CONSISTENT EMBEDDED GAN FOR IMAGE TO IMAGE TRANSLATION

SREERAG P
S7 CSB 31

GUIDE : SARITH DIVAKAR M
ASSISTANT PROFESSOR, CSE

OUTLINE

- Introduction
- Image to Image translation
- Generative Adversarial Networks (GANs)
- Existing System
- Proposed System
- Consistent Embedded GANs
- Implementation
- Experiment
- Conclusion
- References

INTRODUCTION

- Nowadays, Image-to-image translation tasks have attracted much attention in many computer vision articles due to its extraordinary performance.
- 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.
- Consistent Embedded GANs (CEGANs) tackle the problems in GANs.

IMAGE TO IMAGE TRANSLATION

- It aims to learn a mapping that can convert an image from a source domain to a target domain.
- The task preserves the main presentations of the input images.
- Examples :
  - Converting real-world scenes into cartoon images
  - Adding color to grayscale images
  - Filling missing image regions

GENERATIVE ADVERSARIAL NETWORKS (GANs)

- Goal :To generate samples that can confuse the discriminator to distinguish between generated samples and real samples.
- Consist of two Convolutional Neural Networks (CNNs) :
  - Generator G :To produce samples
  - Discriminator D :To classify the samples

GENERATIVE ADVERSARIAL NETWORKS (GANs) (Cont.)

- Mainly considers the error relationship between the generated image and the noisy image.
- Leads to noise and redundancy in the generated images.
- The quality of generated images are unsatisfactory.
- Generates only a few number of samples.

EXISTING MODEL

- The discriminator attempts to differentiate between real images from the dataset and fake samples produced by the generator.
- Fails to achieve realism and diversity.

