Image-to-Image Translation using GAN's

Anonymous CVPR submission

Paper ID *****

Abstract

In this project, we explore the realm of image-to-image translation by implementing and modifying the Image-to-Image Translation with Conditional Adversarial Networks approach. Our goal is to create a flexible and robust framework that can effectively learn the mapping between input-output image pairs while maintaining desired output properties such as realism, consistency, and adherence to the target domain characteristics. To achieve this, we incorporate four advanced GAN architectures - CycleGAN, Auxiliary Classifier GAN (AC-GAN), Wasserstein GAN (WGAN), and Deep Convolutional GAN (DCGAN) - into our implementation, aiming to surpass the accuracy and performance reported in the original paper. By utilizing the Facades, ADE20K, and Tiny ImageNet datasets for training and evaluation, we seek to demonstrate the advantages and impact of our chosen architectures in addressing various image-to-image translation tasks, ultimately pushing the boundaries of image-to-image translation research.

1. Problem Definition

In this project, we intend to use Conditional Adversarial Networks to address the problem of image-to-image translation. The task of converting an input image into a corresponding output image based on a specific mapping, such as semantic segmentation, style transfer, colourization, or any other transformation, is referred to as image-to-image translation.

The goal is to create a flexible and robust framework that can effectively learn the mapping between input-output image pairs from a given dataset while maintaining desired output images properties such as realism, consistency, and adherence to the target domain characteristics. We hope to generate high-quality translated images that capture the essence of the input images and have a visually appealing and natural appearance by leveraging the power of Conditional Adversarial Networks.

2. Datasets

Facades Dataset: The Facades dataset, which can be found at Dataset Link, was created specifically for facade parsing and generation tasks. It contains detailed labeled images of building facades, making it suitable for generating realistic building facades from input images or performing semantic segmentation of facade elements. The dataset contains 606 annotated images, 400 of which are set aside for training, 100 for validation, and 106 for testing. The dataset is approximately 31 MB in size. Despite its small size, its focus on facade images and detailed annotations make it an appealing choice for projects investigating architectural image translation tasks.

The ADE20K Dataset: It is a large-scale dataset designed for scene parsing and is available at Dataset Link. It includes images of various objects as well as their semantic segmentation labels. The dataset contains over 20,000 images covering a wide range of scenes, from indoor environments to outdoor landscapes, with annotations for over 150 object categories. The dataset is approximately 4.5 GB in size. The ADE20K dataset is an excellent choice for our project that require a versatile dataset for image-to-image translation tasks due to its large number of images, diversity of scenes, and extensive annotations.

Considering our objective of image-to-image translation using Conditional Adversarial Networks. We plan to make use of the Facades dataset for architectural image translation tasks, as it is specifically tailored for facade parsing and generation. We plan to employ the ADE20K dataset for a diverse range of scenes and objects.

3. Methodology and Goal

We intend to implement and modify the Image-to-Image Translation with Conditional Adversarial Networks approach to address various image-to-image translation tasks in this project. The original paper proposes using Conditional GANs to learn the mapping between input-output image pairs, leveraging the adversarial training process to produce high-quality translated images that retain the essential features of the input images.

Our proposed methodology entails implementing the original Conditional GAN architecture while also exploring four additional GAN architectures: CycleGAN, Auxiliary Classifier GAN (AC-GAN), Wasserstein GAN (WGAN), and Deep Convolutional GAN (DCGAN). These architectures were chosen based on their utility and performance in image-to-image translation tasks.

One of our primary goals in incorporating these additional architectures is to outperform the original paper's accuracy and performance. We believe that by incorporating these advanced techniques into our implementation, we will be able to achieve better image quality and translation fidelity.

CycleGAN supports unpaired image-to-image translation, allowing it to learn to translate images from one domain to another without the need for explicit pairings. This property allows our implementation to handle datasets with unpaired input-output images, greatly expanding its applicability.

AC-GAN adds an auxiliary classifier to the GAN framework, allowing the model to generate images based on class labels. We anticipate that by incorporating AC-GAN, our implementation will generate images with specific desired properties, further improving output quality.

WGAN improves GAN training stability by using the Wasserstein distance as the optimization objective instead of the traditional Jensen-Shannon divergence. Incorporating WGAN into our implementation should improve the model's convergence and stability during training.

DCGAN is known for its stable training and better-quality image synthesis. By integrating DCGAN into our implementation, we aim to improve the overall quality of the generated images.

To train and evaluate our implementation, we will use the previously mentioned datasets, Facades, ADE20K, and Tiny ImageNet. Facades will allow us to test our model on architectural image translation tasks, whereas ADE20K will provide a diverse set of scenes for more general image-to-image translation problems.

We hope to create a flexible and robust framework for various image-to-image translation tasks by implementing and modifying the original Image-to-Image Translation with Conditional Adversarial Networks approach with these four additional architectures. We anticipate that our implementation will produce high-quality images while retaining desired output properties and demonstrating the advantages and impact of the chosen architectures in addressing the challenges of image-to-image translation. Ultimately, we strive to achieve better accuracy and performance than those reported in the original paper, pushing the boundaries of image-to-image translation research.

References