

Adaptive CSI Feedback for Deep Learning-Enabled Image Transmission

Guangyi Zhang, Qiyu Hu, Yunlong Cai, and Guanding Yu

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China
Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking (IPCAN),
Hangzhou 310007, China

E-mail: {zhangguangyi, qiyuhu, ylcai, yuguanding}@zju.edu.cn

Abstract—Recently, deep learning-enabled joint-source channel coding (JSCC) has received increasing attention due to its great success in image transmission. However, most existing JSCC studies only focus on single-input single-output (SISO) channels. In this paper, we first propose a JSCC system for wireless image transmission over multiple-input multiple-output (MIMO) channels. As the complexity of an image determines its reconstruction difficulty, the JSCC achieves quite different reconstruction performances on different images. Moreover, we observe that the images with higher reconstruction qualities are generally more robust to the noise, and can be allocated with less communication resources than the images with lower reconstruction qualities. Based on this observation, we propose an adaptive channel state information (CSI) feedback scheme for precoding, which improves the effectiveness by adjusting the feedback overhead. In particular, we develop a performance evaluator to predict the reconstruction quality of each image, so that the proposed scheme can adaptively decrease the CSI feedback overhead for the transmitted images with high predicted reconstruction qualities in the JSCC system. We perform experiments to demonstrate that the proposed scheme can significantly improve the image transmission performance with much-reduced feedback overhead.

Index Terms—CSI feedback, deep joint source-channel coding, MIMO, wireless image transmission.

I. INTRODUCTION

According to Shannon's separation theorem [1], [2], modern data transmission is divided into a two-step separated encoding process, namely source coding and channel coding. It has then been proven that the separated source and channel coding approaches its optimum theoretically when the blocklength goes to infinity. However, in the finite blocklength scenarios, the joint source-channel coding (JSCC) has been proven to achieve better performance than the separated scheme [3]. Recently, inspired by the success of deep learning (DL), the autoencoder architecture parameterized by the neural networks (NNs) is used to implement the JSCC system, which outperforms the separated scheme [4]–[9]. It employs deep neural networks (DNNs) to map the input source data directly to channel symbols in a joint manner, which is then decoded by another DNNs at the receiver. Specifically, in [4], the authors firstly proposed the deep JSCC (D-JSCC) technique

for wireless image transmission, where the image pixel values are mapped to the complex-valued channel symbols with a well-designed encoder. The authors in [5] incorporated the channel output feedback into the D-JSCC system to improve the reconstruction performance. Moreover, a unified joint source-channel coding semantic communication system for multi-modal data transmission has been proposed in [7].

Multiple-input multiple-output (MIMO) has been widely deployed in practical communication systems for transmitting various sources, such as image, text, and video. It is regarded as a critical technology for current and future wireless systems, since it can provide high spectral efficiency and reduce the interference by fully utilizing the spatial resources [10]. However, these strengths are highly dependent on the available channel state information (CSI) at the base station (BS), and the user equipment is required to feed the CSI back to the BS through feedback links. Besides, the substantial antennas at the BS for massive MIMO lead to a huge dimension of the CSI matrix, which seriously increases the feedback overhead. Based on compressive sensing (CS), several algorithms have been developed to compress the CSI matrix to reduce the overhead [11], [12]. The authors in [12] use the spatial correlations among nearby antennas to compress the CSI in the spatial-frequency domain. Moreover, there have been also many DL-based methods [13]–[16] to compress the CSI matrix with the NNs, such as CsiNet [15] and CLNet [16]. These methods employ the autoencoder structure that comprises with encoder and decoder. Specifically, the encoder is deployed to sense and compress CSI into a low-dimensional codeword vector. Then, the decoder uses the received codeword vector to recover the original CSI matrix.

Existing D-JSCC methods have made significant performance improvement, however, most of them only focus on the single-input single-output (SISO) channel with additive white Gaussian noise (AWGN). Thus, to design a more realistic and practical system for real communication scenarios, it is necessary to consider the MIMO scenario. Therefore, in this paper, we first propose a DL-based JSCC system for image transmission, and extend it into a more practical scenario by taking the MIMO channels into account. Moreover, we consider the precoding design for the physical layer communication in D-JSCC to improve the performance.

