Deep generative models

Deep learning summative assignment

Michaelmas term, 2022

Introduction

The assignment is to design and train a deep generative model to synthesise high-quality, unique and diverse images subject to some criteria.

You are also to write a short scientific report for the method, experimental results, and limitations in a provided LaTeX template that closely follows parts of the ICLR conference style guidelines. These files must be zipped together like this, replacing the username with your CIS username. You may not submit additional source code files:

```
username.zip
    username-paper.pdf
    username-model-code.ipynb (or .py)
```

To assist in this, the following template reports and starter code are provided to build on:

↗ [Deep Learning Paper Template] - *login with durham email, on 'overleaf.com' click 'make a copy' to edit*

↗ [Google Colab <u>Generative Model</u> Starter Code]

The deep generative model

Using the CIFAR-10 dataset and/or a high-resolution dataset such as STL-10, individual classes of LSUN or the FFHQ dataset, train a deep generative model to synthesise unique images that will be judged on their realism, diversity and uniqueness from the original training data. Then write a short paper up to a maximum of 4 pages using the provided LaTeX template, writing up the methodology, results, and limitations of your approach alongside a short abstract.

The methodology should explain the underpinning theory of your model formally, and the results section should show: (1) a unique batch of 64 non cherry-picked samples (2) interpolations between 8 pairs of your samples, and (3) some cherry-picked examples (a selection of the best images you have seen your model generate).

Use any adversarial training (e.g. GAN) method	-8 marks
Train only on CIFAR-10	-4 marks
Train with STL-10, LSUN or FFHQ resized to 48x48 pixels	$+2$ marks
Train with STL-10, LSUN or FFHQ resized to 64x64 pixels	$+2$ marks
Train with STL-10, LSUN or FFHQ resized to 96x96 pixels	$+2$ marks
Train with LSUN or FFHQ at 128x128 pixels or more	$+2$ marks
Manually edit (paint) any images or outputs	-100 marks
Use or modify someones code without referencing it	-100 marks
Use pre-trained weights from another model	-100 marks
Every page over 4 pages in the paper (excluding references)	-10 marks

Table 1: Penalties and bonuses stack, and are added onto the final mark. Therefore if you successfully train at 128x128 image resolution, you also get the bonuses for 96x96, 64x64 and 48x48 automatically.

The report should be written like an academic paper, with formal mathematical notation that should try to follow the ICLR guidelines (see the template for more information). Therefore your discussions should be short, clear, and concise—*less is more*. Where appropriate, it is recommended to include a high-level architectural diagram in the paper to help explain your approach.

You can use any generative model architecture that you like, and you can use any sampling strategy. However, there are penalties and bonuses that will influence your design, summarised in Table 1.

Please state at the end of the paper the total bonus or penalty you are expecting to receive according to this table. For example, if you successfully train a GAN using data from FFHQ resized at 128x128, you can expect to receive no bonus or penalty because: -8 marks (as its a GAN), then 2+2+2+2 = +8 marks for FFHQ at 128x128 resolution. If you submit a paper that is 7 pages long, you will receive an additional -30 mark penalty.

Generative model marking scheme

There is intentionally a lot of freedom with this assignment; where the licence permits you can reference and extend existing code from GitHub, and you can even train on other imaging datasets outside of those suggested. Essentially, you will be marked on the quality of the work, comprehension of the research field, presentation of the underpinning theory, the degree of mathematical rigour and the thoroughness of the scientific analysis—as if it were submitted and reviewed as a real conference paper.

The paper and submitted code will be marked as follows:

- **[40 marks]** Scientific quality and mathematical rigour of the paper and solution
    - Strategy and presentation of the underpinning mathematical theory
    - Architectural design, sophistication, appropriateness and novelty
    - Clarity, simplicity and frugality of both the scientific writing and the implementation
- **[30 marks]** Realism and resolution of the samples
    - Is the sampled batch of images blurry?
    - Do the image objects have realistic shapes and textures? Do they look real?
    - Do the interpolations look like linear alpha blendings or are all midpoints realistic?
- **[30 marks]** Diversity and uniqueness of the sampled batch of model outputs
    - How different are the images from their nearest neighbours in the dataset?
    - How diverse are the samples within the batch of 64 provided?
    - Do all the images look similar? Is there any mode collapse?

PyTorch Training

This assignment can be completed entirely using Google Colab, or using NCC as outlined in the first practical. If using NCC, please carefully read the documentation and respect other users on the job queuing system. It is recommended to first get a simple model working, such as a 4-2-1 pattern DCGAN on CIFAR-10, and conduct the interpolation experiments with it—before exploring state-of-the-art papers.

Feedback Opportunities

There will be opportunities for feedback on your progress, writing, and solution in the second half of the practicals. There is also a list of common questions and answers in the assessment folder on blackboard.

Closing Comment

I hope that you enjoy this coursework and the freedom you have with it. If you are struggling, please ask questions where we can discuss any issues, such as with programming or relevant theory.