IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 25, 2019, accepted February 13, 2019, date of publication February 21, 2019, date of current version March 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900498

Multiple Description Coding Based on Convolutional Auto-Encoder

HONGFEI LI, LILI MENG, JIA ZHANG, YANYAN TAN, YUWEI REN, AND HUAXIANG ZHANG
School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China
Institute of Data Science and Technology, Shandong Normal University, Jinan 250014, China

Corresponding author: Lili Meng (mengll_83@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572298, Grant 61601269, Grant 61602285, Grant 61601268, Grant 61702310, and Grant U1836216, and in part by the Key Research and Development Foundation of Shandong Province under Grant 2017GGX10117 and Grant 2017CXGC0703.

ABSTRACT Deep learning, such as convolutional neural networks, has been achieved great success in image processing, computer vision task, and image compression, and has achieved better performance. This paper designs a multiple description coding frameworks based on symmetric convolutional auto-encoder, which can achieve high-quality image reconstruction. First, the image is input into the convolutional auto-encoder, and the extracted features are obtained. Then, the extracted features are encoded by the multiple description coding and split into two descriptions for transmission to the decoder. We can get the side information by the side decoder and the central information by the central decoder. Finally, the side information and the central information are deconvolved by convolutional auto-encoder. The experimental results validate that the proposed scheme outperforms the state-of-the-art methods.

INDEX TERMS Convolutional auto-encoder (CAE), multiple description coding (MDC), predictive coding, quality metric.

I. INTRODUCTION

The aim of image compression is to remove the redundant information of the image by using the relevance. The compressed image can be stored and transmitted at low bit rates. In recent years, deep learning has been widely used in lossless and lossy image compression, and has achieved good performance. Deep convolutional neural networks have become a universal tool for successfully solving advanced computer vision tasks. Recently, CNNs have been applied to the field of low-level computer vision and image processing to solve the relatively shallow regression problem of most networks. Machine learning methods are applied to lossy image compression and have achieved promising results with auto-encoders. Auto-encoders has been used in dimensionality reduction, compact representation of images. Parameters in the Auto-encoders can be optimized by the minimized loss function, and it is desirable to achieve better compression performance than traditional image compression algorithms, including JPEG [1] and JPEG2000 [2].

In the training process of the deep learning model, a prediction result is obtained from the input to the output, which will get an error compared with the real result. This error will backpropagation at each level of the model, and the representation of each layer will be adjusted according to this error until the model converges or achieves the desired effect. To solve this problem, end-to-end models are designed. In [3], a complete image compression method for end-to-end optimization of rate and transmission performance is proposed, which is based on nonlinear transform coding. This compression method improves the rate and outperforms the performance of JPEG and JPEG 2000. In [4], an efficient end-to-end compression framework based on two CNNs is proposed, the first CNN is used to generate a compact intermediate representation for encoding using an image encoder, the second CNN is used to reconstruct the high-quality decoded image.

Different from the above end-to-end model, some neural network methods for image compression are proposed. The work [5] proposes a general framework for variable rate image compression and a novel architecture based on convolution and deconvolution LSTM recurrent networks, which provides better visual quality than (headerless) JPEG, JPEG2000 and WebP for compressing 32 x 32 small thumbnail images. In [6], a set of full-resolution lossy image compression methods based on recurrent neural networks are

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

proposed. These work have achieved promising coding performance, but they do not use a real entropy encoder to generate the final code. In [7], a novel $l2$-layer deep convolutional neural network is proposed for compression artifact suppression in JPEG images with hierarchical skip joins and training using multi-scale loss functions.

And some approches are proposed to take advantage of the auto-encoder for image compression. In [8], a lossy image compression method based on machine learning is proposed, which is superior to all existing codecs in real-time operation. In [9], a new method of optimizing auto-encoders for lossy image compression is proposed, which uses residual networks and sub-pixel convolution. This method is inspired by the work of [10]. In [11],the development of the deep convolutional auto-encoder in the Caffe deep learning framework is introduced, but the proposed model does not include the pooling/unpooling layers. In [12],a comprehensive performance comparison was performed on three overall compression architectures based on convolutional autoencoders (CAE) to generative adversarial networks (GAN) and super-resolution (SR) for Image Compression. In [13], a lossy image compression architecture using deep convolutional auto-encoder to achieve efficient coding is proposed.

MDC encodes the source by multiple independent descriptions for transmission, and any independent code stream can reconstruct the source, whose quality is within the acceptance range at the decoder. Moreover, the more descriptions obtained by the decoder, the better the quality of the recovered data. Therefore, it has strong robustness. MDC method is also gradually being used in the processing of video, image and various multimedia signals [14]–[21], and has achieved promising results. MDC can solve the problem of packet losses in a communication network [23]. The MDC method can prevent the image and video quality from being seriously degraded due to packet error or loss.

