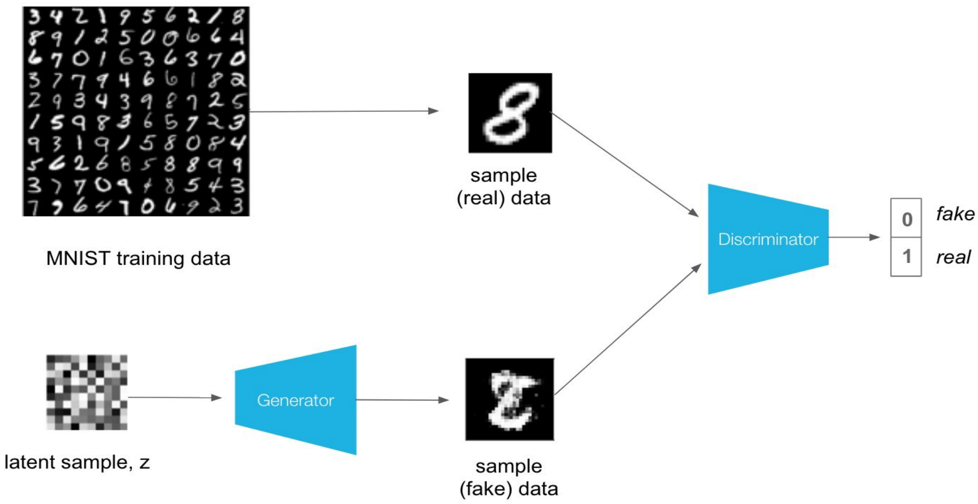
Assignment 4 Report

(RAVI RANJAN, MT19AI032)

Question 1:-

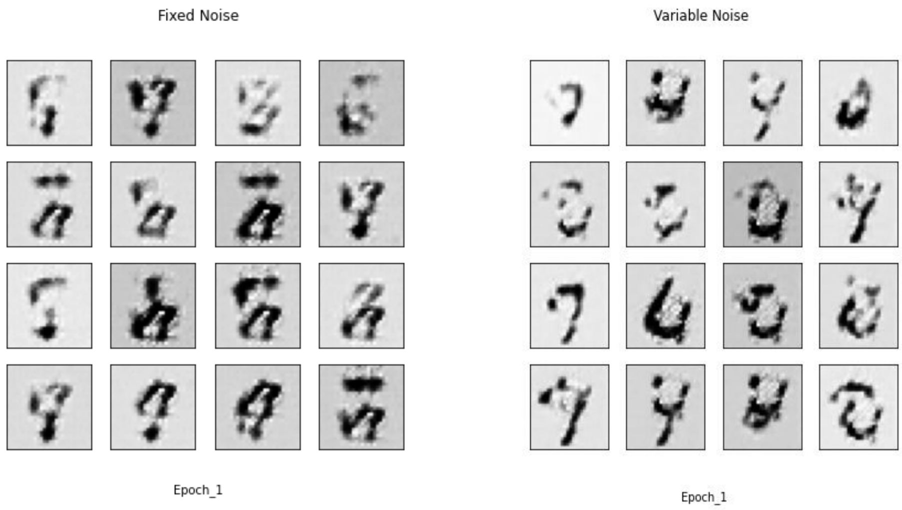
Training of DCGAN for MNIST Images

GAN is a generative adversarial network trained in min-max form, where generator and discriminator are trained against each other in order to generate realistic images. Discriminator function is to distinguish between real and the fake samples (generated by a generator). Below is the architecture diagram of the DCGAN.

