Deep Joint Source-Channel Coding for Image Transmission with Visual Protection

Jialong Xu, *Member, IEEE,* Bo Ai, *Fellow, IEEE,* Wei Chen, *Senior Member, IEEE,* Ning Wang, *Member, IEEE,* and Miguel Rodrigues, *Fellow, IEEE*

***Abstract*—Joint source-channel coding (JSCC) has achieved great success due to the introduction of deep learning (DL). Compared to traditional separate source-channel coding (SSCC) schemes, the advantages of DL-based JSCC (DJSCC) include high spectrum efficiency, high reconstruction quality, and relief of "cliff effect". However, it is difficult to couple existing secure communication mechanisms (e.g., encryption-decryption mechanism) with DJSCC in contrast with traditional SSCC schemes, which hinders the practical usage of this emerging technology. To this end, our paper proposes a novel method called DL-based joint protection and source-channel coding (DJPSCC) for images that can successfully protect the visual content of the plain image without significantly sacrificing image reconstruction performance. The idea of the design is to use a neural network to conduct visual protection, which converts the plain image to a visually protected one with the consideration of its interaction with DJSCC. During the training stage, the proposed DJPSCC method learns: 1) deep neural networks for image protection and image deprotection, and 2) an effective DJSCC network for image transmission in the protected domain. Compared to existing source protection methods applied with DJSCC transmission, the DJPSCC method achieves much better reconstruction performance.**

***Index Terms*—Visual protection, image transform, joint source-channel coding, deep learning.**

I. Introduction

THE modular design principle based on Shannon's separation theorem [1] is the cornerstone of modern communications and has enjoyed great success in the development of wireless communications. However, the assumptions of unlimited codeword length, delay, and complexity in the separation theorem are not possible in real wireless environments, leading to sub-optimal separate source-channel coding (SSCC). Moreover, for time-varying channels, when the channel quality is worse than the target channel quality, SSCC cannot decode any information due to the collapse of channel coding; when the channel quality is better than the target quality, separate coding cannot further improve reconstruction quality. This is the famous "cliff effect" [2], which increases the cost of SSCC during wireless transmission. In recent years, joint source-channel coding (JSCC) has been theoretically demonstrated to have better error exponents than SSCC in discrete memoryless source channels [3], which motivates the development of various JSCC designs over the years. Benefiting from the data-driven nature, deep learning (DL)-based JSCC (DJSCC) successfully reduces the difficulty of coding design existing in traditional JSCC for variant types of sources and channels [4]–[6], balances performance and storage requirements [7]–[9], cooperates with orthogonal frequency division multiplexing (OFDM) widely employed in wireless communication systems [10], and matches semantic communications [11].

The important step to put DJSCC into practice is to protect the source information from eavesdroppers. Note that DJSCC is empowered by data-driven manner, which learns the effective encoding function and decoding function from scratch, causing the correlation between the source signal and the channel input symbols. To conduct a secure DJSCC communication, the source signal or the channel input symbols should be protected during DJSCC transmission. Considering that the channel input symbols belonging to the physical layer can be operated only by the wireless service provider in practical wireless scenarios, in this paper, we design to protect the source signal owned by the users, who can freely manipulate the source signal in the application layer. As illustrated in Fig. 1, the image owner intends to transmit a plain image to the image recipient through the network that contains an untrusted wired network and a wireless transmission service. To protect the visual content of the plain image, the image owner transforms the plain image into a protected image before providing it to the wireless service provider. Then the visually protected image is transmitted by the wireless service provider through DJSCC transmission. After DJSCC wireless transmission, the protected image decoded by the DJSCC decoder with some distortion is transmitted to the image recipient through the untrusted network (e.g., Internet). The image recipient transforms the distorted protected image to the plain image. Even if the protected image or the distorted protected image was leaked or stolen during the wired network

This work is supported by the Natural Science Foundation of China (62122012, 62221001); the Beijing Natural Science Foundation (L202019, L211012); the Fundamental Research Funds for the Central Universities (2022JBQY004). *(corresponding authors: Bo Ai; Wei Chen)*

Jialong Xu is with State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, China, and also with Frontiers Science Center for Smart High-speed Railway System, China (e-mail: jialongxu@bjtu.edu.cn).

Bo Ai is with State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, China, also with Beijing Engineering Research Center of High-speed Railway Broadband Mobile Communications, China, and also with School of Information Engineering, Zhengzhou University, China (e-mail: boai@bjtu.edu.cn).

Wei Chen is with State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, China, and also with Key Laboratory of Railway Industry of Broadband Mobile Information Communications, China (e-mail: weich@bjtu.edu.cn).

Ning Wang is with School of Information Engineering, Zhengzhou University, China (e-mail: ienwang@zzu.edu.cn).

Miguel Rodrigues is with the Department of Electronic and Electrical Engineering, University College London, London, WC1E 7JE, U.K. (e-mail: m.rodrigues@ucl.ac.uk).

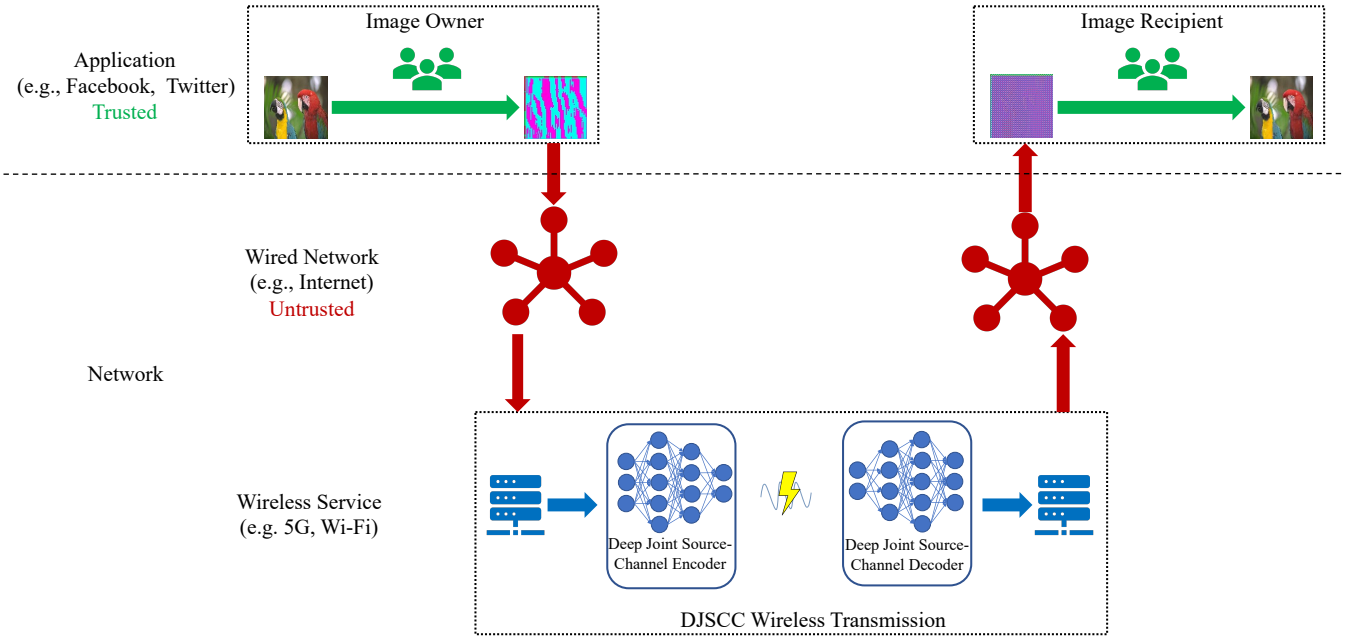


Fig. 1. The DJSCC based wireless communication system.

transmission process, the visual content of the plain image cannot be acquired directly by eavesdroppers.

One might like to use existing protection methods in SSCC-based communication systems, which can be conducted either before the source encoder or after the source encoder. However, a major issue with these protected methods of SSCC is that the change in the visual structure of the plain image leads to the degradation of the transmission of DJSCC. It is worth noting that the potential method applied in this scenario is not limited by an information-theoretic framework [12].

In this paper, we design a DL-based joint protection and source-channel coding method that can generate the visually protected image suitable for the DJSCC transmission. To the best of our knowledge, this is the first scheme to couple source protection with DJSCC based wireless transmission. The major contributions are summarized as follows.

- We design a unified framework which consists of the protection module, the deprotection module, the feature extraction module and the DJSCC module to protect the visual content of the image source taking into account DJSCC transmission, which can overcome the problem of "cliff effect", avoid the latency brought by channel mismatch, and most importantly, lead to considerable performance.
- By using the feature extraction module, we redesign the loss function of the proposed end-to-end framework. During the training stage, the proposed framework learns an effective method to provide visual protection for the plain image, an effective protection domain for the subsequent DJSCC transmission, an effective DJSCC transmission method, and an effective method to reconstruct the plain image. The strength of visual protection can be adjusted to satisfy different levels of protection requirements.
- We propose two design principles for the protection module and the deprotection module. These principles can quickly guide the concrete design of the protection network and the deprotection network and meet the real communication scenario requirements, e.g., the storage overhead and computational complexity.

The rest of this paper is organized as follows. Section II presents related work on deep joint source-channel coding and image protection. Then, the proposed method is presented in Section III. In Section IV, the proposed method is evaluated on datasets with low resolution and high resolution. Finally, Section V concludes this paper.

