Abstract—Shannon’s separation theorem lays the foundation for traditional image compression and transmission schemes, which consist of JPEG-type image compression methods and the usual channel coding schemes such as Turbo and LDPC codes. One of the advantages of the separate design is that each of the two components: channel coding and source coding can be handled independently without considering the other, which is the base of decades-long technologies.

I. INTRODUCTION

Shannon’s separation theorem lays the foundation for traditional image compression and transmission schemes, which consist of JPEG-type image compression methods and the usual channel coding schemes such as Turbo and LDPC codes. One of the advantages of the separate design is that each of the two components: channel coding and source coding can be handled independently without considering the other, which is the base of decades-long technologies. The potential shortcomings of separate processing are the following: Firstly, although the separation theorem establishes that separated processing is optimal in the asymptotic sense, it may not be optimal considering the latency and complexity at the receiver in the finite blocklength regime [1] in practical real-time applications; Secondly, in the specific circumstances such as an image or video transmission, deep learning methods have the ability to extract the feature vector of data, which is more compact and has a higher resistance to the noise due to the fact that a sufficiently trained decoder has immunity against the weak disturbance on the feature vector. Thirdly, the traditional scheme is vulnerable to the “cliff effect” and depends heavily on the channel condition [2] [3]. These factors motivate a joint optimal design of the source and channel coders in image transmission systems. Especially, these days have witnessed the emergence of image transmission schemes [4]–[6] based on deep neural networks (DNN), which have already achieved impressive results for the image compression problem [7] [8].

Generally, the performance of DNN-based source coding is better or at least comparable with the traditional digital schemes with separated channel coding and source coding when the channel state is relatively steady, i.e., the channel is AWGN type or under slow fading [4] [5]. It even behaves well and shows its adaptability to minor changes of channel states when no feedback information is provided. However, the performance still needs improvement compared with the traditional separate schemes when the latter kind is equipped with channel state feedback in the situation where the noise or the channel is time-varying [6]. The reason is that deep source coding can not adapt itself to the channel variation since the optimization is always performed during a training process on parameters of the neural networks before the adoption of the system, where the channel noise is modeled as random variables (or random vectors) from some distributions. To apply DNN-based source coding system in real applications with varying noise or channel states, it is necessary to introduce some elements to adaptively deal with the varying channel, and the new elements should be able to update themselves from the feedback information on the channel state or the noise timely. As the extraction of the feature vectors and recovery of the images are finished by neural networks, which are usually nonlinear functions and complicated to analyze, it is advantageous to regard the feature vectors as the source and consider the optimization by deploying a linear encoder&decoder pair on the input and output feature vectors of the channel [9]. In this way, the optimizations of the neural networks and the linear pre-processing and post-processing are separated, the first of which focuses on the feature extraction of images, and the second focuses on dealing with the channel variation.

In the beginning, the problem of reliable transmission of vector signals from the source to users rises in the transmission of color TV signals, measurements in multi-point telemetering systems and the optimization have been solved in one dimension [10]. A lot of research has been done on the design of the optimal transmission system to cope with different channel conditions [11] [12]. Among them, optimal linear coding for vector channels is first investigated in [13], where there is only information on the covariance matrix of the source vectors and the noise. In this general case, the optimal encoder and decoder depend on the eigenvalues of the covariance matrices of the source and the noise. Typically, when the source is Gaussian distributed, the optimal linear coding schemes were established and the theoretical performance bounds have been given [14]. As image transmission is considered in our setting, we will propose a structure of neural network, which consists of a classifier cascaded with a parallel architecture of encoders and decoders, which can sufficiently utilize the advantages of the existing optimal transmission strategy of the Gaussian source to decrease the MSE between the input feature vectors and their corrupted output ones.

In this work, we integrate NN design and linear encoder and decoder pair to build a transmission system of image data sets. The scheme first divides the image set into several groups with their own labels, each group is equipped with a sub-encoder and sub-decoder pair at the two ends. The principle of the design is to process separately two different kinds of information in the image: the discrete label information between groups and the continuous shape change in each
group. A linear encoder & decoder pair is incorporated as part of the whole encoder and decoder system. When the channel is varying, the linear encoder and decoder update themselves from the estimation of the channel noise, and the other parts of the system remain invariant. The estimation of the noise is obtained by a multimodal learning algorithm, which integrates the noise information from previous several time slots. The contributions of the work are listed as follows.