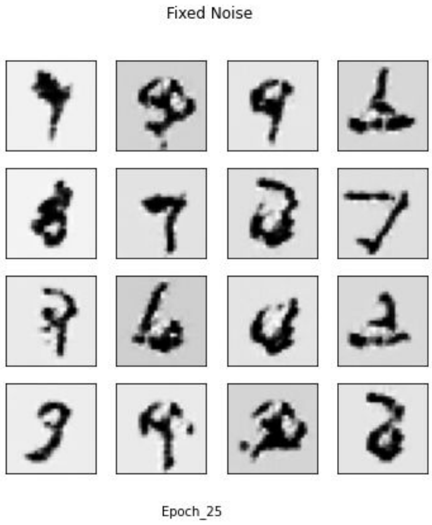


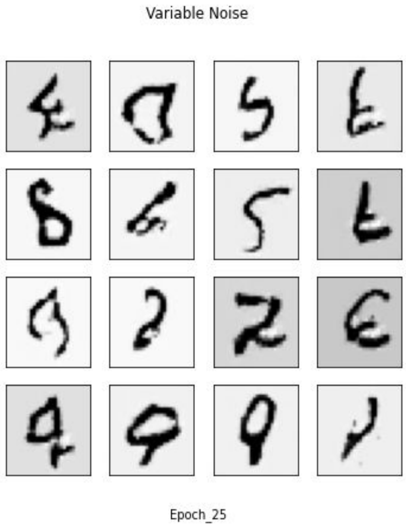
I use Resnet Based Generator and VGG16 based Discriminator, the latent vector z is of size 784 with batch size is 32 and trained for the 50 epochs, with ADAM optimizer. The learning rate is 0.0002 for both generator and discriminator. I use 50k MNIST images in order to train the GAN.Following are the generated images:-

1. After the first epoch

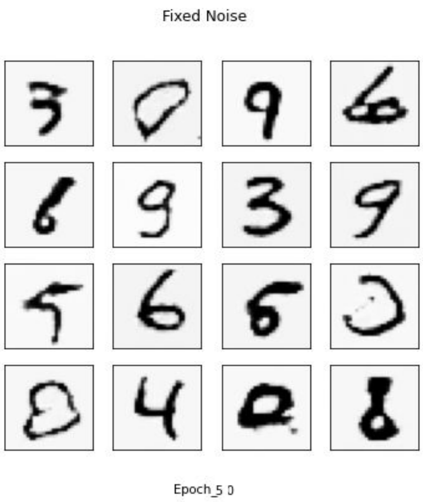


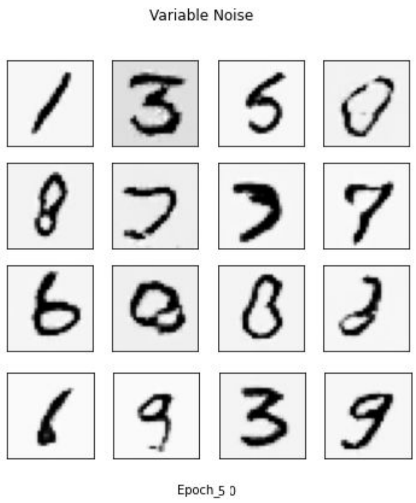
2. After N/2 epochs





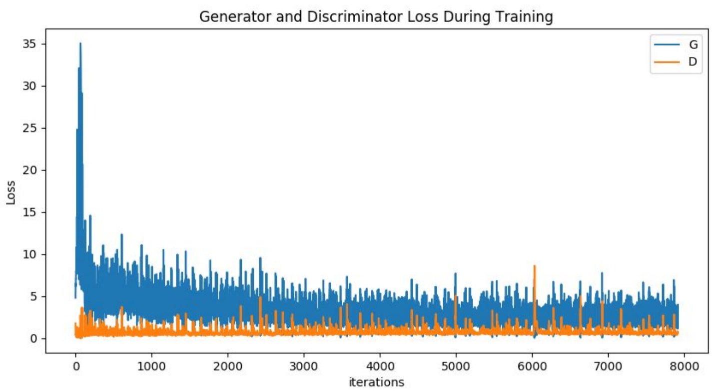
3. After last epoch





We can identify the images after half epochs but for the initial epochs it's hard to identify the generated images.

