Project Submission

Abstract:
The goal of this project is to map a single image from a source domain A, to a target domain B, which contains a set of images.

Problem description:
Many problems in computer vision aim to map between one domain to another. But most works assume large sets for both source and target domains. This project offers a way to do one-shot unsupervised cross domain translation. Which means, that only one image exists in the source domain, and many images exist in the target domain. The source and target sets are completely unpaired.

Chosen method:

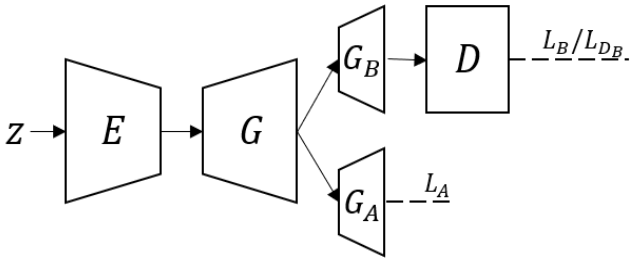
1. Network architecture:
   This project was based on two main works; one is Deep image prior (DIP) by Ulyanov et al. [1], and the other is One-Shot Unsupervised cross Domain Translation (OST) by Benaim et al. [2].
   I used the concept and main architecture of DIP and altered it to match the domain transfer problem. I also used the basic framework of OST.

   The training is done using the following losses functions:

$$L_A = min_\theta \, || \, G_A(E(z)) \, - \, x_A||^2 \quad (DIP \; loss)$$
$$L_B = \, -l(D_B(G_B(E(z))),0)$$
$$L_{D_B} = \, +l(D_B(G_B(E(z))),0) \, + l(D_B(x_B),1)$$

   Where θ are the network parameters, x_A is a sample from domain A and x_B is a sample from domain B.

   This is the main diagram of the project:



   Input: z a random code vector. z has the same spatial size as x_A.
   Output: two images. one from $G_A(G(E(x_\mathrm{A})))$ and the other from $G_B(G(E(x_A)))$.

2. Datasets:
   MNIST ⟷ SVHN
   Winter ⟷ Summer Yosemite

3. Training method:
   Training was divided into two phases:

3.1. The first phase was trained with only images from domain B. I used L_B and L_{D_B} losses and trained the network (with branch B) and the discriminator.
I employed a small number of data augmentation, which included horizontal flips and small rotations.

3.2. The second phase produces the mapping between the single image from domain A to domain B, using a small number of iterations (just like in DIP). In this phase, no augmentation was employed.

The model was trained with adam optimizer. The learning rates for MNIST-SVHN dataset are the same for the encoder-decoder networks, but $G_B$network is updated every 100 iterations. For Winter-Summer dataset the learning rate for updating $G_B$is much smaller than the learning rate for updating the rest of the network (difference of 3 orders of magnitude).

Code explanation:
For this project I used the basic framework from OST. The network was taken from DIP and was altered to match the problem, the encoder stayed the same, but the decoder was split into three parts; one for the shared layers, second for $G_A$and third for G_B. The discriminator architecture is the same as OST.
The architecture for the network from DIP is a U-Net type “hourglass” architecture with/without skip connections (without for MNIST-SVHN, with for Winter-Summer).

MNIST-SVHN Results:
I measured the accuracy for translating MNIST images to SVHN images.

	DIP based	OST	CycleGAN
MNIST to SVHN	47.80	23.50	12.00
SVHN to MNIST	28.00	23.50	10.50

MNIST to SVHN images:

SVHN to MNIST images:

We can see that for a simple domain the mapping using DIP concept shows good results.

Winter-Summer Yosemite Results[1]:
I measured the content loss and style loss between the source image and the translated image.

		DIP based	OST	CycleGAN
Content	Winter to Summer	12.13	6.84	3.74
	Summer to Winter	11.77	10.25	3.07
Style	Winter to Summer	15.04	2.27	2.51
	Summer to Winter	31.51	8.20	3.20

Winter to Summer images:

Summer to Winter images:

Overall, we can see that the network did learn some of the summer/winter features but the quality of the images is not good enough. The images tend to be more blurry than the original image and the results are not good.
I have tried many ways to improve the results, such as changing hyperparameters or changing network architecture, but the results did not improve.

Further research:
Maybe the quality of the images can be improved by using another DIP network that does super-resolution.

[1] All the results were taken from “Bidirectional One-Shot Unsupervised Domain Mapping”.

Conclusion:

This work tries to create a mapping from domain A to B using DIP concept, where A contains only one image and B contains many images. The motivation is that unlike OST, in this work we perform only a few iterations in the second part and thus use less training time.

Based on the results we conclude that while the DIP concept is sufficient for mapping between simple domains, the results are not good enough for complex ones.

Bibliography:

1. Ulyanov, Dmitry, Andrea Vedaldi, and Victor Lempitsky. "Deep image prior." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
2. Benaim, Sagie, and Lior Wolf. "One-Shot Unsupervised Cross Domain Translation." *Advances in Neural Information Processing Systems*. 2018.