Deep Joint Source Channel Coding for Wireless Image Transmission with OFDM

Mingyu Yang, Chenghong Bian, and Hun-Seok Kim
University of Michigan, Ann Arbor, USA

Abstract—We present a deep learning based joint source channel coding (JSCC) scheme for wireless image transmission over multipath fading channels with non-linear signal clipping. The proposed encoder and decoder use convolutional neural networks (CNN) and directly map the source images to complex-valued baseband samples for orthogonal frequency division multiplexing (OFDM) transmission. The proposed model-driven machine learning approach eliminates the need for separate source and channel coding while integrating an OFDM datapath to cope with multipath fading channels. The end-to-end JSCC communication system combines trainable CNN layers with non-trainable but differentiable layers representing the multipath channel model and OFDM signal processing blocks. Our results show that injecting domain expert knowledge by incorporating OFDM baseband processing blocks into the machine learning framework significantly enhances the overall performance compared to an unstructured CNN. Our method outperforms conventional schemes that employ state-of-the-art but separate source and channel coding such as BPG and LDPC with OFDM. Moreover, our method is shown to be robust against non-linear signal clipping in OFDM for various channel conditions that do not match the model parameter used during the training.

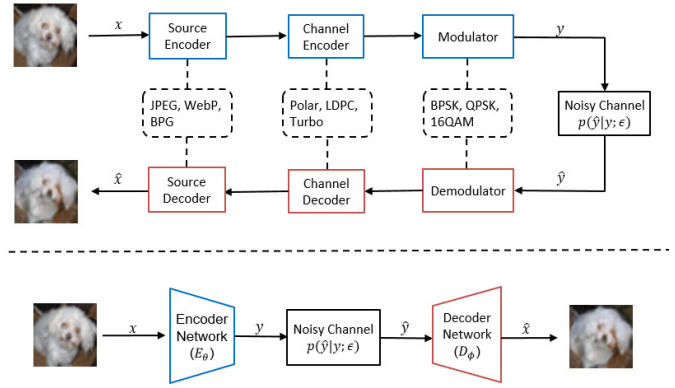*Index Terms*—Joint source channel coding, deep neural networks, OFDM, model-driven machine learning

Fig. 1. Block diagrams of wireless image transmission schemes. Top: Traditional separate source and channel coding scheme. Bottom: Deep learning based joint source channel coding (JSCC) scheme.

I. Introduction

Shannon's separation theorem [1] states that, as the size of the transmitted message goes to infinity for memory-less channels, it is optimal to split the communication task into (i) removing the redundant information of the source as much as possible and (ii) re-introducing redundancy for message reconstruction in the presence of channel noise. Based on this famous theorem, most modern systems for wireless image transmission first compress the image with a source coding algorithm (e.g., JPEG, WebP, BPG) and then encode the bit stream with a source-independent channel code (e.g., LDPC, Polar, Turbo, etc.) as shown in Figure 1. This separate source and channel coding is attractive for practical communication systems because of the convenient modularity. However, in practice, we often transmit a finite number of bits and that breaks the assumption of Shannon's separation theorem. For a finite length image source, reconstruction quality becomes a joint function of the source coding distortion and channel coding error. Thus it is difficult to design an optimal scheme that balances these two sources of the quality degradation. Besides, the optimal maximum-likelihood detection for the joint source and channel coded information is NP-hard in general. Considering the complexity and power constraints of practical systems, sub-optimal separate source and channel coding is regarded as a standard scheme.

In recent years, deep learning has been successfully applied to a wide range of areas such as computer vision [2] [3] and natural language processing [4] [5], and it achieved significant performance improvement over analytical (non-data-driven) algorithms. Its ability to extract complex features from images has led to applications of deep learning to joint source channel coding (JSCC) over simple channel models such as additive white Gaussian noise (AWGN) [6] [7] [8] and binary symmetric channel (BSC) [9] [10]. These methods demonstrate the feasibility of practical machine learning based JSCC schemes that yield better performance than prior separate coding schemes. However, there is no JSCC work that extends to realistic and challenging multipath fading channels.

