Machine Learning 2

Assignment 4

Question 1 - Generative models - GAN

1.1) Model architecture description and illustration, training procedure(hyperparameters, optimization details, etc.).

This GAN follows the DCGAN design principles for generating 64×64 color images. The **generator** starts with a latent noise vector and uses a series of transposed convolutional layers (with increasing feature map dimensions and spatial upsampling) combined with batch normalization and ReLU activations, culminating in a Tanh activation to output an image with three color channels. The **discriminator** takes an input image and passes it through successive convolutional layers that downsample the spatial dimensions while increasing the number of features; it uses LeakyReLU activations (with batch normalization in intermediate layers) and ends with a convolutional layer that outputs a single scalar value, indicating whether the input image is real or fake.

I trained three GAN models with latent space dimensions of 10, 100, and 500, respectively, to investigate the impact of latent dimensionality on the quality and diversity of the generated images.

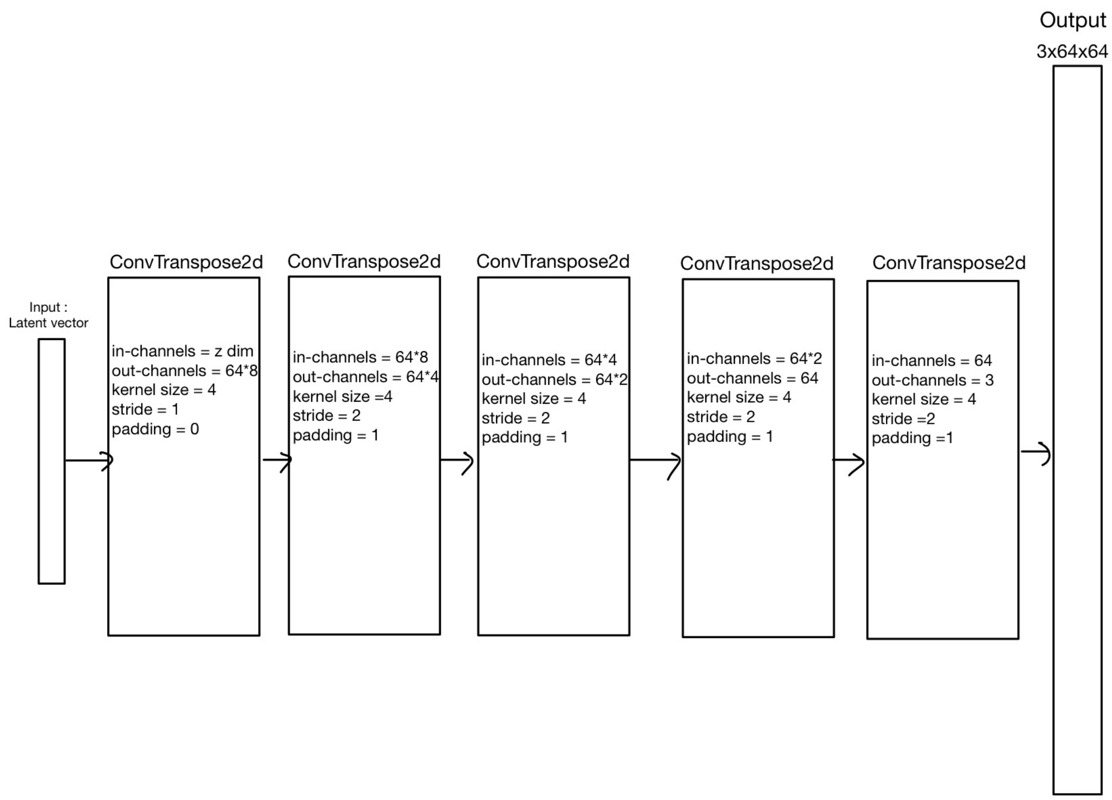
A latent dimension of 10 provides a very compact representation of the data distribution. While this may simplify the model and reduce computational demands, it also limits the generator's capacity to capture intricate variations within the data, often resulting in less diverse and overly simplistic outputs.

