Image Compression Algorithm Based On Variational Autoencoder

Ying Sun’, Lang Li, Yang Ding, Jiabao Bai and Xiangning Xin
Chongqing University of Posts and Telecommunications, Chongqing, China

*Corresponding author e-mail: s180131045@stu.cqupt.edu.cn

Abstract. Variational Autoencoder (VAE), as a kind of deep hidden space generation model, has achieved great success in performance in recent years, especially in image generation. This paper aims to study image compression algorithms based on variational autoencoders. This experiment uses the image quality evaluation measurement model, because the image super-resolution algorithm based on interpolation is the most direct and simple method to change the image resolution. In the experiment, the first step of the whole picture is transformed by the variational autoencoder, and then the actual coding is applied to the complete coefficient. Experimental data shows that after encoding using the improved encoding method of the variational autoencoder, the number of bits required for the encoding symbol stream required for transmission or storage in the traditional encoding method is greatly reduced, and symbol redundancy is effectively avoided. The experimental results show that the image research algorithm using variational autoencoder for image 1, image 2, and image 3 reduces the time by 3332, 2637, and 1470 bit respectively compared with the traditional image research algorithm of self-encoding. In the future, people will introduce deep convolutional neural networks to optimize the generative adversarial network, so that the generative adversarial network can obtain better convergence speed and model stability.

Keywords: Variational Autoencoder, Traditional Autoencoder, Image Compression Algorithm

1. Introduction
Although the recent wave of artificial intelligence research led by deep learning has achieved very good results in the field of supervised learning [1-2], Unsupervised learning, as a method that can truly allow computers to learn from unlabeled real data from the real world, can avoid tedious and unavoidable data labeling work in supervised learning. If you want a computer to better understand the complex real world, the best way is to let the computer generate a representation of the real world in a certain way. The first thing needed to accomplish the above goals is the generative model [3-4].

In order to enhance the spatial resolution, Takeuchi et al. used the translation and aliasing properties of the Fourier transform in the frequency domain to achieve image super-resolution reconstruction based on the relative displacement complementary information between multiple
frames of images [5]. Since then, image super-resolution reconstruction technology has gradually become a hot spot in the field of image processing. X et al. used the back propagation algorithm in the autoencoder, which can realize the processing of high-dimensional complex data through encoding and decoding, which greatly promotes the development of neural networks, which can be used for dimensionality reduction and feature learning [6].

The most prominent performance in generative models in recent years is the variational autoencoder and the generative confrontation network. As an extension of the autoencoder, the former is a good combination of deep learning ideas and statistical learning. Through the encoder network, the high-dimensional distribution of the image can be used to reduce the dimensionality, and then the decoder network can be used to achieve low-dimensional the data distribution automatically generates an image similar to the original image [7-8]. The generative adversarial network uses the idea of zero-sum game, using generators and discriminators to alternate training, and finally can realize the operation from random noise to real image generation [9-10].

2. Research on Image Compression Algorithm Based on Variational Autoencoder

2.1. The Difference between Variational Autoencoder and Traditional Autoencoder

Although the recent wave of artificial intelligence research led by deep learning has achieved very good results in the field of supervised learning. Unsupervised learning, as a method that can truly allow computers to learn from unlabeled real data from the real world, can avoid tedious and unavoidable data labeling work in supervised learning. If you want a computer to better understand the complex real world, the best way is to let the computer generate a representation of the real world in a certain way. The first thing needed to accomplish the above goals is the generative model. The most prominent performance in generative models in recent years is the variational autoencoder and the generative confrontation network. Variational autoencoder, as an extension of autoencoder, combines deep learning ideas with statistical learning. Through the encoder network, the high-dimensional distribution of the image can be used for dimensionality reduction operations, and then its decoder network can be used Realize the operation of automatically generating an image similar to the original image from the low-dimensional data distribution. The generative adversarial network uses the idea of zero-sum game, using generators and discriminators to alternate training, and finally can realize the operation from random noise to real image generation.

Traditional autoencoders are unable to generate data autonomously due to the confusion and unknown output vectors of the hidden layer. Based on the previous years, Diederik PKingma and Max Wllig added a hidden variable z to the hidden layer of the traditional autoencoder, and then passed. This latent variable is used to automatically generate data to form a Variational Auto Encoder (VAE). Variational autoencoder, as a generative model that combines the characteristics of deep learning and statistical learning, makes good use of the very powerful capabilities of deep models when applied to nonlinear fitting.

The problem we want to solve in VAE is when there is now a data composed of N independent and identically distributed continuous or discrete random variables. Suppose these data are generated by a continuous random variable with hidden variable z through some random processes, and these processes include two steps:

(1) The data is generated by some prior distributions.
(2) The data is generated by some conditional distributions. Assuming that the sum comes from the sum of the distribution, and then no matter θ or z, their probability density functions are all differentiable. However, the real parameter θ* is unknown except for the value of the hidden variable.