II. Related Work

A. Deep Joint Source Channel Coding

The initial DJSCC work proposed a recurrent neural network (RNN) for text transmission over binary erasure channels [13]. From then on, DJSCC attracted increasing interest, especially for image compression and transmission. Compared to the SSCC scheme (e.g., JPEG/JPEG2000 for image coding and LDPC for channel coding), the DJSCC scheme designed in [4] has better image restoration quality, especially in the low signal-to-noise ratio (SNR) regime. To well adapt to variable bandwidth and deal with coding for distributed sources [14], DJSCC schemes with adaptive-bandwidth image transmission and distributed transmission are proposed in [7] and [15], respectively. Taking into account the classical feedback scenario [16], image transmission with channel output feedback is proposed in [8]. However, all the aforementioned schemes are trained and deployed under the same channel conditions (the single SNR) to ensure optimality, demanding the use of multiple trained networks to suit a range of SNR that leads to considerable storage requirements in transceivers. To overcome this challenging problem in DJSCC, [9] proposed

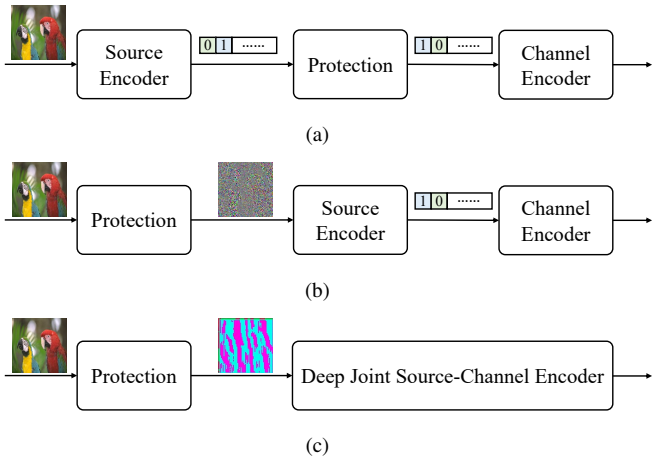


Fig. 2. Different strategies for image protection in SSCC and DJSCC. (a) Protect the data encoded by the source encoder before the channel encoder in SSCC, (b) Protect the image source before the source encoder in SSCC, and (c) Protect the image source before the source encoder in DJSCC.

a single network for DJSCC that can adapt to a wide range of SNR conditions to meet the memory limit of the device in real wireless scenarios. Furthermore, [17] proposed a DJSCC method based on the maximization of mutual information between the source and the received noisy codeword for the binary erasure channel and the binary symmetric channel. [18] and [19] model their DJSCC systems via a variational autoencoder and manifold variational autoencoders for a Gaussian source, respectively. So far, by using the data-driven method, DJSCC successfully reduces the difficulty of coding design in traditional JSCC, making it a promising technology in low-latency and low-power scenarios.

B. Image Protection

In SSCC, the protection operation can be executed between the source encoder and the channel encoder, as shown in Fig. 2(a), or before the source encoder, as shown in Fig. 2(b). The first strategy applies the source encoder to compress the image to the binary data, and then uses bit-oriented based encryption methods, e.g., the data encryption standard (DES) [20], advanced encryption standard (AES) [21], and Rivest–Shamir–Adleman (RSA) [22], to protect the binary data. The second strategy is fit for the typical scenario, where the image provider only takes care of protecting the image content and the telecommunications provider has an overriding interest in improving spectrum efficiency. To directly protect the image content, various pixel-based and space-based image protection methods have been developed to transform a plain image to a visually protected image, e.g., Arnold [23], Hill algorithm [24] and 3D chaotic map-based method [25]. Taking into account the compression needs in the second strategy, dedicated protection and compression methods are proposed to improve the compression ratio in the protection domain [26]–[28]. As shown in Fig. 2(c), to protect the visual content of the plain image for DJSCC transmission, the image protection should be executed in front of the deep joint source-channel encoder, which is similar to the position of the protection module in Fig. 2(b). The protection methods originally designed for the second strategy in SSCC may be applied in DJSCC. However, these image protection methods change the visual structure of the plain image, break the coherence in adjacent pixels, and may cause performance degradation of DJSCC transmission.

DL has led to state-of-the-art performance in various image processing tasks, motivating the application of DL to protect the image source. Image protection methods proposed by [29], [30] are designed for DL-based classification with acceptable classification accuracy. However, these image protection methods designed for the classification task are not suitable for DJSCC. Protected images for the classification task only reserve some specific semantic information relevant to the image class, while the pixel-based information of the image is discarded, which causes the performance degradation for the image reconstruction task.

III. Deep Joint Protection and Source-Channel Coding

Based on the scenario described in Section I, our motivation is to successfully protect the visual content of the plain image without significantly sacrificing the image reconstruction performance and efficiently transmit the visually protected image through the DJSCC. In this Section, a DL-based joint protection and source-channel coding (DJPSCC) method is proposed for these purposes.

A. System Model

Considering a visually protected DJSCC transmission system, as shown in the lower part of Fig. 3, a plain image is represented by $\boldsymbol{x} \in \mathbb{R}^n$, where $\mathbb{R}$ denotes the set of real numbers and $n = h \times w \times c$. Here, h, w and c denote the height, the width and the number of channels of an image, respectively. The protection module transforms the plain image into a visually protected image, which is expressed as:

$$\boldsymbol{y} = e_{\boldsymbol{\mu}}(\boldsymbol{x}) \in \mathbb{R}^n, \tag{1}$$

where $e_{\boldsymbol{\mu}}(\cdot)$ represents an protection module parameterized by the set of parameters $\boldsymbol{\mu}$. The protected image $\boldsymbol{y}$ and the plain image $\boldsymbol{x}$ have the same size.

After the protection process, the protected image $\boldsymbol{y}$ is encoded by the joint source-channel encoder as:

$$\boldsymbol{z} = f_{\boldsymbol{\theta}}(\boldsymbol{y}) \in \mathbb{C}^k, \tag{2}$$

where $\mathbb{C}$ denotes the set of complex numbers, k represents the number of channel input symbols, and $f_{\boldsymbol{\theta}}(\cdot)$ represents a joint source-channel encoder parameterized by the set of parameters $\boldsymbol{\theta}$. The real and imaginary parts of $\boldsymbol{z}$ are mapped to the in-phase components I and the quadrature components Q of the transmitted signals, respectively. During transmission, the power constraint $\frac{1}{k}\mathbb{E}(\boldsymbol{z}\boldsymbol{z}^*) \leq 1$ must be satisfied, where $\boldsymbol{z}^*$ is the complex conjugate transpose of $\boldsymbol{z}$.

The transmitted signals are corrupted by the wireless channel. We adopt the well known additive white Gaussian noise (AWGN) model given by:

$$\hat{\boldsymbol{z}} = \eta(\boldsymbol{z}) = \boldsymbol{z} + \boldsymbol{\omega}, \tag{3}$$

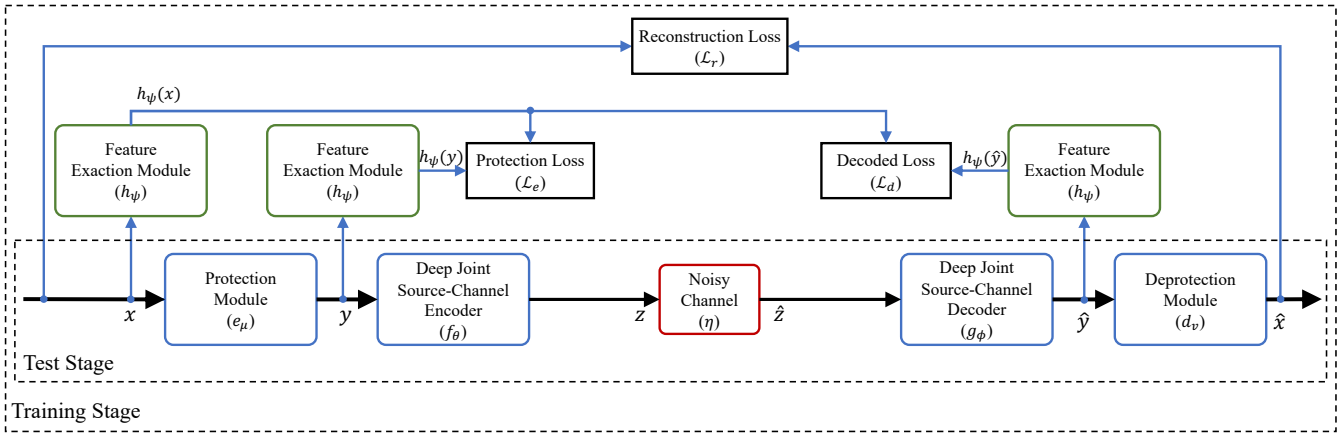


Fig. 3. The system model of the proposed DJPSCC method.

where $\hat{\boldsymbol{z}} \in \mathbb{C}^k$ is the channel output and $\boldsymbol{\omega} \in \mathbb{C}^k$ denotes the additive noise modeled by $\boldsymbol{\omega} \sim \mathbb{CN}(0, \sigma^2 \boldsymbol{I})$, where σ^2 represents the average noise power and $\mathbb{CN}(\cdot, \cdot)$ denotes a circularly symmetric complex Gaussian distribution.

In turn, the channel output symbols $\hat{\boldsymbol{z}}$ are decoded by the joint source-channel decoder as:

$$\hat{\boldsymbol{y}} = g_{\boldsymbol{\phi}}(\hat{\boldsymbol{z}}) \in \mathbb{R}^n, \tag{4}$$

where $g_{\boldsymbol{\phi}}(\cdot)$ represents a joint source-channel decoder parameterized by the set of parameters $\boldsymbol{\phi}$. The decoded image $\hat{\boldsymbol{y}} \in \mathbb{R}^n$ should be a visually protected image, with the same size as the protected image $\boldsymbol{y}$.

Similarly to the protection step, the deprotection module is employed to convert the decoded image to the deprotected image as follows:

$$\hat{\boldsymbol{x}} = d_{\boldsymbol{\nu}}(\hat{\boldsymbol{y}}) \in \mathbb{R}^n, \tag{5}$$

where $d_{\boldsymbol{\nu}}(\cdot)$ represents a deprotection module parameterized by the set of parameters $\boldsymbol{\nu}$ and the deprotected image $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ is a restored version of the plain image. The bandwidth ratio R is defined as k/n, where n is the source size (i.e., image size) and k is the channel bandwidth (i.e., channel input size).

