NORTHEASTERN UNIVERSITY

CS 7180: Advanced Perception (Fall 2023)

Image Enhancement

Submitted By:
Tarun Saxena (002979327)
Anson Antony (002756978)
Amal Mathew (002755808)

Date: *21 September 2023*

Abstract

As a broad framework for image-to-image translation challenges, we present conditional adversarial networks (cGANs). Based on learned mappings between input and output domains, a cGAN model is trained to translate an input image into a corresponding output image. A generator network is used to generate output images, while a discriminator is used to discriminate between generated and real images. Experiments show that pixel labeling, image colorization, photo production, and map-to-aerial photo translation perform better than baselines. The results reveal that conditional adversarial learning is extremely effective for a wide range of image-to-image translation tasks, making it a versatile solution for cross-domain picture mapping.

1. Introduction and prior work

Image-to-image translation is an active research problem in computer vision that involves learning mappings between different image representations. This project is inspired by the pix2pix paper by Isola et al. [1], which presents an effective general-purpose approach for image-to-image translation using conditional adversarial networks (cGANs). The key insight of this work is training a cGAN model that learns to translate an input image from a source domain to a corresponding output image in a target domain. This allows the model to transform sensory data from the input representation to the output representation.

Artistic style transfer, which renders one image in the manner of another image, is another pertinent topic. Neural style transfer, which uses a trained network to match the Gram matrix statistics between input and output images, is one of the notable methodologies. These techniques, however, have prioritized texture transfer above complete translations.

The authors use generative adversarial networks (GANs), which are composed of a generator network that creates fake images and a discriminator network that can tell fake images from real ones. In a minimax game, the two networks are trained jointly until the artificial images cannot be distinguished from genuine ones. Conditional GANs discover a mapping from input picture y to output image x from observed image x and random noise z. Class labels are just one of the many inputs that can condition the mapping.

The fundamental finding of the authors is that conditional GANs are capable of learning mappings across two visual domains, which allows them to convert sensory data from one domain to another. By extending conditional GANs by conditioning the mapping on input images rather than class labels, they have developed an approach they call pix2pix. This enables general-purpose picture-to-image translations by converting an input image into a corresponding output image. Any image from the source domain can then be mapped to the target domain using the trained model.

To develop the image translation system, we draw on additional deep learning and GANs resources in addition to the seminal pix2pix paper. currently published papers that offer examples of pix2pix implementations [2] and GAN tutorial articles that describe the training process [3]. In order to rebuild and improve the picture translation approach employing cGANs for a new domain application, the project directly draws on these earlier studies. Our contributions include improving the training approach, modifying the framework for translating X-domain photos to Y-domain images and assessing a unique dataset.

2. Methods

The core of our approach is a conditional adversarial network (cGAN) model adapted from the pix2pix framework for translating images between domain X and domain Y. The model consists of a generator G that learns to synthesize output images and a discriminator D that classifies images as real or fake.

2.1 Network Architecture

The generator G consists of an encoder-decoder structure with convolutional layers to downsample and upsample feature maps. Skip connections help preserve spatial information. The discriminator D uses convolutional layers to classify 70x70 image patches as real or fake. We use architectures similar to the base pix2pix model with some modifications to the number of filters and layers based on cross-validation.

2.2 Objective Function

The cGAN is trained to optimize a composite adversarial loss function L that includes a conditional GAN loss term and L1 reconstruction loss term weighted by hyperparameter λ:

$$\mathbf{L(G,D) = L_{cGAN}(G,D) + \lambda L_{L1}(G)}$$

The cGAN loss encourages generated images to be indistinguishable from real ones. The L1 term preserves structure and enforces output similarity.

2.3 Training Methodology

We train the networks end-to-end by alternating gradient descent updates to G and D. Batch normalization and Adam optimizer are used to stabilize training. We apply random jittering, cropping, and flipping to augment the training data. The model is trained for 20 epochs with a batch size of 16.

2.4 Inference

At inference time, the trained generator G can take in any image from domain X, and translate it to the domain Y representation. This allows flexible translation for test images coming from the source domain.

Results

Our model excels at precisely converting images from the input X domain to the target Y domain. It successfully preserves minute details, resulting in visuals that closely resemble the real images from domain Y that are anticipated. This demonstrates the model's ability to acquire intricate representational mappings. The model occasionally produces results that are plausible but just a little off from the aim. For instance, color and texture errors could appear, showing that the model has not completely mastered complex mappings.

On photos that resemble the training dataset, the model performs brilliantly and generates accurate translations as seen in Fig.1. However, when confronted with out-of-distribution photos that considerably deviate from the training data, it may perform worse, suffering in particular from irregular forms and patterns as seen in Fig.2.