In this paper, we propose a deep learning based JSCC scheme for wireless image transmission under multipath fading channels. Inspired by prior works [6] [7] [8], we employ a pair of encoder and decoder that directly encodes images as complex valued channel input samples and recovers them from noisy channel outputs. Moreover, adopting the idea of model-driven deep learning approach [11], we extend the JSCC framework to combine trainable CNN layers as well as non-trainable but differentiable layers representing a multipath channel model and OFDM baseband processing blocks. As the authors of other prior work reported in [12] [13], incorporating the domain expert knowledge improves the training
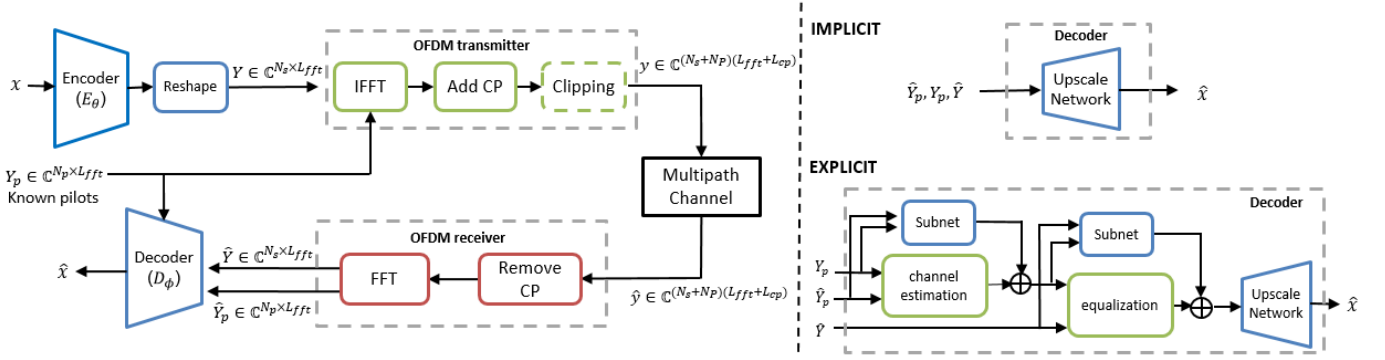
Fig. 2. Left: Flow diagram for the proposed JSCC framework with OFDM extension. Right: Two decoder structures considered in this paper. IMPLICIT: directly concatenate everything and feed them to the decoder network. EXPLICIT: introduce explicit channel estimation and equalization methods with additional residual connections.

speed and enhances the performance over a conventional auto-encoder approach without an explicit domain-knowledge-driven structure. We show that as we impose additional domain-knowledge-driven structure such as channel estimation and equalization into the decoder, its performance further improves. In the proposed scheme, OFDM processing blocks (FFT, cyclic prefix insertion, etc.) and multpath channel model are both implemented as differentiable layers so that the end-to-end framework can be trained with multipath fading as well as non-linear signal clipping at the transmitter. Our results show that the proposed JSCC method outperforms the baseline separate source and channel scheme with a state-of-the-art image compression method and error correction code applied to OFDM. Furthermore, we show that our proposed method is robust for a wide range of SNR and channel conditions that the neural network has not seen during the training process.

II. RELATED WORK

A. Deep Joint Source Channel Coding

Deep JSCC was first studied in [6] where the authors incorporate both source and channel coding into an auto-encoder structure, and jointly learn source compression and error correction coding through the back-propagation optimization. The proposed deep JSCC algorithm outperforms conventional separate schemes under AWGN and Rayleigh flat fading channel. It can be further extended with feedback [8] or progressive transmission [7]. Similarly, deep JSCC on BSC and binary erasure channels have been studied in [9] [10]. Due to their discrete nature, the authors perform the end-to-end optimization in a variational manner and optimize a lower bound of mutual information between the source images and noisy channel outputs. In additional to images, JSCC has been studied for Gaussian source [14] and text source [15].

B. Model-driven Machine Learning for Communications

Recent years, deep learning methods have been widely applied to wireless communications [16] [17] [18]. However, as discussed in [11], these data-driven methods simply treat standard neural networks as a black box, which require a large amount of data and long training time. To address

this issue, model-driven deep learning frameworks combine standard neural networks with analytical/mathematical models and signal processing blocks guided by the expert knowledge to shorten the training time and also to improve the performance. Combining machine learning with expert domain knowledge has shown great success in many applications such as radio transformer network (RTN) [12], symbol modulation and demodulation for OFDM [13], channel estimation [19], and peak-to-average power Ratio (PAPR) reduction [20].

