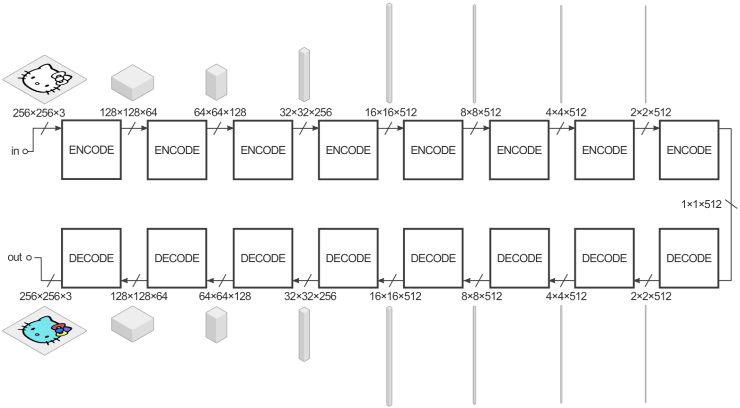
Generator and Discriminator



Generator and discriminator architecture

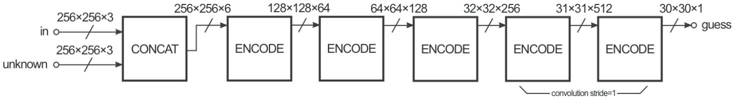


Generator architecture

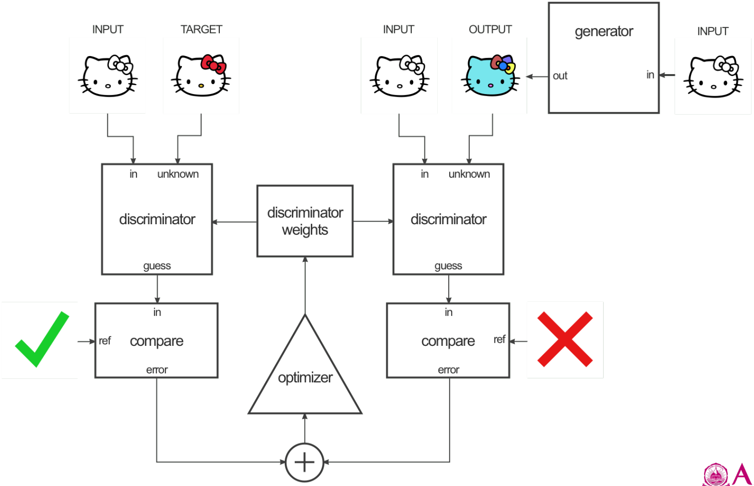


Implementation and Analysis

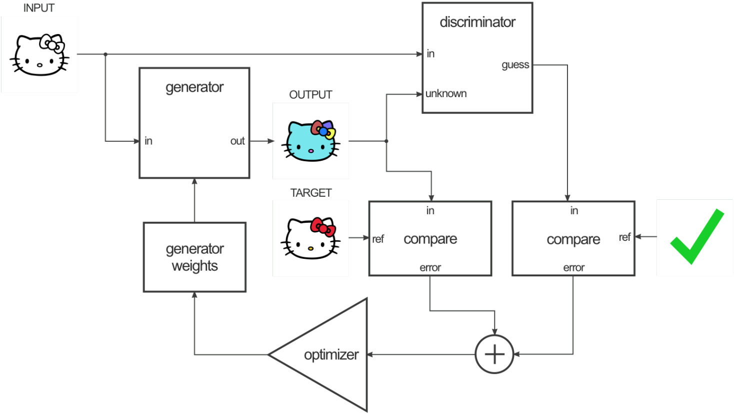
Discriminator architecture



Training discriminator



Training generator

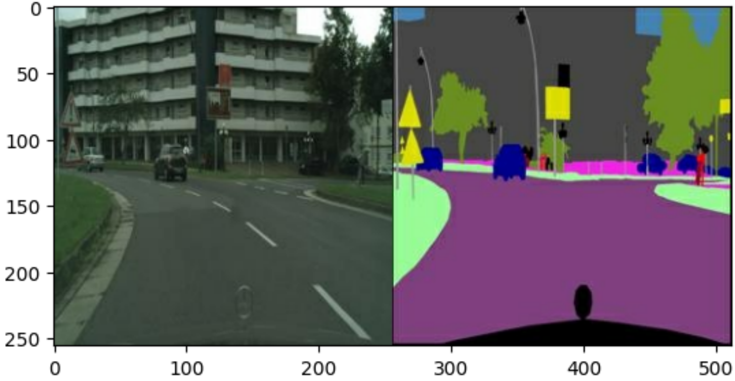


Implementation and Analysis

Generator and discriminator model implemented using 2 neural networks:

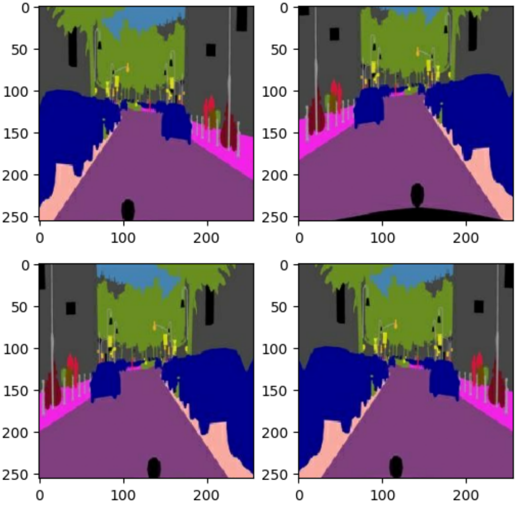
- Generator: U-Net
- Discriminator: PatchGAN

Initial image and segment map:



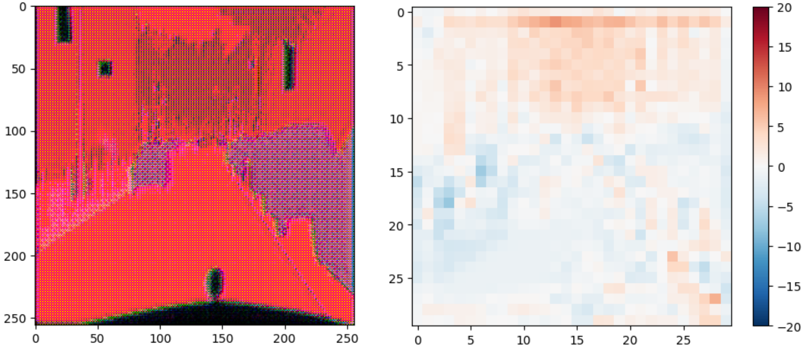
Implementation and Analysis

Post preprocessing:

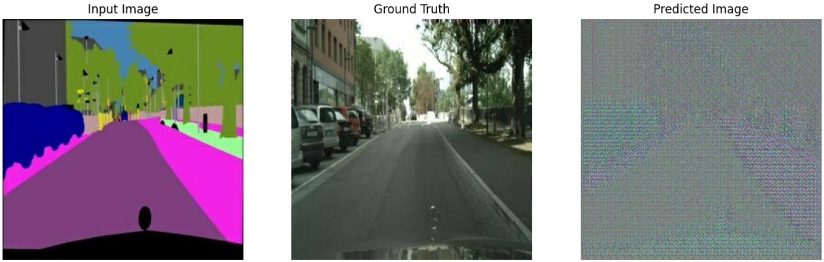


Implementation and Analysis

Generator and discriminator test output



Testing the generator function



Implementation and Analysis

Training the generator-discriminator model for 39k steps

Database details

Cityscapes

- This is a large-scale dataset that contains 5,000 high-quality images with fine annotations and 20,000 images with coarse annotations from 50 different cities.
- The images are captured in various weather conditions and seasons, and cover diverse urban scenarios.
- The dataset provides pixel-level labels for 30 classes, such as car, person, road, sky, etc. The dataset also provides instance-level labels for 8 classes, such as car, person, bicycle, etc.

Final Results and Comparison with Base Paper

Final Results and Comparison with Base Paper

Evaluation Metrics

- Average Mean Square Error: 0.360052078962232605
- Average Structural Similarity Index: 0.0017919301753863692
- Average Peak Signal to Noise Ratio: 10.456946840958546
- FID Score: 4.81634709357915e+106
- KID Score: 0.9804575443267882
- Mean Intersection over Union: 0.3335916286838868

Conclusion

- Image-to-image transformation using pix2pix on Cityscapes dataset showcases the effectiveness of GANs in urban scene translation.
- The model, trained on paired images, adeptly converts input images into realistic urban scenes, demonstrating its potential for applications in computer vision, urban planning, and autonomous systems.
- Further advancements in cGANs, such as pix2pix, are expected to refine and extend the capabilities of image transformation tasks.

Future scope

- Unsupervised Learning and Semi-Supervised Learning: Further exploration of GANs in unsupervised and semi-supervised learning scenarios could lead to breakthroughs in leveraging unlabelled data more effectively for training models.
- 3D and Volumetric Data Generation: Extending GANs to handle three-dimensional and volumetric data will open up new possibilities in applications such as medical imaging, computer-aided design, and virtual reality.
- Real-Time Applications: Improving the efficiency and speed of GANs will be essential for real-time applications, such as video generation, interactive content creation, and live simulations.

References

- 1. P. Isola, J. -Y. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.
- 2. T. Park, M. -Y. Liu, T. -C. Wang and J. -Y. Zhu, "Semantic Image Synthesis With Spatially-Adaptive Normalization," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 2332-2341, doi: 10.1109/CVPR.2019.00244.
- 3. S. Zhang et al., "Painting 3D Nature in 2D: View Synthesis of Natural Scenes from a Single Semantic Mask," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 8518-8528, doi: 10.1109/CVPR52729.2023.00823.a

AMRITA VISHWA VIDYAPEETHAM

References

- Y. Zeng et al., "SceneComposer: Any-Level Semantic Image Synthesis," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 22468-22478, doi: 10.1109/CVPR52729.2023.02152.
- C. Bao et al., "SINE: Semantic-driven Image-based NeRF Editing with Prior-guided Editing Field," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 20919-20929, doi: 10.1109/CVPR52729.2023.02004.