To the best of our knowledge, the data-driven D-JSCC model for image transmission generally achieves different

This work was supported by the National Natural Science Foundation of China under Grants 61971376, U22A2004, and 61831004, the Fundamental Research Funds for the Central Universities226-2022-00195, and the Defense Industrial Technology Development Program under Grant JCKY2020210B021.

performance among the samples. It is mainly because the complexity, e.g., complex texture, varies among different images. We notice the images with higher performance are generally with a higher tolerance for the disturbance, i.e., these images are more robust against noise. Intuitively, this phenomenon inspires us to determine the optimal tradeoff between allocated transmission resources and image reconstruction quality for the D-JSCC with MIMO. Hence, if the reconstruction quality can be evaluated in advance, the resource allocation strategy can be determined. To this end, we first develop a lightweight model to predict the performance on each image. It takes the images as input, and outputs the reconstruction quality, e.g., peak signal-to-noise (PSNR). Then, by taking the source image's predicted reconstruction quality into account, we develop an adaptive scheme to adjust the number of feedback bits. In particular, the receiver would determine the compression ratio of the CSI matrix based on the predicted reconstruction quality of the transmitted images, and sends the compressed CSI through the feedback link to the transmitter for precoding. Our simulation results show that the proposed scheme can significantly improve the image transmission performance and effectiveness with much-reduced feedback overhead.

The rest of this paper is structured as follows. Section II introduces the framework of D-JSCC with MIMO. The proposed adaptive CSI feedback scheme based on the predicted performance is presented in Section III. Simulation results are presented in Section IV. Finally, Section V concludes this paper.

II. DEEP JOINT SOURCE CHANNEL CODING WITH MIMO

In this section, we propose the framework of the practical JSCC system with the MIMO channel.

A. System Model

The proposed D-JSCC is an end-to-end communication system developed to incorporate the channel coding and source coding. In particular, the encoding, decoding, and transmission procedures are parameterized by the NNs, and the system is optimized in a back-propagation manner with the data-driven method.

As shown in Fig. 1, the D-JSCC mainly consists of three parts: encoder, communication channel, and decoder. The encoder and decoder are represented by the DNNs. The input image is represented by a vector, $\mathbf{s} \in \mathbb{R}^{N \times 1}$, where N is the length of the vector. Denote the encoding function of the encoder as $\mathcal{F}_\theta: \mathbb{R}^{N \times 1} \rightarrow \mathbb{C}^{K \times 1}$, with θ denoting its trainable parameters. The encoder encodes \mathbf{s} directly into the complex channel symbol vector, which is given by

$$\mathbf{z} = \mathcal{F}_\theta(\mathbf{s}) \in \mathbb{C}^{K \times 1}, \quad (1)$$

where K is the number of transmitted symbols. Then, the symbol vector \mathbf{z} is constrained to satisfy the average power constraint before transmission.

Subsequently, \mathbf{z} is transmitted through the MIMO channel. In particular, we consider a frequency-division duplexing (FDD) system, and the images are transmitted by the base

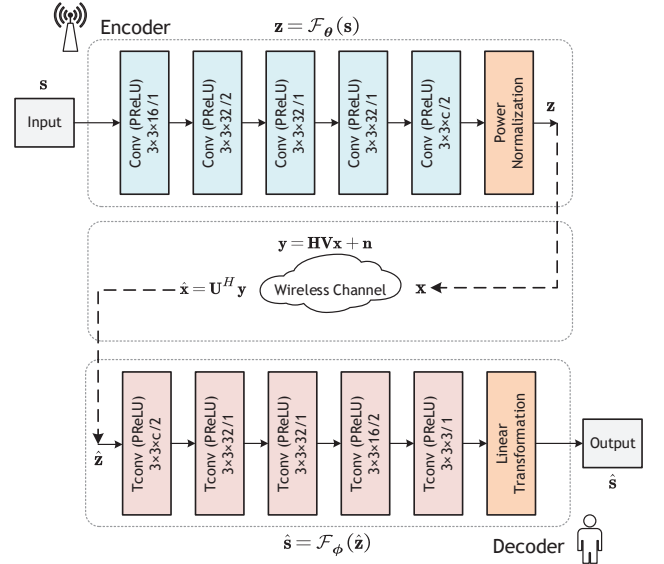


Fig. 1: The architecture of D-JSCC with MIMO channel.

station (BS), which is equipped with N_t transmit antennas. The receiver is equipped with N_r receive antennas.