In fact, the real challenge is: the marginal likelihood integral is difficult to handle, and the true posterior density is not easy to handle, which makes it impossible to use the EM algorithm and other algorithms. For example: when this is a neural network that contains a nonlinear hidden layer, when faced with a large-scale data set, the batch processing overhead will be very large, so it is always hoped to replace the original data with the smallest possible data Perform parameter updates. If a
sampling-based solution is adopted, it will involve cyclic sampling of each data point, and the cost will still be very high. Therefore, it is proposed in the variational autoencoder: effective maximum a posteriori estimation and maximum likelihood estimation for the \( \theta \) parameter, that is, approximate a posteriori inference for the hidden variable \( z \) of the given observation value \( x \) when the parameter \( \theta \) is selected, Perform effective approximate marginal inference on variable \( x \).

2.2. Image Quality Evaluation Measures

In an image coding system, the peak signal to noise ratio (PSNR) is commonly used to measure its performance, and the formulas are as shown in (1) and (2).

\[
MSE = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [x(m,n) - \hat{x}]^2
\]

(1)

\[
PSNR = 10 \log_{10} \frac{255^2}{MSE}
\]

(2)

Among them, MSN is the mean square error; \( x(m,n) \) is the pixel value of the original image; \( (m,n) \) is the pixel value after decompression.

The image super-resolution algorithm based on interpolation is the most direct and simple method to change the image resolution, and its computational complexity is the smallest. This type of method mainly uses the weighted average of known pixels to calculate the value of unknown pixels. That is, the two-dimensional convolution value between the sampling value of the interpolation image and the interpolation function is used. Traditional interpolation methods include Nearest Neighbor Interpolation, Bilinear Interpolation, Bicubic Interpolation and Cubic Spline Interpolation.

Nearest neighbor interpolation is also called zero-order interpolation, and it is the simplest method among all interpolation methods. This method directly sets the gray value of the point to be interpolated to the gray value of the point closest to it. The corresponding formula is:

\[
I(i+u,j+v) = I(i,j)
\]

(3)

Where \( u, v \) are floating point numbers in the interval \([0,1)\), and \( I(i,j) \) is the pixel value of the image at \((i, j)\). Although this method is simple in operation, the high-resolution image obtained has very obvious aliasing and the visual effect is very bad. Therefore it is rarely used.

The bilinear interpolation method is an improvement of the nearest neighbor interpolation method introduced above, and is also called first-order interpolation. This type of method first performs an interpolation operation in the X and Y directions, and then performs a weighted average operation based on the gray values of the four neighboring points of the point to be interpolated. The corresponding formula is as follows:

\[
I(i+u,j+v) = (1-u)(1-v)I(i,j) + (1-u)vI(i,j+1) + u(1-v)I(i+1,j) + uvI(i+1,j+1)
\]

(4)

Bicubic interpolation is also called cubic convolution interpolation. It is different from the first two methods that only consider the pixel values in the \(2\times2\) neighborhood. This method is to interpolate the pixel values in the \(4\times4\) neighborhood of the interpolation point. The calculation formula of the gray value of the point to be interpolated is expressed as:

\[
I(i+u,j+v) = A \times B \times C
\]

(5)

2.3. Image Quality Evaluation Standard

There are two main methods for evaluating the quality of reconstructed high-resolution images: subjective image quality evaluation methods and objective image quality evaluation methods.

The method of subjective evaluation of image quality is mainly a certain evaluation scale, combined with the feeling in the heart, to evaluate the reconstructed high-resolution image. Generally, a more formal subjective evaluation also requires scoring the image and then performing
normalization processing. The result obtained is the subjective evaluation of the reconstructed high-resolution image. The advantages of subjective evaluation methods are intuitive and clear. Its shortcomings are also obvious, that is, it does not use a mathematical model to describe the reconstructed image in detail, and it lacks stability, relies heavily on the external environment, and is easily affected by the observer's observation conditions and visual characteristics. The method of objectively evaluating image quality is a quantitative description of image quality. This kind of method can be divided into two kinds of methods according to whether the reference image is used or not. The first type of method does not require a reference image, but is based on the content of the image to be evaluated, with the help of some
Indexes such as average gradient, information entropy and contrast ratio are evaluated. Because this kind of method can only be used in some specific occasions, it has certain limitations and cannot be widely used. The second type of method is to select a reference image as a standard image, use the standard image as a benchmark, and use some mathematical models to obtain the evaluation index of the image to be evaluated, and to evaluate the image to be evaluated.
3. Image Compression Algorithm Experiment Based on Variational Autoencoder
3.1. The Image Compression Steps of the Variational Autoencoder Transform in the Experiment
The entire image is first transformed by a variational autoencoder, and then the actual encoding is applied to the complete coefficients. Variational autoencoder compression technology is one of the lossy compression technology, generally there are three processes:
(1) Transformation: Transform the transformed data into a wavelet coefficient matrix.
(2) Quantization: The coefficients are quantized into a finite alphabet, this step is not reversible.
(3) Encoding: The symbols obtained after quantization are further compressed to minimize the bit rate.
3.2. Experimental Algorithm
Variational autoencoders can be regarded as a mixture of neural networks and Bayesian networks. In variational autoencoders, the nodes corresponding to the hidden encoding can be regarded as random variables, and other nodes are regarded as ordinary neurons. In this way, The encoder becomes a variational inference network, and the decoder can be seen as a generating network that maps hidden variables to observed variables.
4. Discussion on Image Compression Algorithm Based on Variational Autoencoder
Among the four symbols generated by the encoding of the variational self-encoder, the probability of each symbol is not equal. Zero tree roots have the highest probability of occurrence, occupying more than 50% of the roots, and its continuity is very strong. It can be seen from Table 1 that after encoding with the encoding method improved by the variational autoencoder, the number of bits required for the encoding symbol stream required for transmission or storage of the traditional encoding method is reduced, symbol redundancy is avoided, and the symbol redundancy can be effectively improved. Image compression ratio and coding efficiency reduce algorithm complexity. Table 1 and Figure 1 respectively show the data of the number of coding bits required by the variational autoencoder and the traditional coding:

