David Pettersson
david.pettersson@code.berlin
Semester 04
SE_15 Advanced Machine Learning

Design decision

Problem formulation

Our machine learning, in this case, is a bit further away than the usual machine learning problems. It is an artistic endeavour more than a business one. Hence the problem can be formulated as follows: how do I generate images that are related to the dataset in order to use them in a further context? And in order to be as complete as possible in the frame of this module, we can add the question: how important are metrics in this particular case?

The inputs to the model are random tensors in the latent space that generate images. The decision to use a certain random tensor compared to another one is purely based on an aesthetical point of view. This decision will be made between me and the artist I am collaborating with.

The output of this model will be the generated sequence of images - further processed into various video sequences which will be used in the context of an art installation. As outlined on the roadmap, the creation of another ML model trained on categorising the images will add another layer to the whole system.

Another problem, not solved because of purely hardware problems isHow can we train on full HD (1920x1080) images and later on generate them?" As outlined in the data point below, this was not possible as the used GPU didn't have enough memory to be able to do that.

Error metric

Being in this particular case, the metrics are considered purely on a curiosity level. I have included the ones that the StyleGAN2 model uses, to be able to judge if the generated pictures are close to the original photographs or not. Here are the metrics which have been computed from the last iteration of the network training.

- The first one is the **FID - or Fréchet Inception Distance (FID score)**. It is used to determine a visual similarity between two datasets of images. The FID score for this model shows us the disparity between our dataset and 'reality' represented by the 'inception_v3' model. We can assume that the score would lower with an augmentation of training time. The current score is **23.1470**

- The second metric is the **PPL or Perceptual Path Length**. Meaning that it measures the difference between consecutive images when interpolating between two random inputs. The lower the value, the more perceptually smooth the latent space is. The current score is **71.5810**
- The last metric is Precision and Recall. Precision is the ability of a classification model to identify only the relevant data points. Recall is the ability of a model to find all the relevant cases within a dataset. The current scores are **precision 0.0172 and recall 0.2132**

We can see here that the generated pictures are far from being realistic on a purely numerical and analytical level, which is aligned with the main goal of this project. As a reminder, our problem statement is to purely generate images, with the only condition that they have to be related to our initial dataset in a visual manner. The metrics are on the lower side, but the subjectivity (and the generated visuals) are telling another story.

It's also a lesson in learning to read the outcome of the model. Sometimes, metrics are not enough to be able to assess properly if our model is a failure or success. The flexibility of the human mind allows us to consider this model as a success in resolving our stated problem.

If the generation of non-existing faces or anything of that order being less abstract, more feature prone and more sensitive to the human eyes, these metrics would've been the compass from which I would've changed and tuned my model.

Data

The initial dataset was only composed of a certain type of wave images. To be more precise, they were waves videos captured by the artist Diane Drubay. She owns a remarkable collection of nature-based images and videos. In this particular case, we used pictures shot from the sea, to have only waves in the frame.

The videos have been cropped into a square format (1024x1024) and decomposed into JPEG sequences. This allowed us to have enough data to feed to the algorithm. The first try was done with only a couple of videos, which were all the same in their colors and general shapes.

The result of that training was an image synthesizing network, yes - but general diversity was affected and was not very interesting.

We went again in her footage and selected a more diverse variety of images in terms of colors, shapes, and angles - and used this as a training dataset. The result is the one that is being shown in this project.

Having high-resolution images as training dataset brings its own complications. The higher the resolution, the more memory the GPU needs for it to be able to store all the weights in memory. The chosen resolution was the maximum size we can use with the available (free) services.

Machine learning model

The ML Model chosen for this project is the StyleGAN2 from Nvidia, which learns the features of the incoming dataset and generates images based on these features. This model was chosen because of the easy availability of the training methods (RunwayML and Google Colab) and also because of the nature of the network proposed by Nvidia.

Hyper-parameters

As the initial training was made on a closed box in the form of RunwayML, I was not able to fine-tune the hyperparameters - which is a weakness in this case.

Results

In the frame of our problem, the whole project was a complete success as the generation of synthesized images closely related to our dataset has been possible to do. They are not realistic from a purely metrics side, but this was not the end goal. If the creation of waves that are realistic to hyper-realistic would be the goal, the tuning and optimisation of the hyperparameter would allow for a greater result. The training time is also a major factor in this case.

Coming back to those training times, the initial training from Nvidia has been done on high-end business multi-GPU solutions which are not accessible from a simple student perspective. So the fact that I was able to train the model on a free base, and obtain satisfactory results is another success. But there is also a possibility of using TPUs on Google Colab which would be another way to have faster inference of the networks. But that would require a complete rewrite of the model as the training on GPUs and TPUs is not done in the same way.

The fact also that this model is able to generate new waves on its own gives an impression of close collaboration between the machine and the user. Or even that the machine is able to dream up its own waves without our help (except the initial spark driving the NumPy arrays under the hood).

The interpretation of the results is not objective, but purely subjective. It lies under an artistic point of view. Hence the project was a success, and will be further developed as the art installation it is supposed to be.