B. The Proposed Method

In sharp contrast to DJSCC methods [4], [7]–[9], we require our DJPSCC method to address two issues:

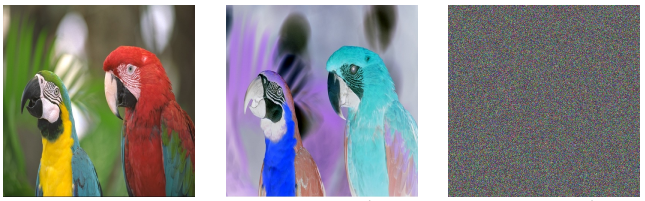
1) Protect the visual content of a plain image.

2) Extract effective features from the protected image for subsequent DJSCC transmission.

The classical full-reference metric of image similarity is peak signal-to-noise ratio (PSNR) between the original image and the restored image, which is defined as:

$$\text{PSNR} = 10\log_{10}\frac{\text{MAX}^2}{\text{MSE}}(\text{dB}). \tag{6}$$

where MAX is the maximum possible value of the image pixels and MSE is the abbreviation of mean square error between the original image and the restored image. Although the prediction of PSNR performance is not always consistent with visual quality perceived by the human visual system,



Plain Image (a) | PSNR:6.179 dB Inverted Image (b) | PSNR:9.957 dB Shuffled Image (c)

Fig. 4. PSNR comparison. (a) plain image, (b) inverted image (the intensity values of the plain image are subtracted by 255), (c) shuffled image (the intensity values of the plain image are randomly shuffled in space and channel dimension).

its simplicity makes it widely used in the field of image processing [31]. However, as illustrated in Fig. 4, PSNR is not a good metric to assess visual security due to the excessive difference between the plain image and the visually protected image. An example is that the image in Fig. 4(b) has a lower PSNR than that in Fig. 4(c), while the visual content (e.g., the birds and the leaf) in Fig. 4(b) are more easily identified than the image in Fig. 4(c).

In recent years, various visual security metrics (VSMs), including handcraft-based VSMs [32]–[35] and DL-based VSMs [36], [37], are designed to assess the visual security of the image. Here, we employ a feature extraction network to measure visual security. The feature extraction method has been used successfully to measure the similarity between two images [38] and the difference between two images [29].

Concretely, in the training stage, the features of the plain image $\boldsymbol{x}$, the protected image $\boldsymbol{y}$, and the decoded image $\hat{\boldsymbol{y}}$ are extracted by the feature extraction module $h_{\boldsymbol{\psi}}$ in Fig. 3, where ψ is the set of parameters of the feature extraction module. Note that both the protected image $\boldsymbol{y}$ and the decoded image $\hat{\boldsymbol{y}}$ are in the protected domain. The feature loss $\mathcal{L}_e$ between the plain image $\boldsymbol{x}$ and the protected image $\boldsymbol{y}$ is expressed as:

$$\mathcal{L}_e = \frac{1}{m}\|h_{\boldsymbol{\psi}}(\boldsymbol{x}) - h_{\boldsymbol{\psi}}(\boldsymbol{y})\|_2^2, \tag{7}$$

and the feature loss $\mathcal{L}_d$ between the plain image $\boldsymbol{x}$ and the

decoded image $\hat{\boldsymbol{y}}$ is expressed as:

$$\mathcal{L}_d = \frac{1}{m}\|h_{\boldsymbol{\psi}}(\boldsymbol{x}) - h_{\boldsymbol{\psi}}(\hat{\boldsymbol{y}})\|_2^2, \tag{8}$$

where $h_{\boldsymbol{\psi}}(\boldsymbol{x}) \in \mathbb{R}^m$, $h_{\boldsymbol{\psi}}(\boldsymbol{y}) \in \mathbb{R}^m$, and $h_{\boldsymbol{\psi}}(\hat{\boldsymbol{y}}) \in \mathbb{R}^m$ are the features of the plain image $\boldsymbol{x}$, the protected image $\boldsymbol{y}$, and the decoded image $\hat{\boldsymbol{y}}$ extracted by the feature extraction network and $m = h_f \times w_f \times c_f$. Here, h_f, w_f, and c_f denote the height, width, and number of channels of an extracted feature, respectively. For simplicity, MSE is adopted in this paper to characterize the strength of visual security. Other forms of the feature loss can also be applied in the proposed DJPSCC method. However, since we focus on the mechanism design, the design of the feature loss is beyond the scope of this work. It is worth noting that once the feature extraction module is chosen, its parameters are fixed during the training stage.

Moreover, the reconstruction loss $\mathcal{L}_r$ between the plain image $\boldsymbol{x}$ and the deprotected image $\hat{\boldsymbol{x}}$ is expressed as:

$$\mathcal{L}_r = d(\boldsymbol{x}, \hat{\boldsymbol{x}}) = \frac{1}{n}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2, \tag{9}$$

where x_i and $\hat{x}_i$ represent the i-th pixel value of the plain image $\boldsymbol{x}$ and the deprotected image $\hat{\boldsymbol{x}}$, respectively.

Unlike image-to-image translation tasks [30], [38] that minimize feature loss in the training stage, the proposed method maximizes $\mathcal{L}_e$ and $\mathcal{L}_d$ to provide visual protection in protected images $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$. The total loss applied to train the proposed DJPSCC method is expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_r - \lambda_e\mathcal{L}_e - \lambda_d\mathcal{L}_d, \tag{10}$$

where $\lambda_e \in \mathbb{R}^+$ and $\lambda_d \in \mathbb{R}^+$ are the weights of $\mathcal{L}_e$ and $\mathcal{L}_d$, respectively. It should be emphasized that the feature loss $\mathcal{L}_e/\mathcal{L}_d$ represents the difference between the plain image and the protected/decoded image. A smaller value of $\mathcal{L}_e/\mathcal{L}_d$ means a small visual difference between the plain image and the protected/decoded image, while a larger value means a big difference between the plain image and the protected/decoded image and is considered to protect the visual content of the plain image. During the training stage, when the total loss $\mathcal{L}_{total}$ is minimized, its component $\mathcal{L}_e/\mathcal{L}_d$ loss is maximized to increase the visual protection ability.

Under a certain bandwidth ratio $R = k/n$, the proposed DJPSCC method learns the parameters of the protection module $\boldsymbol{\mu}$, the deep joint source-channel encoder $\boldsymbol{\theta}$, the joint source-channel decoder $\boldsymbol{\phi}$, and the deprotection module $\boldsymbol{\nu}$ by minimizing the total loss as follows:

$$(\boldsymbol{\mu}^*, \boldsymbol{\theta}^*, \boldsymbol{\phi}^*, \boldsymbol{\nu}^*) = \underset{\boldsymbol{\mu},\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\nu}}{\arg\min}\, \mathbb{E}_{p(\sigma^2)}\mathbb{E}_{p(\boldsymbol{x},\hat{\boldsymbol{x}})}(L_{total}), \tag{11}$$

where $\boldsymbol{\mu}^*, \boldsymbol{\theta}^*, \boldsymbol{\phi}^*, \boldsymbol{\nu}^*$ are the optimal parameters, $p(\boldsymbol{x}, \hat{\boldsymbol{x}})$ represents the joint probability distribution of the plain image $\boldsymbol{x}$ and the deprotected image $\hat{\boldsymbol{x}}$, σ^2 is the average noise power, and $p(\sigma^2)$ represents the probability distribution of the channel noise. Note that the probability distribution of the channel noise instead of the fixed channel noise is adopted in this paper with consideration of the storage overhead and the difficulty to acquire the signal-to-noise ratio (SNR) in the image owner/recipient. In addition, an empirical average instead of a statistical average is adopted in the training stage. During the training stage, the proposed DJPSCC method learns: 1) an effective method to provide visual protection for the plain image, 2) an easy-to-be-extracted image domain for the subsequent DJSCC transmission, 3) an effective DJSCC transmission method, and 4) an effective method to reconstruct the plain image.

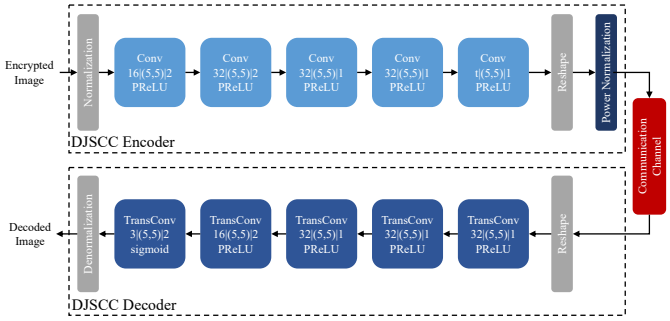


Fig. 5. The architecture of the DJSCC network [4] adopted in this paper. The notation $K|(F,F)|S$ in a convolutional/transposed convolutional layer denotes that it has K filters with size F and stride down/up S.

After training the DJPSCC network, to provide security guarantees against the eavesdropper, the protection module $e_{\boldsymbol{\mu}}$ and the deprotection module $d_{\boldsymbol{\nu}}$ are distributed securely to the owner of the image and the recipient of the image using the security protocol, e.g., the Secure Sockets Layer (SSL) protocol, respectively. The deep joint source-channel encoder and decoder are distributed to the DJSCC transmission service provider. However, the protected/decoded image is correlated with the plain image because of the inherent transform operation existing in the protection stage, causing a potential weakness when attacked by applying a generative adversarial network with abundant data (e.g., protected images and irrelevantly plain images) and sufficient computational resource. To enhance the security of the proposed method, multiple DJPSCC networks could be trained with different initialized parameters, and the network itself could be the key to secure transmission.

In the test stage, the plain image $\boldsymbol{x}$ is first converted to the protected image $\boldsymbol{y}$ by the image owner using Eq. (1). The protected image $\boldsymbol{y}$ is then sent to the DJSCC transmission service provider. DJSCC transmission is executed using Eq. (2) and Eq. (4) and the decoded image $\hat{\boldsymbol{y}}$ is obtained. The DJSCC transmission service provider sends the decoded image $\hat{\boldsymbol{y}}$ to the image recipient, which uses Eq. (5) to deprotected the decoded image $\hat{\boldsymbol{y}}$. The test stage process is illustrated in the lower part of Fig. 3.