III. PROPOSED METHOD

A. Deep JSCC in multipath fading channels

The general framework for deep learning based JSCC is shown in Figure 1. In our approach, a pair of encoder E_θ and decoder D_ϕ directly map the source input x to complex-valued (modulated) baseband samples y and estimate the source signal $\hat{x}$ from noisy multipath fading channel outputs $\hat{y}$. Here, θ and ϕ represent the parameters for the encoder and decoder neural networks respectively. The channel can be represented with the conditional probability $p(\hat{y}|y; \epsilon)$ where ϵ denotes the parameters that describe the channel statistics. The channel is defined in the continuous time domain assuming interpolation and Nyquist sampling for the channel inputs and outputs.

We parameterize the multipath fading channel using a channel transfer function;

$$\hat{y} = h(y; \sigma_0^2, \ldots, \sigma_{L-1}^2, \sigma^2) = h * y + w \qquad (1)$$

where $*$ denotes the convolution operation, $h \in \mathbb{C}^L$ denotes the sample space channel impulse response, and L is the number of multipaths. $w \sim \mathcal{CN}(0, \sigma^2 I_k)$ represents the additive Gaussian noise. Each path experiences independent Rayleigh fading satisfying $h_l \sim \mathcal{CN}(0, \sigma_l^2)$ for $l = 0, 1, \ldots, L - 1$. The power for each path follows an exponential decay profile $\sigma_l^2 = \alpha_l e^{-\frac{l}{\gamma}}$ where α_l is a normalization coefficient to satisfy $\sum_{l=0}^{L-1} \sigma_l^2 = 1$. γ is the time decay (or delay spread) constant. Note that the channel transfer function is fully differentiable given realizations of random channel parameters. It means that the gradients from the decoder can propagate back to the encoder through a particular realization of the multipath

fading channel when the end-to-end system is trained with gradient descent methods. For the gradient descent training of the encoder and decoder pair, we use random realizations of the channel model described in this section.

B. JSCC with OFDM Extension and Signal Clipping

We extend the JSCC framework to an OFDM-based system to efficiently mitigate the multipath fading channel with simple single-tap frequency domain equalization. We assume that each source image x is transmitted in a single OFDM packet that contains N_p pilot symbols and N_s information symbols. For channel estimation, we adopt block-type pilots by sending known symbols on all subcarriers. For simplicity, we assume that the transmitter and receiver are synchronized in time (within the cyclic-prefix length L_{cp}) without any carrier frequency offset.

The OFDM extension to the JSCC framework is shown in Figure 2. With OFDM, the encoder output becomes the frequency domain symbols $Y \in \mathbb{C}^{N_s \times L_{fft}}$ fed into the OFDM transmitter. The pilot symbols $Y_p \in \mathbb{C}^{N_p \times L_{fft}}$ are known to both the transmitter and receiver. Here, L_{fft} denotes the number of subcarriers in an OFDM symbol. We apply inverse discrete Fourier transform (IDFT) to each OFDM symbol and append the cyclic-prefix (CP).

One notable drawback of OFDM is the high PAPR that causes excessive power consumption at the power amplifier [21]. A number of solutions have been proposed to reduce the PAPR. Among them, signal clipping is one of the simplest and extensively studied methods [22] [23]. Suppose we have the original TX sample $y_{preclip}[n] = A[n]e^{j\phi[n]}$ where $A[n]$ is the amplitude and $\phi[n]$ is the phase. The clipping operation can be expressed as:

$$y[n] = \begin{cases} A[n]e^{j\phi[n]} & \text{when } A[n] \le \rho\sqrt{P_s} \\ \rho\sqrt{P_s}e^{j\phi[n]} & \text{otherwise} \end{cases} \quad (2)$$

where ρ denotes the clipping ratio (CR) and P_s is the input signal power. In the proposed JSCC, this clipping operation is considered as an additional non-linear activation function as shown in Fig. 3 and the gradients propagate through this clipping layer during the JSCC encoder-decoder pair training.