Let $\mathbf{V} \in \mathbb{C}^{N_r \times d}$ denote the precoder that BS uses to transmit the signal \mathbf{s} , where d denotes the number of data streams. It can be obtained by applying the singular value decomposition (SVD) or zero-forcing (ZF) precoding algorithms with the channel matrix, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, which is obtained via the feedback link from the receiver. Note that the image is encoded by the encoder into a K -dimension vector, i.e., the complex channel symbols. We split them into a number of signals, whose dimension all equal to d . Take one of the split signal, $\mathbf{x} \in \mathbb{C}^{d \times 1}$, as an example, the received signal can be denoted as

$$\mathbf{y} = \mathbf{H}\mathbf{V}\mathbf{x} + \mathbf{n}, \quad (2)$$

where $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ is the AWGN.

At the receiver, we consider linear receive precoding, thus the estimated signal is obtained by

$$\hat{\mathbf{x}} = \mathbf{U}^H \mathbf{y}, \quad (3)$$

where the receive precoder, \mathbf{U} , is obtained by employing the precoding algorithms with \mathbf{H} , and H denotes the conjugate transpose. Correspondingly, after receiving these split signals, we obtain the received symbol vector, $\hat{\mathbf{z}}$, which will be further processed by the decoder. The decoder employs the decoding function, $\mathcal{F}_\phi: \mathbb{C}^{K \times 1} \rightarrow \mathbb{R}^{N \times 1}$, to map $\hat{\mathbf{z}}$ into an estimate of the original signal for reconstruction, which is given by

$$\hat{\mathbf{s}} = \mathcal{F}_\phi(\hat{\mathbf{z}}) \in \mathbb{R}^{N \times 1}, \quad (4)$$

where ϕ denotes the trainable parameters of the decoder.

III. ADAPTIVE CSI FEEDBACK FOR D-JSCC

In this section, we design the adaptive CSI feedback scheme based on the predicted reconstruction quality of the image.

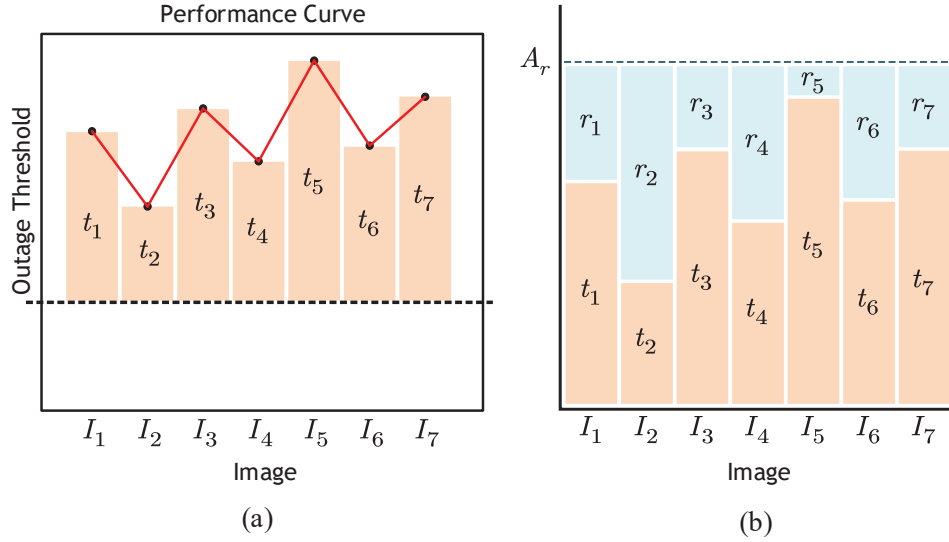


Fig. 2: Visualization of the transmission resources allocation in D-JSCC. (a) The outage threshold and tolerance; (b) Illustration of the water-filling in D-JSCC.

A. Reconstruction Performance Prediction

We have observed that D-JSCC typically achieves different reconstruction qualities on different input images. As for semantic communication, guaranteeing the average reconstruction quality, i.e., average PSNR, is not always suitable, since some of the images would be reconstructed with a rather lower PSNR than the others. In this case, the model performance would not be acceptable for some images.

Concretely, as shown in Fig. 2 (a), we consider the case that all of the transmitted samples are required to be reconstructed to surpass the given minimum threshold, which can be considered as the outage threshold in D-JSCC. The images with higher reconstruction qualities are generally with a higher tolerance for the disturbance, i.e., these images are more robust. This inspires us that there is a trade-off between the allocated transmission resources and the reconstruction qualities based on the predicted reconstruction quality of the image, and to re-determine the transmission resource allocation for the D-JSCC system. More clearly, as shown in Fig. 2(a), the tolerance of the image I_i is denoted as t_i , which can be estimated by the gap between the performance of the image and the given outage threshold. Then, we can allocate the transmission resources, r_i , in a similar way to the water-filling method, which is shown in Fig. 2(b). Specifically, the image with higher tolerance would be allocated with less resources.