Conversely, a latent dimension of 500 offers a highly expressive space that can potentially encapsulate a greater variety of features and subtle details from the data. However, this increased capacity comes at the cost of added complexity, which can lead to training instabilities such as mode collapse, where the generator may repeatedly produce a limited set of outputs.

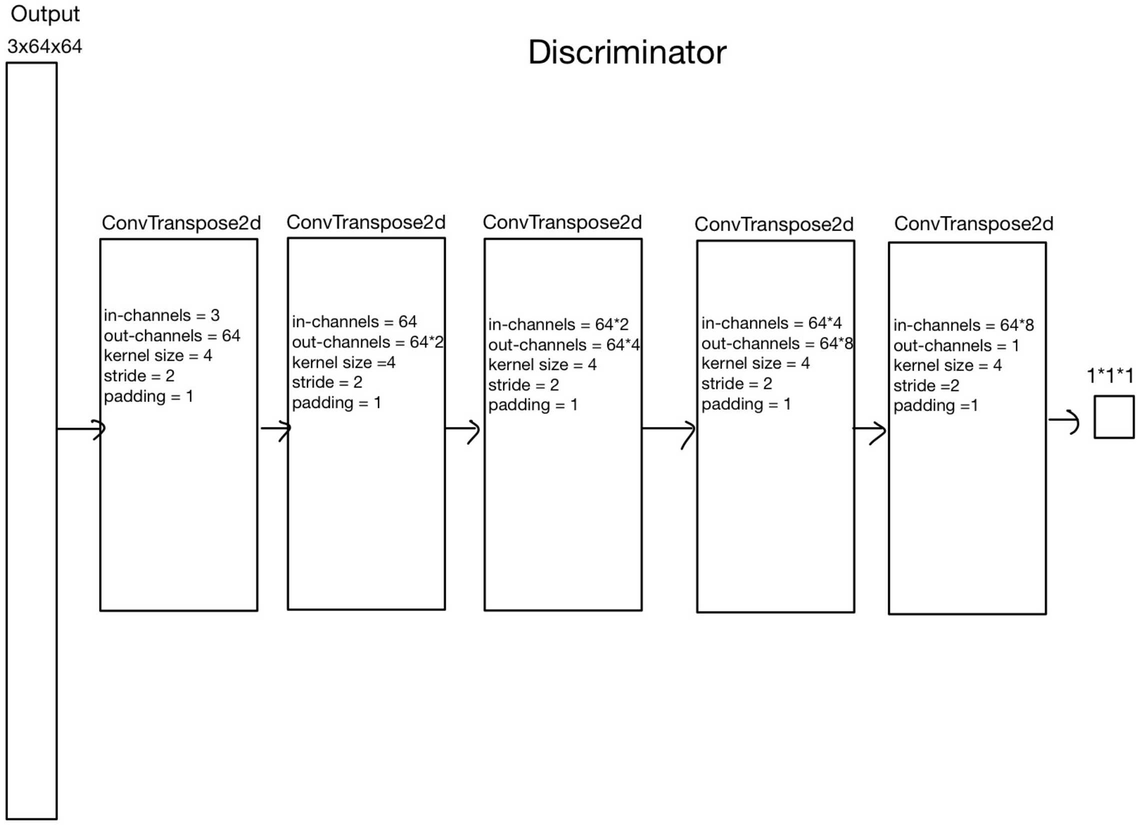
The model with a latent dimension of 100 represents a balanced configuration. It provides sufficient capacity to generate high-quality and diverse images while maintaining a relatively stable training process. This configuration is often seen as a trade-off between the extremes of over-simplification and excessive complexity.

Comparing these three configurations is particularly interesting, as it highlights the trade-offs between model capacity, diversity of generated samples, and training stability.