After clipping, the transmit signal $y \in \mathbb{C}^{(N_s+N_p)(L_{fft}+L_{cp})}$ propagates through the multipath channel described in (1). Once the receiver obtains the noisy channel output $\hat{y}$, it removes the CP, and applies DFT to produce the frequency domain pilots $\hat{Y}_p$ and data symbols $\hat{Y}$. If there was no clipping (either deliberate or amplifier-inherent), the following holds for pilot and information symbols:

$$\hat{Y}_p[i,k] = H[k]Y_p[i,k] + W[i,k] \quad (3)$$

$$\hat{Y}[j,k] = H[k]Y[j,k] + V[j,k] \quad (4)$$

where H denotes the channel frequency response for the k_{th} subcarrier, and both W and V denote noise samples. The decoder is trained to estimate transmitted source $\hat{x}$ given Y_p,

$\hat{Y}_p$ and $\hat{Y}$. The loss function of this learning problem can then be expressed as:

$$L(\theta, \phi) = \mathbb{E}_{p(x,\hat{Y},\hat{Y}_p)}[d(D_\phi(E_\theta(x), \hat{Y}, \hat{Y}_p), x)] \quad (5)$$

where $\mathbb{E}_p[\ ]$ is the expected value over distribution p, and d denotes the distortion calculated by x and $\hat{x}(= D_\phi(E_\theta(x), \hat{Y}, \hat{Y}_p))$. The goal is to find θ^* and ϕ^* that minimize the loss function $L(\theta, \phi)$.

It must be noted that including OFDM signal processing into the proposed JSSC framework is possible because IDFT/DFT and CP removal/insertion can be treated as linear layers in the neural network with fixed parameters. During back-propagation, gradients can pass through these (I)DFT/CP linear layers. Similarly, the proposed JSSC framework learns the encoder and decoder networks in the presence of signal clipping at the transmitter which is similar to a clipped ReLu activation function.

C. Decoder Design with Domain Knowledge

To estimate the source information, one possible approach is to concatenate $\hat{Y}$, Y_p and $\hat{Y}_p$ to directly feed them into a deep ('upscale') neural network. This is illustrated in Figure 2 as the *IMPLICIT* method. This approach relies on the deep neural network to learn the underlying relationships in (3) and (4) to perform the (sub)optimal estimation of the source. We name this method *IMPLICIT* because it implicitly learns traditional channel estimation and equalization processes in the decoder network. However, this method treats the entire decoder as a blockbox, and hence it faces the problems of slow convergence and sub-optimal performance as argued in [11]. As an alternative solution, we propose a decoder structure shown in Figure 2 bottom right where explicit signal processing kernels such as channel estimation and equalization are included as pre-processing steps in front of the neural network based decoder. Due to the signal clipping and other non-linear activations in the transmitter, the manually-designed channel estimation and equalization methods may not be optimal. Thus, we further introduce two residual connections with light-weight neural networks ('subnets') so that they can learn and compensate residual errors in the channel estimation and equalization. We name this method *EXPLICIT*. For simplicity, we use the per-channel MMSE channel estimation method [13] which does not need second order statistics of the channel and matrix inversion. For equalization, we adopt a conventional MMSE equalizer.

IV. TRAINING AND EVALUATION

We test our proposed method using CIFAR-10 and CelebA datasets. The CIFAR-10 dataset contains 60,000 32×32-pixel images whereas CelebA contains more than 200,000 celebrity images. We scaled and cropped CelebA images to 64×64-pixels. For testing, we take 10,000 images from each test dataset (unused for training) and transmit each image 5 times through different random realizations of the multipath channel. We use both PSNR and SSIM to evaluate the reconstruction quality. We test our JSCC method on CIFAR-10 using the
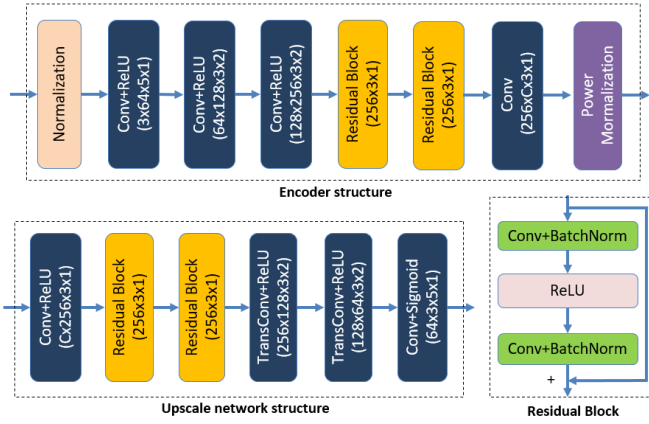
Fig. 3. Network structure for the proposed method. There is a batch normalization layer between each convolutional layer and ReLU activation function (although not shown in the figure).