In [14], a multiple description scalar quantizer(MDSQ) method is proposed, which involves an index assignment problem. This is the first practical MDC scheme. Then applied to the entropy-constrained multiple description scalar quantizers coding scheme [15]. In [16], a two-stage MDSQ scheme is proposed, whose product of central and side distortion is closer to the multiple description rate-distortion bound, and its simple structure can achieve asymptotic distortion products with a smaller gap from the rate-distortion boundary. In [17], using multiple description uniform scalar quantization (MDUSQ), robust and progressive transmission over unreliable channel is performed. In [18], error resilient data compression algorithms based on wavelets, multiple description scalar quantizers and erasure-resilient codes are designed. Reference [19] describes a new algorithm for an optimal generalized multiple description vector quantization(MDVQ),and it performs well in the case of extensive packet loss. In [20], the problem of lattice vector quantizer design is solved for the two-channel multiple description, and the label problem is solved.In [21], a multi-description image coding scheme based on multi-description lattice vector quantization is proposed to achieve better rate and central/side distortion performance. In [22], presenting a construction of multiple description trellis-coded quantizer(MD-TCQ), the complexity of the scheme is almost independent of the rate. In [24], two multi-description coding schemes with random and uniform offset quantization are proposed, the proposed scheme is applied to multi-description image coding based on overlapping transform. Reference [25], [27] are improvements to multiple description coding with randomly offset quantizers(MDROQ).In [28],a novel multi-description video coding scheme based on the characteristics of the human visual system (HVS) is proposed.

In [26], according to image's context features, a new standard-compliant MDC framework based on deep convolutional neural network is proposed, this is the first work using convolutional neural network for multiple description coding.

In this paper, we propose a MDC framework based on symmetric CAE for image compression. Our main contributions are lists as follows:

1) We designe a symmetric CAE network with convolution/deconvolution filter pairs to generate feature maps with low dimensions.
2) MDC is used to replace the traditional codec, avoiding the delay caused by retransmission, and solving the packet losses problem, thus ensuring the real-time transmission of information, realizing the robust transmission of image data and improving the image quality of reconstruction.
3) In this paper, we combine MDC and CAE for image compression to achieve high quality coding efficiency.

The remainder of this paper is organized as follows. The section 2 introduces the framework of this paper, including the CAE network and the MDC framework. The experimental results are given in the section 3. In the section 4, we conclusion this paper.

II. THE PROPOSED FRAMEWORK

In this section, the proposed scheme is introduced firstly, then CAE network and the MDC framework are presented.

A. ARCHITECTURE OF THE PROPOSED SCHEME

CAE is mainly used for image reconstruction, such as image compression and denoising. The MDC is used to replace the traditional compression codec, which is possible to prevent the image and video quality from being seriously degraded due to packet loss or bit error. In this paper, we propose a MDC framework based on symmetric CAE for image compression. The combination of CAE and MDC for image compression can improve the coding efficiency, and obtain a more accurate reconstructed image. As far as we know, there is only one article combining CNN and MDC for image compression [26]. Our framework has two components: symmetric CAE network and MDC framework, as depticted in the Fig. 1.

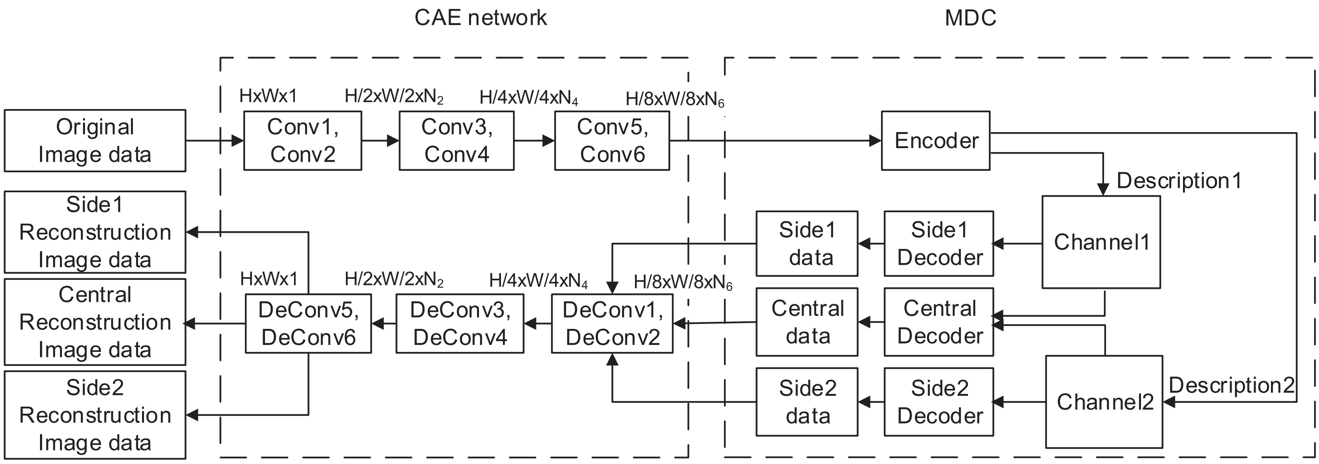


FIGURE 1. MDC framework based on symmetric CAE.

At the encoder, the input image is subjected to a convolution operation of CAE to obtain a set of convolution features, and the extracted features are passed through an MDC encoder, and are divided into two descriptions and transmitted to the decoder by different channels. At the decoder, if all the descriptions are received, we can obtain the central reconstructed data by the central decoder; If only one description is received, we can get the side reconstructed data by the side decoder. Finally, the obtained central data and side data are processed by the CAE deconvolution operation, and we can obtain high quality central reconstructed images and side reconstructed images.

