Item Recommendation based on Multimedia Variational Autoencoder

Yu Ran1,*, Ziyu Chen2, Yilin Mu3
1School of Information Technology & Management, University of International Business and Economics, Beijing, 100029
2College of Fisheries and Life Science, Dalian Ocean University, Dalian, Liaoning, 116023
3School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou, Gansu, 730050

*Corresponding author email: yuran@uibe.edu.cn

Abstract. At present, with the rapid development of science and technology and the rapid development of computer related technology, with the development of this series of technologies, multimedia communication has rapidly become the means of communication and communication. Digital image technology, a very important form of information representation in multimedia communication, has now become a mainstream way. The technology provides more convenience for people's lives, but also has many problems, such as security, transmission and storage. For example, some images will involve personal privacy, business secrets, and even government secrets. In addition, the rapid transmission of information in limited bandwidth is also very important. Especially in the presence of cloud storage, it can not only store large amounts of information, such as files, videos and images, but also provide a large enough online space to store shared data. In this case, the real-time and security requirements of information transmission are higher and stricter. It is of great theoretical and practical significance to compress the image and reduce the bandwidth and ensure the security of image encryption. Image encryption is also called digital image encryption. Image compression is mainly based on the use of different technologies to change the pixel value and location of the image to protect the image. Image compression mainly compresses the image by removing redundant or unrelated information. In this paper, the concept of the variational self encoder is introduced firstly, and then the image compression and encryption based on the variational self coder generation model are studied and summarized briefly.

Keywords: Multimedia Variational Autoencoder; image compression; editor decoder.

1. Introduction

With the rapid development of Internet in the information technology era in twenty-first Century, people's lives and work have become more convenient and fast [1]. Many important information is communicated through multimedia. And it has become the mainstream of information expression.
However, it also brings severe challenges to data storage, transmission and security [2]. Data will be attacked, leaked and spread arbitrarily in the process of transmission. In order to ensure data security and real-time, image compression and encryption technology has become an urgent problem and research hotspot [3]. To protect the image security, we can use image encryption and image hiding as the two main means. Image encryption mainly uses different technologies to change the pixel value and location of images to protect images. Image encryption is widely applied to biomedical institutions, e-commerce, network transmission, and military [4]. Image hiding technology is embedding image information into other carriers. It not only protects the content of the secret image, but also conceals the communication process itself, avoiding the attacker's attention [5]. Image encryption can be divided into: spatial frequency domain encryption and compression encryption according to whether compressed data is compressed. Compression encryption is the first compression of the image in the encryption operation [6]. Image information is fixed, but different image representations lead to different changes in the amount of data stored in the image. Therefore, in a large number of data representation, some data are necessary, some are useless or the information represented by the data is irrelevant or redundant [7]. The main purpose of image compression is to compress images by removing redundant or unrelated information. And it can store and transmit compressed digital data on low bandwidth network in real time [8].

2. Variational self encoder
Variational self encoders (VAE) is a generative model proposed by Diederik P. Kingma and Max Welling in 2013. It is an unsupervised learning generation model, which can classify images, reduce dimension and visualize [9]. The reason why variational self encoders are popular is that they are built on standard approximation functions (neural networks). Moreover, we can use random gradient descent training. At present, the more commonly used depth generation models are GAN, DBN and VAE, which are widely used in machine learning, and are important tools in machine learning [10]. These complex data include handwritten numerals, face recognition, home numbers, CIFAR images, scene physical models, segmentation. And predict the future from static images.

3. Research on image compression and encryption based on variational self coder generation model
The variable auto encoder is a very important generation model. It is proposed by Diederik P. Kingma and Max Welling. It can design complex data generation models and fit them into large data sets. The most advanced machine learning results can also be obtained in image generation and reinforcement learning. The basic structure is like encoder and decoder, but their mathematical theory bases are different. Since the encoder generates an implicit vector each time, it changes the self generator to generate two vectors of Cheng Junzhi (MU) and standard deviation (sigma) every time. The basic framework is shown in Figure 1. Since the encoder has the function of compressing the reduced dimension image data, it also changes itself from encoder to image compression. In the fourth chapter, the effect of the variable self encoder compression image is verified in the fourth chapter and compared with the mean square error (MSE) and the peak signal-to-noise (PSNR) of the compressed image of the multilayer self encoder (stack self encoder).
The generative model is a model that involves a wide range of machine learning fields. It defines the distribution of X in some potential high dimensional spatial data points as P(x). As a popular data, we can create a model class. Each point is composed of thousands or hundreds of thousands of dimensions (pixels), and the work of the generating model is to try to capture the correlation between pixels, for example, the color similarity of adjacent pixels can be organized into objects. To be exact, capturing these correlations depends on what we want to do with the model. We can use numerical value to calculate P(x) in the generation model. In the case of images, the value of X should be very high when it looks like real images, and the image that looks like random noise should get a very low probability.

3.1. A neural network for variational automatic encoders

The encoders and decoders in variational auto encoders are made up of multilayer neural network models. The neural network of encoder and decoder has many choices. For example, CNN, MPL and RNN., because the variational self encoder is a multilayer neural network, we can compress the image at different layers to reduce the bandwidth of data broadcasting. In this chapter, we use a relatively simple neural network. That is, the basic neural network framework of MLPs (multi-layered perceptrons) as shown in Figure 3-2 VAE. For encoder, we use Gauss output MLPs, and for decoder, we use MLPs output by Bernoulli.