TABLE I
CIFAR-10 AND CelebA EVALUATION (TOP: PSNR, BOTTOM: SSIM)

Dataset	CIFAR-10, CPP=0.21			CelebA, CPP=0.05		
SNR	0dB	10dB	20dB	0dB	10dB	20dB
Direct w/o OFDM	21.05	24.93	26.21	21.86	24.76	25.73
IMPLICIT	21.82	26.7	30.54	22.27	26.54	29.37
EXPLICIT	**22.31**	**27.91**	**31.69**	**22.92**	**27.27**	**30.31**

Dataset	CIFAR-10, CPP=0.21			CelebA, CPP=0.05		
SNR	0dB	10dB	20dB	0dB	10dB	20dB
Direct w/o OFDM	0.697	0.84	0.871	0.683	0.788	0.811
IMPLICIT	0.709	0.893	0.952	0.702	0.85	0.908
EXPLICIT	**0.727**	**0.915**	**0.962**	**0.73**	**0.867**	**0.924**

network structure shown in Figure 3, which follows the design principles in [24]. Because of the image size difference, networks for CelebA have one more down-sampling module at the encoder and one more upsampling module at the decoder. The subnets in *EXPLICIT* method follow a *Conv-BatchNorm-ReLu-Conv-BatchNorm* structure just like the residual block in the main network shown in Figure 3.

We adopt mean-squared-error (MSE) as our distortion measure (5) for the network training. For both datasets, we use a batch size of 128 and train the neural networks using ADAM with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. For CIFAR-10, we apply a learning rate of 10^{-3} and train for 400 epochs with linear learning rate decaying for the last 200 epochs. For CelebA, the learning rate is set to 5×10^{-4} and we train the end-to-end system for 60 epochs with linear learning rate decaying for the last 30 epochs.

The OFDM and channel parameters are set to $L_{fft} = 64, L_{cp} = 16, N_p = 2, L = 8$, and $\gamma = 4$. Note that the coding rate of our method depends on the number of samples after encoding. For a source image of the size $H \times W \times C$ pixels (C is the number of color channels), the coding rate is $\frac{(N_p+N_s)(L_{fft}+L_{cp})}{HWC}$ (channel-use per pixel, CPP). The length of CP has to be larger than the delay spread of the channel to avoid inter-symbol interference.
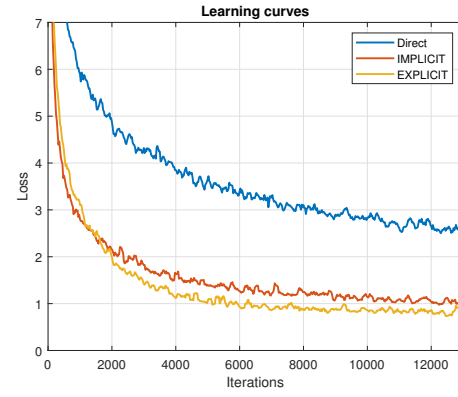

Fig. 4. Learning curves for three methods with a SNR of 20dB and $N_s = 6$ for CelebA dataset.

A. Effect of applying domain knowledge

To quantify the effect of incorporating domain knowledge into our framework, we compare our method with a neural network-only scheme which we call 'direct' where symbols are transmitted through the multipath channel without OFDM layers. All compared methods are tested with both CIFAR10 and CelebA under different SNRs and the coding rate is kept the same for each dataset (CPP is 0.21 and 0.05 for CIFAR-10 and CelebA, respectively). The results are shown in Table I. Including OFDM (I)DFT/CP processing layers, our *IMPLICIT* and *EXPLICIT* method attain superior PSNR and SSIM than the direct method without OFDM layers especially for high SNRs. As we introduce more domain knowledge into *EXPLICIT* configuration, the system experiences additional performance gain. Note that the subnets in *EXPLICIT* (Fig. 2) only introduce 0.2% additional parameters than IMPLICIT method to obtain the gain shown in Table I.

As shown in Figure 4, significant training efficiency gain is obversed for *IMPLICIT* and *EXPLICIT* compared to the 'direct' method. With OFDM (I)DFT/CP layers, *IMPLICIT* and *EXPLICIT* attain significantly (2.5×) lower loss function values with faster (5×) training convergence time for CelebA dataset. The training time is similar for *IMPLICIT* and *EXPLICIT* schemes but *EXPLICIT* method converges to a 20% lower loss function value for CelebA training.

