Introducing Sharpness measure in Variational Autoencoders and Variational Autoencoder Generative Adversarial Networks

Team Members

1. EE15B094 - Karna Datta Sai Krishna Vardhan Reddy
2. ME15B070 - Tanneru Sree Harsha

Introduction

Images generated using Variational Autoencoder are generally not sharp as they try to learn an explicit distribution of the input with a multi-dimensional Gaussian distribution. We propose a sharpness loss term motivated by [1] *"A New Method for Full Reference Image Blur Measure "* to improve sharpness as a part of loss function and present the results on MNIST dataset.

While training models like Variational Autoencoder, element wise differences between original image and generated image are used as a similarity measure. This similarity measure is not invariant to image translation which is not desirable. A similarity measure which measures the difference in features of the image is desired. The paper [2] *"Autoencoding beyond pixels using a learned similarity metric"* proposes a way to tackle this problem by training a Generative Adversarial Network in series with a Variational Autoencoder. We use the intermediate layer output of Discriminator as a measure of similarity. We present the results of VAE-GAN with sharpness loss on MNIST dataset.

Dataset

All experiments have been performed on MNIST dataset: 10 different digits and 5000 32 x 32 images per digit.

Proposed Sharpness Loss term

[1] *"A New Method for Full Reference Image Blur Measure "* proposes a new algorithm to determine the blur in the image via detection of edges in the image or in other words, determining the blur's percentage based on edges' sharpness with respect to the reference image, as the sharpness of the edges is the best criteria to indicate the amount of the existing blur in any image. It proposes a method to calculate the blur in the image based on the edges' sharpness, since the sharpness is inversely proportional to the blur. The algorithm is as follows:

1. Calculate intensity variation of the 8 neighbour pixels for each pixel of both images i.e., reference and the degraded (reconstructed image for our purpose). This is computed by taking the difference between the center pixel and its 8 neighbour pixels for each pixel.
2. Now take the maximum difference among 8 neighbour pixels and the center pixel, for all pixels and for both reference and the reconstructed.
3. Now take the average of maximum difference values of all central pixels in each image and be it Z_1 and Z_2 respectively for reference and reconstructed.
4. $Blur\ percentage\ = \frac{Z_1 - Z_2}{Z_1} \times 100$

With the above defined measure for sharpness (or inversely blur), we propose a sharpness loss for the generated images similar to the losses such as style loss(using gram matrix in Neural Style Transfer) and content loss. We experimented with different ways of implementing this sharpness loss.

Implementation details (PyTorch)

The challenging task here is to implement step 1 given an original image and a reconstructed image and making sure that loss back propagates through this. We model the sharpness loss as a group of convolutions. We defined a 3×3 filter with 8 channels and initialized it with non learnable weights. All 8 channels have the central value as -1 and zeros surrounding it except a 1 at 8 possible locations for 8 channels as shown below.

1	0	0
0	-1	0
0	0	0

0	1	0
0	-1	0
0	0	0

0	0	1
0	-1	0
0	0	0

0	0	0
1	-1	0
0	0	0

0	0	0
0	-1	1
0	0	0

0	0	0
0	-1	0
1	0	0

0	0	0
0	-1	0
0	1	0

0	0	0
0	-1	0
0	0	1

Same goes for other channels but one at different location, so that, when convolution happens it outputs 8 channels. For example, with the above mentioned filter, the first channel of the output contains intensity variation of the central pixel with its top left pixel, for all central pixels. Therefore we have intensity variations of a central pixel with its 8 neighbours in 8 channels.

Now we experimented with different types of losses whose results are shown in the following figure:

a) L_2 loss on the average of maximum of intensity variations around a central pixel
   $sharpness\ loss = (Z_1 - Z_2)^2$, where Z_1, Z_2 are defined earlier.

b) L_2 loss on the maximum of intensity variations around a central pixel.
   $sharpness\ loss = \sum_{for\ each\ pixel,\ j} (m_{j1} - m_{j2})^2$, where 1, 2 denote reconstructed and original

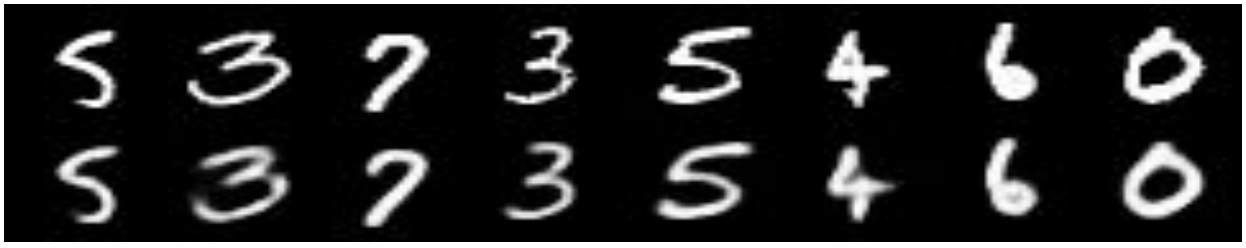
images and m_j denotes maximum intensity variation across 8 neighbouring pixels of jth central pixel.

c) L_2 loss on the mean of squares of intensity variations around a central pixel.

d) We also experimented inducing this sharpness loss at a later time(after some epochs) allowing the network to reconstruct the image.