IV. Experimental Results

The proposed DJPSCC is a general framework that can be employed in most existing DJSCC architectures. To demonstrate the effectiveness of the DJPSCC, the first DJSCC architecture proposed in [4] is adopted in subsequent experiments. As shown in Fig. 5, the DJSCC encoder consists of the normalization layer, five alternant convolutional layers and PReLU layers, the reshape layer, and the power normalization

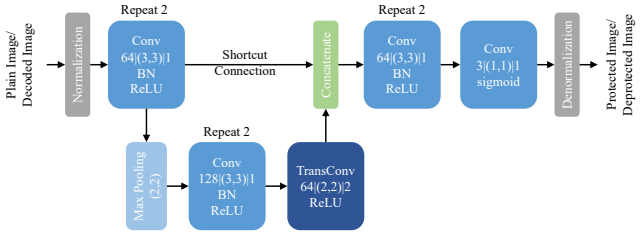


Fig. 6. The shallow version of U-Net [40] with one shortcut connection (UNet-S1) is adopted as the protection/deprotection network in this paper.

layer. The DJSCC decoder consists of the reshape layer, five alternant transposed convolutional layers and activation layers (i.e., four PReLU layers and one sigmoid layer), and the denormalization layer. The normalization layer converts the input image with the pixel value range [0, 255] to the image with the pixel value range [0, 1], and the denormalization layer performs the opposite operation. The notation $K|(F,F)|S$ in a convolutional/transposed convolutional layer denotes that it has K filters with size F and stride down/up S. The power normalization layer is used to satisfy the average power constraint at the transmitter. The channel number of the last convolutional layer in the DJSCC encoder is t, which is relevant to the channel bandwidth.

Although the power of the proposed DJPSCC is based on the ingenious design of the loss function and the end-to-end training strategy, the architectures of the protection module and the deprotection module still affect the reconstruction performance of the proposed DJPSCC. Here, we adopt two principles in designing the protection/deprotection module: 1) Shortcut connections in the protection/deprotection network could enhance the reconstruction performance of the proposed DJPSCC, while 2) A deeper protection/deprotection network would degrade the reconstruction performance of the proposed DJPSCC. The effectiveness of the two principles will be demonstrated in IV-C.

Tensorflow [39] and its high-level API Keras are used to implement the proposed DJPSCC method. DJPSCC is trained with a uniform distribution within the SNR range [0, 20] dB. The following experiments are run on a Linux server with twelve octa-core Intel(R) Xeon(R) Silver 4110 CPUs and sixteen GTX 1080Ti GPUs. Each experiment was assigned six CPU cores and a GPU.

A. DJPSCC Validity on the CIFAR-10 Dataset

We first consider the performance of the proposed method on the CIFAR-10 dataset, which consists of 60000 $32\times32\times3$ color images associated with 10 classes where each class has 6000 images. Note that the goal of our proposed method is to generate visually protected images for the untrusted transmission channels and reconstruct the plain image at the receiver, so the class label of each image is useless in the following experiments. The training dataset and the test dataset contain 50000 images and 10000 images, respectively.

U-Net [40] owns shortcut connections from its contracting path to its expansive path since it meets our first principle

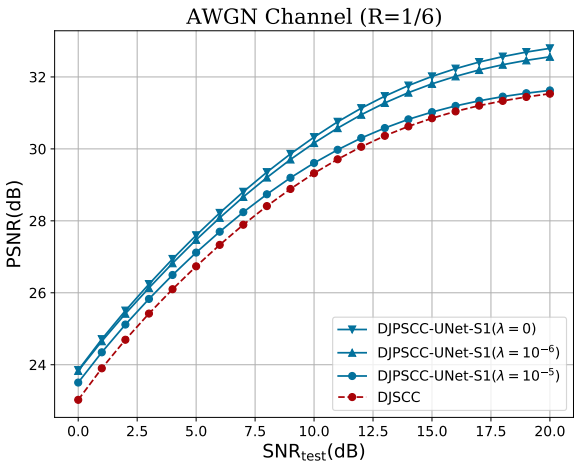


Fig. 7. Reconstruction performance of DJPSCC-UNet-S1 and DJSCC on CIFAR-10 test dataset with bandwidth ratio R=1/6.

for the protection/deprotection module. As shown in Fig. 6, a shallow version of U-Net with one shortcut connection (UNet-S1) is designed as the protection/deprotection network. The VGG16[1] [41] pretrained on ImageNet dataset is adopted as the feature extraction network. All networks were trained for 500 epochs using Adam Optimizer with an initial learning rate of 10^{-3}. Once learning stagnated for 10 epochs, the learning rate was reduced by a factor of 10. The performance of the DJPSCC networks was evaluated at specific $\text{SNR}_{\text{test}} \in$ [0,20] dB on the CIFAR-10 test dataset. To alleviate the effect of the randomness caused by the wireless channel, each image in the CIFAR-10 test dataset is transmitted 10 times. PSNR is used in the evaluation of the reconstruction performance between the plain image and the deprotected image. For simplicity, we allocate the same loss weight for visually protected images as $\lambda_e = \lambda_d = \lambda$.

Fig. 7 compares the reconstruction performance of DJPSCC-UNet-S1 with different loss weights (e.g., $\lambda = 0, 10^{-6}, 10^{-5}$) at bandwidth ratio $R = 1/6$. The reconstruction performance of the DJSCC without the protection module and the deprotection module is also plotted as a reference. With increasing λ, DJPSCC pays more attention to visual protection tasks, which cause a degradation in reconstruction performance. Since DJPSCC-UNet-S1 network is deeper and with more parameters than the DJSCC network, the reconstruction performance of DJPSCC-UNet-S1($\lambda = 0$) without feature loss constraint is almost 2 dB better than that of the DJSCC in $\text{SNR}_{\text{test}} \in [0, 20]$ dB. Although the feature loss is imposed on DJPSCC, benefited from the powerful ability of DJPSCC-UNet-S1, the reconstruction performance of DJPSCC-UNet-S1($\lambda = 10^{-5}$) is still better than that of DJSCC.

Table I evaluates the visual security of DJPSCC-UNet-S1 using the VSMs, i.e., LFBVS [33] and correlation (COR) [42]. Note that the PSNR and the structural similarity index (SSIM) are frequently used as reference in visual security evaluation, so the PSNR and the SSIM are also listed in

[1]https://keras.io/api/applications/vgg/#vgg16-function

TABLE I
VISUAL SECURITY EVALUATION OF DJPSCC-UNET-S1 ON THE CIFAR-10 TEST DATASET

Method	Protected Image				Decoded Image				Deprotected Image			
	LFBVS	COR	PSNR(dB)	SSIM	LFBVS	COR	PSNR(dB)	SSIM	LFBVS	COR	PSNR(dB)	SSIM
DJPSCC-UNet-S1 ($\lambda = 0$)	0.529	0.085	10.034	0.113	0.604	-0.012	8.091	-0.004	**0.198**	**0.984**	**30.330**	**0.947**
DJPSCC-UNet-S1 ($\lambda = 10^{-6}$)	0.684	-0.006	6.280	0.004	0.674	**-0.058**	6.897	**-0.019**	0.202	0.983	30.162	0.944
DJPSCC-UNet-S1 ($\lambda = 10^{-5}$)	**0.721**	**-0.082**	**4.980**	**-0.031**	**0.675**	-0.010	**5.420**	0.001	0.209	0.981	29.686	0.939

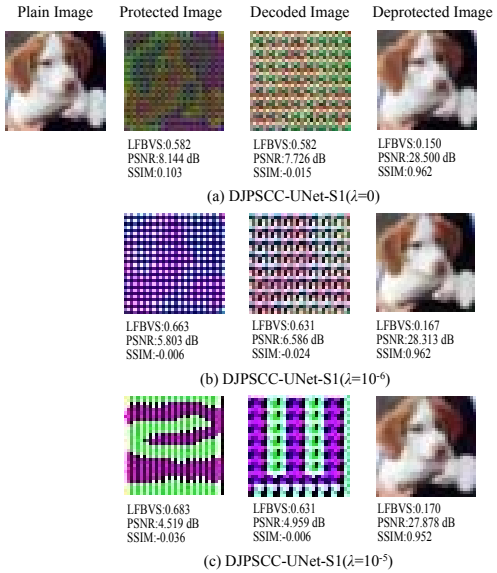


Fig. 8. Visually protected images and reconstruction images generated by DJPSCC-UNet-S1. The image in the first column is the plain image. The images in the second column are the protected images transformed by the image owner. The images in the third column are the decoded images decoded by the DJSCC decoder at $\mathrm{SNR} = 10$ dB. The images in the last column are the deprotected images transformed by the image recipient. (a) $\lambda = 0$, (b) $\lambda = 10^{-6}$, (c) $\lambda = 10^{-5}$.

Table I. During the evaluation, $\mathrm{SNR}_{\mathrm{test}}$ is 10 dB. A high score of the COR, the PSNR and the SSIM reflects high similarity between the visually protected image and the plain image, while a high score of the LFBVS reflects high visual security when comparing the visually protected image with the plain image. The range of PSNR, SSIM and LFBVS is $[0, +\infty]$, $(-1, 1]$ and $[0, 1]$, respectively. With increasing λ, the PSNR and SSIM of the protected image gradually decrease and the LFBVS gradually increases. Evaluation of the decoded image with increasing λ reveals a similar trend, except for a small inconsistency in the COR metric and the SSIM metric. Although the SSIM and the PSNR can exhibit satisfactory performance in predicting image quality, they are not appropriate for evaluating the visual security of protected images. For example, SSIM failed to measure badly blurred images [43]. In this situation, the LFBVS metric is more accurate than the SSIM metric, since the LFBVS is specially designed to measure the visual security between two images. In the following experiments, DJPSCC-UNet-S1($\lambda = 10^{-5}$) is chosen for the subsequent comparison.

