Winged Horses with a 2-channels encoder Variational Autoencoder

Anonymous authors
Paper under double-blind review

Abstract

Under the circumstance that the generator produces a single creature with good effect, it is required to generate a hybrid or exotic creature only by the automatic encoder framework. According to the assignment, The professor tends to ask us build a autoencoder-framework generator. In my work, I just attemped a lot of ways in the autoencoder framework. After trying some data process methods, some network architechtures and loss functions, I finally decided to use a 2-channels encoder in Variational Autoencoder Framework and modify its original loss function. In this modified method, I can easily get the Pegasus. They are winged horses with white wings and white feather. But it is a pity that the generated animals are a little vague because of the disadvantage of autoencoder models I got in the paper and the blog.

1 Introduction to some attempts

As the ancients said, Rome was not built in a day. I may not acquire a proper model by the first time. I tried a lot of methods, including the original autoencoder as the .ipynb file shown, the original Variational Autoencoder(just 1 encoder channel), the Variational Autoencoder with the Transfer Learning and the final model: 2-channels encoder Variational Autoencoder. Each method will be shown below.

1.1 The original Variational Autoencoder

Variational Autoencoder enforces the data representation compresses into the normal distribution. The loss function consists of the reconstruction loss and the KL divergence(Kingma & Welling, 2013). The theory is shown the url below and the lecture slides:

https://www.jeremyjordan.me/variational-autoencoders/

However, the model can only learn the mixture representation in CIFAR-10. It only understand how do the horsesr and birds look like, but doesn't know how do the horses assemble with wings in the birds dataset. So the result is the animals it generates are the mixture of the vague horses and birds.

1.2 Variational Autoencoder with the Transfer Learning

According to the shortage of the last method, I combined the transfer learning with the VAE(Variational Autoencoder). I trained the model which has trained with the horses dataset in CIFAR-10(It can generate horses) by the birds dataset. I hope the horse model can learn spread wings of birds and merge them with themselves by the transfer learning. However, I finally discovered that the horse encoder(the encoder which know how horses encode) can easily change by the bird encoder, that is, it changes to know how birds encode, but forgot horses. So I attempted the final model: 2-channels encoder Variational Autoencoder.

2 Data process in the final method

In deep learning, both supervised and unsupervised data have an important influence on the results. The assignment hopes to generate white flying horses, but after visualizing the CIfar-10 data set, I find that there are too many brown and black pictures in the horse data. Some pictures are only horse heads, which cannot well express all the characteristics of the horse.On the bird side, there's a lot of ostriches or ostrich necks, and there's a lot of birds that don't have extended wings, so they don't have wing characteristics.This data, if not selected, will have an adverse effect on the results of the model, which is likely to fail to learn the white winged horse. As shown below:

Figure 1: Horses examples I selected.

Figure 2: birds(wings) examples I selected.

Therefore, for the purpose of modeling, and on the condition that no additional data sets are added to the assignment as required, I selected the cifar-10 data sets of full-body white horses in different postures and birds with outstretched wings to form a separate data set.On the other hand, the data set selected was small, so I chose to add the data using the data enhancement method of mirroring symmetry and contrast enhancement.The increase of data volume can reduce the probability of model overfitting and increase its generalization.

See the Python script dataProcess.py for details on how to process the data.

3 Final method

After trying different methods in the first part, we found that under the appropriate data set, we also needed multiple feature encoders to compress the horse and wing features into a normal distribution, so I needed to put the picture pair of horse and bird into two feature encoders respectively in the training process.Intuitively, the final generation results also need to be compared with the pictures invested, namely horse and bird pictures, so I also modified the loss function corresponding to VAE. sThe details are as follows:

3.1 Building 2 encoders for each animal: horse & bird respectly

Firstly, describe the network structure inside all encoders and decoders.The internal network structure is the same as that of the automatic encoder in Starter, i.e. the encoder is composed of block and maximum pooling layer, and the decoder is composed of UpSample and block.

Secondly, the differences of the final model are explained. The final model constructs two encoders in terms of encoders, respectively learning the expression of horse's whole body feature and the expression of bird's wing feature, splices the expression of the two, and finally generates the log value of normal distribution's mean value and variance.The meaning of this structure is to compress the characteristics of both the horse and the corresponding wing into the same compression vector.In this way, when tested and sampled from the compression vector, a hybrid expression of the two features can be obtained simultaneously, that is, a winged horse can be intuitively generated.The two encoders learning the two features separately prevents the encoders from confusing the features in the learning process, and enables the encoders to perform their respective functions and learn the expression of the two features independently and clearly.The frame is shown in the figure below:

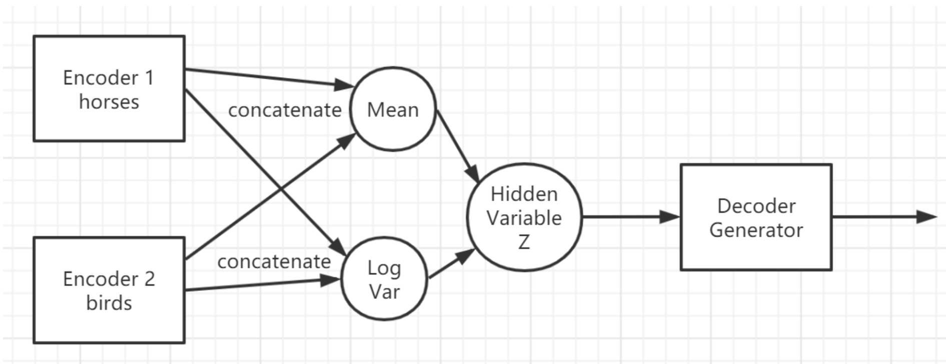

Figure 3: The structure of the model with the original network.

3.2 The modification of its loss function

Since the input data is image pair and the final output image is the feature combination of the two images, the loss function needs to consider both the horse image and the bird image.Since the two types of images only need to learn the central object, and the wings are always a little more fuzzy than the horse's features, we set the weight of reconstruction error with the horse and the weight of reconstruction error with the bird as 0.46 and 0.54 respectively.According to many experiments, this weight ratio is reasonable and can produce a better flying horse.The modified loss function formula is shown in the figure below:

$$\begin{aligned} J(x_n, x, y_n, y, \mu, log\sigma^2) =& \Sigma_B^i - [0.46 w_i [x log x_n + (1-x) log(1-x_n)] \\ &+ 0.54 w_i' [y log y_n + (1-y) log(1-y_n)]] \\ &+ \frac{1}{2} \Sigma_B^i log\sigma^2 - \sigma^2 - \mu^2 + 1 \end{aligned} \tag{1}$$

4 Results & Analysis of The Pegasus generator

After the model is built, I trained the model according to the training method of the deep learning model. The results and analysis of the final model are as follows:

4.1 Results of the generator

During the test, 64 variables were sampled from the normal distribution of the size dimension of hidden variables, and these 64 variables were input into the generated pegasus in the generated model as shown in the Figure 4.

4.2 The analysis of the generator

In the 64 pictures, it can be found that most of the pictures are white horses with wings, although VAE generated pegasus is slightly blurred, but can be seen from the horse's head, body and wings, all the overall shape.However, it is difficult to accurately determine the

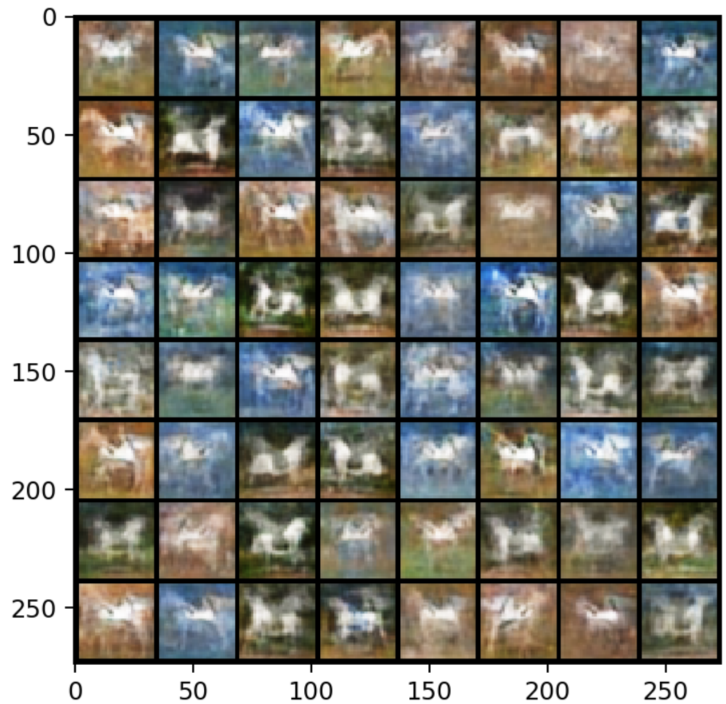

Figure 4: The results of the model generating.

weight ratio between the wings and the horse, so it can be seen that in some pictures, the head of the horse is relatively shallow, while the wings of the horse are relatively deep. In other words, the body of the horse is sometimes not as clear as the wings.However, on the whole, the VAE model I modified to generate the pegasus was more successful.

5 Conclusion & Future work

In this assignment, I tried a variety of automatic encoder methods. After only using and processing CIFAR-10 data, I classified the data according to categories and selected only the data with spread wings and white horses in the bird and horse dataset.The generalization and robustness of the model are enhanced by data augment of the selected data. After the two features of encoder construction respectively to modify VAE and its loss function, finally build a model that can successfully generate white horse with wings, but according to the VAE method as explained in many papers, my method to generate Pegasus is relatively vague.Therefore, the future work is to try to use GAN models, hoping to generate more clear data.

References

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.