MDC assumes that there are multiple channels between the source and the receiver, and multiple descriptions are transmitted to the receiver by different channels, ensuring that acceptable image quality is still obtained when only some descriptions are received. It can be seen that the more descriptions obtained, the better the reconstructed image quality. Since a part of the information can be used to reconstruct acceptable-quality image, MDC is widely used in the field of image processing and image compression. CAE can extract the compression features of the image and input the features into the MDC for further compression, thereby improving compression efficiency. By combining MDC with CAE, higher quality reconstructed images can be obtained.

B. ARCHITECTURE OF CONVOLUTIONAL AUTO-ENCODER NETWORKS

CAE can be used to learn the missing part of the image or denoising for image reconstruction. CAE defines the task of the filter: the best filter is learned by minimizing reconstruction errors for the model. Using the convolution and pooling operations of CNNs to achieve unsupervised feature extraction of feature invariant extraction, the process of its implementation is consistent with the idea of an auto-encoder. Deconvolution is usually used to map low-dimensional features into high-dimensional inputs, which is the opposite of the convolution operation, that is, the forward propagation process of the convolutional layer is the back propagation process of the deconvolution layer.

CAE is a neural network that uses a backpropagation algorithm [29] to make the output value equal to the input value. Firstly, it compresses the input image into a latent spatial representation, and then the output is reconstructed by this characterization.

In this paper, a symmetric CAE network is designed (as shown in Fig. 2). In order to obtain a compressed representation of the input image, the CAE encoding/decoding process requires downsampling/upsampling operations. *Ni* represents the number of filters for convolution or deconvolution. In [10], it points out that the super resolution can be achieved more efficiently by convolving the image and then upsampling, rather than upsampling and then convolving the image. Continuous downsampling reduces the quality of

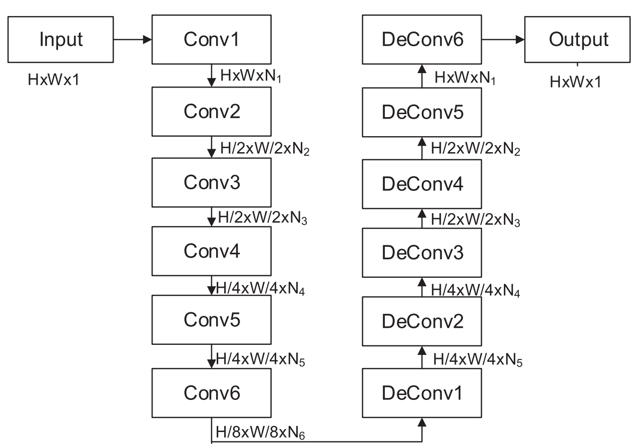


FIGURE 2. The framework of CAE network. Instead of the pooling layer, the convolution layer with stride set to 2 is utilized to preserve image information as much as possible.

the reconstructed image, so this paper uses a convolution/deconvolution filter pair for downsampling or upsampling.

The number of filters in the convolutional layer is set to {32, 32, 64, 64, 64, 32}, and the number of deconvolution layer filters is inverse to the convolutional layer. Instead of the pooling layer, the convolution layer with stride set to 2 is utilized to preserve image information as much as possible. All convolutional layer activation functions use the ReLU function.

When performing a convolution operation, convolution operation between each input data having a depth D, $X = X_1, \ldots, X_D$, and a set of convolution kernels $K_1^{(1)}, \ldots, K_n^{(1)}$ to produce a set of feature maps. In order to improve the generalization ability of the network, each convolution is activated by the nonlinear function $a()$, and the resulting network can learn some nonlinear characteristics of the input data:

$$f_m = a(X * K_m^{(1)} + b_m^{(1)}), \quad m = 1, \ldots, n \tag{1}$$

where X is input data, $K_m^{(1)}$ is m-th convolution kernel, $b_m^{(1)}$ indicates the bias of the m-th feature map. The reconstructed image $\hat{X}$ is the result of convolution between the dimension of the feature F and the deconvolution filter $K^{(2)}$.

$$\hat{X} = a(F * K_m^{(2)} + b^{(2)}) \tag{2}$$

The mean square error(MSE) between the original image X and the reconstructed image $\hat{X}$ is expressed as

$$L(x, \hat{x}) = \frac{1}{2}||X - \hat{X}||_2^2 \tag{3}$$

The CAE consists of two parts: (1) convolutional layer, represented by the encoding function $y = f(x)$,which can be used to compress the input into a latent spatial representation. (2) deconvolutional layer, represented by the decoding function $\hat{x} = g(y)$,which can be used to reconstruct the input from the latent spatial representation. Therefore, the entire auto-encoder can be described by the function $g(f(x)) = \hat{x}$, where the output $\hat{x}$ is close to the original input x. The loss function of CAE is defined as

$$\begin{aligned} L &= \frac{1}{2}\,||x - \hat{x}||^2 + \lambda||y||^2 \\ &= \frac{1}{2}\,||x - g(f(x))||^2 + \lambda||f(x)||^2 \end{aligned} \tag{4}$$

λ controls rate-distortion tradeoff. In this paper, the CAE network is optimized with the Adam algorithm [30]

C. FRAMEWORK OF MULTIPLE DESCRIPTION CODING

MDC is used to replace the traditional compression codec, which can solve the problem of packet losses in a communication network. In the MDC framework applied in this paper, the input source is divided into M subset, and M descriptions can be got. The general expectation distortion expression of MDC, which can be written as

$$D = \sum_{k=0}^{M} p_k D_k \tag{5}$$

(a) (b) (c) (d)

FIGURE 3. The data-set is used for our testing.

where p_k is the probability of received k descriptions, and D_k is the corresponding expected distortion. When $k = 0$, D_k is the variance of input.