B. Comparison with a separate coding scheme

In this section, we compare our proposed JSCC method with a baseline separate source and channel coding scheme. The baseline separate coding scheme uses the state-of-the-art BPG [25] as the image codec and LDPC in the IEEE 802.11n WiFi standard [26] as the channel code. We test with three different LDPC codes $(972, 1944), (1296, 1944), (1458, 1944)$ which correspond to rates $1/2$, $2/3$ and $3/4$ respectively. We use BPSK, QPSK, 16QAM, and 64QAM for OFDM modulation applied to BPG and LDPC coded bit sequences. Then, we enumerated all possible modulation and coding rate combinations to identify the optimal configuration for the baseline separate coding scheme. The channel estimation for this baseline is based on per-channel MMSE, which is also
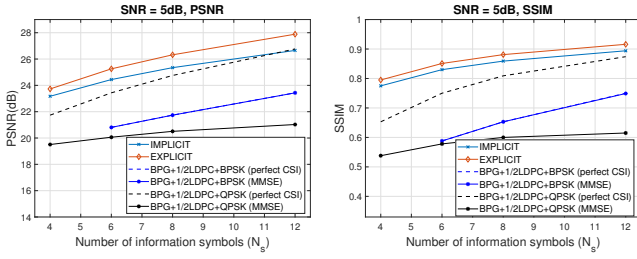
Fig. 5. Performance of the proposed method on CIFAR-10 with respect to the number of information symbols (N_s) for a fixed SNR of 5dB

Fig. 6. Performance of the proposed method on CIFAR-10 with respect to the number of information symbols (N_s) for a fixed SNR of 15dB

used in our *EXPLICIT* scheme. The estimated channel is then used to calculate the log likelihood ratio (LLR) for LDPC decoding in the baseline separate coding scheme.

We compare PSNR and SSIM of the proposed JSSC and the separate coding baseline at the same/similar rate and SNR. Because BPG cannot achieve arbitrary compression rates, we exhaustively search for the BPG parameter to obtain a rate closest to the target. For fair comparison, we did not include the BPG headers in the BPG rate computation. When BPG-LDPC encoded packets fail to reconstruct images due to fatal bit errors, we simply assume they achieve the average PSNR/SSIM of decodable images, giving an advantage to the baseline separate coding scheme.

Figure 5 and 6 compare the performance of deep JSCC schemes and baseline approaches with respect to the number of information symbols N_s in different SNR regimes using CIFAR-10 dataset. Smaller N_s indicates more aggressive (lower) rate to send an entire image with a fewer number of data symbols. For comparison, we also show the performance of baseline separate coding schemes with perfect channel state information (CSI) for error-free channel estimation. Note that, for the $\leq$5dB SNR case, BPG with 1/2 LDPC+BPSK cannot achieve the rate of $N_s = 4$ for the full image. Figure 5 and 6 show our deep JSSC consistently outperforms baseline separate coding schemes especially for the low SNR and low rate regime. Note that the performance of the baseline method approaches to that of the deep JSSC for some high rates when perfect CSI is given for their advantage. For the deep JSSC (both *IMPLICIT* and *EXPLICIT*), we never provide perfect CSI to the decoder. With imperfect CSI (same as in JSSC), the perfomrnace of the baseline scheme is significantly worse (e.g., >3dB PSNR) at the same low rate target. It is known that SSIM is a better metric than PSNR to quantify human perception of an image. Our deep JSSC provides a significant SSIM gain compared to the baseline although it is trained to minimizes the MSE loss function, not SSIM.