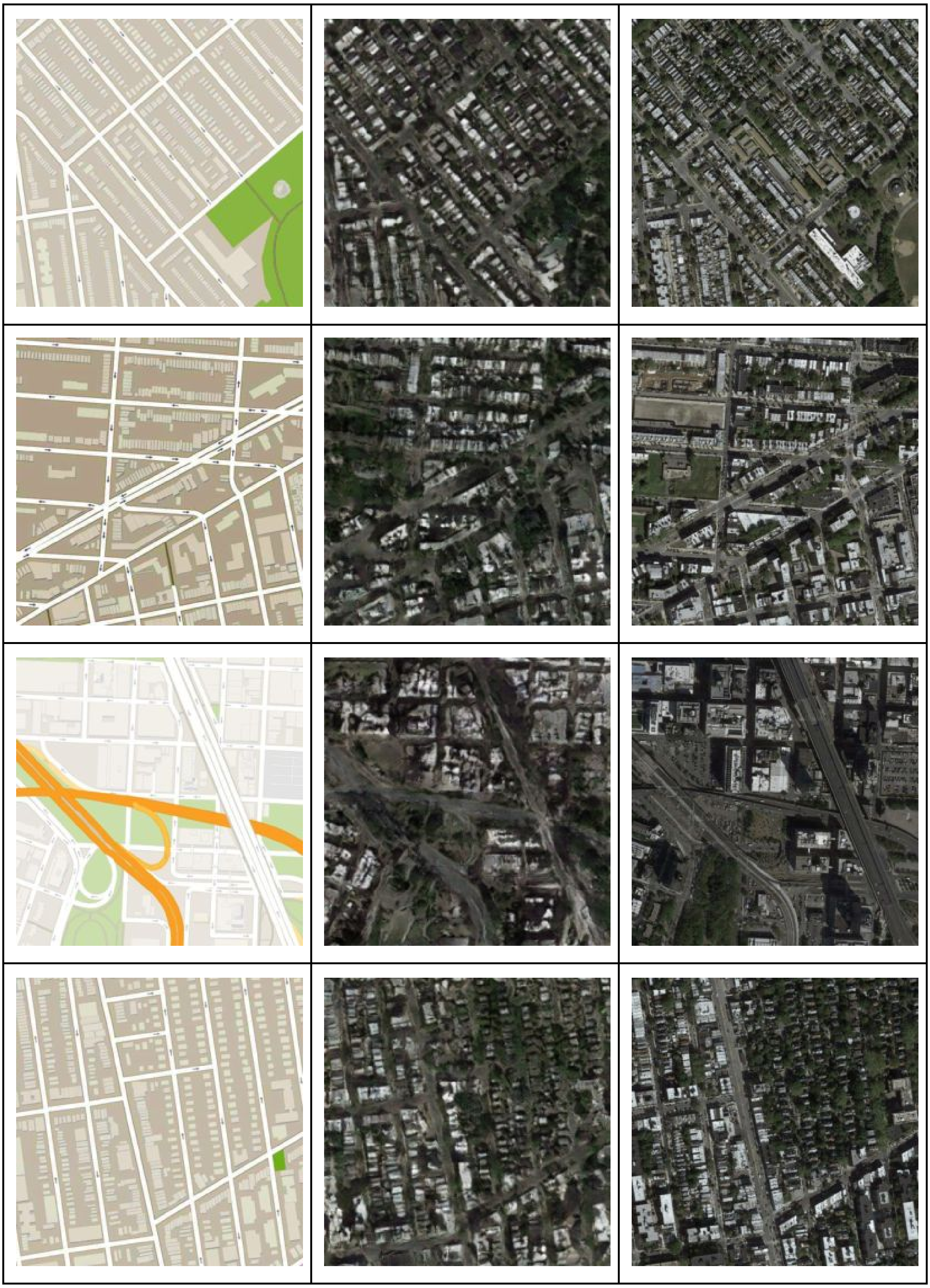
Loss Plot

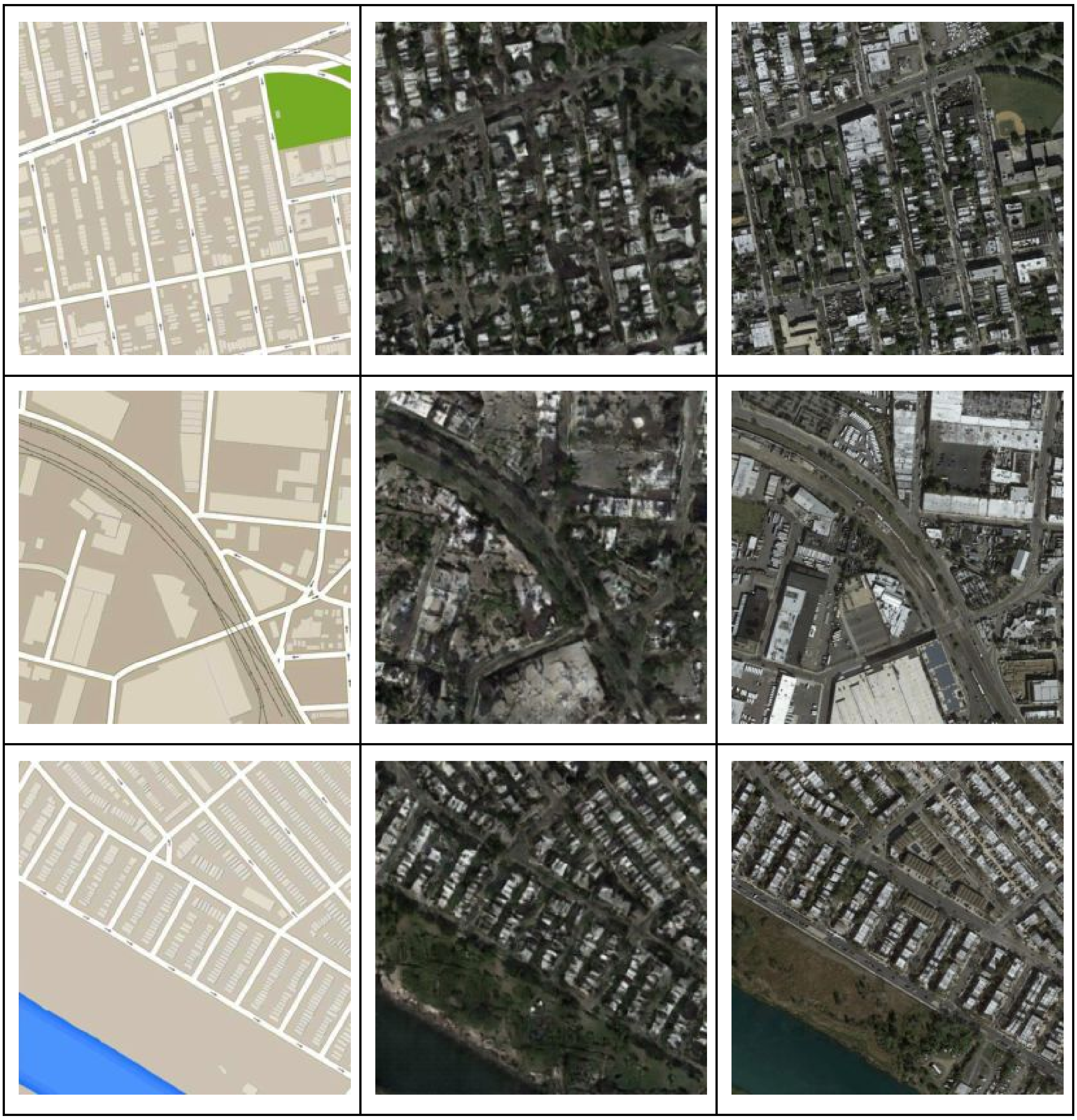