The encoder is a multivariate covariance structure of Gauss distribution.

\[
\log q(z \mid x) = \log N(z; \mu, \sigma^2 I) \\
\text{among}, \mu = W_1h + b_2 \\
\log \sigma^2 = W_2h + b_3 \\
h = \tanh(W_1x + b_1)
\]

Among them, W1; W2; W3; B1; B2; B3 is the weight and offset of MLP, and is also the variational parameter of pH.

The decoder is a multivariate Bernoulli whose probability is a fully connected neural network and a hidden layer calculated by Z:

\[
\log p(x \mid z) = \sum_{i=1}^{D} x_i \log y_i + (1 - x_i) \cdot \log(1 - y_i) \\
\text{among}, y = f_{\sigma}(W_4 \tanh(W_3z + b_4) + b_5)
\]

Where \( f_{\sigma} \) is a sigma activation function, theta = \{W and b\} are the weights and offsets of MLP.
3.2. Image compression based on variational auto coder generation model

The VAE neural network can be seen as a multilayer model. Putting an image on the first level can output data from different layers and reconstruct the original image. If the output data dimension of any layer is smaller than the size of the original image, the representation in this layer is a compressed table. Because the model has more than one hidden layer neuron and each layer is smaller than the input layer, this model can achieve multiple levels of compressed images. Therefore, multiple compression ratios can be obtained by using this model. In the generative model, \( f(\cdot) \) is a nonlinear activation function. The experiment uses sigmoid function as activation function. By training VAE model, we can get a compression representation from a hidden layer. This is the form of image compression. Second hidden layers are selected to reconstruct the compression effect of the image.

3.3. Image encryption based on variational automatic encoders

The steps of image encryption based on variational self encoders: First, two standard images are trained by VAE respectively, and then the trained model data are separated and encrypted on the generating model. Figure 3 shows the flow of the algorithm.

3.3.1. Training generation model. In training generation mode, the more complex the relationship between dimensions is, the more difficult the model is to train. The training module for training VAE can be trained.

The MPL network is used to adjust the weight values to reconstruct the image. The specific steps are as follows:
(1) First grayscale the image and normalize the data to input VAE.
(2) Data can output mean and variance by encoder.
(3) The implicit variable Z can be derived from the formula, and finally the coding probability is obtained, \( q_z(z|x^{(i)}) \).

(4) The sampled Z input decoder is decoded through the Bernoulli network to output the reconstructed data, \( p_{\text{z}}(x^{(i)}|z) \). Probability.

(5) Using random gradient iterative training VAE network to adjust the weight offset by formula and repeat the previous steps until the clear image is reconstructed and the minimum loss function is calculated.

3.3.2. Encryption and decryption of generation model. The VAE neural network shows that the variational self-assembly device is a multi-layer unsupervised neural network. It can learn the basic characteristics of the image by training the unsupervised neural network and train the birth model. As long as the characteristics of the model image can be generated, a clear image can be generated. This experiment uses the weight and offset of the generation model to change the main data of the generating model. Let the generating model generate unrecognized noise images, that is, image encryption. Experiments show that the last layer of the decoder (generation model) has a greater impact on encryption. This paper chooses the final data generated by the model to encrypt. The W of the two image data trained by the variational self encoders (VAE), the B corresponding to the division to change the data in the W model, and the B. The generated data is generated on the generating model to generate encrypted images, such as the 4 encryption flow chart.

![Image Encryption Flow Chart](image1.png)

**Figure 4** image encryption flow chart

When decrypting, we only need to load the inverse operation of the data to the clear image generated by the variational self coder.

![Image Decryption Flow Chart](image2.png)

**Figure 5** image decryption flow chart

4. Conclusion

By studying the basic framework of the variational automatic encoder model and the research of neural network, we have learned how to compress image to reduce the network bandwidth and transmit the data in real time. The frame of the variational self encoder encryption frame clearly illustrates the process of image encryption based on the variant self encoder generating mode, and the specific steps of the training variant self encoder generation model are understood. We use gradient descent to update the weight and offset to adjust the error, get the minimum error value and generate...
the weight and offset in the network model when the reconstructed image is clear. We use the weight and offset data of the two trained images to divide the data of the variant encoder to generate the encrypted image. After loading the data into the variational self coder generation model, we can decipher the data and generate clear images.

References

[1] Zhao Lijun. Image coding and enhancement based on optimized learning [D]. Beijing Jiaotong University, 2019.

[2] Wang Yang. Video coding technology based on deep learning [D]. Harbin Institute of Technology, 2019.

[3] duchuan. Feature learning method based on deep network generation [D]. Xi'an Electronic and Science University, 2019.

[4] He Yijiang, Du Junping, Kou Feifei, Liang Meiyu, Wang Wei, Luo. Image self coding algorithm based on deep convolution neural network [D]. Journal of Shandong University (Engineering Edition), 2019, 49 (02): 61-66+73.

[5] Pan Zhiyong. Image classification method based on structured topic model [D]. Harbin Institute of Technology, 2019.

[6] Huang Guojie, Jin Hui, Yu Yibiao. Enhanced variational self encoder for non parallel corpus speech conversion [J]. signal processing, 2018, 34 (10): 1246-1251.

[7] Yang Guangyi. Image quality assessment and its application in image denoising [D]. Wuhan University, 2018.

[8] Song Xiaodan. Research on image, video, communication and storage technology for cloud mobile media computing [D]. Xi'an Electronic and Science University, 2017.

[9] Ding Shouhong. Research on intelligent processing technology of visual media based on structural analysis [D]. Shanghai Jiao Tong University, 2016.

[10] Coding theory, channel theory and technology, encoder and decoder [J]. Electronic Science and technology digest, 2000 (10): 44-50.