At the encoder, the features extracted from the CAE are divided in to M subset. For one description, uses a smaller quantization step size to quantify a subset, and the other subsets are sequence predicted by the encoded subsets in the same description, and prediction redundancy is encoded with a larger quantization step size.

As shown in Fig. 1, we apply two descriptions coding. For Description 1, subset 1 is quantized with a smaller quantization step size q_0, subset 2 is predicted with reconstructed subset 1, and prediction redundancy e_i is quantized with a larger quantization step size q_1. For Description 2, subset 2 is quantized with a smaller quantization step size q_0, subset 1 is predicted with reconstructed subset 2, and prediction redundancy e_i is quantized with a larger quantization step size q_1. In the i-th description, $\hat{e}_i$ is used to identify reconstruction redundancy, therefore, the reconstructed value in the i-th description is

$$\hat{y}_i = \bar{y}_i + \hat{e}_i \tag{6}$$

y is the feature extracted from the CAE, $\bar{y}_i$ represents the predicted value of y in the i-th description.

Description 1 and Description 2 are transmitted to the decoder by two separate channels, respectively. If all the descriptions are received completely, the central decoder is used to obtain high quality reconstruction data; If only one description is received, the missing information can also be recovered by the side decoder using the received information, we can get the acceptable information. The more descriptions received by the decoder, the better the quality of the reconstructed source.

III. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed scheme, the objective and visual experimental results are given in this section.

A. DATASETS FOR TRAINING AND TESTING

Our whole framework is implemented on the Tensorflow [31] platform. The 400 images with size 180 × 180 from [32] are used as our training data-set,which are cropped, flipped, and rotated to get the final number 3,200 of images with size of 128 × 128 used for our training data set. In our experiments, H and W are set to 128; therefore the input image in the CAE is split into non-overlapping 128 × 128 patches,which can be compressed independently. The test image is given in Fig. 3.

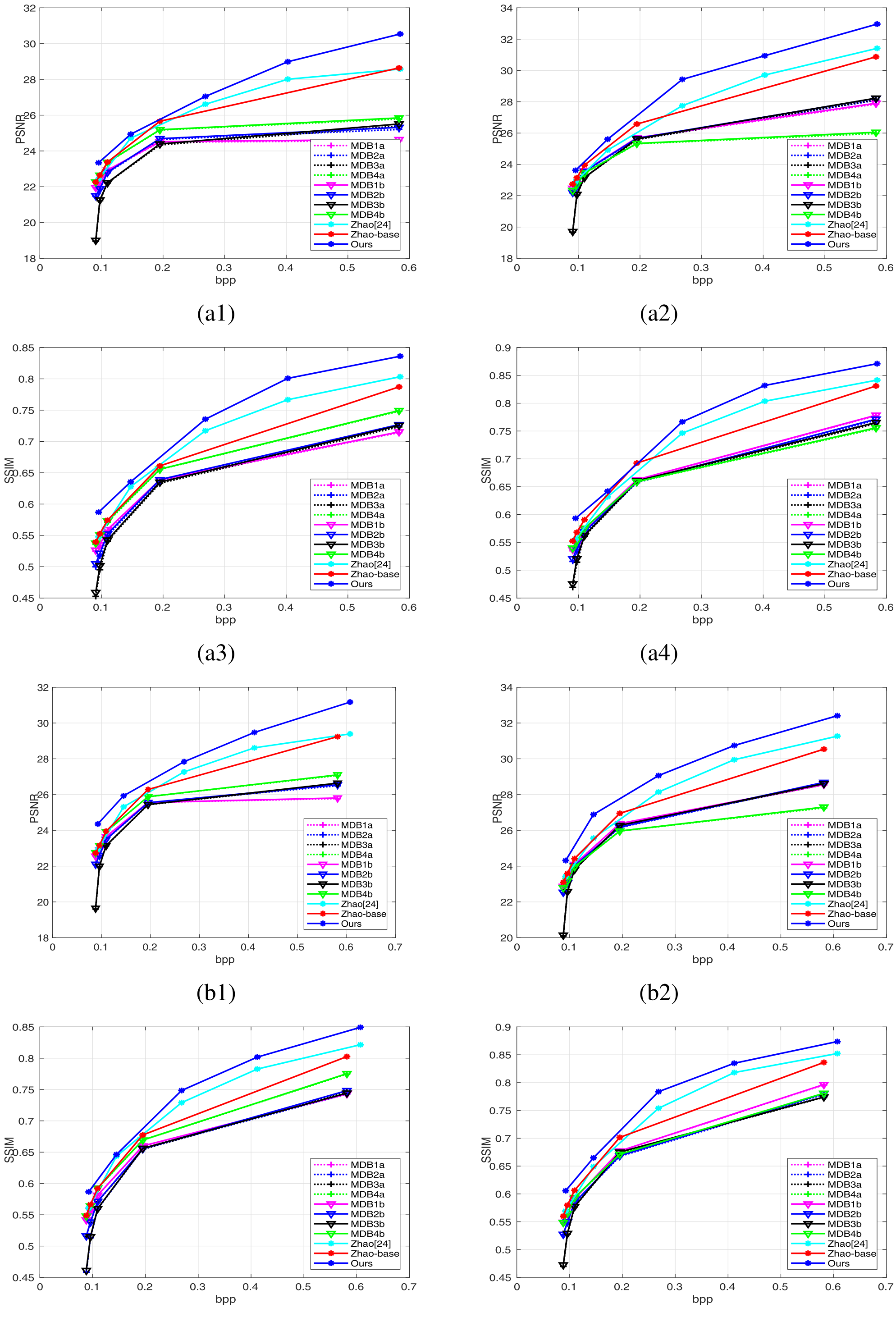


FIGURE 4. Objective metrics comparison for side reconstruction and central reconstruction of PSNR and SSIM of the image(a) and (b) in Fig.3 with the state-of-the-art approaches.

