Image-to-Image translation: from the natural landscape to artistic generated images.

Arpit Garg
The University of Adelaide
Adelaide
a1784072@student.adelaide.edu.au

Dr. Rafael Felix Alves
Supervisor, Australian Institute for Machine Learning
Adelaide
rafael.felixalves@adelaide.edu.au

Abstract

Generalized Adversarial Networks(GANs) consists of two neural network models. One is a generative network that opposes another discriminative network to generate indistinguishable images. The discriminative network regularly tries to classify the original images and generated images [3]. In this paper, we used GAN as a solution for the image to image translation. Its primary objective is to develop the mapping of images from one domain to another. We are trying to present an approach to translate images from its natural behaviour collection X to artistic collection Y without any paired dataset available. It is required to understand the shared information between both domains to learn these mappings. In the absence of paired training data, qualitative results are exhibited on various computer vision tasks specifically on semantic segmentation, colourization, and style transfer. Quantitative comparisons with different prior models illustrate the efficiency of our approach [16].

1. Introduction

In this project, our main aim is to perform image to image translation where input images consist of original landscape images, and the output images must be in artistic images domain. For achieving this, we are using the concept of Generative Adversarial Network, where we have two network one tries to generate similar images as the artistic images from natural landscape images called generator and other tries to classify the discrepancy between the initial picture and the image created known as the discriminator. By adding semantics segmentation, style transfer and, colourization as conditions to both input and output it is possible to learn the mapping and direct the image generation in a particular direction [8] [10]. Deriving paired training data could be expensive and complicated. Acquiring paired dataset for artistic work is more complicated and needs output to include artistic authoring that is a highly complex task. In certain instances, the outcomes expected are not excellently-defined. Hence, we use an algorithm that can learn to map without matched data between domains. It is assumed that there exist underlying relationships between domains. We aim to implement a model that can learn the common unique characteristics from natural landscape images and artistic images and try to translate the collections using these characteristics. We are utilizing the cycle-consistency property for translation. It means if we convert one natural landscape image to an artistic image and then back to a natural landscape image, we should get the same original input back. Our tasks include to make a mapping of the natural to artistic image collection and couple it with inverse mapping from artistic to natural domain. The objective for this image to image translation includes the cycle consistency loss with adversarial losses.

We use our approach for style transfer, colourization and photo enhancement applications also. We also compare our approach with previous approaches of the image to image translations that are focused on the manual factorization of styles and contents.

2. Related Work

Generative Adversarial Network(GANs) also obtained promising outcomes in the generation of images by transforming unsupervised problems to supervised problems. The whole system is based on the game theory, where one network is working against another. Generator networks used to generate fake images and Discriminator used for classifying original and fake images [3]. If the process is unconditioned, then there is no direction, and random images could be generated. To solve this problem, conditional GANs came into existence where several conditions are provided to both input and output to generate the images based on the given conditions. These conditions could be based on class labels [8]. The idea of adversarial loss helps the network to generate similar images from original images. This loss is a powerful tool, and the objective is to optimize this function. The principle of image-to-image transferal in this paper is derived from the frameworks "pix2pix" by Isola et al. [4]. To learn mapping across various contexts, it uses a conditional generative network [8]. Recently, CoupledGAN used weight-sharing techniques to learn common characteristics between domains. This framework is further extended by variational encoders [6]. Such systems employ adversarial networks with external terminologies to implement the output as close to the input under predefined circumstances, such as class labeling. All the prior-mentioned approaches focus on task-specific predetermined requirements for the similarity between the input and output. Our approach does not rely on these conditions, which makes our approach a general-purpose solution.

Using the concept of transitivity to regularize the structured data for visual tracking has been a standard procedure [12]. In recent years, high-order period continuity is used in the structuring of motion [14] also. Zhou et al. [15] and Godard et al. [2] have done comparable experiments and used transitivity by cycle-consistency loss for CNN training. Neural Style Transfer synthesizes images by combining the material of an image with a different one using gram-matrix statistics of pre-trained features [1]. Our primary purpose is to explain how to map between two sets of images rather than single images. We use the concept of cycle-consistency referring to these works and compare them with the mentioned approaches.

3. Methodology

We aim to know the mapping function across natural landscape X and artistic domain Y, where training samples $\{x\}_{i=1}^{N}$ where x_i ϵ X and $\{y\}_{j=1}^{M}$ where y_j ϵ Y. We are using cycle consistency, so our model includes two mappings G : X $\rightarrow$ Y and it's inverse F : Y $\rightarrow$ X. Accordingly, we have two discriminators D_X, which targets to differentiate between original images x and translated images F(y) and D_y which attempts to identify the difference between images y and G(x). Our goal requires adverse losses for matching the distribution of produced and target pictures [2], and cycle consistency losses to prevent the contradiction of learned mappings [16].