Fig. 8 shows the visualization of the plain image, the protected images transformed by the owner of the image, the decoded images decoded by the DJSCC transmission service provider, and the deprotected images transformed by the image recipient at $\mathrm{SNR} = 10$ dB with different loss weights. The plain image comes from the CIFAR-10 test dataset. The outline of the dog can be vaguely identified in the protected image and the decoded image when $\lambda = 0$. With increasing λ, the outline of the dog gradually disappears in the protected image and the decoded image. The most successful visual protection exists in the decoded image with $\lambda = 10^{-5}$. All deprotected images with different λ can reconstruct the main visual content conveyed by the plain image. Although there is no explicit visual security in the loss function when $\lambda = 0$, the visual content is weakly protected to some extent due to the transformation provided by the protection module. If strong visual security is needed, a trade-off between the reconstruction performance and the visual protection performance exists in the DJPSCC method. That is, with increasing λ, the visual protection ability of the DJPSCC increases, while the quality of the DJPSCC reconstruction decreases.

This is the first work to design the DJPSCC framework that can protect the visual content of the plain image for DJSCC transmission. Here, we compare DJPSCC with one SSCC-based visual protection method, i.e., encryption then compression (EtC) [28], and two visual protection methods designed for DL-based classification task, i.e., the learnable image encryption (LE) method [44] and the pixel-based image encryption (PE) method [45]. The EtC method is a block scrambling-based encryption scheme with JPEG compression, which can securely transmit plain images through an untrusted channel, e.g., the untrusted wired network in Fig. 1. In the SSCC-based wireless communication system, the EtC can be regarded as the source coding module. During wireless transmission, the channel coding module and the modulation module should be followed to provide reliable transmission. To avoid choosing the concrete adaptive modulation and coding (AMC) strategy, the Shannon capacity—an upper bound of the transmission rate—is assumed for the EtC method to provide error-free transmission. The LE method and the PE method are originally designed for DL-based classification tasks. According to [29], the classification accuracy of ResNet-20 [46] based on the LE method and the PE method are 87.02% and 86.99% on the CIFAR-10 test dataset, respectively. The results show that the PE method and the LE method are compatible with DL. Since the DJSCC method is also a DL-based method, we combine the PE/LE method with the DJSCC method as the compared method.

Fig. 9 compares the reconstruction performance of the DJPSCC method with that of the aforementioned methods for the bandwidth ratio $R = 1/6$. The SSCC-based EtC method, the PE method combined with the DJSCC method,

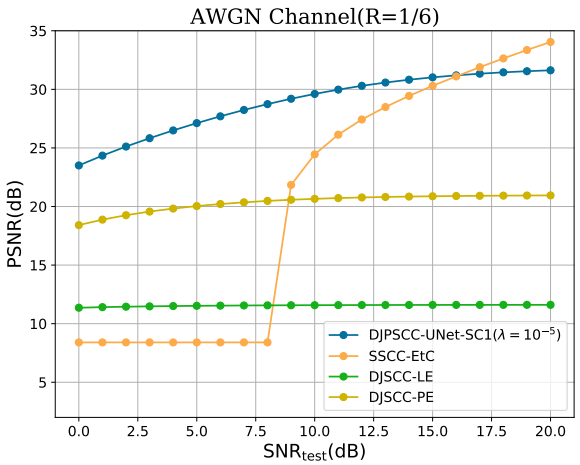


Fig. 9. Reconstruction performance of DJPSCC-UNet-SC1($\lambda = 10^{-5}$), SSCC-EtC, DJSCC-LE and DJSCC-PE on the CIFAR-10 test dataset.

and the LE method combined with the DJSCC method are named as SSCC-EtC, DJSCC-PE, and DJSCC-LE, respectively. The initial flat curve of SSCC-EtC, which is around 9.5 dB when $\mathrm{SNR}_{\mathrm{test}} \leq 8\mathrm{dB}$, is due to the fact that the EtC method completely breaks down in this SNR region, i.e., the minimum bit length of the EtC scheme is greater than the maximum transmission rate calculated by Shannon capacity. With the $\mathrm{SNR}_{\mathrm{test}}$ increases from 8 dB to 16 dB, the reconstruction performance of SSCC-EtC is still worse than that of DJPSCC-UNet-SC1($\lambda = 10^{-5}$), while the gap gradually decreases. When $\mathrm{SNR}_{\mathrm{test}} \geq 16\mathrm{dB}$, the reconstruction performance of SSCC-EtC is superior to that of DJPSCC-UNet-SC1($\lambda = 10^{-5}$). It is worth noting that DJPSCC-UNet-SC1($\lambda = 10^{-5}$) is with the first DJSCC architecture proposed in [4], the performance of which is inferior to that of the ADJSCC architecture proposed in [9]. By using an advanced DJSCC architecture, the DJPSCC may have a better result and be superior to the SSCC-EtC in high SNRs. The reconstruction performance of the DJSCC-LE is around 11.5 dB when $\mathrm{SNR}_{\mathrm{test}}$ is in the range from 0dB to 20dB, which is much lower than that of DJPSCC-UNet-SC1($\lambda = 10^{-5}$). The reconstruction performance of the DJSCC-PE is much better than that of the DJSCC-LE. However, it is still 5 dB lower than the reconstruction performance of DJPSCC-UNet-SC1($\lambda = 10^{-5}$) at $\mathrm{SNR}_{\mathrm{test}} = 0$ dB and the performance gap between the DJSCC-PE and DJPSCC-UNet-SC1($\lambda = 10^{-5}$) is further widened with increasing SNR.

Fig. 10 shows the corresponding visual performance for the aforementioned protection methods at $\mathrm{SNR} = 10$ dB. Based on the EtC encryption strategy, visually protected images from SSCC-EtC are grayscale-based images with $3 \times h \times w$ pixels. Since the LFBVS, the PSNR, and the SSIM require two images with the same shape, these evaluation scores are absent for the protected image and the decoded image of the SSCC-EtC. Visually protected images of the SSCC-EtC are shown as chaotic block images, while visually protected images of the DJSCC-LE and the DJSCC-PE are shown as noisy images. All protection methods can protect the visual content of the plain

Fig. 10. Visually protected images comparison generated by DJPSCC-UNet-SC1($\lambda = 10^{-5}$), SSCC-EtC, DJSCC-LE and DJSCC-PE with the bandwidth ratio R=1/6 and SNR=10 dB. The image in the first column is the plain image. The images in the second column are the protected images transformed by the image owner. The images in the third column are the decoded images and the images in the last column are the deprotected images. (a) DJPSCC-UNet-SC1($\lambda = 10^{-5}$), (b) SSCC-EtC, (c) DJSCC-LE, (d) DJSCC-PE.

image. However, when comparing the reconstruction quality of the deprotected images, DJPSCC-UNet-SC1($\lambda = 10^{-5}$) shows the best performance.

B. DJPSCC Generality on CIFAR-10 Dataset

We have mentioned that the DJPSCC framework is a general framework. In addition to the architecture of the UNet-SC1, other network architectures can be applied as the protection/deprotection network in the DJPSCC. Moreover, the feature extraction module can also adopt other network architectures instead of VGG16. To demonstrate the generality of the proposed method, we design new protection/deprotection architectures following the proposed principles of the protection/deprotection module and adopt a new feature extraction network for concrete DJPSCC networks to execute the following experiments.

The architecture of the dense block proposed in [47] uses multiple shortcut connections, satisfying the first principle about the shortcut connection. As shown in Fig. 11, we design a Dense Block-B based protection/deprotection network named as DenseNet-S1. ResNet[2] [46] pretrained on the ImageNet dataset is adopted as the feature extraction network. Besides the convolution neural network based architectures,

[2]https://keras.io/api/applications/resnet/#resnet50-function

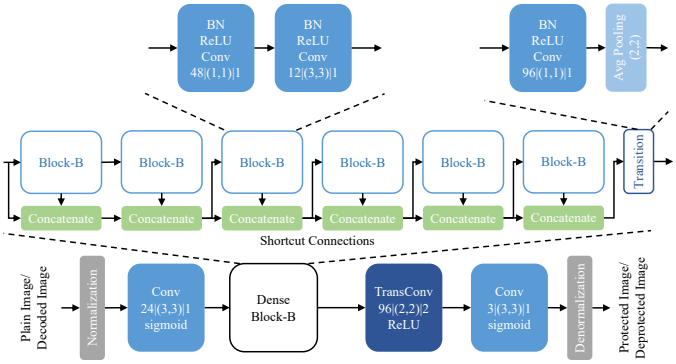


Fig. 11. A protection/deprotection architecture based on one dense block (DenseNet-S1).

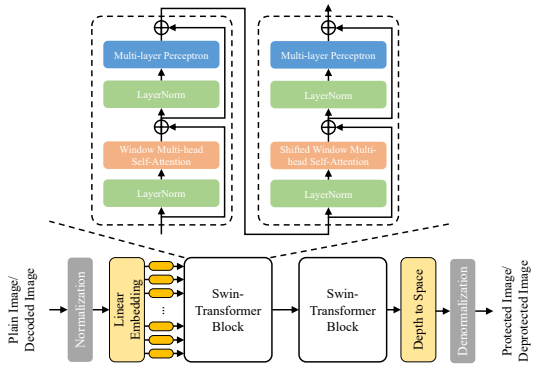


Fig. 12. A protection/deprotection architecture based on Swin-Transformer block (SwinTNet).