Please note that the test image is not included in the training datasets. The learning rate of training is initially set as 0.0001.

B. EXPERIMENTAL RESULTS

We use four state-of-the-art artifact removal techniques [33]–[36] and super-resolution techniques based on very deep convolutional neural networks [37] to form four baselines "MDB1a-MDB4a". At the same time, the super resolution of [38] is combined with artifact removal [33]–[36] to construct the other four baselines "MDB1b-MDB4b". Based on the image context feature, the new CNN network based on the MDC framework [26], which is denoted as "Zhao[24]". In addition, in order to fully prove the efficiency of the proposed method, Zhao's

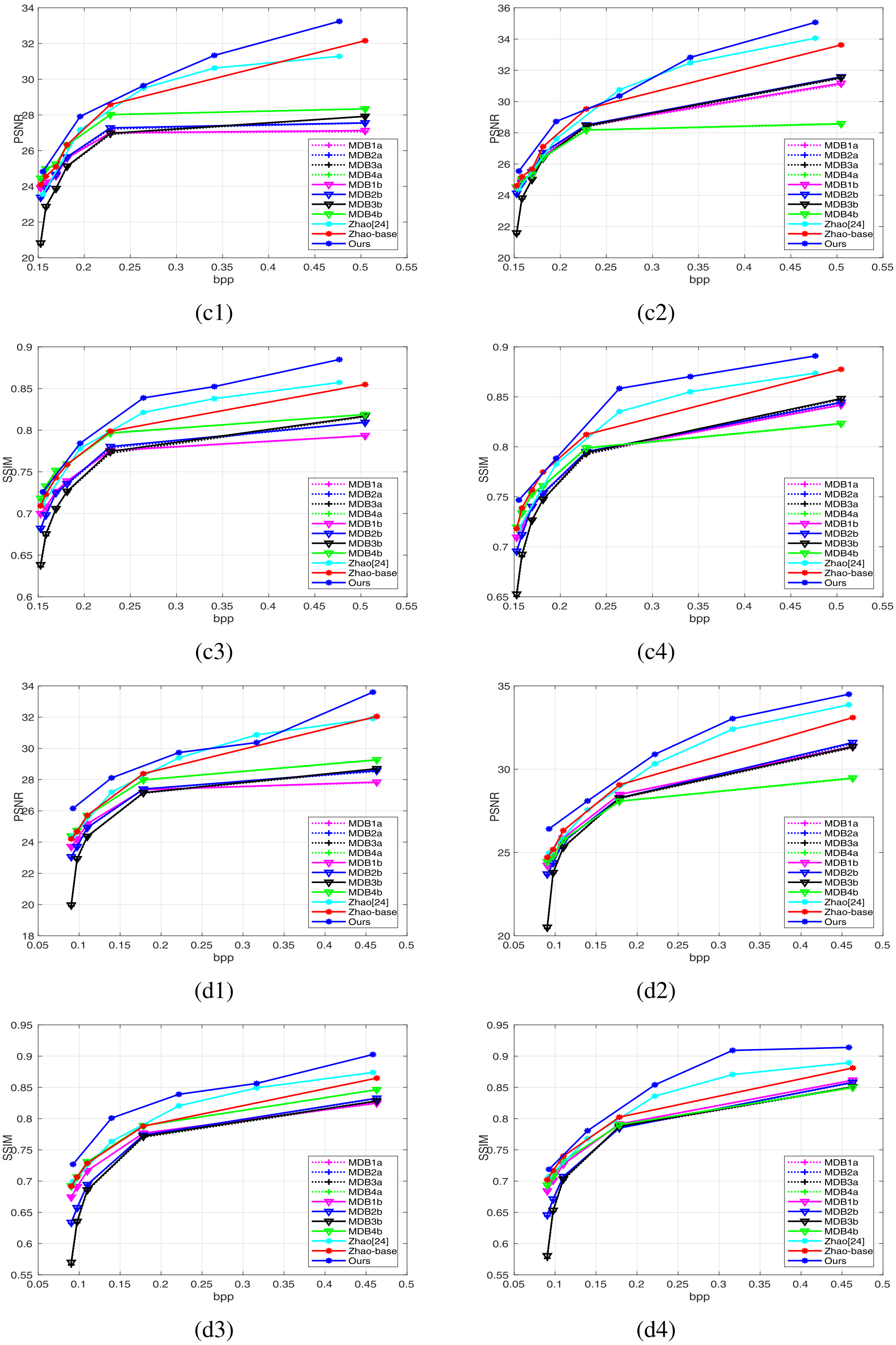


FIGURE 5. Objective metrics comparison for side reconstruction and central reconstruction of PSNR and SSIM of the image(c) and (d) in Fig.3 with the state-of-the-art approaches.

replaces the multiple description generating network MDGN with poly-phase downsampling technique in [39], forming a baseline model, which is expressed as "Zhao-base". For the sake of convenience, "Ours" denotes our proposed framework. To verify the efficiency of the framework, we use Peak Signal to Noise Ratio (PSNR) and structural similarity index measure (SSIM) to measure the objective quality.

