IMAGE TO IMAGE SYNTHESIS USING CYCLEGANS

Anshula Awasthi[1],Raj Dilip Mesta[2], Chaitanya Sanjeev Wagh[3] ,Yuvraj Singh[4]

[1,2,3,4]Department of Computer, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India.

Abstract— In the artificial intelligence sector, deep learning has achieved great success, and several deep learning models have been created. One of the deep learning models is Generative Adversarial Networks (GAN), and it has become a recent research hotspot. Generative modeling is an unsupervised task in machine learning involving the automatic discovery and learning of the regularities or patterns in input data in such a way that the model can generate or produce new examples that could have been plausibly extracted from the original dataset. GANs have now been widely studied due to the immense potential for applications, including image and vision computing, video and language processing, etc. In this report, we have presented a CycleGAN model to translate photos of horses to zebras, and back again.The whole project is divided into three tasks. 1.To load and prepare the image translation dataset for modeling 2.to train a pair of CycleGAN generator models for translation 3.To load saved CycleGAN models and use them to translate photographs.

Index Terms— Deep Learning; Artificial intelligence; Generative Adversarial Networks (GAN); CycleGAN;

1 INTRODUCTION

Image-to-image translation involves creating, with a particular alteration, a new synthetic version of a given image, such as converting a zebra landscape into a horse. Usually, training a model for translation of image-to-image involves a large dataset of paired examples.. These datasets, including images of paintings by long-dead artists, can be difficult and costly to plan and in some cases impossible.

CycleGAN is a methodology that requires automated training without paired examples of image-to-image translation models. The models are trained in an unsupervised manner employing a collection of images from the source and target domain that don't have to be related in any way. The CycleGAN extends the GAN architecture[1]. Two generator models and two discriminator models are trained simultaneously in this.

Using a training set of aligned image pairs, image-to-image conversion is a class of vision and graphics problems where the objective is to learn the mapping between an input image and an output image.[2] Examples of image-to-image translation include: Translating zebra landscapes to horse landscapes (or the reverse). Translating paintings to photographs (or the reverse). Translating horses to zebras (or the reverse). Traditionally, a dataset consisting of paired examples includes training and an image-to-image translation model. That is, a huge dataset of several input image X examples (e.g. zebra landscapes) and the same image with the desired change that can be used as a predicted output image Y . These datasets, such as images of different scenes under different circumstances, are difficult and costly to prepare. It can be difficult and costly, however to obtain paired training data. It can be even more difficult to obtain input-output pairs for graphics tasks such as artistic stylization, because the desired output is extremely complex, usually requiring artistic authoring. For many tasks, like object transfiguration (e.g., zebra <-> horse), the desired output is not even well-defined.[3]. This is known as the unpaired image-to-image conversion problem.

2 MOTIVATION

Deep learning models are getting more attention in the area of computer vision and recognition tasks.Deep learning based methods are able to exploit the hidden relations in an image data and learn better features for representation of raw input data[4].

The CycleGAN does not require a dataset of paired images, unlike other GAN models for image translation. For example, if we are interested in translating photographs of oranges to apples, we do not require a training dataset of oranges that have been manually converted to apples. This enables a translation model to be built on issues where there might be no training datasets, such as translating paintings into photographs. Therefore using CycleGANs are more beneficial than traditional GANs.

3 Related Work

Neural Style Transfer is another way to perform image-to-image translation, which synthesizes a novel image by combining the content of one image with the style of another image (typically a painting) by matching the Gram matrix statistics of pre-trained deep features[5]. Our main focus, on the other hand, is learning the mapping between two domains, rather than between two specific images, by trying to capture correspondences between higher-level appearance structures. Therefore, our method can be applied to other tasks, such as painting→ photo, object transfiguration, etc. where single sample transfer methods do not perform well

Cycle Consistency The idea of using transitivity as a way to regularize structured data has a long history. In visual tracking, enforcing simple forward-backward consistency has been a standard trick for decades[6]. In the language domain, verifying and improving translations via "back translation and reconciliation" is a technique used by human translators. More recently, higher-order cycle consistency has been used in structure from motion , 3D shape matching, co-segmentation, dense semantic alignment, and depth estimation.here in this project we use a cycle consistency loss as a way of using transitivity to supervise CNN training. In this work, we are introducing a similar loss to push G and F to be consistent with each other[7].