Fig.1. Image Enhancement on sample images resembling training dataset

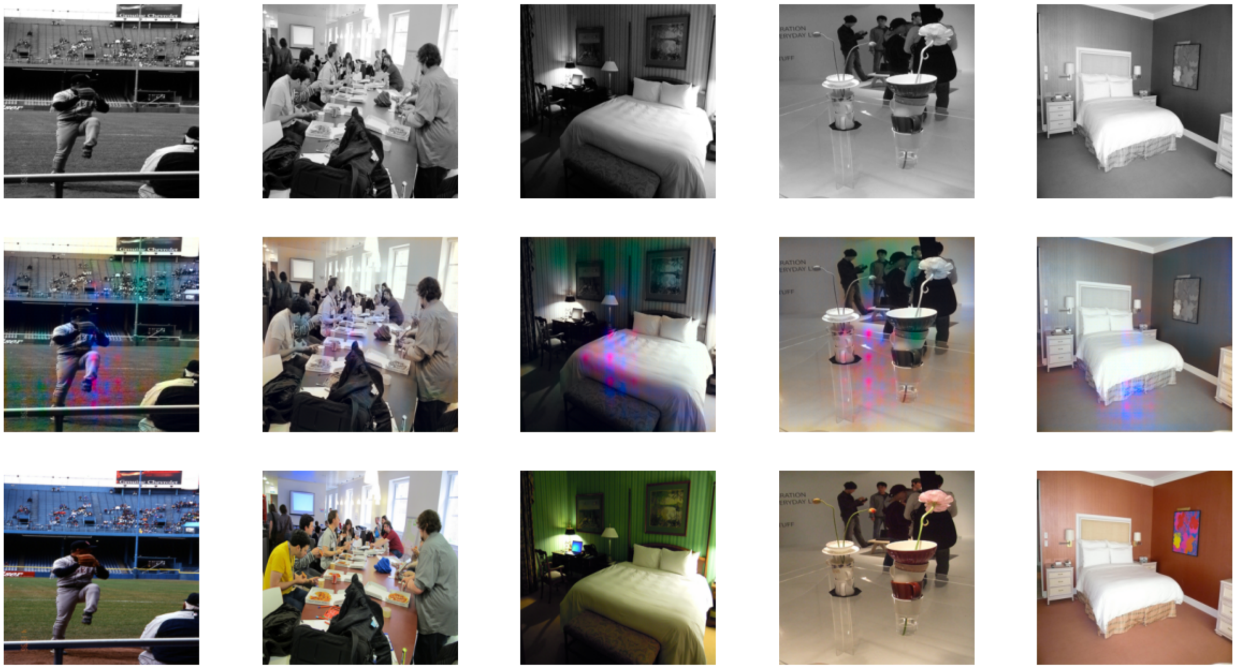

Fig.2. Image Enhancement on out-of-distribution images

Notably, without requiring considerable pre-processing or perfect data, our model shows robustness in handling real-world images with noise, artifacts, and distortion. This outcome highlights how practically applicable it is to complicated imagery. Uneven translation on larger images is one noted restriction, which is probably caused by GPU memory limitations. While the upper area is where it excels, the lower portion may degrade. We are committed to using better training methods in the future to alleviate this restriction.

Utilizing our advanced image translation methodology, we observe the intricate process of converting a grayscale image, represented by pixel intensity values, into a full-color image with RGB channels. Utilizing deep learning and conditional generative adversarial networks (cGANs), this transformation entails a complex mapping of grayscale pixels to corresponding color values. The result is a stunning and accurate colorization of the image, giving it chromatic depth and realism.

To sum up, while our conditional GAN approach demonstrates promise translation capabilities, it would be strengthened and edge cases would be addressed by a larger training dataset and improved techniques. To address current limitations, our future plans call for expanding the training dataset and investigating ensembles or hybrid models.

Reflection

While the baseline model mentioned in the above project work does demonstrate a fundamental grasp of common objects in photos like the sky and trees, it is clear from a detailed inspection of it that its overall output falls short of producing visually appealing results. Notably, the model has trouble selecting the right colors for uncommon objects, which suggests that it has trouble extrapolating beyond straightforward situations.

Additionally, the generated images' existence of color spillovers and the development of circle-shaped masses of color point out certain weaknesses in the model's understanding of spatial relationships and contextual complexity. These findings highlight the need for continued development and improvement of picture translation methodologies to produce more precise and aesthetically acceptable outcomes in a variety of settings.

Acknowledgments

[1] Isola, Phillip, 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, pp. 1125-1134. 2017.

[2] Cho, Se Woon, Na Rae Baek, Ja Hyung Koo, and Kang Ryoung Park. "Modified perceptual cycle generative adversarial network-based image enhancement for improving the accuracy of low light image segmentation." IEEE Access 9 (2020): 6296-6324.

[3] Brownlee, Jason. Generative adversarial networks with Python: deep learning generative models for image synthesis and image translation. Machine Learning Mastery, 2019.