Among all experimentations, Loss defined in (b) gave best result with it being present from the beginning. The results can be seen in the following images corresponding to the 4 experiments above respectively. In each image, first row images are original from dataset and second row images are generated. We first show the images without our loss on the VAE and then for above four experiments.

Naive VAE



(a) L2 loss on average of maximum intensity variation around each pixel over all pixels

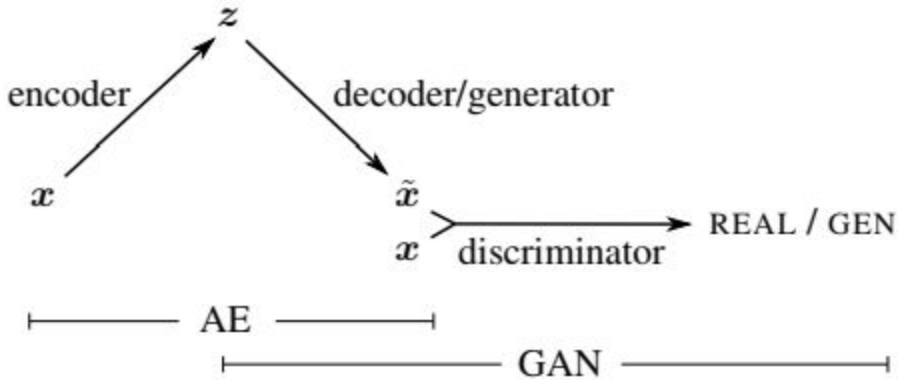
(b) Average of L2 loss of maximum intensity variation around a pixel compared to original image for all pixels

(c) L2 loss on mean of sum of squares of intensity variations of each pixel around 8 corner pixels

(d) Sharpness Loss as in (b) induced after 15 epochs of training

The generated images in (b) are better compared to naive VAE (1). The improvement can be seen when one observes first '3' in the images.

Variational Autoencoder Generative Adversarial Network



A Variational Autoencoder Generative Adversarial Network is formed by fusing generator network of GAN and Decoder network of VAE into one. A Variational Autoencoder consists of two networks. An Encoder network which encodes the image x to a latent representation z and a Decoder network which decodes the latent vector back to the image space.

$$z \sim Enc(x) = q(z|x)$$
$$\bar{x} \sim Dec(z) = p(x|z)$$
$$L_{VAE} = L_{prior} + L_{reconstruction}$$
$$L_{reconstruction} = -E_{q(z|x)}[log\ p(x|z)]$$
$$L_{prior} = D_{KL}(q(z|x) \,||\, p(z))$$

A Generative adversarial network consists of a generator network which maps the latent vector to image space while the discriminator network networks assigns a probability of the image real.

$$\bar{x} = Gen(z)$$

$$y = Disc(\bar{x}) \in [0, 1]$$
$$L_{GAN} = -[log(1 - y) + log(Disc(x))]$$

The reconstruction loss calculated by VAE is not translation invariant. The paper [2] *"Autoencoding beyond pixels using a learned similarity metric"* proposes a better way to measure image similarity by using the intermediate layer output of discriminator as presents a better way to measure reconstruction error by replacing the reconstruction error term in VAE with reconstruction error term in GAN. We measure the reconstruction error at an intermediate layer l in the discriminator.

$$p(Disc_l(x)|z) = N(Disc_l(x) \mid Disc_l(\hat{x}), I)$$
$$L_{reconstruction} = -E_{q(z|x)}[log\ p(Disc_l(x)|z)]$$

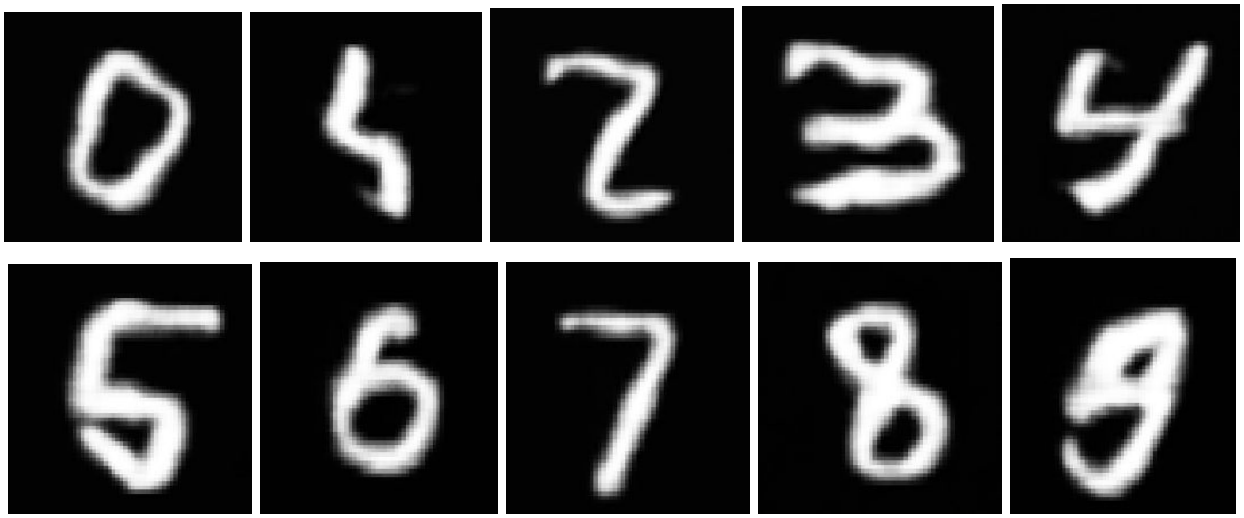
The total loss function is $L_{total} = L_{prior} + L_{reconstruction} + L_{GAN}$