In Fig. 4, (a1,b1), (a3,b3) are the side reconstruction PSNR and the central reconstruction PSNR of the image (a) and (b)

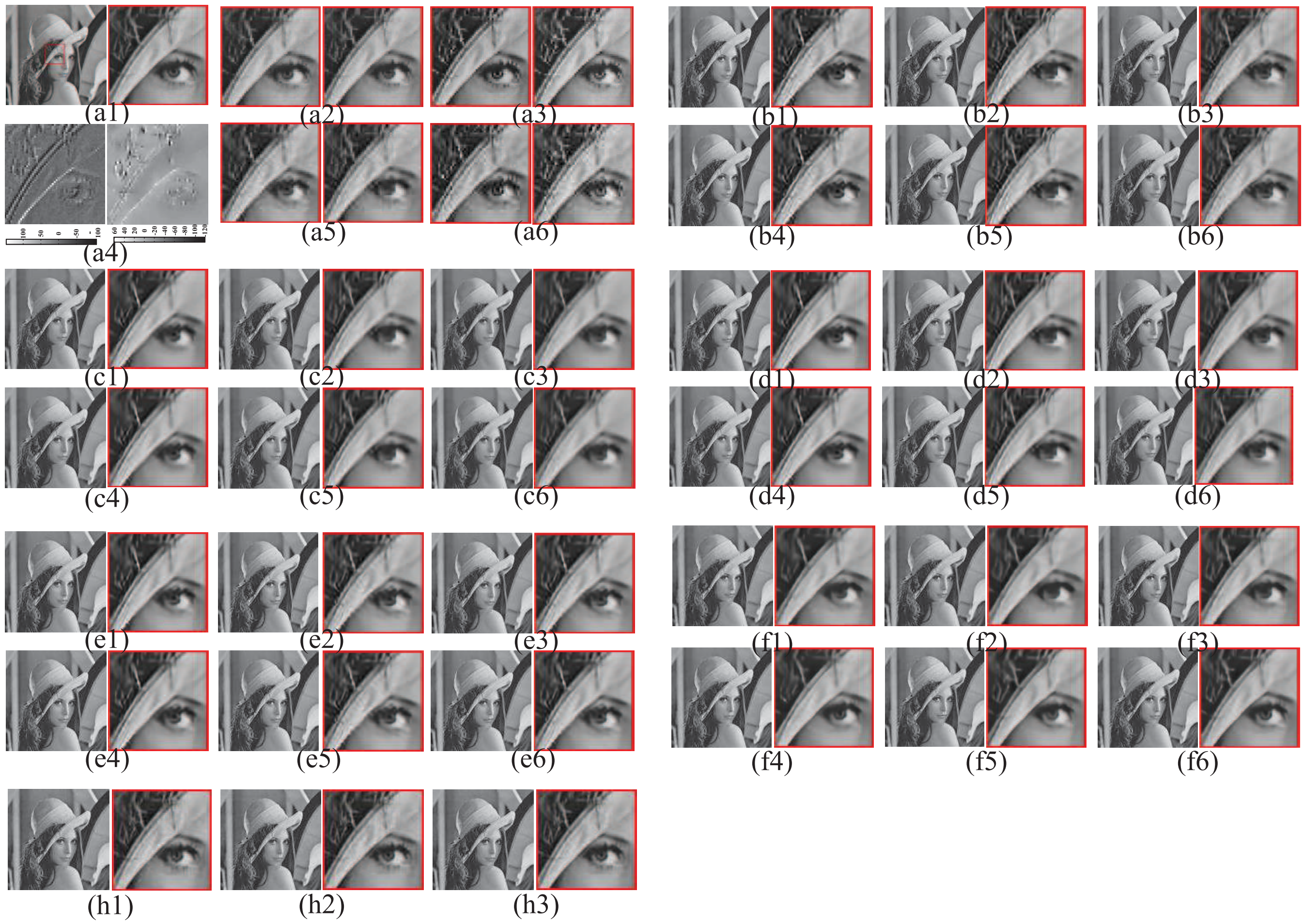


FIGURE 6. **Comparison of visual quality of different methods of image (d) in Fig. 3. Where (a1) represents the original image of the input, (a2) represents multiple descriptions created by poly-phase down-sampling techniques,and (a3) represents multiple description networks generated by Zhao's proposed MDGN, (a4) represents differences between a pair of images in (a2) and (a3), respectively, (a5)and (a6) represent compressed images of (a2) and (a3),respectively; (b-h) description the reconstructed image, where the (b1-h1, b2-h2) and (b4-f4, b5-f5) represents side reconstructed images, the (b3-h3) and (b6-f6) represents central reconstructed images; (b1-b3) MDB1a (27.808/0.824/0.232(s) and 31.319/0.861/0.463(c)), (b4-b6) MDB1b (27.843/0.825/0.232(s) and 31.367/0.862/0.463(c)); (c1-c3) MDB2a (28.524/0.832/0.232(s) and 31.526/0.857/0.463(c)), (c4-c6) MDB2b (28.579/0.833/0.232(s) and 31.601/0.858/0.463(c)); (d1-d3) MDB3a (28.642/0.827/0.232(s) and 31.288/0.850/0.463(c)), (d4-d6) MDB3b (28.700/0.828/0.232(s) and 31.352/0.851/0.463(c)); (e1-e3) MDB4a (29.244/0.846/0.232(s) and 29.428/0.850/0.463(c)), (e4-e6) MDB4a (29.270/0.846/0.232(s) and 29.471/0.850/0.463(c)); (f1-f3) Zhao-base (32.040/0.865/0.232(s) and 33.098/0.881/0.463(c)), (f4-f6) Zhao[24] (31.913/0.874/0.229(s) and 33.865/0.889/0.458(c))); (h1-h3) Ours (33.597/0.902/0.229(s) and 34.496/0.914/0.458(c)).**

in Fig.3, respectively. (a2,b2), (a4,b4) are the side reconstruction SSIM and the central reconstruction SSIM of the image (a) and (b) in Fig.3, respectively. And in Fig. 5, (c1,d1), (c3,d3) are the side reconstruction PSNR and the central reconstruction PSNR of the image (c) and (d) in Fig.3, respectively. (c2,d2), (c4,d4) are the side reconstruction SSIM and the central reconstruction SSIM of the image (c) and (d) in Fig. 3, respectively.

From the Fig. 4 and Fig. 5, it can be seen that our method has more PSNR and SSIM gain than MDB1a-MDB4a, MDB1b-MDB4b, Zhao[24] and Zhao-base. In (c1-c4) of Fig. 5, Ours may have a litter lower PSNR and SSIM than MDB4a, MDB4b, Zhao-base or Zhao[24] at low bit rates.

In this paper, we compare the visual quality of the proposed method with MDB1a-MDB4a, MDB1b-MDB4b, Zhao-base and Zhao[24], as shown in Fig. 6. Among these images, MDB1a (27.808/0.824/0.232(s) and 31.319/0.861/0.463(c)) represents PSNR/SSIM/bpp metrics for side reconstruction and center reconstruction based on the MDB1a method. In the Fig. 6, the visual quality comparison of the proposed method with state-of-the-art methods of (d) in Fig.3 is given. It can be seen that the side reconstruction image and the central reconstruction image of the proposed method have more detail preservation than the other methods.

From the above comparison of objective and visual quality, we can see that our method is better than the state-of-the-art methods. It is shown that the combination of CAE and MDC for image compression can improve the image reconstruction efficiency and obtain high-quality side reconstructed images and central reconstructed images.

