Due Date : December 8th, 2023 at 11 :00 pm

Question 1. (**VAE**)

1.

2.

3. After having trained the model, we ask you to provide the following additional results from your experiments. Please provide

    (a) Training samples :

(a) Epoch 15

(b) Epoch 20

(c) Epoch 24

(d) Epoch 39

(e) Epoch 44

(f) Epoch 49

FIGURE 1 – VAE training samples at various epochs

(a) General Reconstruction on 24 epochs

(b) General Reconstruction on 49 epochs

Figure 2 – VAE reconstructions

(b) The samples also appear somewhat squished during training. This distortion might be due to the way the VAE is encoding and decoding the data, possibly hinting at an imbalance in the representation learning in the latent space.

In the training samples of epoch 25 with the VAE, the samples exhibit a degree of blurriness and wobbliness. This suggests that the model might still be refining its ability to capture fine details, which is common in early learning stages or could indicate a need for model adjustment. Indeed, at the epoch 50, the wobbliness is reduced but the bluriness is still a bit present.

The digits line width varies, with some appearing bold and others thinner and slightly transparent. Notably, this inconsistency diminishes when training is extended to 50 epochs, leading to a more uniform line width in the numbers.

There's a specific issue in reconstructing the digit '6', where its orientation is opposite sometimes. This could point to a challenge the model faces in learning certain orientations. Even humans will mistake the '6' and '9' depending on the orientation. Although the model seems to have a better understanding at the 50th epoch.

Furthermore, the samples appear somewhat 'squished' during training. But the issue is not noticeable in the reconstruction. The reconstructions are much clearer and readable but do maintain a varying line width leading to bluriness.

(c) Interpolation on 50 epochs :

(a) Interpolation 1

(b) Interpolation 2

(c) Interpolation 3

(d) Interpolation 4

(e) Interpolation 5

(f) Interpolation 6

FIGURE 3 – VAE latent space interpolations

We observe that the interpolation between two points is smooth, with each image displaying minor, gradual changes that cumulatively lead to a different representation of a digit. Although it may not be immediately clear what the starting point is, such as in the third image, the transition progresses smoothly, ultimately resembling recognizable digits like ’9’ or ’1’.

Question 2. (**GAN**)

1.
2. After having trained the model, we ask you to provide the following additional results and insights from your experiments. Please provide
   (a) Training samples :

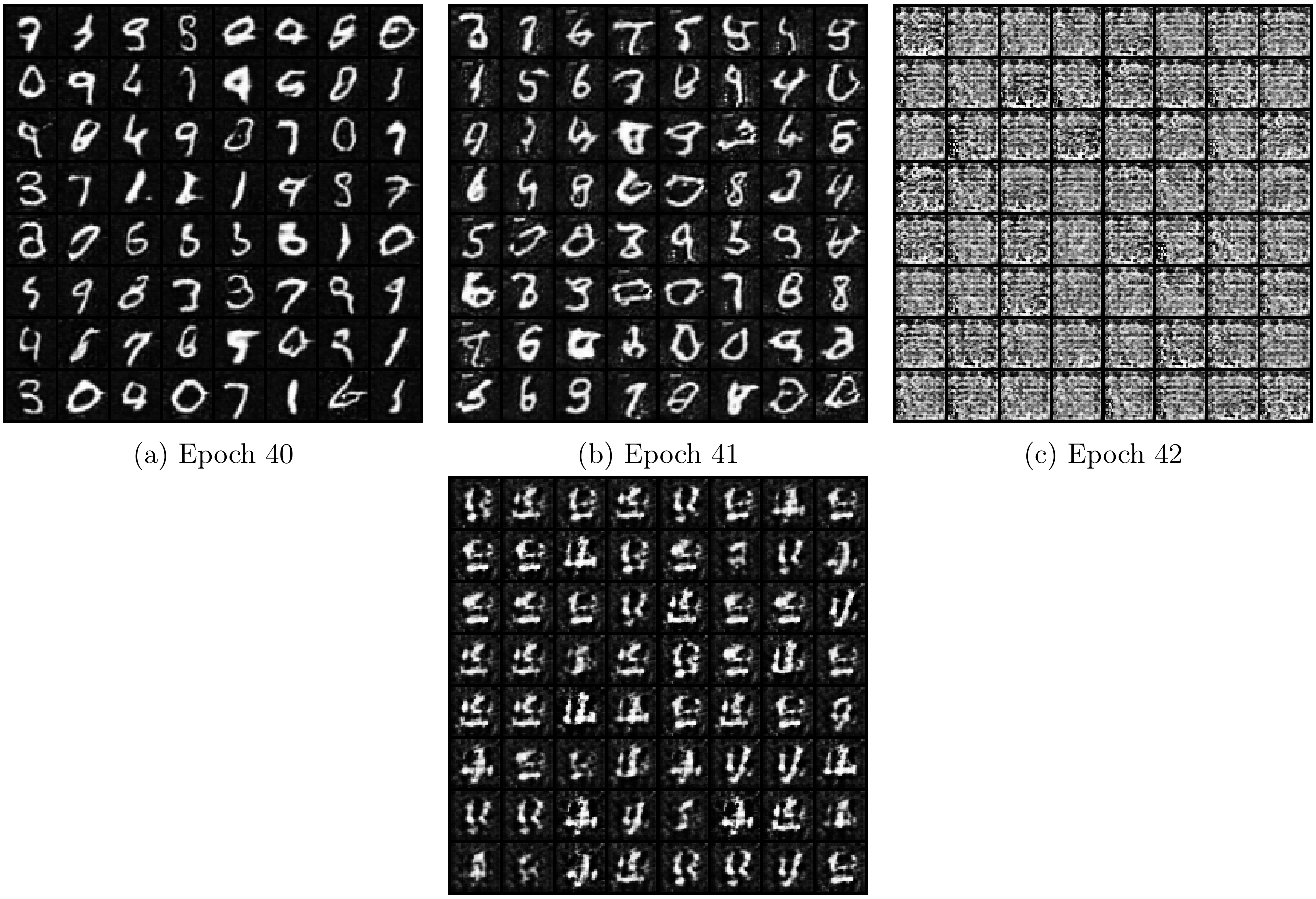
(a) Epoch 15

(b) Epoch 20

(c) Epoch 24

FIGURE 4 – GAN training samples at various epochs (out of 25)

