Assignment 5

Deep Learning

Submitted By:

Gyanendra Chaubey (M23AIR005)
Shripad Pate (M23MCA007)

Submitted To:

Dr. Deepak Mishra
Assistant Professor
Dept. of CSE

1. Methodology:

In this work, first training of dermoscopic lesion images has been done using vector quantized variational autoencoders. Then, an auto regressive model has been trained from the latent which are generated from the code books. In this work, after training the VQ-VAE model, encoder, codebook, and decoder has been saved. Then, quantized latent vectors are generated using the training data from the encoder and codebook. After generating the quantized latents it is given to auto regressive model to train it and save its model. After, saving the model, inference can be taken by passing an image to the encoder, then generated latent to codebook to get quantized latent then pass this quantized latent to autoregressive model and get another latent to pass to the decoder to get the output image. Then calculated the reconstruction loss and perplexity to get the accuracy of the model.

1.1 Model Architecture VQ-VAE

In this architecture, we have built an encoder decoder architecture with the quantizer which finds the distance between the codebook and the feature map generated by the encoder. Then, quantized vector generated from the generator is passed to the decoder to get the output image. In this architecture, quantizer which find the distance between feature map and the code book discretizes latent so that model can be trained efficiently. After training the VQ-VAE model its weights are saved for inference. The architecture that has been used in this work for VQ-VAE is given in Fig. 1.

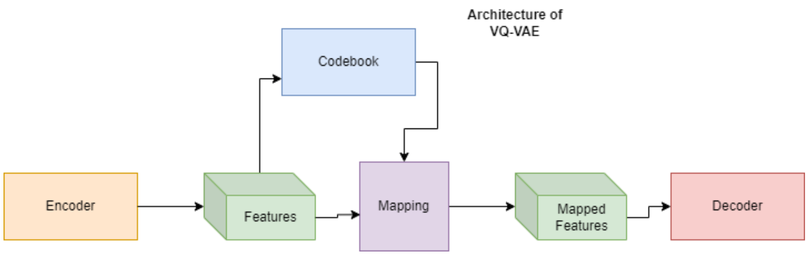


Fig. 1 Architecture of VQ-VAE

1.2 Architecture of Autoregressive Model

For autoregressive model, Gated Pixel CNN is used to train the latent which are closest of the code book. These latent vectors are generated from the trained VQ-VAE and passed as input to the Gated Pixel CNN. After training the Gated Pixel CNN, its weights are saved to generate the latent for the decoder. The architecture of the Gated Pixel CNN is given in Fig. 2.

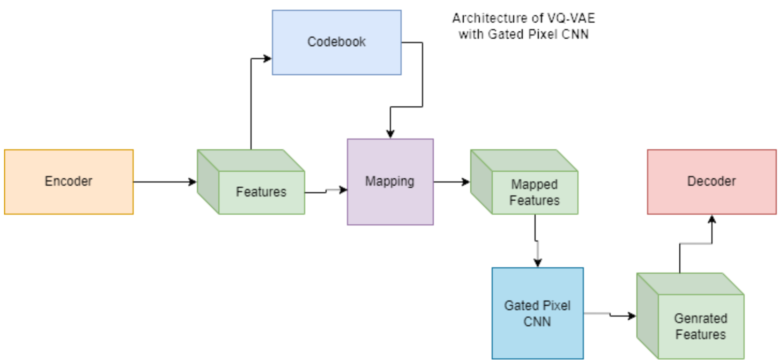


Fig. 2 Architecture of Auto Regressive Model

1.3 Loss and Score functions for VQ-VAE and Autoregressive

In this work for training the VQ-VAE, reconstruction loss has been used. For checking the output certainty or uncertainty Perplexity score has been used. Reconstruction loss is the mean square error (MSE) between the latent generated and the input. Perplexity explains about the certainty and uncertainty about an image. In the Pixel Gated CNN also we have used the Binary Cross Entropy (BCE) loss to find the error in the given latent with labels and generated latent with labels. The equation for the MSE loss, Perplexity and BCE loss is given as Eq.1, 2, and 3 respectively.

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (1)$$

$$Perplexity = 2\exp\left(\frac{1}{N}p_i\log(p_i)\right) \qquad (2)$$

$$BCE(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}(y_i * \log(\hat{y})) + (1 - y_i) * (1 - \log(\hat{y})) \qquad (3)$$

2. Results

The results shown in the below approaches are the best obtained results after several iterations of modifications either in the architecture or in the hyper parameters of the model. For each of the model's loss functions are plotted.

2.1 VQ-VAE

VQ-VAE has been trained for the 10 epochs (about 12k iterations), the MSE loss and perplexity of the model is given as below in Fig. 3 and Fig. 4 respectively. We can see that, the reconstruction loss is around 0.15 initially, and decreases in each of the epochs iterating over 10 epochs and reach to approximating 0.