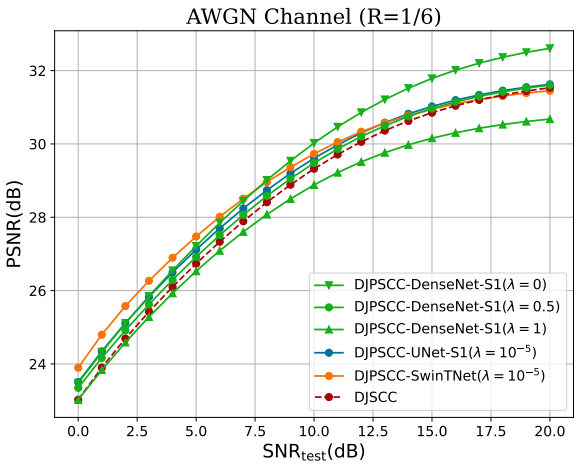


Fig. 13. Reconstruction performance of DJPSCC-DenseNet-S1 and DJPSCC-SwinTNet on CIFAR-10 test dataset with bandwidth ratio R=1/6.

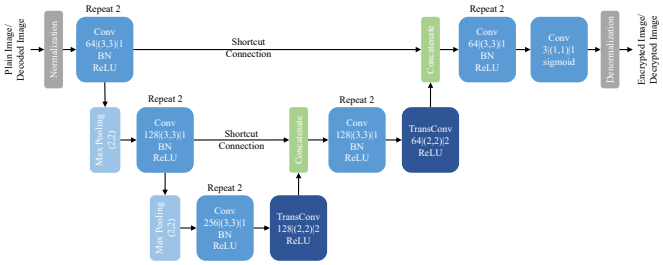


Fig. 14. A shallow version of U-Net with two shortcut connections (UNet-S2).

different backbones, e.g., transformer-based architectures, can also be employed to verify the generality of the proposed method. To compare the influence brought by different backbones, a Swin-Transformer based protection/deprotection network illustrated in Fig. 12 is also used as the backbone of protection/deprotection backbone. The training conditions of DJPSCC-DenseNet-S1 and DJPSCC-SwinTNet are the same as the conditions trained for DJSCC-UNet-S1.

Fig. 13 and Table II compare the reconstruction performance and the visual security performance of DJPSCC-DenseNet-S1 with the bandwidth ratio R=1/6, respectively. The reconstruction performance of DJPSCC-UNet-S1($\lambda = 10^{-5}$), DJPSCC-SwinTNet($\lambda = 10^{-5}$) and the DJSCC are plotted as reference in Fig. 13. Similarly to DJPSCC-UNet-S1, with an increase of λ, the visual security performance of DJPSCC-DenseNet-S1 improves while the reconstruction performance of DJPSCC-DenseNet-S1 degrades. The reconstruction performance of DJPSCC-DenseNet-S1($\lambda = 0.5$) is near that of DJPSCC-UNet-S1($\lambda = 10^{-5}$). Note that λ in DJPSCC-DenseNet-S1 is not of the same order of magnitude as λ in DJPSCC-UNet-S1 due to the different feature extraction networks chosen in the two DJPSCC methods. Fig. 13 also provides a comparison of the reconstruction performance brought by different backbones. For the same visual security level, the reconstruction performance of DJPSCC-SwinTNet is almost the same as that of DJPSCC-UNet-S1 when $\mathrm{SNR}_{\mathrm{test}} \geq 10\mathrm{dB}$. Although the reconstruction performance of DJPSCC-SwinTNet is better than that of DJPSCC-UNet-S1, the maximum reconstruction performance gap is barely less than 0.5 dB. The same results can be observed when comparing the reconstruction performance between DJPSCC-UNet-S1($\lambda = 10^{-5}$) and DJPSCC-DenseNet-S1($\lambda = 0.5$), if we consider that the feature extraction modules, i.e., VGG16 for DJPSCC-UNet-S1 with $\lambda = 10^{-5}$ and ResNet for DJPSCC-DenseNet-S1 with $\lambda = 0.5$, provide similar visual security levels. These observations may reflect the fact that the reconstruction performance of the proposed method is more dependent on the trade-off parameters λ than the concrete protection/deprotection network architectures.

C. DJPSCC Ablation Study on CIFAR-10 Dataset

To demonstrate the efficacy of these principles in the design of the protection/deprotection module, we construct three types of protection/deprotection architectures. As shown in Fig. 14, the first architecture named UNet-S2 is a deeper version of UNet-S1 with two shortcut connections. The second architecture is the UNet-S1 without the shortcut connection, which is called UNet-NS1. The UNet-S2 without shortcut connections named UNet-NS2 is the third architecture of the protection/deprotection network. Similarly, DenseNet-S2, DenseNet-NS1 and DenseNet-NS2 represent a deeper dense block based network with shortcut connections, a DenseNet-

TABLE II
VISUAL SECURITY EVALUATION OF DJPSCC-DENSENET-S1 ON THE CIFAR-10 TEST DATASET

Method	Protected Image				Decoded Image				Deprotected Image			
	LFBVS	COR	PSNR(dB)	SSIM	LFBVS	COR	PSNR(dB)	SSIM	LFBVS	COR	PSNR(dB)	SSIM
DJPSCC-DenseNet-S1 ($\lambda = 0$)	0.573	-0.004	10.514	0.024	0.592	0.048	8.829	0.041	**0.202**	**0.983**	**30.023**	**0.944**
DJPSCC-DenseNet-S1 ($\lambda = 0.5$)	0.701	-0.004	6.483	**0.006**	0.704	0.000	5.921	0.004	0.214	0.981	29.478	0.937
DJPSCC-DenseNet-S1 ($\lambda = 1$)	**0.702**	**-0.010**	**5.748**	0.041	**0.720**	**-0.005**	**5.598**	**0.002**	0.224	0.978	28.879	0.929

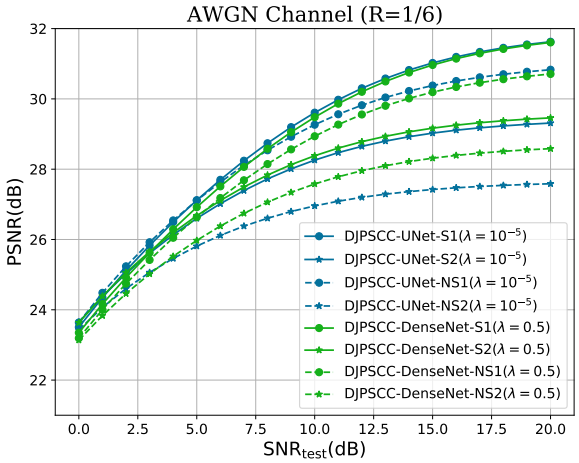


Fig. 15. Reconstruction performance of DJPSCC-UNet-S1, DJPSCC-UNet-S2, DJPSCC-UNet-NS1, DJPSCC-UNet-NS2, DJPSCC-DenseNet-S1, DJPSCC-DenseNet-S2, DJPSCC-DenseNet-NS1 and DJPSCC-DenseNet-NS2 evaluated on the CIFAR-10 test dataset with bandwidth ratio R=1/6.

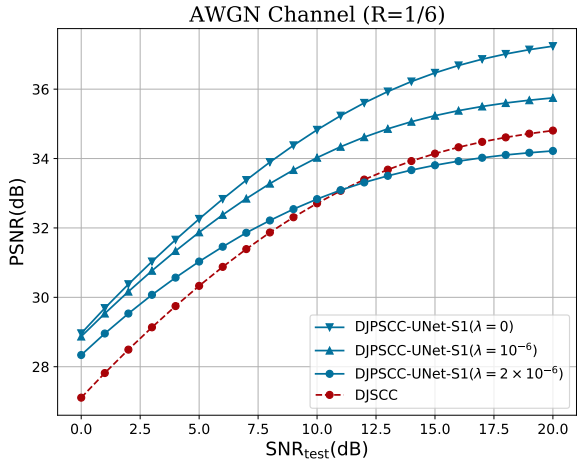


Fig. 16. Reconstruction performance of DJPSCC-UNet-S1 and DJSCC trained on Imagenet dataset and evaluated on Kodak dataset with R=1/6.

S1 without shortcut connections and a deeper network without short connections, respectively. The training conditions of the aforementioned networks are consistent with the training conditions of DJPSCC-UNet-S1 described in Section IV-A. In the training stage, VGG16 and ResNet are adopted as the feature extraction network for UNet-based networks and DenseNet-based networks, respectively. To acquire similar visual protection abilities for the aforementioned methods, the loss weight λ for UNet-based networks is set to 10^{-5} and that for DenseNet-based networks is set to 0.5.

Here, we only compare the reconstruction performance as shown in Fig. 15. DJPSCC-UNet-S1 performs better than DJPSCC-UNet-NS1 at high SNRs. This performance gain comes from the shortcut connection used in DJPSCC-UNet-S1. In conventional DL-based tasks, e.g., classification tasks and semantic segmentation tasks, the shortcut connection is widely used to improve the performance. In DJPSCC, the shortcut connection in the protection/deprotection architecture has the same utility. However, the performance of DJPSCC-UNet-S2 is inferior to that of DJPSCC-UNet-S1, which is in contrast to the intuition that increasing the depth of the network can improve performance [48]. Indeed, in conventional DL-based tasks, the shortcut connection can facilitate the training of deep networks and promote networks to learn more complex feature patterns to improve the performance. In the DJPSCC, there is a wireless channel naturally placed between the DJSCC encoder and the DJSCC decoder. It is impossible to create a shortcut connection across the wireless channel, which limits the efficiency of backward propagation in deeper networks during the training stage and causes performance degradation for deeper networks. In addition, the visual protection task and the reconstruction task executed by the proposed DJPSCC belong to low-level tasks. It is unnecessary to employ deeper protection/deprotection architectures to extract high-level semantic features. Similar results are revealed when comparing the DenseNet-based networks. Again, the shortcut connections in the protection/deprotection module can promote the reconstruction performance of the proposed DJPSCC, while deepening the protection/deprotection module brings an opposite result.