Therefore, if we can predict the reconstruction quality of the image in advance, it is possible to determine the resource allocation strategy, and improves the efficiency of the model. To achieve this, we firstly develop a performance evaluator. It takes the transmitted images as input and outputs the predicted PSNR value. Compared with the D-JSCC, it is a light model that consists of a few convolutional layers. We model the prediction task as a regression problem, specifically, the PSNR values achieved by the D-JSCC are set as the

labels to train the evaluator. We exploit the convolutional layer for the architecture of the performance evaluator, which is parameterized by \mathcal{E}_π with π denoting its trainable parameters. Denote the output predicted PSNR and true achieved PSNR for image I_i as $\hat{\gamma}_i$ and γ_i , respectively. The training target for the proposed performance evaluator can be formulated into a regression problem, that is

$$\min_{\pi} \mathcal{L} = \frac{1}{N} \sum_{i=1}^N (\hat{\gamma}_i - \gamma_i)^2 = \frac{1}{N} \sum_{i=1}^N (\mathcal{E}_\pi(I_i) - \gamma_i)^2, \quad (5)$$

where N denotes the size of the training batch.

B. Source-Related Adaptive CSI Feedback

We have designed the D-JSCC with MIMO in Section II, and we aim at adaptively adjusting the overhead of MIMO channel feedback for D-JSCC based on the predicted reconstruction quality. We assume that perfect CSI has been acquired through pilot-based training. As shown in Fig. 3, the transmitter firstly decides on the compression ratio based on the predicted PSNR value of the image, and sends the compression ratio to the receiver. Then, the estimated real channel, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, will be compressed for feedback based on the compression ratio. Then, the transmitter needs to recover the CSI matrix with the compressed representation to obtain the recovered CSI matrix, $\hat{\mathbf{H}}$. The difference between the recovered CSI matrix, $\hat{\mathbf{H}}$, and real CSI matrix, \mathbf{H} is measured by the normalized mean squared error (NMSE) [15], which can be computed as

$$\text{NMSE} = \mathbb{E} \left\{ \frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_2^2}{\|\mathbf{H}\|_2^2} \right\}. \quad (6)$$

In particular, the higher compression ratio will lead to the worse reconstruction quality, i.e., a higher NMSE. In traditional MIMO communication systems, the accuracy of the CSI

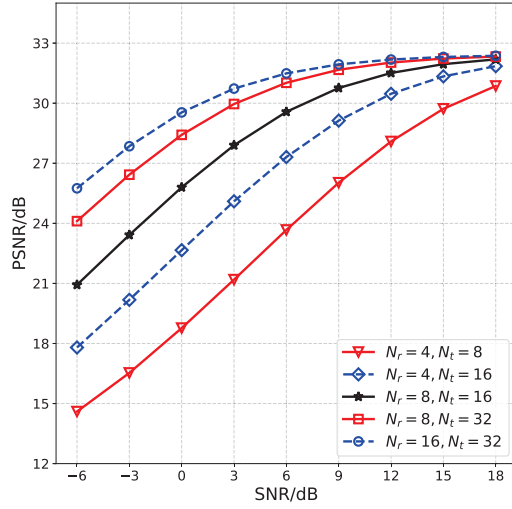


Fig. 4: The performance of proposed model with different numbers of antennas versus SNR.

and transmit array response vectors, respectively. Consider a uniform linear array, the response vector is given by

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}} \left[1, e^{-j2\pi \frac{d}{\lambda} \sin(\phi)}, \dots, e^{-j2\pi \frac{d}{\lambda} (N-1) \sin(\phi)} \right]^T, \quad (10)$$

where N and ϕ are the numbers of antenna elements and azimuth angle, respectively. Moreover, d and λ represent the adjacent distance antennas and carrier wavelength, respectively. In the simulation, we choose $N_{cl} = 2$ clusters and $N_{ray} = 4$ rays. To quantize the CSI matrix, we employ the non-uniform quantization method, Lloyd-Max algorithm, which quantizes the parameter of the CSI matrix with a generated table. In each transmission, we randomly generate a channel matrix for each image, where we consider a block-fading scenario.

