CycleGANs to generate UKiyo-e art

Wang, Rovina, Yashraj, Iniyan

Motivation and About the Project

The Project is based on the work done by Jun-Yan Zhu etal on "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks"

Cycle GAN algorithm learns to automatically translate an image between tow domains by training on two unordered image collections e.g., Monet Painting and landscape photo, Zebra and Horse from ImageNet

Traditional image-to-image translation uses a training set of aligned image pairs to learn the mapping between an input image and an output image

Motivation for the project arises from the fact that the paired training data will not be available for most of the cases. CycleGAN is an approach for learning to translate an image between domains in the absence of paired examples. It overcomes the need for a corresponding image in the target domain by a two-step transformation of source domain image

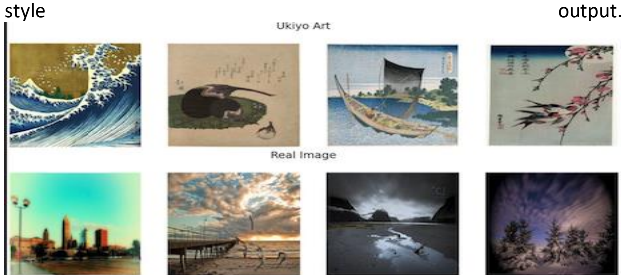
1. Mapping the image to target domain
2. Mapping back to the original Image

Data and Labels

The data we are using is taken from link. We have two types of images

1. Ukiyo Art
2. Reality Images

The goal is to pass reality images through a generator and get Ukiyo style output.



References

- Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks
- https://github.com/junyanz/CycleGAN
- Understanding and Implementing CycleGAN in TensorFlow
- A Gentle Introduction to CycleGAN for Image Translation
- CycleGAN in PyTorch
- Distribution Matching Losses Can Hallucinate Features in Medical Image Translation

Model

The CycleGAN has two generators and two discriminators. One generator takes images from the first domain as input and generates images for the second domain, and the second generator takes images from the second domain as input and generates images for the first domain. Discriminator models are then used to determine how plausible the generated images are and update the generator models accordingly.

But as the authors note, this alone is not enough to ensure that the translation is exact. There is an additional parameter the checks if the image output from the first generator that is fed as input to the second, results in an output that matches the first generator's input. This is called cycle consistency. The reverse is also true: that an output from the second generator can be fed as input to the first generator and the result should match the input to the second generator.

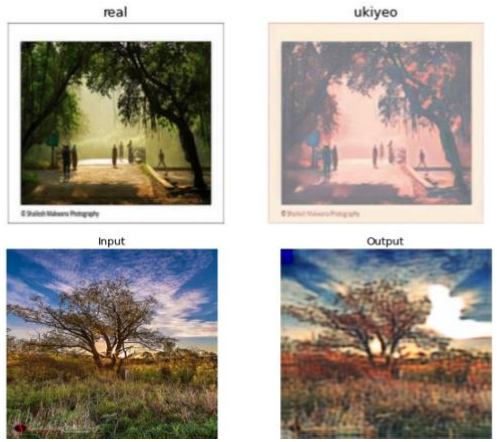
… we exploit the property that translation should be "cycle consistent", in the sense that if we translate, e.g., a sentence from English to French, and then translate it back from French to English, we should arrive back at the original sentence

…Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, 2017.

An additional loss term is added to encourage cycle consistency and then to measure the difference between the generated output of the second generator and the original input image, and the reverse. This acts as a regularization of the generator models.

We replaced the generator architecture with U-Net, If we closely look at the above architectures of the two models, their structure is kind of similar. Both have Encoder, Transformation (Bottleneck in case of UNet) and Decoder blocks. Both takes an image as an input and also outputs an image. So, we wanted to see how the model would perform if we replaced the Generators architecture with the UNet's architecture. Only the generator architecture is different and everything else is same, i.e., no changes in the discriminators, the loss functions and the training steps.

Results (Top: Generator Replaced with Unet, Bottom: Generator and Discriminator)



Conclusion and Future Work

we learnt how to develop a model that can learn how to translate the style of one set of images to the style of another, in the absence of paired training examples.
CycleGANs "exploit the property that translation should be "cycle consistent"
The biggest advantage of CycleGANS is that

- The method does not rely on any "task-specific pre-defined similarity function between the input and output", and neither does it assume that the input and output have to lie in the same low-dimensional embedding space.
- Instead, the model learns the mapping between two domains by trying to capture relationships between higher-level features and it can be applied to problems such as paintings to photos transformations.

But the disadvantages are,

- Results obtained from CycleGANs are far from uniform.
- Tasks that involve geometric changes result in translations don't perform well enough.
- And a gap between the results that are achievable with paired training data and those achieved by unpaired data using CycleGAN exists.
-

An interesting research work showed that given how the models work through matching the translation output to the distribution of the target domain, there can be problems is the data provided in the target domain if it has an over or under representation of some classes (e.g. healthy or sick). This can lead to what they call "hallucinated image features", such as tumors, which should not be used for direct interpretation (e.g., by doctors).

We trained our model to generate Ukiyo style images successfully. But given the computational limits, our results cannot be compared with that of the author.
We can benefit from a longer epoch training than what we use. Nonetheless, CycleGANs don't fail to impress. Another point worth remembering is that the replication of results is not nearly possible unless averaged over very long training periods due to the stochastic nature of the training process.