IV. CONCLUSION

In this paper, we propose a MDC framework based on CAE for image compression. Firstly, a symmetric CAE network using downsampling/upsampling pairs is designed to replace the conventional transforms. Secondly, MDC is used to replace the traditional compression codec, avoiding the delay caused by retransmission, and solving the packet loss problem, thereby ensuring the real-time performance of information transmission.The experimental results show that the proposed framework can achieve better coding efficiency.

REFERENCES

[1] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consum. Electron.*, vol. 38, no. 1, pp. 18–34, Feb. 1992.

[2] M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard," *Signal Process., Image Commun.*., vol. 17, no. 1, pp. 3–48, Jan. 2002.

[3] J. Ballé, V. Laparra, and E. P. Simoncelli. (2016). "End-to-end optimized image compression." [Online]. Available: https://arxiv.org/abs/1611.01704

[4] W. Tao *et al.*, "An end-to-end compression framework based on convolutional neural networks," in *Proc. IEEE Data Compress. Conf.*, Apr. 2017, p. 463.

[5] G. Toderici *et al.*, "Full resolution image compression with recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2016, pp. 5306–5314.

[6] L. Cavigelli, P. Hager, and L. Benini, "CAS-CNN: A deep convolutional neural network for image compression artifact suppression," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2016, pp. 752–759.

[7] G. Toderici *et al.* (2015). "Variable rate image compression with recurrent neural networks." [Online]. Available: https://arxiv.org/abs/1511.06085

[8] O. Rippel and L. Bourdev, "Real-time adaptive image compression," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2922–2930.

[9] L. Theis, W. Shi, A. Cunningham, and F. Huszár. (2017). "Lossy image compression with compressive autoencoders." [Online]. Available: https://arxiv.org/abs/1703.00395

[10] W. Shi *et al.*, "ar, Ferenc, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[11] V. Turchenko and A. Luczak, "Creation of a deep convolutional auto-encoder in caffe," in *Proc. 9th IEEE Int. Conf. Intell. Data Acquisition Adv. Comput. Syst., Technol. Appl. (IDAACS)*, vol. 2, 2017, pp. 651–659.

[12] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto. (2018). "Performance comparison of convolutional autoencoders, generative adversarial networks and super-resolution for image compression." [Online]. Available: https://arxiv.org/abs/1807.00270

[13] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep convolutional autoencoder-based lossy image compression," in *Proc. IEEE Picture Coding Symp. (PCS)*, San Francisco, CA, USA, Jun. 2018, pp. 253–257.

[14] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[15] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple-description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 245–250, Jan. 1994.