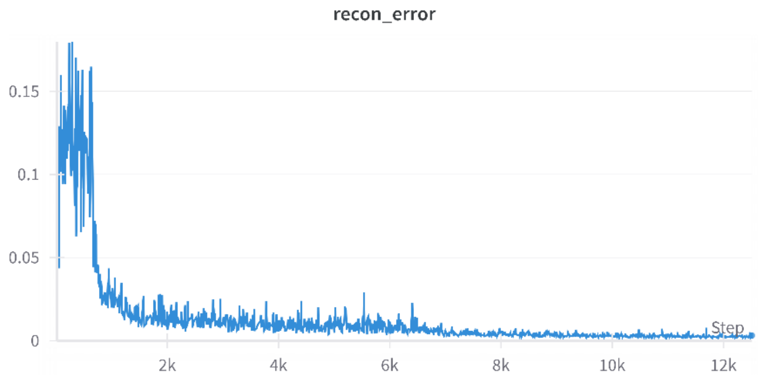


Fig. 3 Reconstruction loss of the VQ-VAE

The Perplexity has been observed increasing in the Fig. 5 on increasing the training epochs of the model.

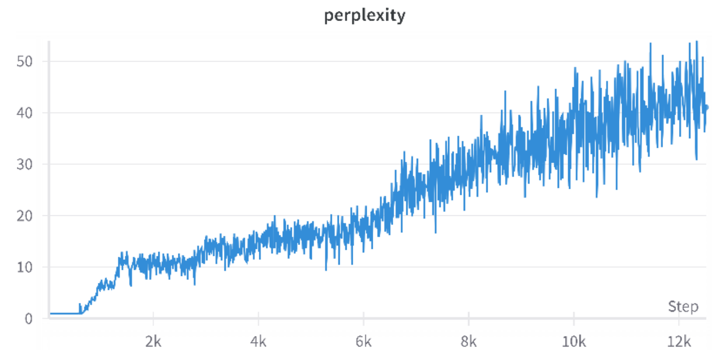


Fig. 4 Perplexity of the VQ-VAE

We have also drawn the normalized mean square plot and perplexity separately which can be seen in Fig. 5.

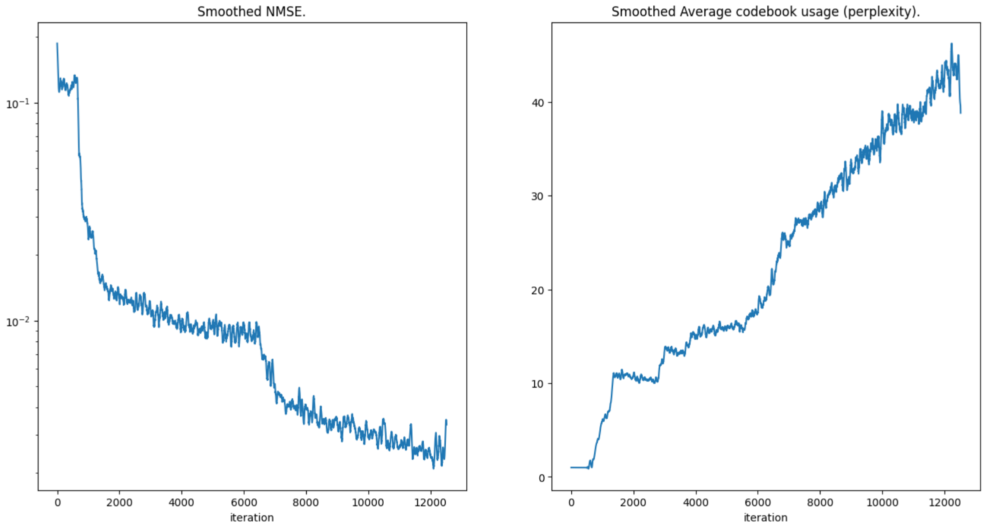
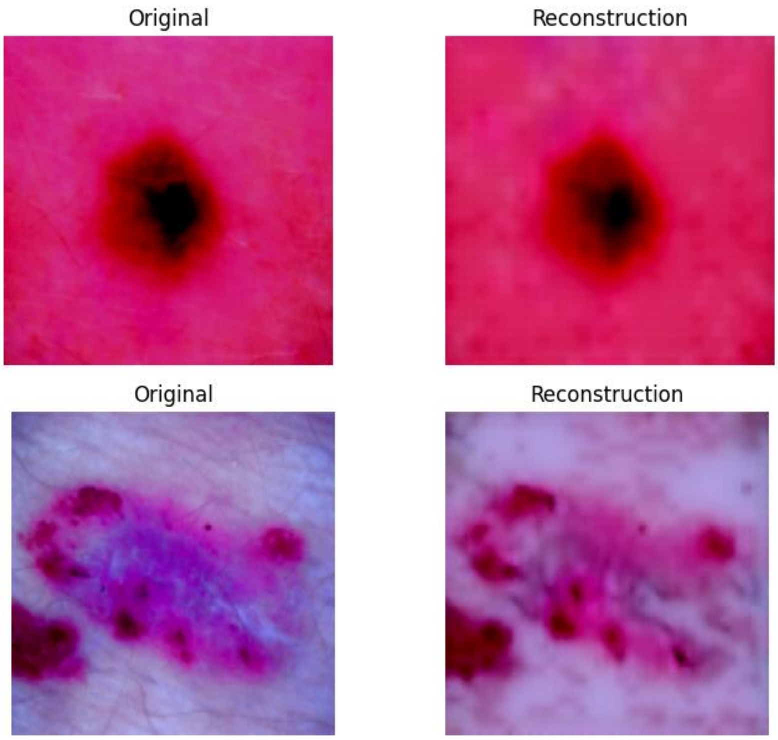