For the evaluation in Figure 7, we fix the rate to $N_s = 6$ for the the CIFAR-10 dataset (0.21 channel-use per pixel) and evaluate the reconstructed image quality for a wider SNR range. For the baseline method, we choose at each SNR point the optimal combination of the OFDM modulation QAM size and LDPC rate that yields the best performance. For 0dB SNR, the baseline method without perfect CSI failed to provide
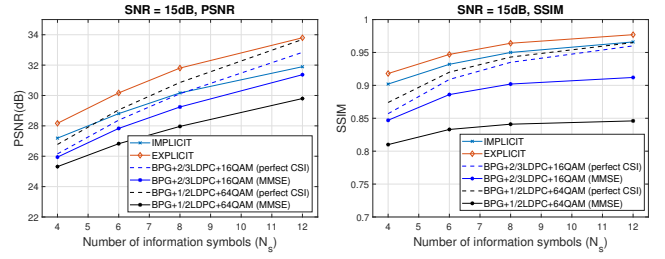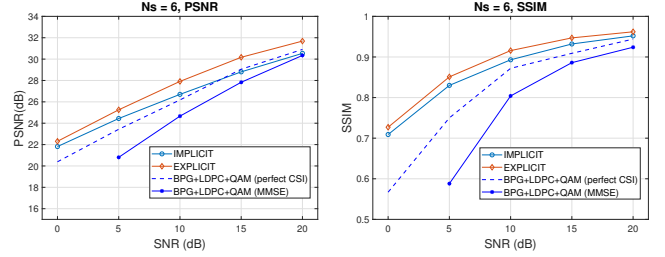

Fig. 7. Performance of the proposed method on CIFAR-10 with respect to SNR for a fixed number of information symbols ($N_s = 6$)

reasonable PSNR/SSIM at the given rate. It is observed that the proposed *EXPLICIT* JSCC method consistently outperforms the baseline with >3dB SNR gain for the same PSNR. The SNR gain is even higher when the image quality is measured in SSIM.

C. Robustness Analysis

Finally, we evaluate robustness of the JSCC with respect to the signal clipping. We also evaluate the impact of mismatched channel statistics such as SNR or the number of multipaths used during the JSCC training vs. actual testing.

Figure 8 shows the effect of clipping to the performance of the *EXPLICIT* JSCC and baseline method with various clipping ratios (CR) of $\rho = 1$, 1.4, and ∞ as defined in (2). Note that for the baseline, we adopt the methods in [22] [23] to calculate the LLR including the non-linear distortion introduced by clipping. It is observed that our JSCC provides a graceful degradation in PSNR/SSIM than the baseline method as more aggressive clipping is applied. And same as the previous section, our method outperforms the baseline in all range of SNRs in the presence of clipping.

Figure 9 left shows the JSCC performance when it is trained for a specific SNR (5dB or 15dB) and tested with a wide range of SNR from 0dB to 20dB. It is observed that the model trained with 5dB performs better for the low SNR regime while the model trained with 15dB performs better in the high SNR region. More balanced JSCC performance is obtained when the model is trained with random SNR values covering a wide range. The *EXPLICIT* JSCC model in Figure 9 right was trained with an $L = 8$ multipath channel model and tested with a wide range of L but the same noise level. Two curves show the performance at 5dB and 15dB SNR cases.
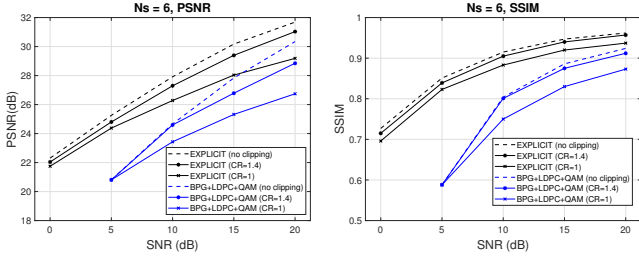
Fig. 8. Performance of the proposed method on CIFAR-10 with respect to SNR for different clipping ratios. $N_s = 6$.
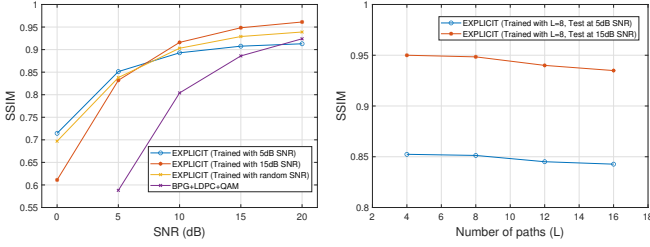

Fig. 9. Robustness test for the JSCC with CIFAR-10 images and $N_s = 6$. Left: training vs. testing SNR mismatch scenarios, Right: multipath channel model parameter (L) mismatch scenarios with SNR = 5dB or 15dB.

It is confirmed that our JSCC method is relatively insensitive to the mismatch in the number of multipaths of the channel model for the training vs. testing.

