SEMINAR REPORT

ON

CONSISTENT EMBEDDED GAN FOR IMAGE TO IMAGE TRANSLATION

Submitted by

SREERAG P

(KSD18CS083)

To the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

LBS COLLEGE OF ENGINEERIING

KASARAGOD – 671542, KERALA

NOVEMBER 2021

DECLARATION

"*I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of this Institute or other Institute of higher learning, except where due acknowledgement has been made in the text.''*

PLACE: KASARAGOD

DATE: 26/11/2021

NAME: SREERAG P

REG NO: KSD18CS083

CERTIFICATE

This is to certify that the seminar report entitled **CONSISTENT EMBEDDED GAN FOR IMAGE TO IMAGE TRANSLATION** *submitted by* **SREERAG P** *to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the degree of* **Bachelor of Technology in Computer Science and Engineering** *is a bonafide record of the work done by him under my supervision and guidance*.

Mr. Sarith Divakar M	**Mr. Sudheesh KP**	**Mrs. Smitha Mol MB**
Assistant Professor	Assistant Professor	Department of CSE
Department of CSE	Department of CSE	(Head of Department)
(Guide)	(Seminar coordinator)	

Place: Kasaragod
Date: 26/11/2021

ACKNOWLEDGEMENT

It is a really a momentous opportunity and privilege to express my deep sense of gratitude to all those who have us to accomplish this task. I express humble God Almighty for his incessant blessing on us during this seminar.

I have taken efforts in this seminar. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend our sincere to all of them.

I sincerely thank our principal **Dr. MOHAMMED SHEKOOR T** for providing me facilities in order to go ahead with our seminar development. I express our sincere gratitude to head of the department **Dr. SMITHAMOL M B** and I also express heartiest gratitude to seminar coordinator **Mr. SUDHEESH KP** in computer science and engineering for their valuable advice and guidance. I would always oblige for the helping hands of all other staff members of the department and all my friends and well-wishers who directly or indirectly contributed to this venture.

Last but not least, I am indebted to God almighty for being the guiding light throughout this work and for helping me to complete the same within the stipulated time.

SREERAG P

ABSTRACT

Generative Adversarial Networks (GANs) have achieved remarkable progress in image-to-image translation tasks. However these methods have the common problem that lacking the ability to generate both perceptually realistic and diverse images in the target domain. To solve this problem, the paper propose a model named Consistent Embedded Generative Adversarial Networks (CEGAN) for image-to-image translation. It aims to learn conditional generation models for generating perceptually realistic outputs and capture the full distribution of potential multiple modes of results by enforcing tight connections in both the real image space and latent space. To achieve realism, unlike existing GANs models that their discriminators attempt to differentiate between real images from the dataset and fake samples produced by the generator, the discriminator in the proposed model distinguishes the real images and fake images in the latent space to reduce the impact of the redundancy and noise in generated images. By this way, the model produces more diverse and realistic results in the target domain.

CONTENTS

List of Figures

List of Abbreviations

GAN : Generative Adversarial Network

CE-GAN : Consistent Embedded GAN

CNNs : Convolutional Neural Networks

1. INTRODUCTION

Nowadays, image-to-image translation tasks have attracted much attention in many computer vision articles due to its extraordinary performance [1]–[3]. It aims to learn a mapping that can convert an image from a source domain to a target domain, while preserving the main presentations of the input images. For instance, networks have been used to translate real-world scenes into cartoon images, add color to grayscale images and fill missing image regions.

The goal of generative adversarial networks GANs [4] is to generate samples that can confuse the discriminator to achieve the purpose of falsehood to be dressed up as truth. It has achieved impressive success in image editing, super resolution, representation learning, and image generation. Particularly, the GANs are extensively studied in image-to-image translations.

Generative Adversarial Networks (GANs) have achieved remarkable progress in image-to-image translation tasks. However it lacking the ability to generate more realistic and diverse outputs in the target domain. In this work, the focus is to learn conditional generation models for generating perceptually realistic outputs and model a distribution of potential multiple modes of results by enforcing tight connections in both real image space and latent space.