| Image | Encoding of traditional autoencoder | Variational autoencoder encoding | Save (bit) |
|-------|-------------------------------------|---------------------------------|------------|
| Image 1 | 25678 | 22346 | 3332 |
| Image 2 | 24783 | 22146 | 2637 |
| Image 3 | 23457 | 21987 | 1470 |
Figure 1. Data showing the number of coding bits required by the variational autoencoder and traditional coding

According to the similarity between the corresponding sub-image structures of different layers, the fractal coding of the first-level sub-image is constructed from the fractal coding of the second-level sub-image. This paper compares the experimental results with the experimental results of the traditional self-encoder image compression method. The results are shown in Table 2 and Figure 2. It can be seen from the comparison results that the image compression method based on the variational autoencoder only performs the construction process of the fractal code in 5/16 sub-image blocks of the original image, and the construction is in each smaller sub-image Complete, so that the encoding time is reduced by orders of magnitude compared with traditional methods.

Table 2. Comparison of image compression methods of traditional autoencoders and image compression methods of variational autoencoders

| Coding method                                      | Time (s) | PSNR(dB) | Rate(bpp) |
|----------------------------------------------------|----------|----------|-----------|
| Image compression method of traditional autoencoder| 78       | 28.8     | 0.28      |
| Image compression method of variational autoencoder| 2156     | 29.2     | 0.39      |

Figure 2. Comparison of image compression methods of traditional autoencoders and image compression methods of variational autoencoders

5. Conclusion

Variational autoencoder, as a special form of autoencoder model, has quickly become one of the most concerned forms in the field of deep generative models. VAE is a kind of deep hidden space generative model, which has great performance in data generation. The application value of, especially in image generation, has been widely used. This paper conducts experimental research based on the variational autoencoder algorithm. The experimental results show that the image research algorithm using variational autoencoder for image 1, image 2, and image 3 reduces the time of image research by using the variational autoencoder by 3332, 2637, respectively, 1470 bit.

Acknowledgments

Major national science and technology projects, NO.2018ZX03001026-002.
References

[1] Li H. Deep linear autoencoder and patch clustering-based unified one-dimensional coding of image and video. Journal of Electronic Imaging, 2017, 26(5):1.

[2] Wang R. Research on Image Generation and Style Transfer Algorithm Based on Deep Learning. Open Journal of Applied Sciences, 2019, 09(8):661-672.

[3] Hongliang, Wu, Weimin, et al. Research on image retrieval algorithm based on LBP and LSH. AIP Conference Proceedings, 2017, 1864(1):20038-20038.

[4] Cheng Z, Sun H, Takeuchi M, et al. Energy Compaction-Based Image Compression Using Convolutional AutoEncoder. IEEE Transactions on Multimedia, 2020, 22(4):860-873.

[5] Cheng Z, Sun H, Takeuchi M, et al. Energy Compaction-Based Image Compression Using Convolutional AutoEncoder. IEEE Transactions on Multimedia, 2019, PP (99):1-1.

[6] Han C, Duan Y, Tao X, et al. Toward Variable-Rate Generative Compression by Reducing the Channel Redundancy. IEEE Transactions on Circuits and Systems for Video Technology, 2020, PP (99):1-1.

[7] Srairi K, Messaoudi A. Colour image compression algorithm based on the DCT transform using difference lookup table. Electronics Letters, 2016, 52(20):1685-1686.

[8] Li S, Yin H, Fang X, et al. Lossless Image compression Algorithm and Hardware Architecture for Bandwidth Reduction of External Memory. IET Image Processing, 2017, 11(6):379-388.

[9] Sun G, Chu Y, Liu X, et al. A Plant Image Compression Algorithm Based on Wireless Sensor Network. Journal of Computer and Communications, 2019, 07(4):53-64.

[10] Aljanabi A K J. Ultrafast and Efficient Scalable Image Compression Algorithm. Journal of ICT Research and Applications, 2016, 9(3):214-235.