The Generator Architecture :



The Discriminator Architecture :



The Training Procedure :

The training procedure is structured to run over 50 epochs using a batch size of 128, with the Flowers102 dataset resized to 64×64 and normalized appropriately. Both the generator and discriminator are optimized using the Adam optimizer with a learning rate of 2e-4 and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$, which are standard choices for stabilizing GAN training. The loss function employed is the Binary Cross-Entropy with Logits (BCEWithLogitsLoss), where the discriminator is trained by minimizing the combined loss over real images (labeled as 1) and fake images generated by the generator (labeled as 0), while the generator is simultaneously optimized to produce images that the discriminator classifies as real. This setup is repeated for three different latent space dimensions (10, 100, and 500), allowing for a comparative analysis of how latent space capacity influences the generator's expressiveness, image quality, and overall training stability.

1.2) Training convergence plots as a function of training time:

o GAN: discriminator and generator losses

When **Latent Dimension = 10** :

The generated images from the GAN with **z_dim = 10** exhibit several notable characteristics:

1. **Blurriness and Lack of Detail**: The images are clearly blurry, with soft edges and an overall lack of fine-grained details. This suggests that the generator is struggling to represent complex structures.

2. **Recognizable Color Patterns**: While the images do not capture precise floral structures, some display color patterns that resemble flowers, with patches of red, yellow, blue, and green. This indicates that the generator has learned some aspects of the dataset's distribution but fails to generate clear and high-fidelity outputs.

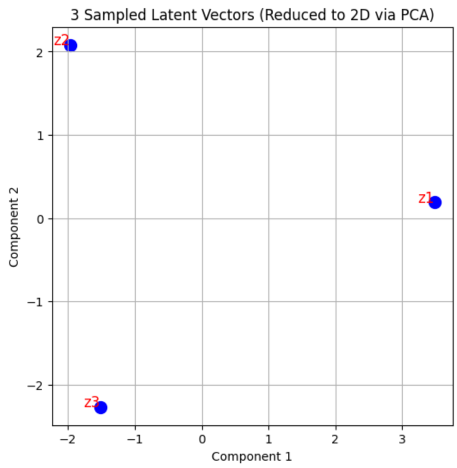
3. **Low Diversity in Structural Representation**: The shapes appear repetitive, with similar blob-like structures across different samples. A small latent space (z_dim = 10) constrains the generator's ability to encode a rich variety of patterns, leading to reduced diversity in the generated samples.

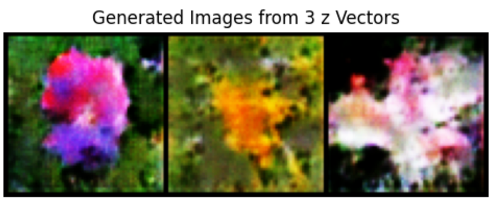
4. **Artifacts and Noise**: Some images exhibit noticeable artifacts, with regions of unnatural texture or color smearing. This could be due to mode collapse, where the generator learns to produce only a limited set of similar outputs, failing to explore the full distribution of possible flower images.

These issues may be caused by the limited capacity of the latent space.

(Note : I understood from the forum that this method was admitted to visualize the latent distribution of the GAN)

Now let's sample 3 vectors from the latent vectors distribution. We will visualize those vectors projected on their two main components, using PCA method and then we will take a look on the corresponding generated images.





Observations on the PCA Projection:

1. **Distinct Separation in Latent Space**: The three sampled latent vectors are well-separated in the 2D PCA projection, suggesting that even with a small latent dimension (z_dim = 10), the latent space is diverse and spans different regions of the data distribution.

2. **Potential for Smooth Transitions**: If additional intermediate points between these vectors were sampled and visualized, it would be interesting to observe whether there is a smooth transition between generated images, indicating a well-structured latent space.