To evaluate the performance of the D-JSCC, the PSNR is adopted. It measures the ratio between the maximum possible power and the noise, which can be calculated by

$$\text{PSNR} = 10 \log_{10} \frac{\text{MAX}^2}{\text{MSE}} (\text{dB}), \quad (11)$$

where $\text{MSE} = d(\mathbf{s}, \hat{\mathbf{s}})$ represents the mean square-error (MSE) between the source image, \mathbf{s} , and the reconstructed image, $\hat{\mathbf{s}}$. Moreover, MAX is the maximum possible value of the pixels, e.g., MAX equals 255 for the images of RGB format.

Fig. 4 presents the performance of the investigated schemes equipped with different numbers of antennas versus the SNR, and we assume that the transmitter obtains the perfect CSI. We train the proposed D-JSCC model with SNR = 6 dB and test it in SNR from -6 dB to 18 dB. It is readily seen that the PSNR achieved by D-JSCC increases with SNR. The system equipped with more antennas generally outperforms the system with fewer antennas. It is mainly because the diversity gain increases with the number of antennas. It demonstrates

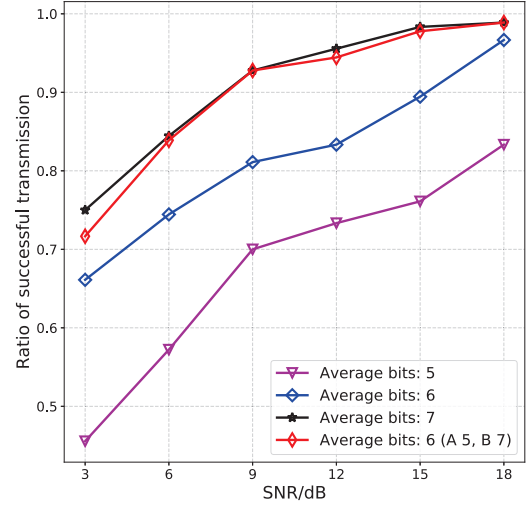


Fig. 5: The performance of the proposed adaptive scheme with different numbers of feedback bits.

that the MIMO system is still effective in improving the performance of DL-based D-JSCC.

Fig. 5 presents the ratio of successful transmission of the proposed adaptive scheme with different numbers of feedback bits. In particular, the ratio is obtained by dividing the number of images that have higher reconstruction quality (PSNR) than the threshold by the total number of testing images. Moreover, the number of feedback bits refers to the number of the average quantization bits for each element in the channel matrix. As for the red curve, we divide the transmitted images into two groups, A and B, according to their predicted PSNR values. The numbers of images in the group with higher predicted PSNR values, A, and the group with lower predicted PSNR values, B, are the same. To make the number of average feedback bits for A and B equal to 6, we set the number of feedback bits for A as 5, and that for B as 7, thus the average number is 6. As for other curves, the numbers of feedback bits for all the images are both fixed at the same value, e.g., 5 for the blue curve. From the figure, we can see that allocating the same feedback bits for each image is not optimal, since the reconstruction qualities of the images are different. In comparison, our proposed adaptive CSI feedback scheme significantly outperforms the average strategy by adaptively adjusting the number of feedback bits.

Fig. 6 shows the ratio of successful transmission of the proposed adaptive scheme versus the outage threshold. In this figure, all the schemes are required to meet the given outage threshold. From the figure, we can see that the performance of these schemes decrease with the threshold. Moreover, our proposed scheme outperforms the scheme with the same feedback bits, as shown by the blue line. It demonstrates the superiority of the proposed adaptive schemes in different threshold requirements.

In Fig. 7, we compare the number of required feedback bits when keeping the same ratio of successful transmission.

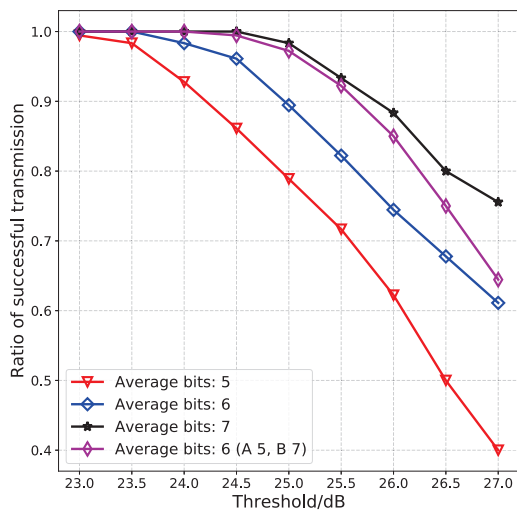


Fig. 6: The performance of the proposed adaptive scheme versus the threshold.

