Generative Adversarial Networks (GANs)

IMPORTANT NOTES:
Study lecture materials at least 1 hour and prepare the questions prior to the tutorial session.
The questions will be discussed in the tutorial session.

1. How does the adversarial training dynamic in GANs resemble a security game, and what challenges does this analogy introduce in terms of model stability and convergence?

   The adversarial setup of GANs mirrors a security game where the generator (G) tries to fool the discriminator (D), akin to an attacker evading detection. This dynamic introduces instability because improvements in one player can destabilize the other. Unlike traditional optimization, the goal is not a single minimum but a Nash equilibrium, which is harder to reach and maintain.

2. Given the minimax nature of GAN training, discuss how the interdependence of generator and discriminator loss functions affects the learning process. What strategies can be used to stabilize training?

   The generator's loss depends on the discriminator's ability to detect fakes, and vice versa. This mutual dependency can cause oscillations or collapse if one player dominates. Stabilization strategies include using feature matching, label smoothing, or loss functions.

3. CycleGAN uses cycle consistency loss to enforce semantic preservation. Propose a scenario where this loss might fail to preserve meaningful content, and suggest how the architecture could be modified to address it.

   Cycle consistency loss is a key component in CycleGAN that ensures if an image is translated from domain X to Y and then back to X, the result should resemble the original image. If domain X and Y have ambiguous mappings (e.g., X: abstract art, Y: real photos), cycle consistency might enforce incorrect semantic preservation. To address this, one could perform the following:

   - Use perceptual loss to compare deep features extracted from a pretrained network instead of comparing raw pixels.
   - Introduce attention mechanisms that allow the model to focus on important regions (e.g., facial features) during translation and reduces the chance of losing critical semantic information.

4. You observe that your GAN consistently generates high-quality but visually similar outputs. Identify the likely cause and propose a multi-step strategy to improve diversity without sacrificing realism.

   This suggests mode collapse. Mode collapse occurs when the generator learns to produce only a limited subset of outputs that successfully fool the discriminator. Instead of capturing the full data distribution, it focuses on a few "safe" modes. Lack of diversity means the model is not learning the true data distribution, which limits its usefulness and generalizability.

   To improve diversity, one could:

   - Use minibatch discrimination to introduce a layer in the discriminator that compares batches of images. This helps the discriminator detect lack of diversity and penalize repetitive outputs.
   - Introduce noise injection or dropout to encourage variability. Use instance noise during training to prevent overfitting to specific patterns.
   - Use multiple discriminators, each focusing on different aspects (e.g., texture, shape), makes it harder for the generator to exploit a single weakness.

5. Compare the implications of using a GAN vs a CycleGAN for translating CT scans to MRI images. What factors would influence your choice of model architecture and loss functions? The choice between GAN and CycleGAN depends on data availability, clinical requirements, and desired output fidelity. For unpaired data like CT scans and MRI images, CycleGAN is powerful but must be carefully tuned to preserve medical semantics using suitable loss functions. Loss functions should include adversarial loss to ensure realism, cycle consistency loss to ensure reversibility, identity loss to preserve anatomical features and perceptual loss to encourage semantic similarity using deep features (e.g., VGG).

6. You are evaluating two GAN models: Model A with low FID but high PPL, and Model B with moderate FID and low PPL. Which model would you choose for a generative art application, and why?

   Model A: Low FID, high PPL

   - Low FID indicate the model is able to generate very realistic images.
   - High PPL means unstable latent space, small changes in input lead to unpredictable or jarring changes in output. Difficult to control or interpolate smoothly between styles or concepts.

   Model B: Moderate FID, low PPL

   - Moderate FID shows that the images may be less sharp or slightly less realistic
   - Low PPL means smooth latent space, this is ideal for interpolation, animation, and style blending. Easier to explore and manipulate for creative purposes.

   Model B is better for art, as smooth transitions in latent space allow creative control. High PPL in Model A may produce erratic changes, limiting usability in style exploration.