D. DJPSCC Validity on Imagenet Dataset

We have demonstrated the validity of the proposed method in a low-resolution image dataset (i.e., CIFAR-10 dataset) in Section IV-A. In this part, DJPSCC-UNet-S1 is trained on higher resolution image dataset (i.e.,ImageNet dataset) and evaluated on Kodak dataset. The Imagenet dataset consists of more than 1.2 million images. In the training stage, each image in the ImageNet dataset is resized to 128×128 and then fed into the proposed network. Adam optimizer with a learning rate of 10^{-4} and a batch size of 32 are used to train the proposed model. The training process is stopped when there is no improvement in the validation loss for five consecutive epochs. In [4], owing to the full convolutional architecture adopted by the DJSCC method, the Kodak dataset of size 512×768 can

TABLE III
VISUAL SECURITY EVALUATION OF DJPSCC-UNET-S1 ON THE KODAK DATASET

Method	Protected Image				Decoded Image				Deprotected Image			
	LFBVS	COR	PSNR(dB)	SSIM	LFBVS	COR	PSNR(dB)	SSIM	LFBVS	COR	PSNR(dB)	SSIM
DJPSCC-UNet-S1 ($\lambda = 0$)	0.593	**-0.041**	12.605	0.167	0.661	0.025	6.737	0.008	**0.068**	**0.994**	**34.836**	**0.946**
DJPSCC-UNet-S1 ($\lambda = 10^{-6}$)	0.766	-0.001	6.197	0.004	0.680	**-0.023**	**6.013**	**0.002**	0.072	0.993	34.035	0.939
DJPSCC-UNet-S1 ($\lambda = 2 \times 10^{-6}$)	**0.786**	-0.004	**5.412**	**0.003**	**0.720**	-0.008	6.303	0.004	0.083	0.993	32.820	0.936

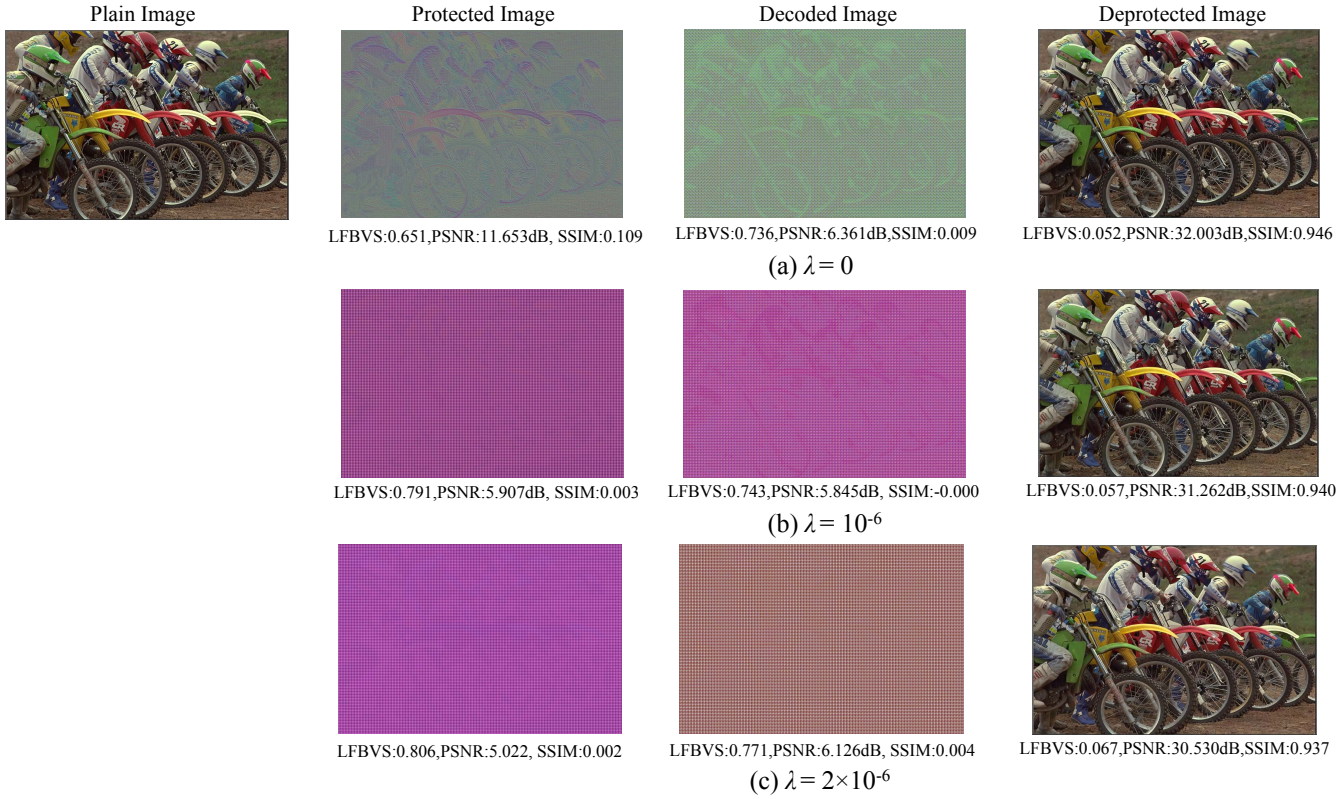


Fig. 17. Visually protected images generated by the proposed method at $\mathrm{SNR} = 10$ dB. The image in the first column is the plain image. The images in the second column are the protected images transformed by the image owner. The images in the third column are the decoded images decoded by the DJSCC decoder. The images in the last column are the deprotected images transformed by the image receiver. (a) $\lambda = 0$, (b) $\lambda = 1e^{-6}$, (c) $\lambda = 2 \times 10^{-6}$.

be directly fed into the DJSCC network and the reconstruction performance is acceptable. [9] further demonstrates the performance of the full convolutional architecture when the test dataset is consistent/inconsistent with the training data set. Due to the full convolution network architecture of the UNet-S1 adopted in the protection/deprotection network, DJPSCC-UNet-S1 architecture is a full convolutional network and can directly deal with the Kodak dataset.

The reconstruction performance, the visual security performance and the image visualization of DJPSCC-UNet-S1 with loss weights $\lambda = 0, 10^{-6}, 2{\times}10^{-6}$ at bandwidth ratio $R = 1/6$ are shown in Fig. 16, Table III and Fig. 17, respectively. Similarly to DJPSCC-UNet-S1 and DJPSCC-DenseNet-S1 trained on the CIFAR-10 dataset, with an increase of λ, the visual security performance of DJPSCC-UNet-S1 trained on the ImageNet dataset improves, while the reconstruction performance of DJPSCC-UNet-S1 trained on the ImageNet dataset degrades. As illustrated in Fig. 17, although DJPSCC-UNet-S1 ($\lambda = 0$) yields the best reconstruction performance, the outline of motorcycles and riders can be seen in the protected image and the decoded image. DJPSCC-UNet-S1($\lambda = 10^{-6}$) protects most of the visual information. However, the shadow of motorcycles and riders can be seen vaguely in the decoded image. DJPSCC-UNet-S1($\lambda = 2 \times 10^{-6}$) provides the best visual protection, since we can see a regular lattice only in the protected image and the decoded image. In real wireless scenarios, if high quality of the deprotected image is required, the transmit power or bandwidth usage can be increased to improve the quality of the deprotected image of DJSCC transmissions.

V. CONCLUSION

We have proposed a novel DJPSCC method to protect the privacy and confidentiality of information. Concretely, we have constructed a general framework including the protection module, the DJSCC module, and the deprotection module, redesigned the loss function by using the feature extraction module, and proposed two principles to guide the concrete de-

sign of the protection/deprotection module. During the training stage, the proposed DJPSCC method can successfully learn an effective protection network, an effective DJSCC network, and an effective deprotection network.

With increasing loss weight λ, the visual protection performance of the proposed DJPSCC network increases and the reconstruction performance decreases. Compared with the SSCC-based image protection method (e.g., the EtC method) and image protection methods for DL (e.g., the LE method and the PE method), the proposed DJPSCC method has shown much better reconstruction performance. The proposed DJPSCC method is a general method to protect the visual content of the plain image transmitted by DJSCC. Following the proposed design principles of the protection/deprotection module in Section IV, multiple protection/deprotection networks can be designed to meet the requirements in real communication scenarios, e.g., storage overhead and computational complexity. In addition, with some appropriate modifications to DJPSCC, the proposed DJPSCC method can be applied to various DJSCC architectures, e.g., DJSCC-l [7], DJSCC-f [8], and ADJSCC [9].

REFERENCES

[1] T. M. Cover, *Elements of Information Theory*. John Wiley & Sons, 1999.

[2] M. Skoglund, N. Phamdo, and F. Alajaji, "Hybrid digital–analog source–channel coding for bandwidth compression/expansion," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3757–3763, 2006.

[3] Y. Zhong, F. Alajaji, and L. L. Campbell, "Joint source–channel coding error exponent for discrete communication systems with markovian memory," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4457–4472, 2007.

[4] E. Bourtsoulatze, D. B. Kurka, and D. 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.

[5] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2434–2444, 2021.

[6] J. Xu, B. Ai, N. Wang, and W. Chen, "Deep joint source-channel coding for csi feedback: An end-to-end approach," *IEEE Journal on Selected Areas in Communications*, 2022, accepted.

[7] D. B. Kurka and D. Gündüz, "Bandwidth-agile image transmission with deep joint source-channel coding," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 8081–8095, 2021.

[8] ——, "Deepjscc-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, 2020.

[9] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2315–2328, 2022.

[10] M. Yang, C. Bian, and H.-S. Kim, "OFDM-guided deep joint source channel coding for wireless multipath fading channels," *IEEE Transactions on Cognitive Communications and Networking*, 2022.

[11] J. Xu, T.-Y. Tung, B. Ai, W. Chen, Y. Sun, and D. Gunduz, "Deep joint source-channel coding for semantic communications," *arXiv preprint arXiv:2211.08747*, 2022.

[12] C. Li, X. Guang, C. W. Tan, and R. W. Yeung, "Fundamental limits on a class of secure asymmetric multilevel diversity coding systems," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 4, pp. 737–747, 2018.

[13] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2326–2330.

[14] R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1111–1120, 1999.