[16] M. Liu and C. Zhu, "Enhancing two-stage multiple description scalar quantization," *IEEE Signal Process. Lett.*, vol. 16, no. 4, pp. 253–256, Apr. 2009.

[17] T. Guionnet, C. Guillemot, and S. Pateux, "Embedded multiple description coding for progressive image transmission over unreliable channels," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2001, pp. 94–97.

[18] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 813–826, May 2000.

[19] M. Fleming and M. Effros, "Generalized multiple description vector quantization," in *Proc. Data Compress. Conf.*, 1999, pp. 3–12.

[20] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, "Multiple-description vector quantization with lattice codebooks: Design and analysis," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1718–1734, Jul. 2001.

[21] H. Bai, C. Zhu, and Y. Zhao, "Optimized multiple description lattice vector quantization for wavelet image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 912–917, Jul. 2007.

[22] H. Jafarkhani and V. Tarokh, "Multiple description trellis coded quantization," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 799–803, 1999.

[23] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[24] L. Meng, J. Liang, U. Samarawickrama, Y. Zhao, H. Bai, and A. Kaup, "Multiple description coding with randomly and uniformly offset quantizers," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 582–595, Feb. 2014.

[25] J. Zong *et al.*, "Adaptive Reconstruction based Multiple Description Coding with Randomly Offset Quantizations," *Multimedia Tools Appl.*, vol. 77, no. 8, pp. 1–21, 2018.

[26] L. Zhao, H. Bai, A. Wang and Y. Zhao. (2018). "Multiple description convolutional neural networks for image compression." [Online]. Available: https://arxiv.org/abs/1801.06611

[27] J. Zong, L. Meng, Y. Tan, and Y. Ren, "Perceptual multiple description coding with randomly offset quantizers," in *Proc. IEEE Int. Symp. Circuits Syst.*, Dec. 2017, pp. 1–5.

[28] H. Bai, W. Lin, M. Zhang, A. Wang, and Y. Zhao, "Multiple description video coding based on human visual system characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 8, pp. 1390–1394, Aug. 2014.

[29] S. Ioffe and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167

[30] D. P. Kingma and J. Ba. (2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[31] M. Abadi *et al.* (2016). "TensorFlow: Large-scale machine learning on heterogeneous distributed systems." [Online]. Available: https://arxiv.org/abs/1603.04467

[32] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.

[33] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive DCT for high-quality denoising and deblocking of grayscale and color images," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1395–1411, May 2007.

[34] H. Chang, M. K. Ng, and T. Zeng, "Reducing artifacts in JPEG decompression via a learned dictionary," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 718–728, Feb. 2014.

[35] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[36] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao, "CONCOLOR: Constrained non-convex low-rank model for image deblocking," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1246–1259, Mar. 2016.

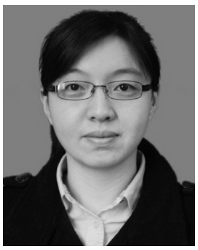
[37] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[38] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[39] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 3, pp. 321–334, Mar. 2005.

HONGFEI LI received the B.S. degree in software outsourcing from Qilu Normal University, China, in 2017. She is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong Normal University. Her research interests include deep learning and image compression.

LILI MENG received the B.E. degree from Shandong University, Jinan, China, in 2005, and the Ph.D. degree from Beijing Jiaotong University, Beijing, China, in 2013, under the supervision of Prof. Y. Zhao. Since 2010, she has been visiting the National Kaohsiung University of Applied Sciences, Taiwan. From 2010 to 2011, she was a Visiting Student with Simon Fraser University, Canada. From 2017 to 2018, she was a Visiting Scholar with Simon Fraser University, Canada. She is currently with the School of Information Science and Engineering, Shandong Normal University, Jinan, China. Her research interests include image/video coding and machine learning.



JIA ZHANG received the B.E. and M.E. degrees from the School of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China, in 2006 and 2009, respectively, and the Ph.D. degree in communication and information systems from Shandong University, in 2013. She is currently with the School of Information Science and Engineering, Shandong Normal University, Jinan, China. Her research interests include MIMO radio techniques, joint resource allocation and optimization in multicell cellular networks, dynamic programming, and interference coordination in heterogeneous cellular networks.

YANYAN TAN received the Ph.D. degree in intelligent information processing from Xidian University, Xi'an, China, in 2013. She is currently a Lecturer with the School of Information Science and Engineering, Shandong Normal University, Jinan, China. Her current research interests include computational intelligence, multi-objective optimization, data analysis, and machine learning.

YUWEI REN received the M.S. degree from Yanshan University, Qinhuangdao, China, and the Ph.D. degree from the Institute of Automation Chinese Academy of Sciences, Beijing, China, in 2007, and 2011, respectively. She is currently with the Department of Information Science and Engineering, Shandong Normal University, China. Her current research interests include SDC and collaborative FTC.

HUAXIANG ZHANG received the Ph.D. degree from Shanghai Jiao Tong University, in 2004. He was an Associate Professor with the Department of Computer Science, Shandong Normal University, from 2004 to 2005. He is currently a Professor with the School of Information Science and Engineering, and also with the Institute of Data Science and Technology, Shandong Normal University, China. He has authored more than 160 journal and conference papers and has been granted nine invention patents. His current research interests include machine learning, pattern recognition, evolutionary computation, cross-media retrieval, and web information processing.