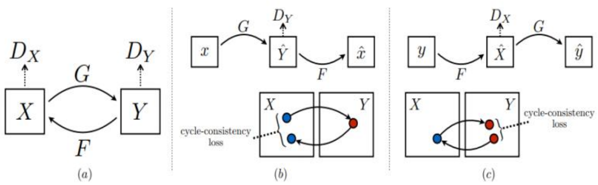

Fig. 3.1 Consistency Cycle Loss

Image-to-Image Translation
The idea of image-to-image translation goes back at least to Hertzmann et al. 's Image Analogies, who employ a nonparametric texture model on a single input-output training image pair. More recent ion using CNNs. More recent approaches use a dataset of input-output examples to learn a parametric translation function using CNNs. Our approach builds on the "pix2pix" framework of Isola et all which uses a conditional generative adversarial network [8] to learn a mapping from input to output images. However, unlike these prior works, we learn the mapping without paired training examples.

Generative Adversarial Networks (GANs)
The key to GANs' success is the idea of an adversarial loss that forces the generated images to be, in principle, indistinguishable from real images[9]. This is particularly powerful for image generation tasks, as this is exactly the objective that much of computer graphics aims to optimize.

4 Gap Analysis

This application is an intersection between generator and discriminator models.The image dataset are of 64*64 pixels and require a huge amount of space to be stored. This can be optimized without reducing the resolution of the output image[10].Two inputs are fed into each discriminator(one is original image corresponding to that domain and other is the generated image via a generator) and the job of discriminator is to distinguish between them, so that discriminator is able to defy the generator and reject images generated by it. While the generator would like to make sure that these images get accepted by the discriminator[11], it will try to generate images which are very close to original images.

5 Proposed Work

Consider the problem where we are interested in translating images from zebra to horse and horse to zebra. We have two collections of photographs and they are unpaired, meaning they are photos of different locations at different times; we don't have the exact same scenes in horse and zebra.We have develop an architecture of two GANs, and each GAN has a discriminator and a generator model, meaning there are four models in total in the architecture. The first GAN will generate photos of horse given photos of zebras, and the second GAN will generate photos of zebras given photos of horses[12].

Each GAN has a conditional generator model that will synthesize an image given an input image. And each GAN has a discriminator model to predict how likely the generated image is to have come from the target image collection[13]. The discriminator and generator models for a GAN are trained under normal adversarial loss like a standard GAN model.

So far, the models are sufficient for generating plausible images in the target domain but are not translations of the input image. Each of the GANs are also updated using cycle consistency loss. This is designed to encourage the synthesized images in the target domain that are translations of the input image.

6 Results and Discussion

As per the proposed workflow, expected results are generated satisfactorily. Cycle consistency loss compares an input photo to the CycleGAN to the generated photo and calculates the difference between the two, e.g. using the L1 norm or summed absolute difference in pixel values[14].

The first GAN (GAN 1) will take an image of a zebra landscape, generate an image of a horse landscape, which is provided as input to the second GAN (GAN 2), which in turn will generate an image of a zebra landscape. The cycle consistency loss calculates the difference between the image input to GAN 1 and the image output by GAN 2 and the generator models are updated accordingly to reduce the difference in the images[15].

This is a forward-cycle for cycle consistency loss. The same process is related in reverse for a backward cycle consistency loss from generator 2 to generator 1 and comparing the original photo of horse to the generated photo of horse.