II. The Transmission Model

In this section, we briefly introduce the deep source coding image wireless transmission scheme. Each image is expressed as a real vector \( x \) with length \( n \). The transmitter maps the input image \( x_t \) to a complex-valued vector which is equivalent to \( y_t \in \mathbb{C}^k \) by an neural network \( f_{\theta} \) at time index \( t \), where \( n \) is denoted as source bandwidth and \( k \) is the channel bandwidth.

The channel is modeled in the form of real vectors as

\[
\hat{y}_t = y_t + z_t,
\]

where \( z_t \sim \mathcal{N}(0, \Sigma_t) \). The random input vectors \( y_t \) should satisfy an average power constraint

\[
\text{tr}\{\text{Cov}(y_t)\} \leq P.
\]

After receiving \( \hat{y}_t \), the receiver maps it to a real vector \( \hat{x}_t \) as the approximation of the original image. Note that the covariance matrix of the noise \( z_t \) varies with respect to the time slot \( t \). The time-varying noisy channel has been reported in the literature, such as [3] [15]. Conventionally, the neural networks at the transmitter and the receiver are jointly trained as the approximation of the original image. Note that the classifiers of the neural networks. The classified image will be sent to the corresponding sub-encoder, and then is compressed as a complex vector \( \hat{y}_t \), whose equivalent real vector is expressed as \( f(x_t) \). The third part of the encoder is a linear filter \( A_t \), which will be responsible for power control and dealing with the varying noise with output \( y_t = A_t(f(x_t)) \).

The Decoder: The structure of the decoder consists of three parts: a linear decoder \( B_t \), a correcting code’s decoder \( \hat{h}_{\phi_t} \) which only deals with the corrupted codewords of the labels \( \hat{e}_t \), i.e. \( \hat{u}_t = h_{\phi_t}(\hat{e}_t) \) and the parallel sub-decoders \( g_{\phi_1}, \ldots, g_{\phi_K} \) that deal with the corrupted feature vectors of the images.

The neural networks of the sub-encoders \( f_{\phi_i}, i = 1, \ldots, K \) and the sub-decoders \( g_{\phi_i}, i = 1, \ldots, K \) include several layers: two normalization layers at the two ends and a full convolution layer in the middle. Note that in the system, only \( A_t \) and \( B_t \) update with respect to time index \( t \). The other parts remain invariant once the training process is over.

B. On the Optimal Linear Encoder \( A_t \) and Decoder \( B_t \)

As the core parts to adaptively cope with the varying noise, the analytical solutions of \( A_t \) and \( B_t \) in terms of the noise statistics are derived in this section. As we have explained in Section arch, \( A_t \) and \( B_t \) depend on the previous channel feedback \( \Sigma_{t-1} \) in the pilot signals, which is an estimation of the matrix \( \Sigma_{t-1}, l = 0, \ldots, L - 1 \), where \( L \) is the length of our considered time slots to estimate the covariance matrix at time slot \( t \). For convenience, we omit the time index \( t \) in the following analysis. Especially, the distribution of the channel noise is expressed as \( z \sim \mathcal{N}(0, \Sigma) \).

First, the output of the neural network at the transmitter is denoted as \( f(x) \), which is modeled as a random vector subjected to Gaussian distribution \( \mathcal{N}(\mu_1, \Sigma_1) \). It is pre-processed by the linear encoder \( A_t \) in terms of the matrix \( A \), and then fed into the channel to get the output \( \hat{y} \):

\[
y = A \cdot f(x),
\]

\[
\hat{y} = y + z,
\]

where the random vectors \( y \) should satisfy an average power constraint (power1). From the assumption on \( f(x) \) of Gaussian distribution with \( \mathcal{N}(\mu_1, \Sigma_1) \), the power constraint can be formulated as the trace of the variance of \( y \), i.e.,

\[
\text{tr}\{\text{Cov}(y)\} = \text{tr}\{A\Sigma_1A^T\} \leq P.
\]

Then, the output vector \( \hat{y} \) is post-processed by the linear decoder \( B_t \) to get an approximation of \( f(x) \). In general, the
optimal encoder matrix $A$ and the linear estimator parameters $(B, b)$ depend on each other. To solve the problem, we first assume $A$ is determined and find $B_t$ as a function of $A$. The target of $B_t$ is to recover the vector $f(x)$, hence $B_t$ can be designed as a pair $(B, b)$ as the best linear estimator of $f(x)$:

$$
\hat{y} = B \cdot \tilde{y} + b,
$$

$$
(B^*, b^*) = \arg\min_{B, b} \mathbb{E} \left[ \|\hat{y} - f(x)\|^2 \right]. \tag{6}
$$

Finally, we have the following theorem for the above optimization problem.