Fig. 5 Normalized Mean Square Error and Perplexity

Some of the outputs of the VQ-VAE are shown in the Fig. 6.



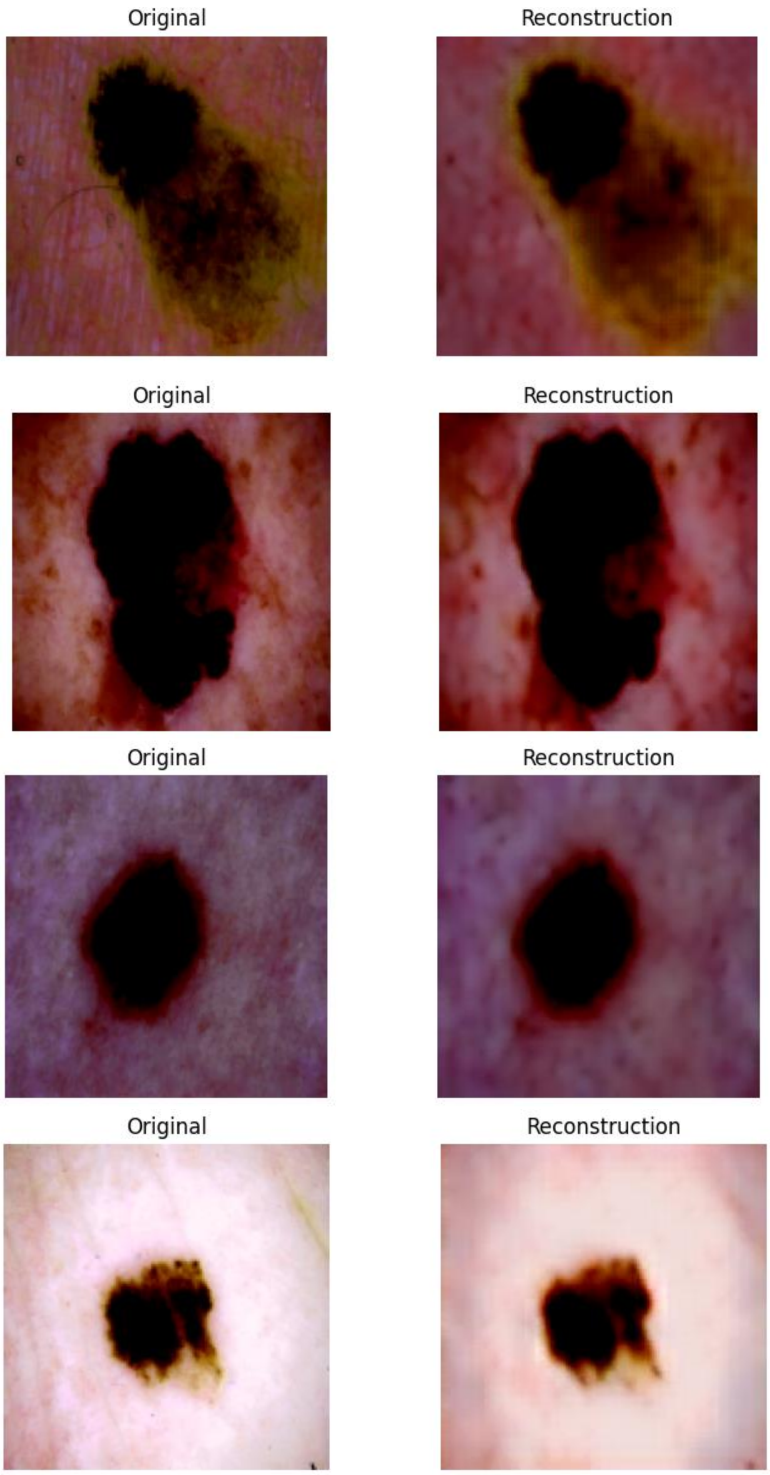


Fig. 6 Original and Reconstructed images from the VQ-VAE

2.2 VQ-VAE with GatedPixel CNN

The autoregressive architecture contains the training of the Gated Pixel CNN for the generation of the latent vectors of the decoder. A single feature map that has served as the input to the Gated Pixel CNN is shown in Fig. 7.