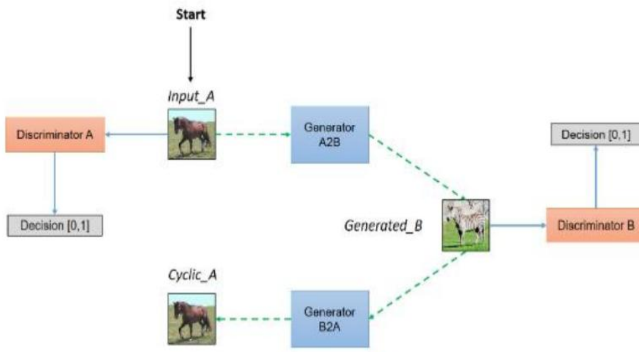

Fig. 1. Horse to Zebra Image Translation using CycleGANs

Obtained Results

We first compare our approach against recent methods for unpaired image-to-image translation on paired datasets where ground truth input-output pairs are available for evaluation.We can observe that the Generators start with a very high error, but with time it starts to produce decent image translations, thus helps in bringing down the error. Both the discriminator errors show very little fluctuation in error. But by the end of 5000 epochs, we can see that both discriminator errors have [16], thus forcing Generators to do more realistic image translations.We then study the importance of both the adversarial loss and the cycle consistency loss and compare our full method against several variants. Finally, we demonstrate the generality of our algorithm on a wide range of applications where paired data does not exist.

7 Conclusion

In this report, we proposed method for image-to-image transformation called as CSGAN. The CSGAN is based on the Cyclic-Synthesized loss.Ideally, the cycled image should be similar to the synthesized image in a domain. The Cyclic-Synthesized loss finds the error between the synthesized and cycled images in both the domains. By adding the Cyclic-Synthesized loss[17] to the objective function (i.e., Cycle-consistency loss), the problem of unwanted artifacts is minimized.The performance of proposed CSGAN is validated over two benchmark image-to-image translation datasets and the outcomes are analyzed with the recent state-of-the-art methods. The thorough experimental analysis, confirms that the proposed CSGAN outperforms the state-of the-art methods.

8 Future Work

The performance of the proposed method is also either better or comparable over other datasets. In future we want to extend our work towards optimizing the generator and discriminator networks and to focus on unpaired datasets i.e., towards unsupervised learning.The technology used in this project can be further modified and used in various other domain such as Object transfiguration, which can be achieved when model is trained to translate one object class from ImageNet to another[18]. Also Collection Style transfer can be achieved which refers to the learning of artistic style from one domain, often paintings, and applying the artistic style to another domain, such as photographs. It can also be modified for Painting to Image Translation in which we introduce an additional loss to encourage the mapping to preserve color composition between the input and output.

References

[1] “Unpaired Image to Image Translation using Cycle Consistent Adversarial Network” , by Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros

[2] “Image-to-Image Translation with ConditionalAdversarial Networks“, Phillip Isola Jun-Yan Zhu Tinghui Zhou Alexei A. Efros

[3]“Generative-Adversarial Networks” by Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza

[4]"Learning from Simulated and Unsupervised Images through Adversarial Training" Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind

[5] "Improved Techniques for Training GANs" by Tim Salimans, Ian Goodfellow, Wojciech Zaremba

[6] J. Donahue, P. Krahenb ¨ uhl, and T. Darrell. Adversarial ¨ feature learning. arXiv preprint arXiv:1605.09782, 2016. 5

[7] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan. Unsupervised pixel-level

domain adaptation with generative adversarial networks. arXiv preprint arXiv:1612.05424, 2016.

[8] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In ICCV, pages 2650–2658, 2015.

[9] L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman. Preserving color in neural artistic style transfer. arXiv preprint arXiv:1606.05897, 2016.

[10] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. CVPR, 2016.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In NIPS, 2014. 2, 3, 4, 5

[12] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504–507, 2006. 4

[13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to image translation with conditional adversarial networks. In CVPR, 2017. 2, 3, 4, 5

[14] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, pages 694–711. Springer, 2016. 2, 3, 4

[15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunning- ´ ham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint arXiv:1609.04802, 2016.

[16] J. Zhao, M. Mathieu, and Y. LeCun. Energy Based generative adversarial network. arXiv preprint arXiv:1609.03126, 2016. 2

[17] T. Zhou, P. Krahenbuhl, M. Aubry, Q. Huang, and A. A. Efros. Learning dense correspondence via 3d-guided cycle consistency. In CVPR, pages 117–126, 2016. 2, 3

[18] J.-Y. Zhu, P. Krahenb ¨ uhl, E. Shechtman, and A. A. Efros. ¨ Generative visual manipulation on the natural image manifold