2. Image to Image Translation

The task of image-to-image translation is to convert an image from a source domain to a target domain while preserving its certain properties. The first unified model for image-to-image translation based on conditional GANs, which has been successfully applied to many applications. For example, in generating high-resolution images. Conditional GANs demonstrated that both generator and discriminator are conditioned on some extra information to generate the output we expected. Similarly, conditional VAE aims to translate the source domain to the target domain by adding random noise to the given image. Potentially, all of the methods defined above could be easily conditioned and have shown promise, while image-to-image conditional GANs have lead to a substantial boost in the quality of the results, such as pixel values, semantic features, class labels, or pairwise sample distances. In addition to this a UNIT framework was proposed, which assumes a shared latent space such that corresponding images in two domains are mapped to the same latent code.

3. Generative Adversarial Networks (GANs)

Generative Adversarial Networks, or GANs for short, are an approach to generative modeling using deep learning methods, such as convolutional neural networks.

Generative modeling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset.

GANs are a clever way of training a generative model by framing the problem as a supervised learning problem with two sub-models: the generator model that we train to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated). The two models are trained together in a zero-sum game, adversarial, until the discriminator model is fooled about half the time, meaning the generator model is generating plausible examples.

3.1 The Generator Model

The generator model takes a fixed-length random vector as input and generates a sample in the domain.

The vector is drawn from randomly from a Gaussian distribution, and the vector is used to seed the generative process. After training, points in this multidimensional vector space will correspond to points in the problem domain, forming a compressed representation of the data distribution. This vector space is referred to as a latent space, or a vector space comprised of latent variables. Latent variables, or hidden variables, are those variables that are important for a domain but are not directly observable.

3.2 The Discriminator Model

The discriminator model takes an example from the domain as input (real or generated) and predicts a binary class label of real or fake (generated).

The real example comes from the training dataset. The generated examples are output by the generator model. The discriminator is a normal (and well understood) classification model.

After the training process, the discriminator model is discarded as we are interested in the gen erator.