Question 2:-

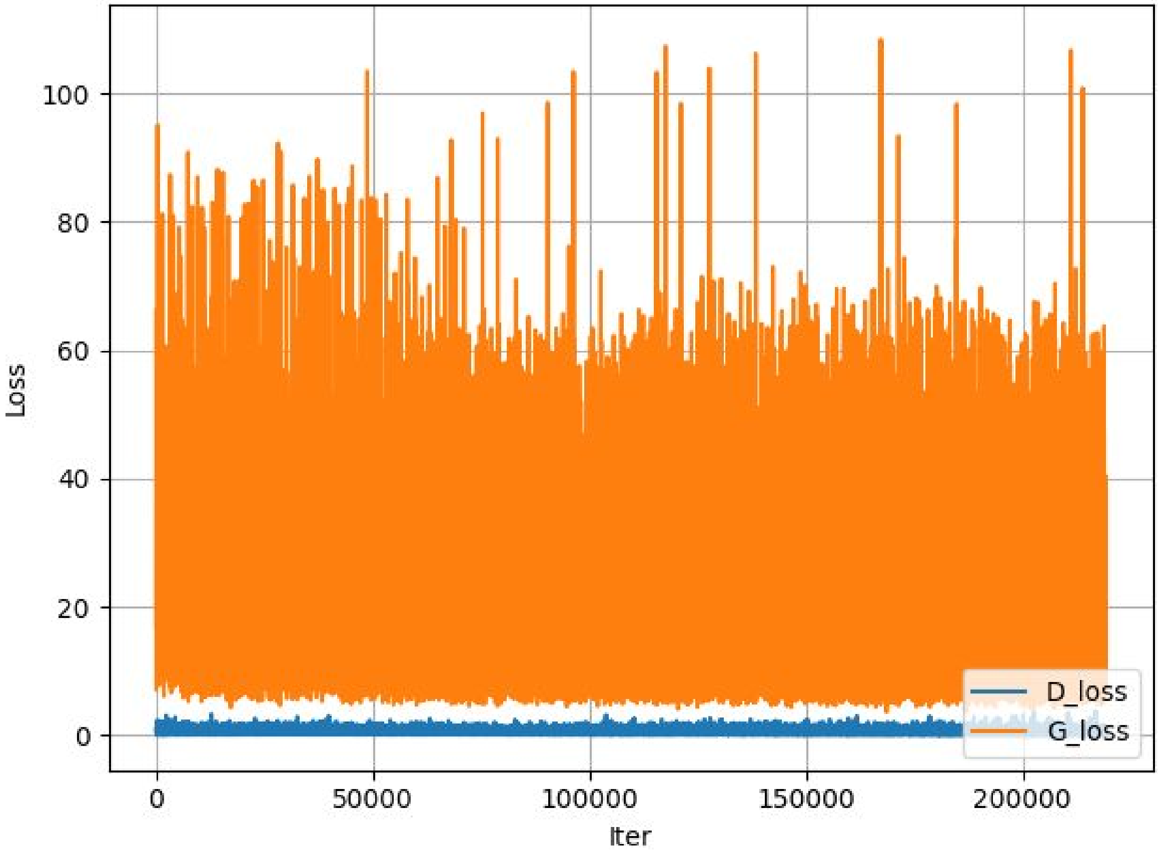
a)