The parameters are updated as follows

$\theta_{Encoder}$ += $-grad(L_{prior} + L_{reconstruction})$

$\theta_{Decoder}$ += $-grad(L_{reconstruction} - L_{GAN})$

$\theta_{Discriminator}$ += $-grad(L_{GAN})$



Results of VAE-GAN on MNIST dataset

Variational Autoencoder Generative Adversarial Network with Sharpness Loss

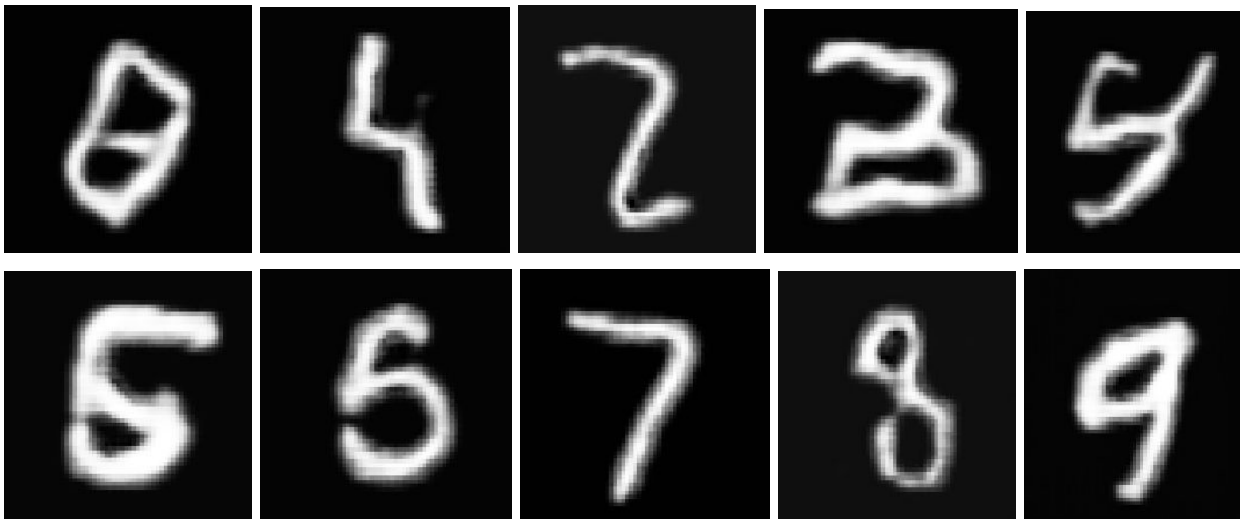
We now trained the network using additional sharpness loss defined below.

$L_{sharpness} = \sum_{for\ each\ pixel,\ j} (m_{j1} - m_{j2})^2$, where 1, 2 denote reconstructed and original images and denotes maximum intensity variation across 8 neighbouring pixels of jth central pixel. The entire network is trained by minimizing total loss function.

$$L_{total} = L_{prior} + L_{reconstruction} + L_{GAN} + L_{sharpness}$$

The parameters are updated as follows

$$\theta_{Encoder} \mathrel{+}= -grad(L_{prior} + L_{reconstruction} + L_{sharpness})$$
$$\theta_{Decoder} \mathrel{+}= -grad(L_{reconstruction} - L_{GAN} + L_{sharpness})$$
$$\theta_{Discriminator} \mathrel{+}= -grad(L_{GAN})$$



Results of VAE-GAN with sharpness loss on MNIST dataset

Conclusion

Sharpness loss as defined in (b) on a simple VAE gave better results. The poor results of (a) and (c) can be attributed to the averaging as it loses information of sharp edge location compared to original image.

Adding sharpness loss term in VAE-GAN resulted in poorer results compared to VAE-GAN without sharpness loss. This might be due to

1. Reconstruction Loss defined using intermediate layer of discriminator is translation invariant where as Sharpness loss is a pixel wise measure.
2. Any changes induced by sharpness loss may result in discriminator predicting the digit as some other digit.

Contribution

Both worked on framing the sharpness loss.
EE15B094 - Code and Experimentations on simple VAE.
ME14B070- Code and Experimentations on VAE-GAN.

References

1. A New Method for Full Reference Image Blur Measure
2. Autoencoding beyond pixels using a learned similarity metric
3. Git hub implementation of VAE-GAN
4. Tutorial on VAE pytorch