**Theorem 1.** The optimal linear estimator for (mse) with given $A$ is

$$
B^* = (\Sigma_1 \cdot A^T)(A \Sigma_1 A^T + \Sigma)^{-1}, \tag{7}
$$

$$
b^* = \mu - (\Sigma_1 \cdot A^T)(A \Sigma_1 A^T + \Sigma)^{-1} A \cdot \mu_1. \tag{8}
$$

Theorem 1 provides a closed form of the optimal linear decoder in terms of several quantities: the mean vector and the covariance matrix of the source (or the feature vectors) in each group, the covariance matrix of the noise and linear encoder matrix $A$. In practice, the covariance matrix of channel noise $\Sigma$ cannot be perfectly known, and is substituted by its estimation $\hat{\Sigma}$ from the pilot signals attached with the previous feature vectors by a multimodal learning algorithm, which will be discussed in the next section.

As we have determined the optimal linear decoder, we will go further to solve the same optimization problem with the given $B^*$ and $b^*$ to get the optimal linear encoder $A$. First, the solutions of Theorem Theorem1 is substituted into the MSE,

$$
\mathbb{E} \left[ \|\hat{y} - f(x)\|^2 \right] = \text{tr}(\mathbb{E}((\hat{y} - f(x))(\hat{y} - f(x)^T))) \tag{9}
$$

$$
= \text{tr}(\Sigma_1) - \text{tr}((\Sigma_1 A^T)(A \Sigma_1 A^T + \Sigma)^{-1} A \Sigma_1),
$$

From the above equation and the power constraint $\|\text{Cov}(y)\|^2 \leq P$, we further have

$$
\max_A \text{tr}\{(\Sigma_1 A^T)(A \Sigma_1 A^T + \Sigma)^{-1} A \Sigma_1\} \tag{10}
$$

$$
\text{s.t. } \text{tr}\{A \Sigma_1 A^T\} \leq P.
$$

In the following, we will find the maximization of (objective2) to find the solution of optimal $A$. Since $\Sigma$ and $\Sigma_1$ are the covariance matrices of the source and the noise, they are positive definite. There are orthogonal matrices $Q_1$ and $Q_2$ that the following diagonalizations hold:

$$
\Sigma = Q_1 \text{diag}(\sigma_1^2, \cdots, \sigma_d^2)Q_1^T, \tag{11}
$$

$$
\Sigma_1 = Q_2 \text{diag}(\lambda_1^2, \cdots, \lambda_d^2)Q_2^T, \tag{12}
$$

where $\lambda_1 \geq \cdots \geq \lambda_d$.

**Theorem 2.** Denote $P_j = \sum_{i=1}^j \sigma_i^2, j \in \{1, \cdots, d-1\}$ and $C_j = \frac{1}{P + \sum_{i=1}^{d-1} \sigma_i^2}, j \in \{2, \cdots, d\}$. The optimal solution of $A$ is

$$
A^* = Q_1 \text{diag}(\sqrt{k_1}, \cdots, \sqrt{k_d})Q_2^T, \tag{13}
$$

where the values of $k_1, \cdots, k_d$ depend on the value of $P$ and are listed as the following cases:

1) when $P \leq P_1$:

$$
k_1 = \frac{P}{2}, \quad i = 2, \cdots, d. \tag{14}
$$

2) When $P_{j-1} < P \leq P_j$ where $j \in \{2, \cdots, d-1\}$:

$$
k_i = \frac{\sigma_i^2}{\lambda_i^2}, \quad i = 1, \cdots, j. \tag{15}
$$

3) When $P > P_{d-1}$:

$$
k_i = \frac{\sigma_i^2}{\lambda_i^2}, \quad i = 1, \cdots, d. \tag{16}
$$

Theorem 2 is a typical water-filling solution on each sub-channel of the vector Gaussian channel to minimize the MSE between the input and the output vectors with fixed noise statistics under a given power constraint. The power allocation on each sub-channel depends on the ratio of corresponding eigenvalues of the covariance matrices of the source and the noise. Similar to the last theorem about the optimal linear decoder, the closed form of the linear encoder also depends on the estimation of the covariance matrix of the noise.