Observations on the Generated Images:

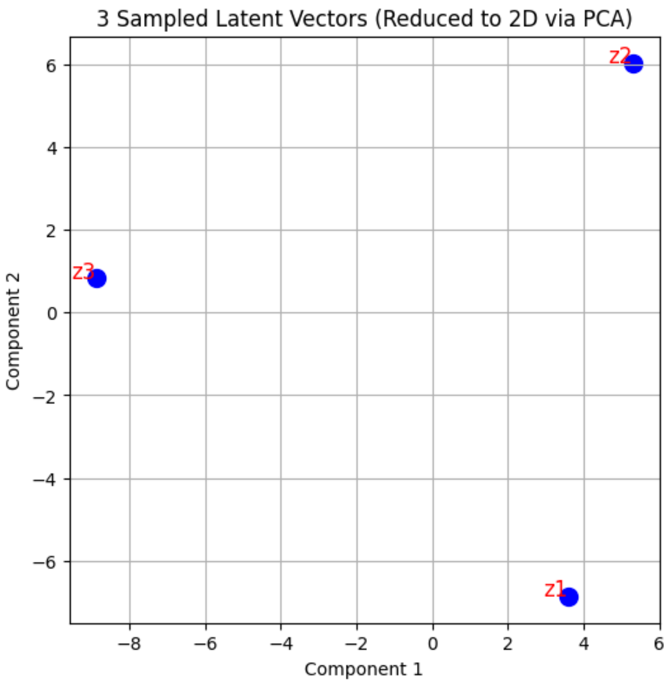
1. **Visual Diversity**: The three generated images exhibit notable differences in color and structure. This suggests that, despite the small latent dimension, the GAN is capable of producing distinct outputs.
2. **Blurred Features and Low Fidelity**: The images remain blurry, lacking fine details and realistic textures. This aligns with previous observations that a low latent dimension (z_dim = 10) limits the generator's capacity to encode complex image features.
3. **Color Consistency Across Samples**: The generated images still capture relevant color distributions, with floral-like hues. However, the lack of precise shape formation suggests that the generator is struggling to learn clear boundaries and structures.

When **Latent Dimension = 100** :

Generated images (z_dim=100)

The generated images from the GAN with **z_dim = 100** exhibit several notable characteristics and some improvements compared to the GAN with z_dim = 10:

1. **Increased Diversity in Color and Structure**: Compared to z_dim = 10, these images display a broader variety of colors and structures. There are more distinct floral shapes and a greater range of colors that align with real flowers.
2. **Slightly Sharper Features**: While the images are still blurry, some shapes appear more defined. This suggests that a larger latent space allows the generator to encode finer details, though it still struggles with high-frequency details.
3. **Better Contrast and Texture**: The textures are less uniform than in the z_dim = 10 case, where blobs of color dominated. The model now captures some of the intricate color variations found in flowers.



Generated Images from 3 z Vectors

Observations on the PCA Projection:

1. **Wide Spread of Latent Vectors**: The three sampled latent vectors are well-separated in the reduced 2D space, with significant distances between them. This suggests that the latent space with z_dim = 100 is more expressive compared to z_dim = 10 and allow a bigger variety of generated outputs.
2. **Potential for Smooth Transitions**: The spread in latent space indicates that interpolation between these vectors could result in diverse image variations. However, the extent to which the interpolation produces realistic transitions would need further testing.

Observations on the Generated Images:

