Deep Learning Feedback

Dr Chris G. Willcocks and Dr Amir Atapour-Abarghouei

Last Modified: March 28, 2023

Individual Feedback and Marks

Here are your (cvhm34) marks and your final grade:

Percentages	
Solution quality:	65.0%
Sample realism:	65.0%
Batch diversity:	65.0%

Student code:	cvhm34
Solution marks:	26/50
Realism marks:	20/30
Diversity marks:	20/20
Bonus/penalty:	4
Final grade:	**70**/100

Comments

The report is reasonably well-written though more details on the underlying theory and further analysis of the performance could have been included. The submission includes a solid DDPM implementation, where the main limitations are in the long training times required to scale this type of architecture to high-resolution images and also the long inference time in sampling the model. Nevertheless the sample quality is good and are generally with reasonable quality but lacking in structure. There are still some issues within the output images particularly with some that seem to have collapsed with considerable amount of noise, repeating patterns and strange textures. The implementation of the interpolation is not working well, especially with the STL-10 data.

General Feedback

Deep Learning

The generative modeling trilemma (ICLR 2022 ↗) essentially captures the challenge of this years assignment: GANs require large datasets, are not diverse, but produce high-quality samples quickly. EBMs such as score-based models produce the highest-quality samples with the best diversity and mode coverage (they can also be trained on tiny datasets) but at the expense of poor scalability and increased sample times. VAEs are as efficient to sample as GANs, have excellent coverage, but produce low-quality blurry samples. Normalising flows are somewhere in-between. Hybrid models can inherit multiple advantages, but at the cost of increased architectural complexity and difficult hyperparameter tuning.

There was a group of people that simply modified DCGAN with added conditionals, a Wasserstein gradient-penalty or spectral normalisation. Three years ago, I would have given high marks for such experimentation as it tended to produce the best results at the time. However now these DCGAN submissions generally (with 1-2 exceptions) only produced relatively satisfactory samples, which shows how fast this field changes.

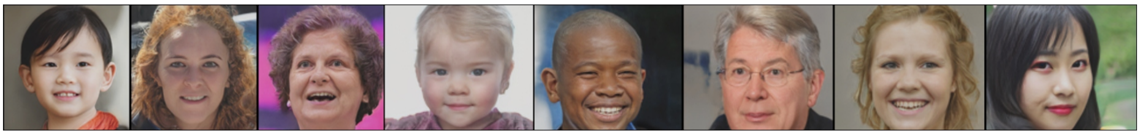
The students that produced the most realistic and diverse high-resolution samples tended to implement state-of-the-art 2021+ literature from peer-reviewed top conferences: ICLR, NeurIPS, CVPR, ECCV and ICML. Their code was often based on the authors original implementations, where notably successful approaches included FastGAN, StyleGAN2, GLOW, and DDPMs. Generally diffusion-based approaches, such as latent diffusion and hybrids such as unleashing transformers produced the best quality samples along with conditional adversarial models like StyleGAN2.

Here's a selection of some of these highly ranked samples (in terms of realism and diversity) from a variety of different approaches (GLOW, StyleGAN2, unleashing transformers, and a score-based model):

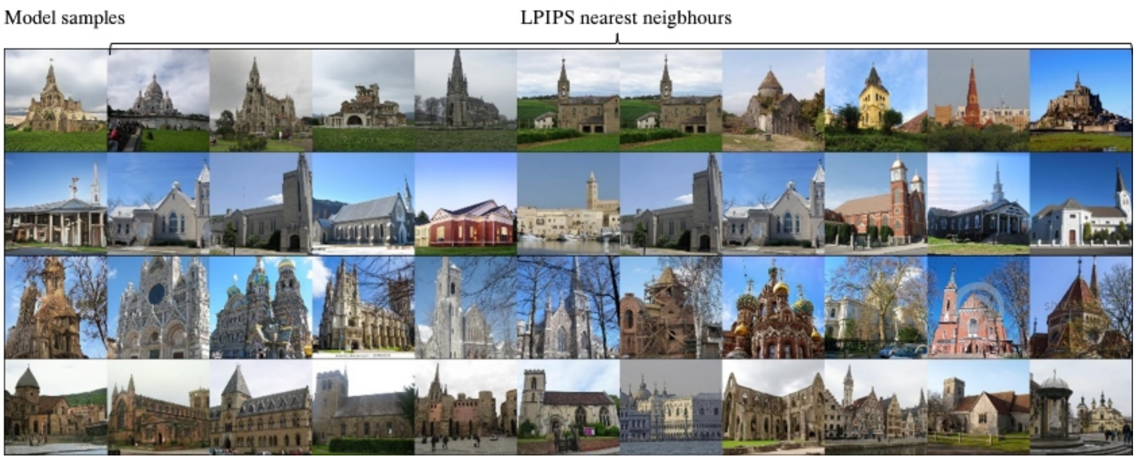
Normalising flow samples.

Score-based generative model STL-10 interpolations.

StyleGAN2 interpolations on FFHQ. Each image looks believable with a smooth transition.



Examples of cherry-picked high-quality and diverse model samples.



Unleashing transformer diffusion samples shown to be perceptually different to the training data.

Closing Comment

We hope you found this assignment rewarding and that it has helped set realistic expectations into what can and cannot be achieved with these large models using the compute available. We also hope that you appreciate the importance in keeping up-to-date with the never-ending stream of new literature in this fast-changing field.