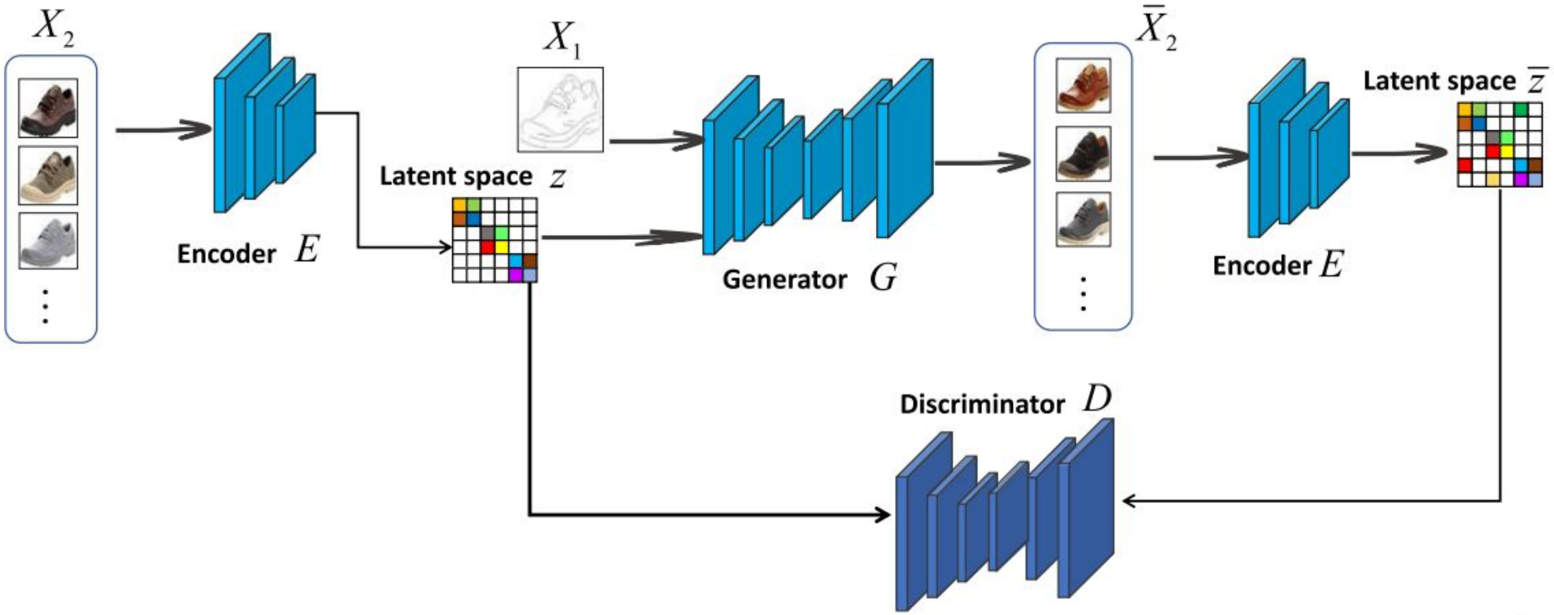
PROPOSED MODEL

- The discriminator distinguishes the real images and fake images in the latent space.
- Reduces the impact of the redundancy and noise in generated images.

CONSISTENT EMBEDDED GANs (CEGANs)

- The model that combines GAN and latent space learning.
- The discriminator distinguishes the real images and fake samples in the latent space instead of the real image space.
- Mapping between real image space and latent space.

IMPLEMENTATION



IMPLEMENTATION (Cont.)

Network configuration :

- CEGAN is constructed with identical network architecture for G, D and E.
- Generator is configured with equal number of downsampling and upsampling layers.
- Discriminator is three fully connected layers, which aims to predict the real or fake latent code rather than images.
- Encoder includes several convolutional layers to downsample the input and a few residual blocks to further process it.

IMPLEMENTATION (Cont.)

Injecting the latent code to generator :

- Encoding the possible multiple outputs in the latent space.
- Combines latent code with the given image and feed it into the generator as input.
- By learning a mapping between real image space and latent space, multiple modes of the images are generated.

EXPERIMENTS

Dataset

- Edges—Shoes: Provided by [2] , which contain images of shoes with binary edge generated by the HED edges detector [3]. All the images are revised to 256×256 for this model training.

EXPERIMENTS (Cont.)

Baselines

- **cVAE-GAN**: Combines a variational autoencoder with a generative adversarial network to translate the images from source domain to target domain.
- **cLR-GAN**: This is another approach to capture image mode in latent space
- **BicycleGAN**: The method realizes bidirectional mapping by combining cVAE-GAN and cLR-GAN.

EXPERIMENTS (Cont.)

Evaluation metrics :

- **AMT Perceptual Study**: In order to compare the faithfulness and realism of translation outputs generated by different methods
- **LPIPS Distance**: LPIPS distance is one of the universal indicators for measuring image translation diversity.
- **FID Score**: FID score is a measure of similarity between two datasets of images. Lower FID values mean better image quality and diversity

EXPERIMENTS (Cont.)

EXPERIMENTS (Cont.)

Results :

Methods	Edges-shoes	
	AMT Fooling %	LPIPS Distance
cVAE-GAN	22.56±2.85	0.171±0.021
cLR-GAN	39.27±1.97	0.121±0.014
BicycleGAN	51.62±3.26	0.159±0.025
CEGAN	**55.12±2.34**	**0.178±0.032**

FID Scores

Methods	Edges-shoes
cVAE-GAN	0.678
cLR-GAN	0.724
BicycleGAN	0.412
CEGAN	**0.397**

CONCLUSION

- In this proposed system, an 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.
- It reduce the impact of the redundancy and noise in generated images, unlike other GANs, the discriminator in our model distinguish the real images and fake images in the latent space.

REFERENCES

1. https://ieeexplore.ieee.org/document/8825805
2. A. Yu and K. Grauman, ‘‘Fine-grained visual comparisons with local learning,’’ in Proc. IEEE CVPR, Jun. 2014, pp. 192–199.
3. S. Xie and Z. Tu, ‘‘Holistically-nested edge detection,’’ Int. J. Comput. Vis., vol. 125, nos. 1–3, pp. 3–18, Dec. 2017.

THANK YOU