CS283: Assignment 5

March 4, 2024

Due: Monday March 18th, 11:59pm
Submit by the **blackboard system**

By turning in this assignment, I agree with the KAUST student protocol and declare that all of this is my own work.

Requirements:

- Put all files in a zip file and the file should be named LastnameFirstname_Assignment5.zip
- It is an individual assignment.
- It is a must to attach all the coding and notebook files with the submission file.
- Late submission will receive credit penalty based on the late policy, that was announced in the class.

Homework Description

In this homework, you will need to answer questions related to zero-shot learning and diffusion models. Diffusion models are a family of probabilistic generative models that progressively destruct data by injecting noise, then learn to reverse this process for sample generation. Current research on diffusion models is mostly based on three predominant formulations:

1. Denoising Diffusion Probabilistic Models [1, 2, 3] (DDPMs)
2. Score-based Generative Models [4, 5] (SGMs)
3. Score Stochastic Differential Equations [6, 7] (SDEs)

1 Diffusion Models: Questions (30 points)

1. **(5 pt):** Detail the forward and reverse processes of Diffusion Model (DDPM) [1], with equations and a simple digram/Figure?
2. **(5 pt):** What are the key differences between DMs [1, 4, 7] and Hierarchical Variational Autoencoders [8, 9]?
3. **(5 pt):** Discuss the advantages of Latent Diffusion Model [10] in detail, and ellaborate why it can achieve these advantages?
4. **(5 pt):** What is the difference between Classifier [11] and Classifier-Free [12] Guidance mechanisms to condition DMs? Discuss advantages and disadvantages.
5. **(5 pt):** Read the Cascaded Diffusion Models paper [13] and discuss its main idea. Then, discuss the main challenge they face while conecting the different stages together and discuss the proposed solution?
6. **(5 pt):** How and why SDE [7] is connected with SBGMs [4]?

2 Diffusion Models: Programming (40 points)

Please, refer to the accompanied files

3 Latent-Diffusion Model: Programming (30 points)

Try to start as early as possible for this part as training the model may take 12 hours. Try to make the model as small as possible to make the training faster.

Hint: 5 millions paramters are enough for the model to generate acceptable images. So try to make the model compact as possible.

1. Download the CelebHQ dataset, and preprocess it to be 256x256 instead of 1024x1024.
2. Train LDM [10] to generate high quality faces.
3. Adapt the Unet archeticture in the LDM code base to be a very small by reducing the number of channels and number of ResNet blocks, to make the training fast. In addition, use the VQ-GAN-4, as recommended in the readme, to encode the image before the Unet block.
4. After training the model succefully for 50-150 epochs, run the sampling code to generate faces.
5. Please submit everything you have done, even if it is not complete, the training codes, sampling codes, training logs, and generated samples.

6. If you couldn't finish this part, submit a report with the codes, detailing your efforts and the challenges you faced to get grades according to your efforts. It is not binary grading, so show me your efforts.

4 [Optional/Bonus] Generative Zero-Shot Learning Questions (15 points)

1. **(7.5 pt):** What are the hallucinated features [14]? What is their purpose? How are they being constructed?

2. **(7.5 pt)** Run the code of CIZSL on one text-based dataset (e.g., CUB-wiki). Please report your performance using the provided hyperparameters (your performance might be slightly different from the reported due to small instability and different hyperparameters). You can find the code here.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[2] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

[3] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.

[4] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

[5] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020.

[6] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in Neural Information Processing Systems*, 34:1415–1428, 2021.

[7] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through

stochastic differential equations. In *International Conference on Learning Representations*, 2020.

[8] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems*, 29, 2016.

[9] Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. Ladder variational autoencoders. *Advances in neural information processing systems*, 29, 2016.

[10] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[11] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

[12] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

[13] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *The Journal of Machine Learning Research*, 23(1):2249–2281, 2022.

[14] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, and Ahmed Elgammal. A generative adversarial approach for zero-shot learning from noisy texts. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1004–1013, 2018.