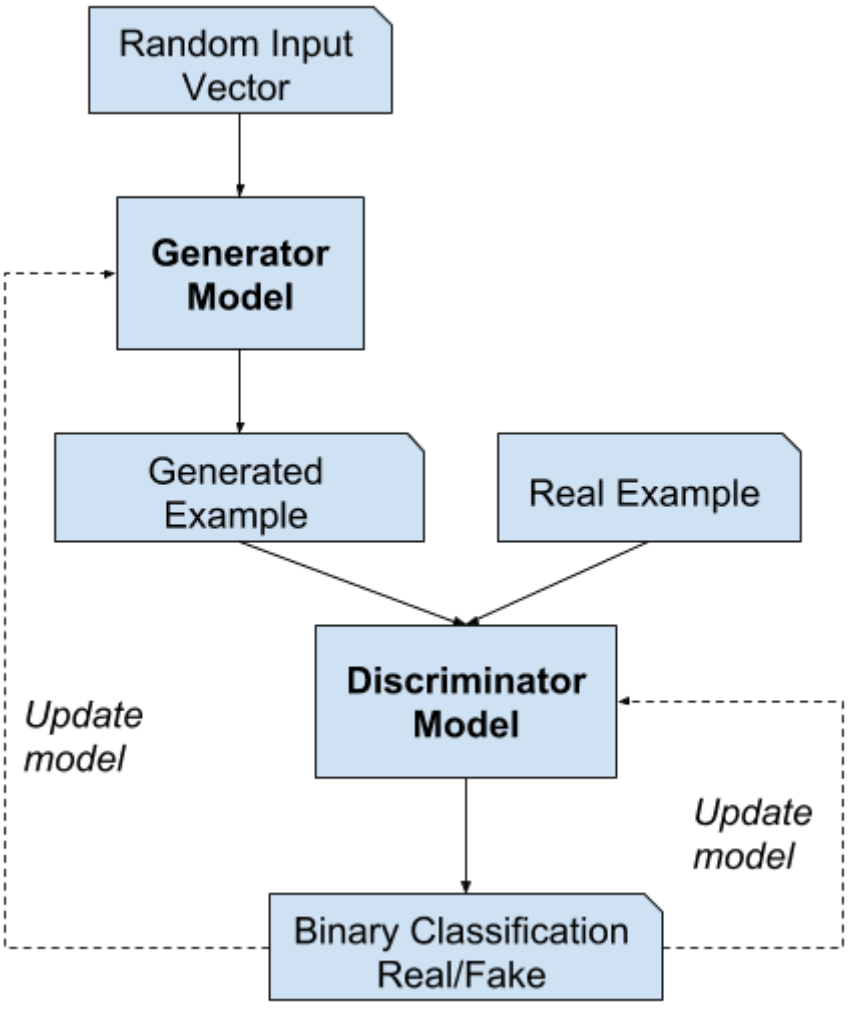

Fig.1. Generative Adversarial Network Model Architecture

GANs are an exciting and rapidly changing field, delivering on the promise of generative models in their ability to generate realistic examples across a range of problem domains, most notably in image-to-image translation tasks such as translating photos of summer to winter or day to night, and in generating photorealistic photos of objects, scenes, and people that even humans cannot tell are fake.

3.3 Limitations of GANs

Most existing GAN frameworks usually consist of two Convolutional Neural Networks (CNNs). One is the generator G which is trained to produce output that confuses the discriminator. The other is the discriminator D which classifies whether the image is from the real target manifold or synthetic. However, in real-world application, the original images are usually high-dimensional images that may well contain redundant features or noise. The traditional GAN structure mainly considers the error relationship between the generated image and the noisy image, which leads to noise and redundancy in the generated images. As a result, the quality of generated images are unsatisfactory.

4. Consistent Embedded GAN Networks (CEGANs)

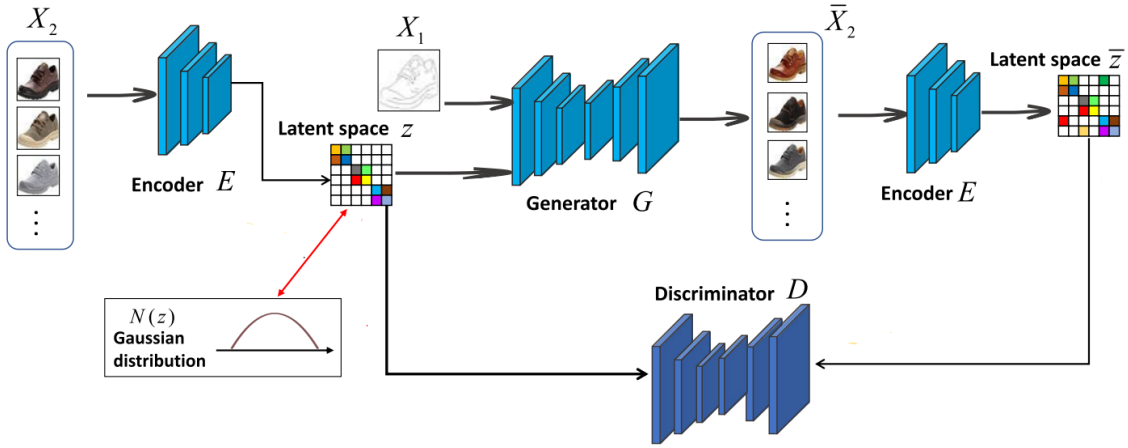

Fig.2. Model Overview

Above figure shows an overview of the proposed model. In the training progress, let x1 ∈ X1 and x2 ∈ X2 be the images from two different image domains, which are a dataset of paired images and are representative of a joint distribution p(x1, x2). We should learn a multi-modal mapping between two image domains, for example, X1 and X2 represent edges and ground truth photographs respectively, and we want to generate a set of photographs about X2 which have different colors and textures according to the edges of X1. To achieve this, we train G to translate an input images x1 into an output images x2 conditioned on the target domain images' latent vector. It is important to note that there could be multiple plausible paired images x2 which would correspond to an input images x1 but the training dataset usually contains only one such pair. However, given a new image x1 ∈ X1 during test time, the model CEGAN would be able to generate a diverse set of output x2 ∈ X2, corresponding to different modes in the distribution p(x2 |x1). We would like to learn the mapping that could sample the output x2 from true conditional distribution given x1, and produce results which are both diversity and realism. In order to achieve diversity, we learn a low-dimensional latent code z that encapsulates the ambiguous aspects of the output mode which are not present in the input image.