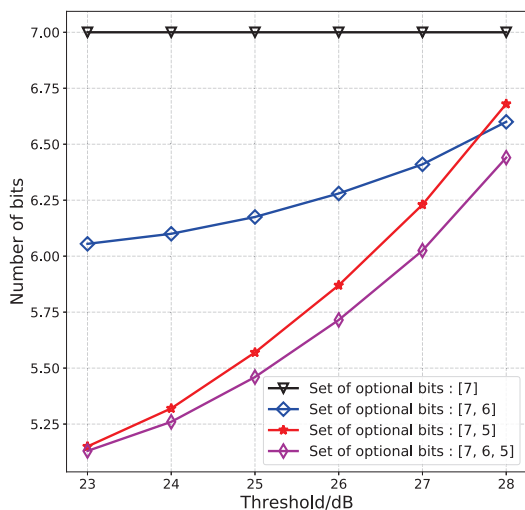


Fig. 7: The number of required bits versus the threshold.

Specifically, all the schemes are required to achieve the same performance as that 7 bits feedback is used for all the test images. We counted the number of required bits when the set of optional bits are given, i.e., all the images need to be transmitted with the number of feedback bits selected from the given set. For example, as for [7, 6], the images should be allocated with 7 or 6 feedback bits for transmission. From the figure, it is readily seen that our proposed adaptive scheme can reduce the feedback overhead while keeping the same ratio of successful transmission. Moreover, with more optional numbers, the feedback overhead can be further reduced.

V. CONCLUSION

In this paper, we have proposed a D-JSCC communication system for image transmission over MIMO channels. Then,

to improve the performance, we have designed the precoding for the D-JSCC. The encoder and decoder are jointly trained in an end-to-end manner. Moreover, we have redefined the outage in the D-JSCC and propose an adaptive CSI feedback scheme. It is able to adjust the compression rate based on the predicted performance of the image. Simulation results show that the proposed scheme can significantly improve the image transmission performance and effectiveness with significantly reduced feedback overhead.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.
- [2] C. Q., T. M. Cover, and J. A. Thomas, *Elements of information theory*. Elements of information theory, 2006.
- [3] F. Zhai, Y. Eisenberg, and A. K. Katsaggelos, "Joint source-channel coding for video communications," *Handbook of Image and Video Processing (Second Edition)*, pp. 1065–1082, 2005.
- [4] E. Boursoulatz, D. Burth Kurka, and D. Gunduz, "Deep joint source-channel coding for wireless image transmission," *IEEE Trans. Cognit. Comm. Netw.*, vol. 5, no. 3, pp. 567–579, Sep. 2019.
- [5] D. B. Kurka and D. Gunduz, "DeepJSCC-f: Deep joint source-channel coding of images with feedback," *IEEE J. Select. Areas Inf. Theory*, vol. 1, no. 1, pp. 178–193, May 2020.
- [6] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, Apr. 2021.
- [7] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, X. Tao, and G. Y. Li, "A unified multi-task semantic communication system for multimodal data," *arXiv preprint arXiv:2209.07689*, 2022.
- [8] M. Yang, C. Bian, and H.-S. Kim, "OFDM-guided deep joint source channel coding for wireless multipath fading channels," *IEEE Trans. Cognit. Comm. Netw.*, vol. 8, no. 2, pp. 584–599, Jun. 2022.
- [9] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications with masked VQ-VAE enabled codebook," *arXiv preprint arXiv:2206.04011*, 2022.
- [10] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Overview of deep learning-based CSI feedback in massive MIMO systems," *arXiv preprint arXiv:2206.14383*, 2022.
- [11] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5117–5144, Apr. 2016.
- [12] P.-H. Kuo, H. T. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 492–497.
- [13] Mashhadi, M. Boloursaz, Yang, Qianqian, and G. Deniz, "Distributed deep convolutional compression for massive mimo csi feedback," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2621–2633, Apr. 2021.
- [14] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Convolutional neural network-based multiple-rate compressive sensing for massive mimo csi feedback: Design, simulation, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2827–2840, Oct. 2020.
- [15] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.
- [16] S. Ji and M. Li, "CLNet: Complex input lightweight neural network designed for massive MIMO CSI feedback," *IEEE Commun. Lett.*, vol. 10, no. 10, pp. 2318–2322, Oct. 2021.
- [17] S. S. Ioushua and Y. C. Eldar, "A family of hybrid analog-digital beamforming methods for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3243–3257, Jun. 2019.