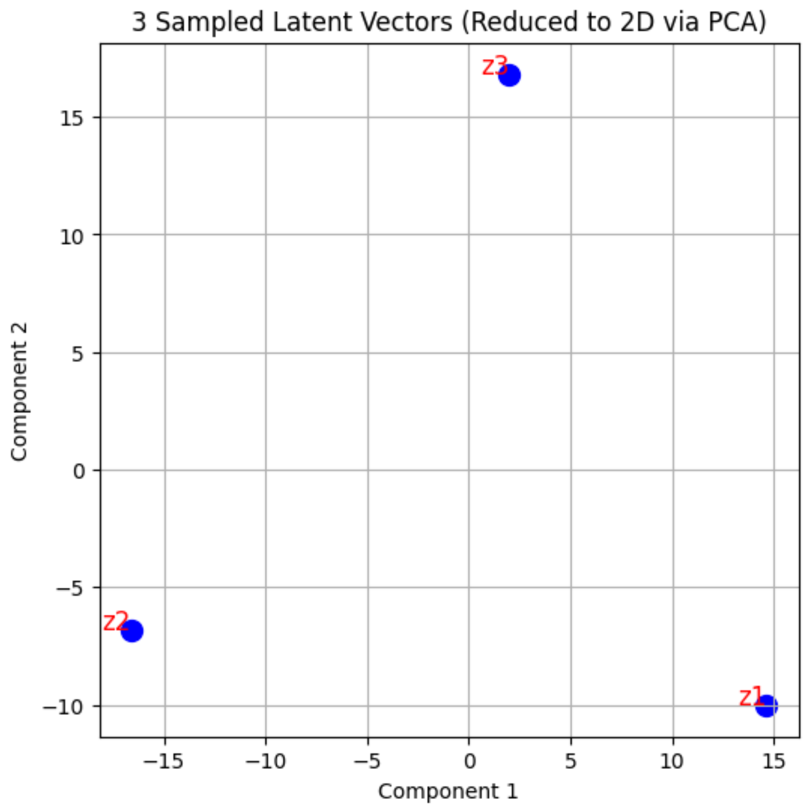
1. **Increased Variability in Structure and Color**: Compared to z_dim = 10, these generated images display more distinct floral structures and better color separation. The model is learning to capture variations in flowers more effectively.
2. **Still Some Blurriness**: While shapes are more recognizable, the generated images still exhibit a lack of sharpness and fine details. This suggests that although the larger latent space allows for more expressiveness, the generator still struggles to capture intricate textures.
3. **Better Contrast and Definition**: The colors are more vibrant, and there is a better distinction between background and foreground elements. Some images resemble flowers more clearly compared to those from z_dim = 10.

When **Latent Dimension = 500** :

Generated images (z_dim=500)

The generated images from the GAN with **z_dim = 500** exhibit several notable characteristics and some improvements compared to the GAN with z_dim = 10 and z_dim = 100:

1. **No Significant Sharpness Improvement**: The images remain blurry and lack well-defined structures. Despite the increased latent dimension, the generator does not seem to utilize the extra information to enhance details.
2. **Increased Noise and Artifacts**: Some images exhibit increased graininess and color smearing, suggesting that a high-dimensional latent space might introduce redundant or unstable variations that do not contribute to meaningful feature representation.
3. **Slightly More Diversity**: There is a noticeable variety in colors and patterns, but the floral structures are not significantly more detailed compared to z_dim = 100.



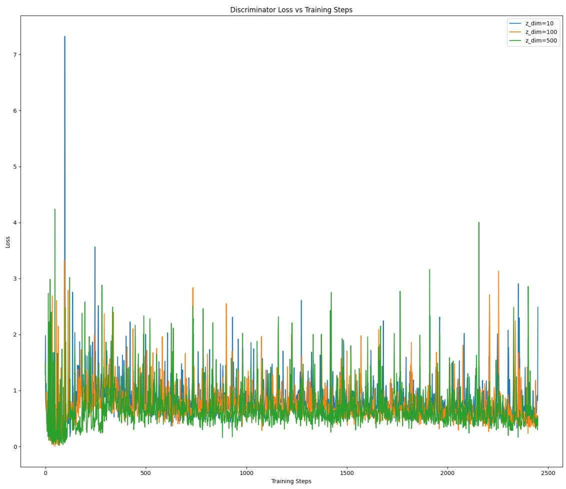
Observations on the PCA Projection:

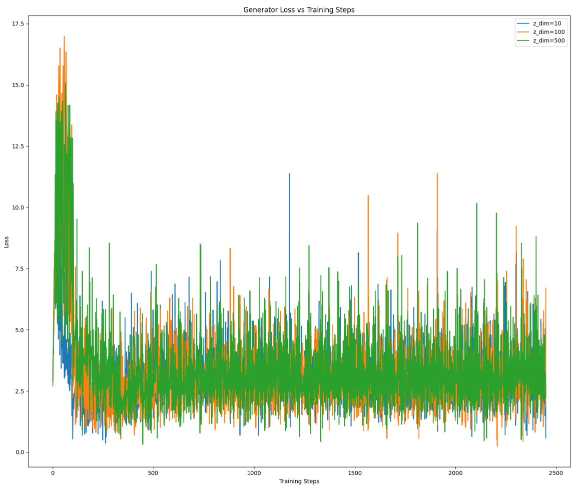
1. **Increased Spread in Latent Space**: Compared to previous experiments with z_dim = 10 and 100, the three sampled vectors in the z_dim = 500 space exhibit a much larger spread in the PCA-projected 2D space. This suggests that with a higher-dimensional latent space, individual latent vectors are mapped to significantly different regions, which may introduce additional variance in the generated images.
2. **Potential Over-Dispersal**: While diversity is desirable, too large a spread can indicate that the latent vectors are not being efficiently utilized by the generator. This could lead to unstable or overly noisy outputs, as observed in the generated images.

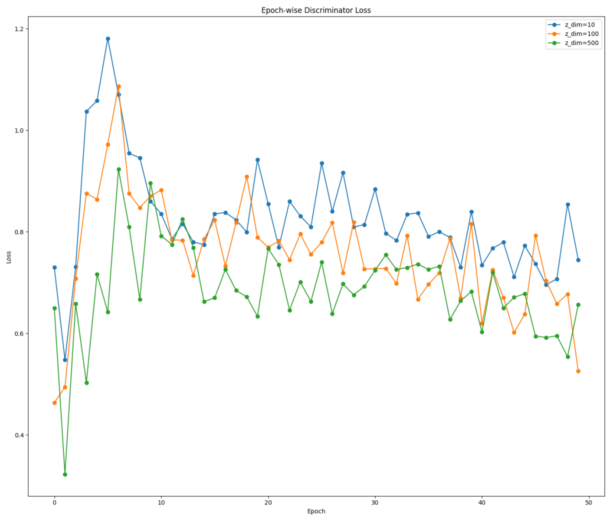
Observations on the Generated Images:

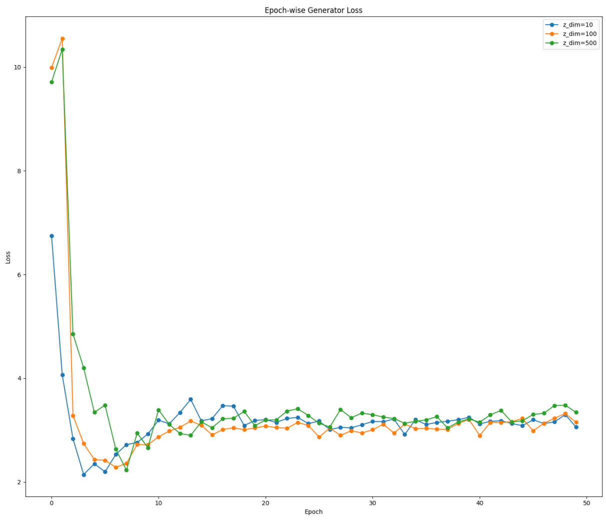
1. **No Significant Detail Improvement**: Despite the larger latent space, the images do not exhibit finer details compared to z_dim = 100. The structures remain blurry, and the generator does not seem to take full advantage of the expanded latent space.

2. **Increased Variability in Colors and Patterns**: The images display more color diversity, with some clear distinctions between different floral regions. However, this does not translate into a substantial improvement in perceptual quality.
3. **Potential Loss of Coherence**: Some images appear more fragmented or noisy, suggesting that the latent space might be too high-dimensional for the generator's capacity, leading to difficulty in learning a structured mapping.