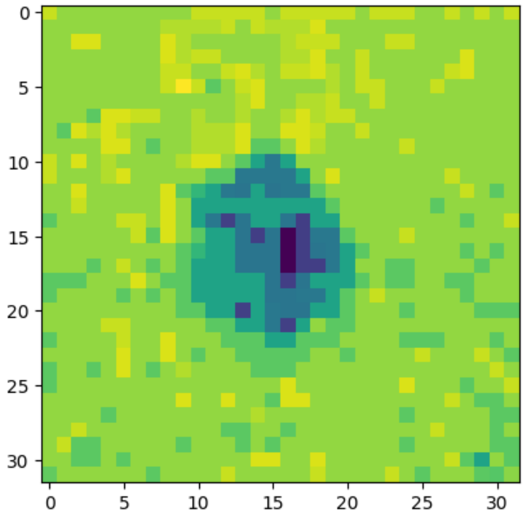


Fig. 7 Input (Quantized Vector) for the GatedPixel CNN

The autoregressive architecture has been trained for the 20 epochs. The training loss for the autoregressive model is shown in the Fig. 8. Initially loss has started decreasing to a sharp way and then gradually decreases its rate for reduction while running up to 20 epochs.

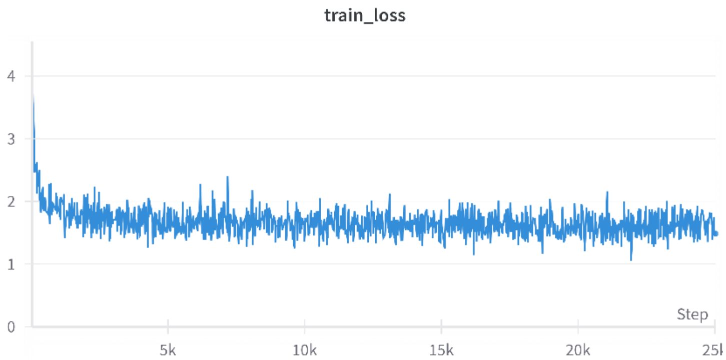


Fig. 8 Train loss for the Gated Pixel CNN Model

Some of the outputs of the autoregressive model are shown in Fig 9.

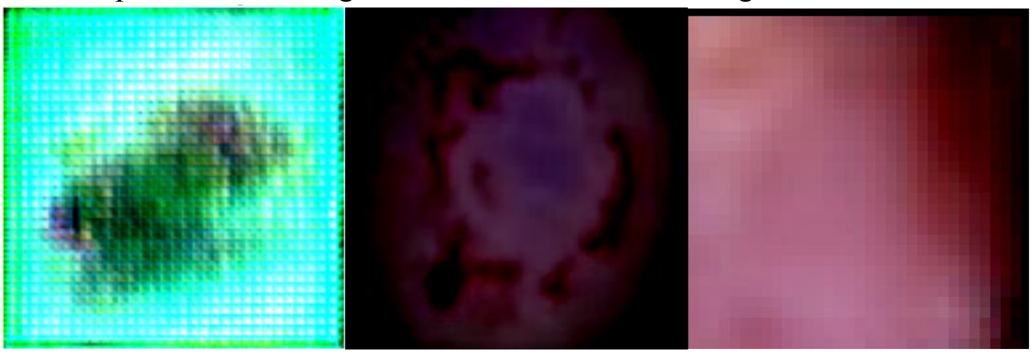


Fig. 9 Samples of the generated images

2.4 Observations

- The transformation of images matters for the training the models, if coarsecutout and RGBshift type of transformations are used they are using more time for reconstruction but learning better images.
- The training of the VQ-VAE is faster than the VQ-VAE with Gated Pixel CNN

- On training a smaller number of epochs autoregressive model was not performing well. Also, it converges slowly.

3. Discussion and Conclusions

From this work, exploration of different autoregressive models has been done that can be used with VQ-VAE to generate images like PixelCNN, GatedPixel CNN, RNN, Transformers, and GPT etc. This assignment helps to understand the concept of VQ-VAE and its integration with some of autoregressive model. In this work, Gated Pixel CNN has been used for generating the latent vectors to feed to the decoder to generate the new images.