Input	Output	Target
		
		
		





b)



c)

Generated images from pre-trained model	Generated images from our trained model
	

Generated images from pre-trained model and Generated images from our trained model are found to be very similar and Average SSIM score is 0.206

Question 3

Paper Title: Contrastive Learning for Unpaired Image-to-Image Translation

Summary of the paper

Contrastive learning for unpaired image to image translation is a self-supervision image translation approach, based on the contrastive learning paradigm. Contrastive learning is a kind of learning where we want to learn distinctiveness i.e. formulate the task of finding similar and dissimilar things and if two things are similar, then we want the encodings for these two things to be similar as well. This paper aims to do image to image translation while preserving the structure of the image. Here the target image properties like colour, pattern, and texture are enforced by the adversarial loss while content is preserved using an alternative of cycle consistency, in which correspondence in content is maintained by maximizing the mutual information between corresponding input and output patches of images. Unlike other papers based on self-supervision and contrastive learning, which uses heavy data augmentation in creating negative and positive examples to apply the contrastive loss, the authors here use the patches of the same image to create a pool of negative pairs, which enable their approach to work even on a single image. Another advantage is that this approach is not resource hungry, as contrastive learning gives better results if we have a large pool of negative pairs and with heavy data augmentation along with large batch sizes to create a large pool of negatives. This approach slows down the training if image size is large. In addition, drawing negatives internally from within the input image forces the patches to better preserve the content of the input.

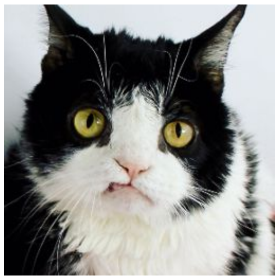
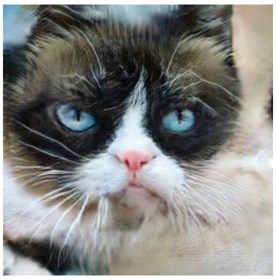
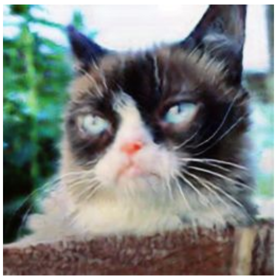
The authors in the paper uses a type of contrastive loss function i.e. InfoNCE loss which aims to learn embeddings that associate corresponding patches while disassociating them from others. The method proposed in the paper requires learning the mapping in one direction and avoids using inverse auxiliary generators and discriminators. Adversarial loss is used to force the output to be visually similar to input and a noise contrastive estimation framework to maximize mutual information between input and output. Here in the context of contrastive learning, an association is to learn between the query (the patches from zebra) and it's positive counterpart (i.e. corresponding patches from horse and zebra) in contrast to other points (i.e. non-corresponding patches) in the dataset, which referred to as negatives. The query, positives and negatives are mapped to K- dimensional vector and an (N+1) way classification problem is setup, where the scaled up distances between are passed as logits and cross-entropy loss is calculated, that gives probability of the positive examples being selected over negatives. The authors employ a multilayer patch-wise learning objective be inspiring from, that not only the whole images but also corresponding patches between input and output images share the contents. Here the aim of authors is to match corresponding input and output patches at a specific location and other patches are leveraged as negative.

In correspondence from all above points the final objective of the paper is to learn embeddings such that the generated images should be realistic while patches in the input and output images should share correspondence. This method learns a cross-domain similarity function and is the first image translation algorithm, to not use any predefined similarity function.

Code is taken from the link given below:

https://github.com/cryu854/CUT

Reproduced results from the implementation of the paper

Fake image	Identical image	Real image A	Real Image B
			
			
			

References

Park, T., Efros, A. A., Zhang, R., & Zhu, J. Y. (2020, August). Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision* (pp. 319-345). Springer, Cham.