CS698X (2020-21-II)

# Project Proposal

## *Energy-based Deep Generative models*

Pulkit Gopalani | Ritik Saxena | Abhishek Shah | Shivam Singhal
180564 | 180617 | 180025 | 180721
{ gpulkit | ritiks | shahabhi | singhals } @iitk.ac.in

## Problem Introduction

Deep Generative models have been successful in modeling complex distributions, such as those involving images ([3],[4],[5]), videos ([6]) and sound ([7]). Variational Auto-Encoders (VAEs) ([1]) are a class of such models, that estimate a data distribution (possibly of images) using an encoder-decoder architecture, with a bottleneck in between for modeling latent variables. Unlike GANs ([3]), a model which simulates a 2-player game between the generator and discriminator for generating samples, VAEs maximize the likelihood of data under the model.

However, VAEs tend to assign high probability density to regions in data space outside the actual data distribution and often fail at generating sharp images. Energy Based Models ([8]) are a class of probabilistic models which assign an "energy" to a region, in place of its probability density. These models assign low energy to high probability regions in the data-space. They require no restrictions on the model architecture, and hence are an attractive alternative to other generative models.

## Prior Work

The paper "VAEBM: A Symbiosis between Variational Autoencoders and Energy-based Models" ([2]) made the following contributions :

- A new generative model using the product of a VAE generator and an EBM defined in the data space.
- How training this model can be decomposed into training the VAE first, and then training the EBM component.
- How MCMC sampling from VAEBM can be pushed to the VAE's latent space, accelerating sampling.

The VAEBM generative model in the paper is defined as

$$h_{\psi,\theta}(\mathbf{x},\mathbf{z}) = \frac{1}{Z_{\psi,\theta}} p_\theta(\mathbf{x},\mathbf{z}) e^{-E_\psi(\mathbf{x})}$$

where $p_\theta(\mathbf{x},\mathbf{z}) = p_\theta(\mathbf{z}) p_\theta(\mathbf{x}|\mathbf{z})$ is a VAE generator and $E_\psi(\mathbf{x})$ is the neural network-based energy function operating only in the $\mathbf{x}$ space.
The likelihood function for the optimization is then given by marginalizing out the latent variable $\mathbf{z}$, i.e.,

$$h_{\psi,\theta}(\mathbf{x}) = \frac{1}{Z_{\psi,\theta}} \int p_\theta(\mathbf{x},\mathbf{z}) e^{-E_\psi(\mathbf{x})} d\mathbf{z} = \frac{1}{Z_{\psi,\theta}} p_\theta(\mathbf{x}) e^{-E_\psi(\mathbf{x})}$$

To maximise the marginal log-likelihood, the paper instead tries to maximise a lower bound to the log likelihood. First they replace $\log p_\theta(\mathbf{x})$ with the variational lower bound and then separate into two terms, the first to corresponding to VAE objective and second corresponding to EBM component.

$$\begin{aligned} \log h_{\psi,\theta}(\mathbf{x}) &= \log p_\theta(\mathbf{x}) - E_\psi(\mathbf{x}) - \log Z_{\psi,\theta} \\ &\geq \underbrace{\mathbb{E}_{\mathbf{z}\sim q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))}_{\mathcal{L}_{VAE}(\mathbf{x},\theta,\phi)} \underbrace{- E_\psi(\mathbf{x}) - \log Z_{\psi,\theta}}_{\mathcal{L}_{EBM}(\mathbf{x},\psi,\theta)} \end{aligned} \tag{1}$$

The first gradient can be estimated easily by sampling from VAEBM model using MCMC. To reduce the computational complexity in estimating the second gradient, the paper proposes a two-stage algorithm. First, $\theta$ and $\phi$ are trained using a reparameterized trick by maximising the $\mathcal{L}_{vae}(\mathbf{x}, \theta, \phi)$. In second stage, these parameters are kept fixed and thus, the $\mathcal{L}_{ebm}(\mathbf{x}, \psi, \theta)$ is optimized only w.r.t. $\psi$. The first stage minimizes the distance between the VAE model and the data distribution. As the pre-trained VAE $p_\theta(x)$ provides a good approximation to $p_d(x)$ and thus , smaller number of expensive updates are needed for training $\psi$. The pre-trained VAE provides a latent space with an effectively lower dimension and thus helps in more efficient MCMC sampling.

## Possible Improvements

After going through the reviews ([10]) of the paper ([2]), we found the following improvements that could be worked upon:

- The EBM in the model currently corrects only the image space $(\mathbf{x})$, we could work on extending the energy function to the latent space $(\mathbf{x}, \mathbf{z})$
- We could investigate better sampling techniques like Hamiltonian Monte Carlo for improved mixing.
- There is apparently mode-collapse for a few cases in the generated images for CIFAR-10, we could also investigate this issue.

## Tentative Plan

We aim at studying and investigating these properties of EBMs, in conjunction with Variational Auto Encoders for generating image samples based on VAEBMs ([2]). We would start by implementing the model in the main VAEBM paper as a baseline, on datasets used in the paper. Time permitting, we also hope to investigate the disentanglement properties of this model, closely related to ([9]), and work on a few possible improvements to the original paper ([2]).

## Proposed Timeline

| Time Period | Proposed Work |
|---|---|
| January 2021 - Mid-Semester Examinations (27 February 2021) | Understanding existing literature relevant to the project |
| Mid-Semester Recess - Last week of March | Baseline implementation (Reproducing results from the paper), Running experiments on different datasets |
| First week of April - End of Semester | Working on the proposed improvements, disentanglement of latent factors and final report submission |

# References

[1] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2014. arXiv: 1312.6114 [stat.ML].

[2] Zhisheng Xiao, Karsten Kreis, Jan Kautz, et al. *VAEBM: A Symbiosis between Variational Autoencoders and Energy-based Models*. 2021. arXiv: 2010.00654 [cs.LG].

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].

[4] Andrew Brock, Jeff Donahue, and Karen Simonyan. *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. 2019. arXiv: 1809.11096 [cs.LG].

[5] Tero Karras, Samuli Laine, and Timo Aila. *A Style-Based Generator Architecture for Generative Adversarial Networks*.

[6] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. *Generating Videos with Scene Dynamics*. 2016. arXiv: 1609.02612 [cs.CV].

[7] Jesse Engel, Kumar Krishna Agrawal, Shuo Chen, et al. *GANSynth: Adversarial Neural Audio Synthesis*. 2019. arXiv: 1902.08710 [cs.SD].

[8] Yann LeCun, Sumit Chopra, Raia Hadsell, et al. "A tutorial on energy-based learning ". In: *PREDICTING STRUCTURED DATA*. MIT Press, 2006.

[9] Hao Wu, Babak Esmaeili, Michael L Wick, et al. "Conjugate Energy-Based Models ". In: *Third Symposium on Advances in Approximate Bayesian Inference*. 2021. URL: https://openreview.net/forum?id=4k58RmAD02.

[10] Zhisheng Xiao. *VAEBM: A Symbiosis between Variational Autoencoders and...* Jan. 2021. URL: https://openreview.net/forum?id=5m3SEczOV8L.