Comments on **losses** plot for both **Generator** and **Discriminator** as a function of **Epochs** :

Observations on Discriminator Loss

1. **Early Instability in Training:**
   - In the first few epochs, the discriminator loss exhibits sharp fluctuations across all z_dim settings. This is expected, as the generator starts by producing poor-quality images, making it easy for the discriminator to distinguish real from fake samples.
   - The loss rises initially, indicating that the discriminator is learning to differentiate better, then stabilizes as the generator improves.

2. **Gradual Stabilization:**

    - After around **10 epochs**, the discriminator loss stabilizes for all latent dimensions, fluctuating around a steady value.
    - The **z_dim = 500 (green)** curve appears slightly lower than the others, suggesting that for a very high-dimensional latent space, the discriminator might be learning more effectively, but this does not translate to better generation quality.

3. **Slightly Higher Loss for z_dim = 10 (Blue Curve):**

    - This suggests that with a smaller latent space, the generator struggles to create diverse images, leading to easier classification for the discriminator, hence a **higher discriminator loss**.

Observations on Generator Loss

1. **Sharp Initial Drop Across All z_dim Values:**

    - The generator loss starts **very high (~10)** in the first epoch, then drops sharply within the first **5-10 epochs.**
    - This is expected as the generator starts from noise and gradually learns to produce samples that fool the discriminator.

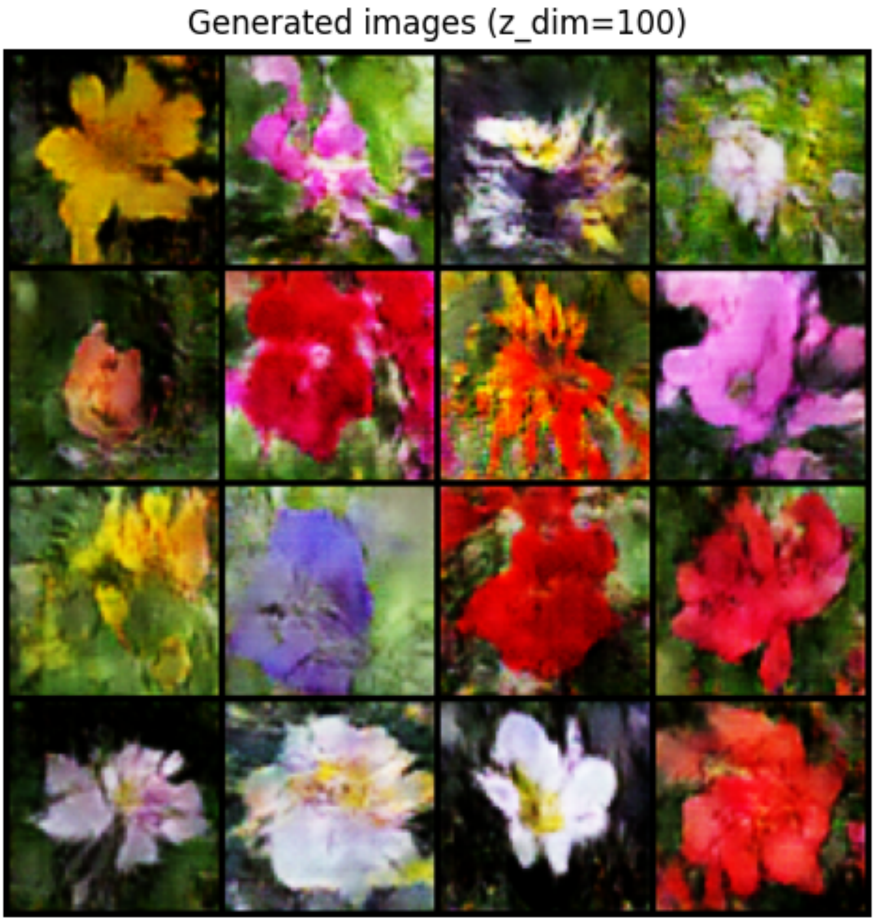
2. **Stable Convergence After 10-15 Epochs:**

    - After the initial drop, generator loss stabilizes at around **2-3**, suggesting that the GAN has reached a form of equilibrium.
    - There is no strong divergence, meaning the adversarial balance between generator and discriminator is maintained, avoiding **mode collapse** or **training failure**.