[15] S. Wang, K. Yang, J. Dai, and K. Niu, "Distributed image transmission using deep joint source-channel coding," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5208–5212.

[16] V. Kostina, Y. Polyanskiy, and S. Verd, "Joint source-channel coding with feedback," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3502–3515, 2017.

[17] K. Choi, K. Tatwawadi, A. Grover, T. Weissman, and S. Ermon, "Neural joint source-channel coding," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1182–1192.

[18] Y. M. Saidutta, A. Abdi, and F. Fekri, "Joint source-channel coding for gaussian sources over AWGN channels using variational autoencoders," in *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 1327–1331.

[19] ——, "Joint source-channel coding of gaussian sources over AWGN channels via manifold variational autoencoders," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2019, pp. 514–520.

[20] R. Davis, "The data encryption standard in perspective," *IEEE Communications Society Magazine*, vol. 16, no. 6, pp. 5–9, 1978.

[21] S. Heron, "Advanced encryption standard (AES)," *Network Security*, vol. 2009, no. 12, pp. 8–12, 2009.

[22] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[23] L. Wu, J. Zhang, W. Deng, and D. He, "Arnold transformation algorithm and anti-arnold transformation algorithm," in *2009 First International Conference on Information Science and Engineering*, 2009, pp. 1164–1167.

[24] B. Acharya, S. K. Panigrahy, S. K. Patra, and G. Panda, "Image encryption using advanced hill cipher algorithm," *International Journal of Recent Trends in Engineering*, vol. 1, no. 1, pp. 663–667, 2009.

[25] A. Kanso and M. Ghebleh, "A novel image encryption algorithm based on a 3D chaotic map," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 7, pp. 2943–2959, 2012.

[26] J. Zhou, X. Liu, O. C. Au, and Y. Y. Tang, "Designing an efficient image encryption-then-compression system via prediction error clustering and random permutation," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 1, pp. 39–50, 2014.

[27] X. Kang, A. Peng, X. Xu, and X. Cao, "Performing scalable lossy compression on pixel encrypted images," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–6, 2013.

[28] T. Chuman, W. Sirichotedumrong, and H. Kiya, "Encryption-then-compression systems using grayscale-based image encryption for JPEG images," *IEEE Transactions on Information Forensics and security*, vol. 14, no. 6, pp. 1515–1525, 2019.

[29] H. Ito, Y. Kinoshita, M. Aprilpyone, and H. Kiya, "Image to perturbation: An image transformation network for generating visually protected images for privacy-preserving deep neural networks," *IEEE Access*, vol. 9, pp. 64 629–64 638, 2021.

[30] W. Sirichotedumrong and H. Kiya, "A GAN-based image transformation scheme for privacy-preserving deep neural networks," in *2020 28th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 745–749.

[31] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, "Subjective and objective quality assessment of image: A survey," *arXiv preprint arXiv:1406.7799*, 2014.

[32] Y. Mao and M. Wu, "A joint signal processing and cryptographic approach to multimedia encryption," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 2061–2075, 2006.

[33] L. Tong, F. Dai, Y. Zhang, and J. Li, "Visual security evaluation for video encryption," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 835–838.

[34] T. Xiang, Y. Yang, H. Liu, and S. Guo, "Visual security evaluation of perceptually encrypted images based on image importance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 4129–4142, 2020.

[35] W. Wen, K. Wei, Y. Fang, and Y. Zhang, "Visual quality assessment for perceptually encrypted light field images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2522–2534, 2021.

[36] G. Yue, C. Hou, K. Gu, T. Zhou, and H. Liu, "No-reference quality evaluator of transparently encrypted images," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2184–2194, 2019.

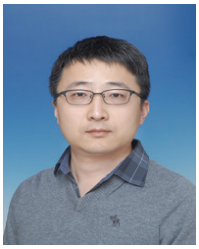
[37] Y. Yang, T. Xiang, H. Liu, and X. Liao, "Convolutional neural network for visual security evaluation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3293–3307, 2021.

[38] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.
[39] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.
[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[42] O. F. Mohammad, M. S. M. Rahim, S. R. M. Zeebaree, and F. Ahmed, "A survey and analysis of the image encryption methods," *International Journal of Applied Engineering Research*, vol. 12, no. 23, pp. 13 265–13 280, 2017.
[43] G.-h. Chen, C.-l. Yang, and S.-l. Xie, "Gradient-based structural similarity for image quality assessment," in *2006 International Conference on Image Processing*, 2006, pp. 2929–2932.
[44] M. Tanaka, "Learnable image encryption," in *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*. IEEE, 2018, pp. 1–2.
[45] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, and H. Kiya, "Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 674–678.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[47] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

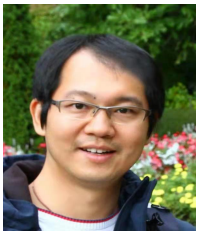
Jialong Xu (Member, IEEE) received the B.E. and M.S. degrees from Engineering University of PAP in 2009 and 2012 respectively. He received his Ph.D. degree from Beijing Jiaotong University in 2022. He joined DOCOMO Beijing Laboratories in 2023 and is now a researcher in the solution department. His research interests include deep learning, wireless coding, information theory, and Native AI for the physical layer.

Bo Ai (Fellow, IEEE) received the M.S. and Ph.D.degrees from Xidian University, Xian, China, in 2002 and 2004, respectively. He was with Tsinghua University, Beijing, China, where he was an Excellent Postdoctoral Research Fellow in 2007. He is currently a Professor and an Advisor of Ph.D.candidates with Beijing Jiaotong University, Beijing, where he is also the Deputy Director of the State Key Laboratory of Rail Traffic Control and Safety. He is also currently with the Engineering College, Armed Police Force, Xian. He has authored or coauthored six books and 270 scientific research papers, and holds 26 invention patents in his research areas. His interests include the research and applications of orthogonal frequency-division multiplexing techniques, high-power amplifier linearization techniques, radio propagation and channel modeling, global systems for mobile communications for railway systems, and long-term evolution for railway systems.

Dr. Ai is a Fellow of The Institution of Engineering and Technology. He was as a Co-chair or a Session/Track Chair for many international conferences such as the 9th International Heavy Haul Conference (2009); the 2011 IEEE International Conference on Intelligent Rail Transportation; HSRCom2011; the 2012 IEEE International Symposium on Consumer Electronics; the 2013 International Conference on Wireless, Mobile and Multimedia; IEEE Green HetNet 2013; and the IEEE 78th Vehicular Technology Conference (2014). He is an Associate Editor of IEEE TRANSACTIONS ON CONSUMER ELECTRONICS and an Editorial Committee Member of the Wireless Personal Communications journal. He has received many awards such as the Qiushi Outstanding Youth Award by HongKong Qiushi Foundation, the New Century Talents by the Chinese Ministry of Education, the Zhan Tianyou Railway Science and Technology Award by the Chinese Ministry of Railways, and the Science and Technology New Star by the Beijing Municipal Science and Technology Commission.



Wei Chen (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2006 and 2009, respectively, and the Ph.D. degree in computer science from the University of Cambridge, Cambridge, U.K., in 2013. He was a Research Associate with the Computer Laboratory, University of Cambridge from 2013 to 2016. He is currently a Professor with Beijing Jiaotong University, Beijing. His current research interests include sparse representation, Bayesian inference, wireless communication systems and image processing. He was the recipient of the 2013 IET Wireless Sensor Systems Premium Award and the 2017 International Conference on Computer Vision (ICCV) Young Researcher Award.



Ning Wang (Member, IEEE) received the B.E. degree in communication engineering from Tianjin University, China, in 2004, the M.A.Sc. degree in electrical engineering from The University of British Columbia, Canada, in 2010, and the Ph.D. degree in electrical engineering from the University of Victoria, Canada, in 2013. He was on the Finalist of the Governor General's Gold Medal for Outstanding Graduating Doctoral Student with the University of Victoria in 2013. From 2004 to 2008, he was with the China Information Technology Design and Consulting Institute as a Mobile Communication System Engineer, specializing in planning and design of commercial mobile communication networks, network traffic analysis, and radio network optimization. He was a Postdoctoral Research Fellow of the Department of Electrical and Computer Engineering with The University of British Columbia, from 2013 to 2015. Since 2015, he has been with the School of Information Engineering, Zhengzhou University, Zhengzhou, China, where he is currently an Associate Professor. He also holds adjunct appointments with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, and the Department of Electrical and Computer Engineering, University of Victoria, Victoria, Canada. He has served on the technical program committees of international conferences, including the IEEE GLOBECOM, IEEE ICC, IEEE WCNC, and CyberC. His research interests include resource allocation and security designs of future cellular networks, channel modeling for wireless communications, statistical signal processing, and cooperative wireless communications.

Miguel Rodrigues (Fellow, IEEE) received the Licenciatura degree in electrical and computer engineering from the University of Porto, Porto, Portugal, and the Ph.D. degree in electronic and electrical engineering from the University College London (UCL), London, U.K. He is currently a Professor of Information Theory and Processing, UCL, and a Turing Fellow with the Alan Turing Institute - the UK National Institute of Data Science and Artificial Intelligence. His research lies in the general areas of information theory, information processing, and machine learning. His work has led to more than 200 articles in leading journals and conferences in the field, a book on Information-Theoretic Methods in Data Science (Cambridge Univ. Press), and the IEEE Communications and Information Theory Societies Joint Paper Award 2011. He is an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, and the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. He was an Associate Editor for the IEEE COMMUNICATIONS LETTERS, and a Lead Guest Editor of the Special Issue on "Information-Theoretic Methods in Data Acquisition, Analysis, and Processing" of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING. He was a Co-Chair of the Technical Programme Committee of the IEEE Information Theory Workshop 2016, Cambridge, U.K. He is a member of the IEEE Signal Processing Society Technical Committee on "Signal Processing Theory and Methods", and the EURASIP SAT on Signal and Data Analytics for Machine Learning (SiG-DML).