According to the latent code z, we could get different styles with the same input. We then learn a deterministic mapping F: (x1, z) → x2. To enable stochastic sampling, we desire the latent code to be drawn from some prior distribution p(z) 1= E(x2). On the other hand to achieve realism, unlike the existing GANs framework that the discriminator D attempts to differentiate between real samples and generated samples, the discriminator D in CEGAN distinguish the real images and fake samples in the latent space. As we all know that during training process, the generate images always contain redundancy features, noise and outlying entries, which would lead to unreliable and inaccurate results. By latent space learning, we encode this images to the low-dimensional latent code z to alleviate such a challenging problem. Then discriminator D classifies whether the latent code is from the real target manifold or synthetic. Furthermore, to further ensure the quality of the generated images, we use a standard Gaussian distribution N (0, I) to constraint the latent distribution.

4.1 Implementation

4.1.1 Network Configuration

CEGAN is constructed with identical network architecture for G, D and E. For generator, it is configured with equal number of downsampling and upsampling layers. In addition, we configure the generator with symmetric skip connections between downsampling and upsampling layers as in BicycleGAN, making it a U-Net [43]. Such a design has been shown to produce strong results in the unimodal image prediction setting since it enables low-level information to be shared between input and output pairs. Without the skip layers, information from all levels has to pass through the bottleneck, typically causing significant loss of high-frequency information. For discriminator, we employ three fully connected layers, which aims to predict the real or fake latent code rather than images or overlapping image patches. Such a configuration is effective in capturing low-dimensional latent code distribution and it

fulfills our needs well. For the encoder, it includes several strided convolutional layers to downsample the input, and a few residual blocks to further process it, followed by a global average pooling layer and a fully connected layer.

4.1.2 Injecting the Latent Code to Generator

To realize the diversity of outputs, we encode the possible multiple outputs in the latent space and combine the latent code with the given image as the input of the generator. By learning a mapping between real image space and latent space, we random sample the ambiguity mapping to express multiple modes. So how to propagate the information encoded by latent code to the image generation process is critical to our applications. There are two common solutions in existing methods. The most simply strategy is to extend a Z-dimensional latent code to an H × W × Z spatial tensor and concatenate it with the H × W × 3 input image. Alternatively, the other method is to add the latent code to each intermediate layer of the network G. In this paper, we chose the former because the experiment results are not much different but the first strategy is easy to implement.

4.2 Datasets

Edges—Shoes : Provided by [5], which contain images of shoes with binary edge generated by the HED edges detector [6]. All the images are revised to 256×256 for this model training.

5. Conclusion

In this paper, a novel image-to-image translation model named Consistent Embedded Generative Adversarial Networks (CEGAN) is proposed to generate both realistic and diversity images. This method captures the full distribution of potential multiple modes of results by enforcing tight connections between the latent space and the real image space. Particularly, to alleviate the impact of the redundancy and noise in generated images, unlike other GANs, the discriminator in the proposed model distinguish the real images and fake images in the latent space.

6. References

[1] R. Zhang, P. Isola, and A. A. Efros, ‘‘Colorful image colorization,’’ in Proc. ECCV, 2016, pp. 649–666.

[2] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, ‘‘Toward multimodal image-to-image translation,’’ in Proc. NIPS, 2017, pp. 465–476.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, ‘‘Image-to-image translation with conditional adversarial networks,’’ in Proc. IEEE CVPR, Jul. 2017, pp. 5967–5976.

[4] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ‘‘Generative adversarial nets,’’ in Proc. NIPS, 2014, pp. 2672–2680.

[5] A. Yu and K. Grauman, ‘‘Fine-grained visual comparisons with local learning,’’ in Proc. IEEE CVPR, Jun. 2014, pp. 192–199.

[6] S. Xie and Z. Tu, ‘‘Holistically-nested edge detection,’’ Int. J. Comput. Vis., vol. 125, nos. 1–3, pp. 3–18, Dec. 2017.

[7] FENG XIONG, QIANQIAN WANG , AND QUANXUE GAO, “Consistent Embedded GAN for Image-to-Image Translation”. Available: https://ieeexplore.ieee.org/document/8825805