3. **No Clear Improvement with Higher z_dim:**

    - Despite increasing the latent space size from **10 to 500**, the generator loss does not exhibit significant differences.
    - This supports previous observations that simply increasing z_dim **does not necessarily lead to better performance**. In fact, **z_dim = 100 (orange curve)** appears to be slightly more stable than z_dim = 500.

When **Latent Dimension = 100** with 100 Epochs :



Observations on the Generated Images:

1. **Noticeable Improvement in Structure and Color Representation**
    - Compared to **z_dim = 100 with 50 epochs**, the images exhibit better-defined floral structures.
    - Some flowers have recognizable petal formations, and the contrast between the flowers and the background is improved.
2. **Increased Color Variety**
    - The model has learned to generate a wider spectrum of flower colors, from reds and yellows to blues and whites.
    - This suggests that the generator is capturing a broader distribution of the dataset.
3. **Remaining Issues: Blurriness and Detail Loss**
    - Despite the increase in training epochs, the images still lack **sharp details**.
    - Many flowers appear smudged, with edges blending into the background.
    - This suggests that the generator is **still struggling to model high-frequency details**, even after prolonged training.

1.3) Summary of your attempts and conclusions. Your conclusions and explanations should be based on the actual results you received during your attempts.

I trained GANs with three different **latent dimensions** (**z_dim = 10, 100, and 500**) and compared their performance over **50 epochs**, then extended training for **100 epochs** with **z_dim = 100** to evaluate the impact of prolonged training. Below is a summary of observations and possible explanations :

Latent Dim and Epochs	Image Quality	Diversity	Sharpness
Z Dim = 10 and 50 Epochs	Poor	Low	Low
Z Dim = 50 and 50 Epochs	Poor/Middle	High	Middle
Z Dim = 100 and 50 Epochs	Poor	High but with noise	Middle
Z Dim = 100 and 100 Epochs	Middle/Good	High	High

Now I will try to give explanations of the results of the experiments.

Below are some potential reasons on why extending the latent variables dimension does not always improve the generated images :

1. **Overparameterization Without Benefit**: When the latent dimension is too large relative to the complexity of the dataset and the model's capacity, it becomes harder for the generator to learn a structured mapping. The added dimensions might not provide useful variations but instead introduce noise, making training more challenging.
2. **Mode Collapse Risk**: A very high-dimensional latent space can make optimization more difficult, potentially leading to mode collapse (where the generator produces repetitive or less meaningful variations).
3. **Insufficient Training Time**: With z_dim = 500, the model might require more epochs to fully utilize the increased dimensionality, as it needs to learn a more complex mapping.
4. **Mismatch Between Generator and Discriminator Capacity**: Increasing the latent dimension without adjusting the generator's architecture (e.g., more layers or wider convolutional filters) might lead to under-utilization of the additional latent features.

Conclusion and directions for Future Works :

I explored the impact of latent dimension size (z_dim = 10, 100, 500). z_dim = 100 provided the best balance between diversity and stability, while z_dim = 10 was too restrictive and z_dim = 500 introduced noise without clear benefits. Increasing training time to 100 epochs improved diversity but did not resolve blurriness, highlighting the importance of architectural choices over sheer latent space size or training duration.

Future work could be attempts to improve the generator's architecture and adapt it to the latent variables dimension, exploring more loss functions and tuning other hyper-parameters

Important notes : 1) Ziv Tamir accepted my request to submit alone.

2) I submitted the GAN with z_dim = 100 and 100 epochs. I combined the Generator and Descriminator in the same pkl file. To reload it, here is the script :

```
models = torch.load("/your_path/HW4_931215248.pkl", map_location="cpu") # or where ever you want
generator = models["generator"]
discriminator = models["discriminator"]
```