V. CONCLUSION

In this paper, we present a deep learning based JSCC scheme for wireless image transmission over multipath fading channels. We extend the JSCC scheme with explicit OFDM layers and signal processing steps driven by the expert domain knowledge. With addition explicit steps, the JSCC training convergence time improves with enhanced performance compared to naive *IMPLICIT* approaches without exploiting domain knowledge. Through extensive experimental simulations, we show that the deep JSCC method outperforms the conventional separate source and channel coding schemes with OFDM especially for low SNR and low rate regimes. The proposed JSCC framework effectively learns source compression as well as protection against noise in challenging multipath channels. Moreover, the JSCC framework incorporates deliberate signal clipping during the training process to significantly reduce the PAPR with graceful performance degradation. Our approach is shown to be robust when the evaluation channel model characteristics are mismatched to those used during the training.

REFERENCES

[1] Claude E Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[3] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[6] Eirina Bourtsoulatze, David Burth Kurka, and Deniz Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.

[7] David Burth Kurka and Deniz Gündüz, "Successive refinement of images with deep joint source-channel coding," in *SPAWC*. IEEE, 2019, pp. 1–5.

[8] David Burth Kurka and Deniz Gündüz, "Deepjscc-f: Deep joint source-channel coding of images with feedback," *IEEE Journal on Selected Areas in Information Theory*, 2020.

[9] Kristy Choi, Kedar Tatwawadi, Aditya Grover, Tsachy Weissman, and Stefano Ermon, "Neural joint source-channel coding," in *ICML*. PMLR, 2019, pp. 1182–1192.

[10] Yuxuan Song, Minkai Xu, Lantao Yu, Hao Zhou, Shuo Shao, and Yong Yu, "Infomax neural joint source-channel coding via adversarial bit flip," *arXiv preprint arXiv:2004.01454*, 2020.

[11] Hengtao He, Shi Jin, Chao-Kai Wen, Feifei Gao, Geoffrey Ye Li, and Zongben Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 77–83, 2019.

[12] Timothy O'Shea and Jakob Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[13] Alexander Felix, Sebastian Cammerer, Sebastian Dörner, Jakob Hoydis, and Stephan Ten Brink, "Ofdm-autoencoder for end-to-end learning of communications systems," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[14] Yashas Malur Saidutta, Afshin Abdi, and Faramarz Fekri, "Joint source-channel coding for gaussian sources over awgn channels using variational autoencoders," in *ISIT*. IEEE, 2019, pp. 1327–1331.

[15] Nariman Farsad, Milind Rao, and Andrea Goldsmith, "Deep learning for joint source-channel coding of text," in *ICASSP*. IEEE, 2018, pp. 2326–2330.

[16] Tianqi Wang, Chao-Kai Wen, Hanqing Wang, Feifei Gao, Tao Jiang, and Shi Jin, "Deep learning for wireless physical layer: Opportunities and challenges," *China Communications*, vol. 14, no. 11, pp. 92–111, 2017.

[17] Zhijin Qin, Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang, "Deep learning in physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 93–99, 2019.

[18] Guan Gui, Hongji Huang, Yiwei Song, and Hikmet Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, 2018.

[19] Xuanxuan Gao, Shi Jin, Chao-Kai Wen, and Geoffrey Ye Li, "Comnet: Combination of deep learning and expert knowledge in ofdm receivers," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2627–2630, 2018.

[20] Minhoe Kim, Woongsup Lee, and Dong-Ho Cho, "A novel papr reduction scheme for ofdm system based on deep learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 510–513, 2017.

[21] Xiaodong Li and Leonard J Cimini, "Effects of clipping and filtering on the performance of ofdm," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*. IEEE, 1997, vol. 3, pp. 1634–1638.

[22] Hideki Ochiai and Hideki Imai, "Performance analysis of deliberately clipped ofdm signals," *IEEE Transactions on communications*, vol. 50, no. 1, pp. 89–101, 2002.

[23] Hun Seok Kim and Babak Daneshrad, "Power optimized pa clipping for mimo-ofdm systems," *IEEE transactions on wireless communications*, vol. 10, no. 9, pp. 2823–2828, 2011.

[24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[25] Fabrice Bellard, "BPG image format," https://bellard.org/bpg/.

[26] tavildar, "Ldpc codes," https://github.com/tavildar/LDPC.