3.1. Adversarial Loss

In our approach, we have two mapping functions, and we require adversarial losses for both. For the mapping function G : X $\rightarrow$ Y the discriminator is D_Y and objective is:

$$\begin{aligned}\mathcal{L}_{\mathrm{GAN}}\left(G, D_Y, X, Y\right) = & \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\log D_Y(y)\right] \\ & + \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\log\left(1 - D_Y(G(x))\right]\right.\end{aligned} \tag{1}$$

Here G tries to generate artistic images G(x) that seems undifferentiated from artistic images from domain Y. D_Y identifies the generated artistic images G(x), and real sample y. G tries to reduce the objective and D attempts to optimize it. We follow the same approach for mapping function F : Y $\rightarrow$ X and it's discriminator D_X.

3.2. Cycle-Consistency Loss

Mappings are developed from adversarial losses, but a network maps the duplicate batch of input data to some random collection in the output domain in the presence of a massive capacity. In order to handle these space of feasible mapping functions, our approach of learned mappings should be cycle consistent. Our image translation cycle must be able to produce input(x) back to the original picture for every input image, i.e. x $\rightarrow$ G(x) $\rightarrow$ F(G(x)) $\approx$ x (forward coherence). For images y from discipline Y, it should maintain backward coherence: y $\rightarrow$ F(y) $\rightarrow$ G(F(y)) $\approx$ y [16]. Based on this, loss for cycle consistency is:

$$\begin{aligned}\mathcal{L}_{\mathrm{cyc}}(G, F) = & \mathbb{E}_{x\sim p_{\text{data}}(x)}\left[\|F(G(x)) - x\|_1\right] \\ & + \mathbb{E}_{y\sim p_{\text{data}}(y)}\left[\|G(F(y)) - y\|_1\right]\end{aligned} \tag{2}$$

3.3. Full Objective

Based on both adversarial and cycle-consistency our final objective is [16]:

$$\begin{aligned}\mathcal{L}\left(G, F, D_X, D_Y\right) = & \mathcal{L}_{\mathrm{GAN}}\left(G, D_Y, X, Y\right) \\ & + \mathcal{L}_{\mathrm{GAN}}\left(F, D_X, Y, X\right) \\ & + \lambda\mathcal{L}_{\mathrm{cyc}}(G, F)\end{aligned} \tag{3}$$

where λ manages the comparative significance of of two objectives. We target to solve:

$$G^*, F^* = \arg\min_{G,F}\max_{D_x, D_Y}\mathcal{L}\left(G, F, D_X, D_Y\right) \tag{4}$$

Our model can be seen as training two autoencoders and should be viewed as an exceptional instance of adversarial autoencoders in which adversarial failure functions as an autoencoder's bottleneck layer one with F $\circ$ G : X $\rightarrow$ X, and another with G $\circ$ F : Y $\rightarrow$ Y. It can then be treated as a particular demonstration of adversarial autoencoders in which adversarial loss functions serve as the bottleneck condition of the autoencoder.

3.4. Pipeline

We obey the design principles of Johnson et al. [5] and Radford et al. [9] where network displayed promising results for super-resolution and neural style transfer [1]. Alternatively, we are not using pooling layers for downsampling and upsampling, but we uses strided and fractionally strided convolutions. Batch-normalisation is used both in the generator and in the discriminator, because it standardises inputs for each mini-batch and stables the learning cycle by reducing the number of training epochs. Exception for the output layer, all non-residual convolutionary layers are followed by non-linearity ReLUs, which then use a scaled tanh to ensure that the output pixels are in the correct range.

We use PatchGANs close to Johnson et al. [5] architecture, which helps to distinguish the false and the actual images. We use patch-level discriminator architecture because it has lower parameters than a full image discriminator and can work in a completely convoluted way on images of different sizes.

We adopt various strategies to the training method from recent experience in order to stabilize our training process. The first approach is, we use least-squares loss in-place of negative log-likelihood for $\mathcal{L}_{GAN}$ (Equation 1) [7]. The impairment becomes more reliable through this loss and delivers good quality outcomes. Moreover, to enhance the stability while training, we are using this loss:

$$\begin{aligned}\mathcal{L}_{\text{LSGAN}}\left(G, D_Y, X, Y\right) = \mathbb{E}_{y \sim p_{\text{data}}(y)}\left[\left(D_Y(y) - 1\right)^2\right] \\ + \mathbb{E}_{x \sim p_{\text{data}}(x)}\left[D_Y(G(x))^2\right]\end{aligned} \tag{5}$$

Furthermore, To avoid a drastic shift in the model from iteration to iteration, the discriminators are fed with a background of generated images rather than only the ones created in the latest iterations of the generator [11].