I also tried training on 50 epochs and got some interesting results after the 40th epoch :



(a) Epoch 40 (b) Epoch 41 (c) Epoch 42

(d) Epoch 49

FIGURE 5 – GAN various training samples after 25 epochs

(b) Here the digits have wobbly edges, significant distortions and are nearly unrecognizable. However, these samples are not as blurry as the ones from our VAE and the model is trying to create new ways of writing the digits, which would explain the difficulty to recognize them. We notice artifacts around the digits, particularly when the model is not fully trained. The model seems to have 'exploded' after 42 epochs suggesting that the training process became unstable. This might be because of the discriminator being too powerful since in our GAN versiion we just alternate between one step of the discriminator and one step of the discriminator but in practice, performing more steps on the discriminator for each step of the generator has shown some improvements.

   Overall, compared to our VAE samples, the VAE samples tend to be more blurred but also more consistent in the quality of reconstruction. The GAN-generated images, especially in the middle epochs, show sharper edges but suffer from instability, leading to anomalies and noise generation observed in the latter epochs.

(c) Interpolation images on 25 epochs :

(a) Interpolation 1 (b) Interpolation 2 (c) Interpolation 3

Figure 6 – GAN latent space interpolations

I would say the interpolation is smoother than VAE because we see less changes through the images. The changes are less significant but we still see a gradual change cumulating to a different representation of a digit.

3. When training the discriminator, we want to update its weights without affecting the other neural network being the generator. When we're using '.backward()' on the loss, it computes the gradients for all tensors that have told to have gradient active. If we didn't use '.detach()', the backward pass would also calculate gradients for the generator's parameters because the fake data it generated was used to compute the discriminator's loss. Meaning that calling '.backward()' on the discriminator's loss would affect the weights of the generator.

   Since the generator and discriminator are in competition, the generator is trying to produce data that is indistinguishable from real data, while the discriminator is trying to get better at distinguishing real data from fake. If the generator's weights were adjusted during the discriminator's training, this would interfere with this adversarial training process.
4. (a) The GAN model is not really designed to reconstruct an input images, they're meant to produce new samples from the learned data distribution. It therefore makes sense that our samples are unrecognizable since the generator's goal is to create images that are realistic enough to fool the discriminator, not to recreate a given input image.

   (b) They do not explicitly model the data distribution, they instead learn to generate saamples from a distribution that approximates the true data distribution. As a result we cannot really use them to calculate approximations of the log-likelihood.

   (c) However, they are effective in learning features that capture the underlying distribution of the data. In our case, where the discriminator is alternated with the generator more frequently, this will leadd to a more robust learning of data representations, as the discriminator is constantly adapting to the improving generator. However, the effectiveness of this approach for representation learning will still depend on the specific architecture and the nature of the data.

Question 3. (**Diffusion Models**)

1.
2. After having trained the model, we ask you to provide the following additional results and insights from your experiments. Please provide
   (a) Training samples :

(a) Epoch 15 (b) Epoch 20 (c) Epoch 24

FIGURE 7 – DDPM training samples at various epochs (out of 25)

(b) The samples produced by the diffusion model appear sharper and more representative, showing less jaggedness compared to those generated by GANs. However, they are not as adept at accurate reconstruction as the VAE, although precise reconstruction is not the primary objective of diffusion models. We observe the recurring challenge in generating certain numbers, particularly with the '6'. These issues might diminish with further training. Overall, the output from diffusion models tends to be less blurry than that of VAEs and sharper than that of GANs.

I tried training on 50 epochs but forgot about it when it was running and didn't run the cells to transfer to my Drive so I lost the sample images :(

Question 4. (**Generative Models ; benefits and pitfalls**)

Now that you have implemented the three popular generative models : VAE, GAN and DDPM, can you comment on the following

1. (6 pts) What are the biggest strengths and draw-backs (in terms of quality and diversity of the generated samples, computational load at training and inference, etc.) of
   — VAE :

     As we saw, they seem to be more stable during training without the saturating gradients problems often seen in GANs. Additionally, they also are good at reconstructing the input data. Another strength, it is much faster to train than GANs and especially diffusion models.

     However, as observed and as it is commonly known with VAEs, the generated samples often lack sharpness and are blurrier compared to GANs or DDPMs. The fact that they are more able to reconstruct the input data might limit the model's "imaginatiion" to create new data.
   — GANs :

     They will be able to generate higher quality and new representation images. This will lead to more diverse distributions of the data while respecting the main idea of what defines the subject in the data.

As we experienced when training longer, GANs often suffer from training instability, leading to issues like mode collapse. Furthermore, the training will be computationally heavier and longer due to the fact we have two networks for our adversarial training process.

— DDPM :

They produce very high-quality samples, surpassing GANs in terms of realism and detail. Furthermore, they are generally more robust to training and to changes in hyperparameters compared to GANs. From what we've seen in class, they are more versatile and can be adapted to differente types of data or tasks.

The biggest drawback I (and my Google Collab credits) felt was the training time. It is computationally demanding and intense on the VRAM. Furthermore, the slow inference time contributes to that training tume, as generating samples involves the sequential process of iterative denoising.