**REFERENCES**

[1] Y. Polyanskiy, H. V. Poor and S. Verdu, “Channel Coding Rate in the Finite Blocklength Regime,” *IEEE Trans. Inf. Theory*, Vol. 56, No. 5, pp. 2307-2358, May 2015.

[2] N. Thomos, N. V. Boulgouris and M. G. Strintzis, “Optimized transmission of JPEG2000 streams over wireless channels,” *IEEE Trans. Image process.*, Vol. 15, No. 1, pp. 54-67, Jan. 2006.

[3] I. Kozintsev and K. Ramchandran, “Robust Image Transmission Over Energie-constrained Time-Varying Channels Using Multiresolution Joint Source-Channel Coding,” *IEEE Trans. Signal Process.*, Vol. 46, No. 4, pp. 1012-1026, Apr. 1998.

[4] E. Bourjoulatze, D. B., Kurka, D. Gündüz “Deep Joint Source-Channel Coding for Wireless Image Transmission,” *IEEE Trans. Cogn. Commun. Neww.,* Vol. 5, No. 3, pp. 567-579, Sep 2019.

[5] D. B. Kurka and D. Gündüz, “Successive Refinement of Images with Deep Joint-Source-Channel Coding,” in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2019, pp. 1-5.

[6] D. B. Kurka and D. Gündüz, “DeepJSVC-f: Deep Joint-Source-Channel Coding of Images With Feedback,” *IEEE Journal On Selected Areas In Information Theory*, Vol. 1, No. 1, pp. 178-193, May 2020.

[7] J. Ballé, V. Laparra, and E. P. Simoncelli, “End-to-end Optimized Image Compression, in “Proc. Int. Conf. Learning Represent.(ICLR), Apr. 2017, pp. 1-27.

[8] J. Ballé, D. Minnen, S. Singh, S. Hwang and N. Johnston, “Variational Image Compression with A Scale Hyperprior,” in Proc. Int. Conf. Learning Represent.(ICLR), 2018.

[9] S.-L. Huang, X. Xu, L. Zheng, and G. W. Wornell, “An Information Theoretic Interpretation to Deep Neural Networks,” in 2019 IEEE international symposium on information theory (ISIT), IEEE, 2019, pp.1984–1988.

[10] C. E. Shannon, “A Mathematical Theory of Communication,”*The Bell System Technical Journal*, Vol. 27, Issue: 3, Jul. 1948.

[11] E. Akyl, K. Rose, and T. Ramstad, Optimal Mappings for Joint Source Channel,” in “2019 IEEE Information Theory Workshop on Information Theory (ITW 2010)”, IEEE Trans. Inform. Theory (ITW 2010), Cairo, pp 1-5, 2010.

[12] M. Joham, W. Utschick, and J. A. Nossek, “Linear Transmit Processing in MIMO Communications Systems,”*IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, 2005.

[13] K. H. Lee, and D. P. Petersen,”Optimal Linear Coding for Vector Channel,” *IEEE Trans Communications.*, Vol. 24, No. 12, pp. 1283–1290, 1976.

[14] T. Basar, B. Sankur, and H. Abut, “Performance Bounds and Optimal Linear Coding for Discrete-Time Multichannel Communication Systems (corresp.),” *IEEE Trans. Inform. Theory*, vol. 26, no. 2, pp. 212–217, 1980.
[15] Y. Hou, R. Liu, BDai and L. Zhao “Joint Channel Estimation and LDPC Decoding Over Time-Varying Impulsive Noise Channels,” *IEEE Trans Communications.*, Vol. 66, No. 6, pp. 2376 - 2383, 2018.

[16] https://imerit.net/blog/22-free-image-datasets-for-computer-vision-all-pbm/

[17] T. Goblick Jr, “Theoretical Limitations on the Transmission of Data from Analog Sources,” *IEEE Trans. Inform. Theory*, Vol. 11, No. 4, pp. 558-567, 1965.

[18] Z. Xuan and K. Narayanan, “Analog joint source-channel coding for Gaussian sources over AWGN channels with deep learning,” in *International Conference on Signal Processing and Communications*, 2020, pp. 1–5.