4. Evaluation

For the evaluation, we report the comparative results for the proposed method and several works from the literature both qualitatively and quantitatively. The removal of GAN failure substantially degrades the results, as does the exclusion of cycle consistency loss. Nevertheless, we believe that both techniques are essential for performance. We also check approach with cycle loss in a single direction: GAN+forward cycle loss, or GAN+backward cycle loss, and finds that it often contributes to the training instability and causes mode failure, particularly for the removed mapping route [16].

1-Nearest Neighbor (1-NN) two-sample check appears to fulfill our most useful properties. In two-sample experiments, the 1-Nearest Neighbor Classifier is used to determine when two distributions are similar. This is a part of the two-sample research category, for which a binary classifier may be endorsed in practice. We are accepting the 1-NN classifier since it does not require hyperparameter tuning and no specialized training. [13].

5. Discussion

About the results, we have found that there is a definite pattern of cycle consistency to work better than standard methods of our proposed image translation system. For particular, the treatment of more complex and extreme transitions seems beneficial. We also tend to consider the disparity between tests of matched training data and unpaired data. Some form of soft semantic regulation can be needed to resolve this uncertainty. In comparison, the integration of weak or semi-supervised data produces more robust results and brings in slightly more effective image translators. A significant concern regarding our method is that the lack of cycle-consistency works in translating pictures from natural landscape to artistic forms and back into identical natural landscape pictures. In order to answer this question, we compare our results with one direction also (GAN+forward or GAN+backward). It always results in instability.

For certain instances, fully unpaired data is freely accessible and can be included. This paper stretches the boundaries on what is achievable in this ”unsupervised” environment. Seeing the wide variety of natural landscapes and artistic datasets accessible in the area, with several classes and proportions seen-unseen. We believe that there is still plenty of space for improvement in our strategy. GANs are typically hard to learn, notably in unbridled and large-scale problems. Potential research in this field can also concentrate on solving these issues.

6. Conclusion

In this proposal, We are proposing a system to produce different artistic style images from natural landscape images. The critical point discussed in this paper is the use of cycle-consistency GANs for image translations of unpaired data. Our suggested solution is guided by a loss of cycle-consistency [16], where we impose that the converted artistic style images map back to identical landscape images. Especially, our approach tries to stabilize the training process of GANs by changing loss functions and controlling the oscillations. However, we focused on neural style transfer and super-resolution techniques to improve the quality of generated images [1]. In addition to that, we use 1-Nearest Neighbor evaluation metrics [13] for qualitative comparison of different baseline image translation approaches.

References

[1] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[2] Clément Godard, Oisin Mac Aodha, and Gabriel J Brostow. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 270–279, 2017.

[3] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.

[4] Phillip Isola, 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*, pages 1125–1134, 2017.

[5] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[6] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[7] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, and Zhen Wang. Multi-class generative adversarial networks with the l2 loss function. *arXiv preprint arXiv:1611.04076*, 5:1057–7149, 2016.

[8] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[9] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[10] Amélie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin Murphy. Xgan: Unsupervised image-to-image translation for many-to-many mappings. In *Domain Adaptation for Visual Understanding*, pages 33–49. Springer, 2020.

[11] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2107–2116, 2017.

[12] Narayanan Sundaram, Thomas Brox, and Kurt Keutzer. Dense point trajectories by gpu-accelerated large displacement optical flow. In *European conference on computer vision*, pages 438–451. Springer, 2010.

[13] Qiantong Xu, Gao Huang, Yang Yuan, Chuan Guo, Yu Sun, Felix Wu, and Kilian Weinberger. An empirical study on evaluation metrics of generative adversarial networks. *arXiv preprint arXiv:1806.07755*, 2018.

[14] Christopher Zach, Manfred Klopschitz, and Marc Pollefeys. Disambiguating visual relations using loop constraints. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1426–1433. IEEE, 2010.

[15] Tinghui Zhou, Philipp Krahenbuhl, Mathieu Aubry, Qixing Huang, and Alexei A Efros. Learning dense correspondence via 3d-guided cycle consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 117–126, 2016.

[16] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

7. Additional Timeline Information:

7.1. Semester 1

- Week 1-2 : Discussion about project topic with supervisor
- Week 3-4 : Proposal Submission
- Week 5-6 : Implement various GAN Networks
- Week 7-8 : Data Collection and setting environment for particular application
- Week 9-10 : Implementing GAN model on particular application
- Week 11-12 : Submission of Project report and Project Presentation

7.2. Semester 2

- Week 1-2 : Continuation from last semester, implementing GAN model and hyperparameter tuning
- Week 3-4 : Visualization of the project
- Week 5-6 : Research Paper work starts
- Week 7-8 : Research Paper and project finalisation
- Week 9-10 : Comparing output and evaluation
- Week 11-12 : Final report and presentation