Rensselaer Polytechnic Institute
Department of Electrical, Computer, and Systems Engineering
ECSE 4966/6966: Computational Creativity, Spring 2025

Homework #1: due Friday, January 31st, at 5 PM.
Show all work for full credit!

Submit your homework writeup using Gradescope, and your images/video/code to the `HW1\yourteamname` folder on Box that you will share with me. Your team only needs to submit one assignment on Gradescope (there is a way to indicate who is on what team when you upload). Note that these homeworks are meant to be true team efforts; that is, all of your members should be participating in solving each problem, as opposed to assigning one problem per person!

1. (40 points) Create an interesting dataset of images that you will use to train a custom VAE in this assignment and custom GANs/diffusion models in HW2. What I'm looking for here is a fun, unique, and personal dataset you collect and curate yourself. What I *don't* want is an off-the-shelf ImageNet/Kaggle dataset that someone else collected using a web scraper. The goal is for you to get your hands dirty and be intentional about what you put into the dataset. Using your own images will also make you think more about the relationship between your inputs and outputs and to shape the results to your liking.

   This video from Eryk Salvaggio does a great job of describing how to collect a good dataset for training, so I suggest you take a look at this first. As the video suggests, I want you to collect at least **500** images (before any sort of data augmentation) for your dataset – and more is better! This sounds like a lot, but keep in mind these are not going to be very interesting pictures. Here are some guidelines:

   - All images should be 1024×1024. If you're taking pictures with your phone, you can put it in square mode to save time. You can also run a batch job in an image editor (e.g., IrfanView or Photoshop) after the fact to do the cropping/resizing.
   - Your images should have similar backgrounds and they should all "feel the same". That is, you don't want extraneous objects or very different lighting/framing/compositions in the dataset. Remember that you're trying to make it easy for an algorithm to identify the critical patterns that define a class of objects.
   - You want both a lot of redundancy (i.e., the same object from different perspectives) and a lot of diversity (i.e., enough different objects that the generative model isn't going to just produce the same image each time). Keep in mind that the balance of the images you put into your dataset is going to translate into what the generative model will produce.
   - Don't try to make the photographs interesting or aesthetic; no fancy composition, no post-production, no filters, nothing artistic. The idea is to think more like a machine than as a photographer.

   After acquiring your base set of images, you should expand the dataset with typical augmentations like flipping the images from left-to-right, zooming in slightly, or rotating. Just make sure that whatever augmentations you apply make sense. For example, a face can be flipped left-to-right but not a sign containing text, and neither can be rotated. It's probably better to save the augmentations in a different directory (and again to automate this process).

   In your write-up, describe your data collection, curation, and augmentation process. Why did you choose these images in the first place? Especially as you put these images through the VAE in the

next problem, discuss whether you had to go back and take more or different kinds of images to get a satisfying result. Obviously you can't include all of your training samples in your writeup, but include a 20–30 image subset to give me an impression of the dataset.

2. (30 points) Train a VAE on the data you collected in Problem 1. Depending on your team's computational setup, you could do this in several different ways. You can find and clone a GitHub repo and run it on your own computer. You can modify a Colab notebook that runs the code in the cloud if your own computer isn't powerful enough. Maybe you can even use RPI's powerful AiMOS supercomputer!

   (a) Describe the pipeline you used to train and apply the VAE. What kind of VAE was it (e.g., vanilla or VQ)? What were the dimensions of your latent space? How long did it take? Did you have to lower the resolution of the data? What problems did you encounter and overcome? What did you learn that you would tell another student trying to do the same thing?

   (b) Pass 10 original images through the encoder and decoder and provide a comparison of the original vs. reconstructed images. Critically assess the results.

   (c) Generate 20–30 examples of new images by sampling from the VAE in latent space, and critically assess the results. What features of your input dataset did the VAE seem to pick up on? Are the images reasonable?

   (d) Demonstrate some type of latent space arithmetic using your VAE. For example, if there is a natural feature in your data, you could estimate the vector corresponding to that feature and move in the positive or negative direction of this vector to generate new samples (see Foster p. 91). Or, you can interpolate between locations in latent space to create a morph between output images (see Foster p. 92).

   Your images in this problem may be blurry; keep in mind that VAEs are only the first image generation method we'll study. The images you create from the same dataset in Homework 2 will hopefully be much sharper.

(continued on next page)

3. (30 points) In the second half of the semester, you will create a final project of your own design, leveraging at least three of the concepts from the class (at the very least something involving both generative images/video and generative text). While we've only just started, I'd like you and your team to brainstorm what you think your final project will look like, based on what you know we'll be discussing over the course of the semester. I'm expecting that the result will be some sort of video (even if still images are involved).

   The project is meant to be creative, not just a lab demonstration of an algorithm. You might consider making a music video, planning a short live-action video, remixing or reprocessing already-shot footage from TV shows or movies, or recording an installation/performance that includes real-time visual effects. The final project should go beyond results from the homeworks, requiring you to write new code to integrate the pieces and create new effects.

   Since this is expected to be a major project that you work on throughout the semester, you will submit a progress report later in the semester to keep you on track. The overall project grade will be computed as 10% proposal (this problem), 20% progress report, 35% final report, and 35% project evaluation.

   - What are the effects that you hope to create in your project? How do you think the effects will support the narrative of the video? (that is, why are the effects you propose suitable to the context?)
   - An initial storyboard that walks us through what we will see as the video progresses. You can use crude stick-figure sketches, still images, frames of video with arrows and such drawn on top of them in a paint program, etc.
   - How are the effects related to one or more of the concepts in the book/course? What other related work is there from the technical or artistic literature that is related to your idea? (research papers, examples of similar effects in videos you've seen).
   - Where will you collect your raw image/video/audio data from?
   - What software do you anticipate using to accomplish your goals? What type of coding will need to be done to accomplish the effect? Note that I don't want to see projects that are entirely composed using RunwayML/MidJourney, especially at the 6966 level; some sort of hand-written code is important to get "under the hood".
   - What is your strategy/plan of action for realizing your conceptual idea? That is, what tasks need to be performed? What is the rough schedule for these tasks? It is important to be realistic in this step. You shouldn't propose a highly complex project if you have no idea how to realize it.

   While I'm not going to hold you to every aspect of your proposal, you should be as careful and realistic as you can, especially as you develop a sense of how difficult real data is to acquire and process. This isn't the kind of project that you can leave until a few days before the due date (and remember it counts for 40% of the grade).

   I'll review the proposals and let you know if I foresee any problems. You might look through the proceedings of recent CVPR or SIGGRAPH conferences to get inspiration, or think about cool effects you've recently seen on Youtube/Instagram/TikTok.

Note that part of your responsibility on each homework is to constructively evaluate your peers' assignments via a Google Form that will be distributed after everyone submits the assignment. This peer assessment forms part of your participation grade, and your homeworks will not be graded until everyone on your team provides a peer assessment! We will work out the timing and kinks of this system after HW 1.