IEEE Open Journal of Vehicular Technology

Received 20 September 2023; accepted 18 November 2023. Date of publication 22 November 2023;
date of current version 13 December 2023. The review of this article was coordinated by Editor Youngwook Ko.

Digital Object Identifier 10.1109/OJVT.2023.3335358

Source Separation in Joint Communication and Radar Systems Based on Unsupervised Variational Autoencoder

KHALED A. ALAGHBARI [1], HENG SIONG LIM [1] (Senior Member, IEEE), BENZHOU JIN [2] (Member, IEEE), AND YUTONG SHEN [2]

[1]Faculty of Engineering and Technology, Multimedia University, Melaka 75450, Malaysia
[2]College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

CORRESPONDING AUTHOR: HENG SIONG LIM (e-mail: hslim@mmu.edu.my)

This work was supported by Multimedia University (MMU), Malaysia.

ABSTRACT Source separation of a mixed signal in the time-frequency domain is critical for joint communication and radar (JCR) systems to achieve the required performance, especially at a low signal-to-noise ratio (SNR). In this paper, we propose the use of a generative model, such as the unsupervised variational autoencoder (VAE), to separate sensing and data communication signals. We first analyse the VAE system using different mask techniques; then, the best technique is selected for comparison with popular blind source separation (BSS) algorithms. We verify the performance of the proposed VAE by adopting different metrics such as the signal-to-distortion ratio (SDR), source-to-interference ratio (SIR), and sources-to-artifacts ratio (SAR). Simulation results show that the proposed VAE outperforms the BSS techniques at low SNR for the case of a mixed signal in the time-frequency domain and at low and high SNR for a mixed signal in the time domain. It enables the JCR system in the challenging first scenario to obtain SDR gains of 11.1 dB and 6 dB at 0 dB SNR for recovering the sensing and data communication signals respectively. Finally, we analyse the robustness of the JCR system in detecting an interference signal operating in the same frequency band, where the simulation result indicates an accuracy of 91% based on the proposed steps.

INDEX TERMS Joint communication and radar sensing (JCR), RadCom, linear frequency modulated (LFM), variational autoencoder (VAE), β-VAE, generative models, deep learning, blind source separation (BSS), FastICA, vehicular communications.

I. INTRODUCTION

Recently, there has been a lot of research and commercial interest in the integration of communication and radar systems (data communication and sensing signals) into a technology called joint communication and radar (JCR) or joint radar-communication (RadCom) system [1], [2]. The applications of JCR can be found in vehicular technology [3], security and military systems, unmanned aerial vehicle (UAV) communication and sensing [4], and future wireless networks [2]. Much attention has been paid to the efficient utilisation of spectral resources and the sharing of frequency bands in JCR systems [1]. The JCR system requires the sensing signal and data communication signal to operate simultaneously and effectively in the same frequency band without causing harmful interference to each other. In this situation, the signals from both systems are often mixed in the time and frequency domains. Hence, the main challenges are to mitigate the mutual interference and efficiently recover the sensing and communication signals.

Blind source separation (BSS) that relies on independent component analysis (ICA) is one approach to solve the issue mentioned previously. There are many popular BSS techniques, such as FastICA [5], FastICA's extension to non-circular complex sources (ExComplexFastICA) [6], complex ICA [7], [8], second-order blind identification (SOBI) [9], and joint approximate diagonalization of eigen-matrices (JADE) [10]. BSS has a wide range of applications for source separation in various fields, such as audio/speech signals, medical

signal processing, and wireless communication signals [11], [12]. In [13], BSS based on modified FastICA was proposed for a coexistence JCR system for the first time in order to blindly separate the sensing signal from the data communication signal at the receiver. However, BSS is very sensitive to signals with rapidly varying amplitude and is not sufficient in situations with complex overlapping subsignals in a highly noisy environment [3].

Autoencoder (AE) is a neural network architecture that aims to reconstruct the input data by learning a compressed representation of it. AE consists of an encoder network that maps the input to a lower-dimensional latent space and a decoder network that maps the latent space back to the original input space. It is commonly used for dimensionality reduction, image denoising, and anomaly detection [14]. Variational autoencoder (VAE) is a type of AE that learns a probabilistic distribution over the latent space using variational Bayesian inference, which enables it to generate new data points by sampling from the learned distribution. β-VAE is a powerful generative model of VAE that uses a hyper-parameter called beta to control the trade-off between reconstruction accuracy and disentanglement of the learned latent representation. By modifying beta, the model learns to disentangle factors of variation in the input data [15], [16].

Recently, β-VAE has been used for the disentanglement (separation) of mixed audio signals. In [17], the VAE framework was proposed for monaural audio source separation. Compared to baseline methods, the suggested framework resulted in reasonable improvements; however, the framework required prior knowledge of the sources in the mixture. The β-VAE with a weak class supervision method was proposed for audio source separation in [18]. The VAE was trained on a dataset of mixed audio signals and corresponding class labels to learn how to separate individual sources. The weak class supervision was used to encourage the VAE to learn meaningful representations of the audio sources, which can improve separation performance. The suggested method outperformed the baseline AE model in terms of the signal-to-distortion ratio (SDR), source-to-interference ratio (SIR), and sources-to-artifacts ratio (SAR). In [19], a hybrid method combining VAE and a bandpass filter was proposed, in which the bottleneck feature was filtered to capture only the frequency range of human speech. In [20], Neri et al. proposed a method for unsupervised BSS using β-VAE, which can learn to disentangle the underlying sources of a mixture signal without requiring any prior information about the sources. The β-VAE method was evaluated on handwritten digits and mixed audio spectrograms datasets and was shown to outperform existing state-of-the-art unsupervised source separation methods.

The implementation of VAE for source separation of mixed signals is limited to a few studies, as discussed previously. To the best of our knowledge, the usage of VAE in the JCR domain has not been considered yet. Our contributions in this paper can be summarized as follow:

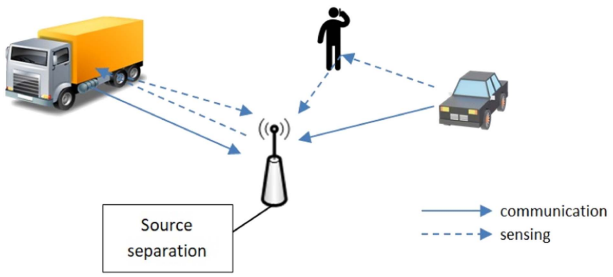

FIGURE 1. Illustration of JCR system with source separation base station.

- We propose the utilization of β-VAE for unsupervised separation of mixed signal consisting of radar sensing and data communication signals.
- An evaluation of two established and three novel masks is conducted, leading to the identification of the best techniques for recovering the sensing and data communication signals.
- The trained VAE model is further employed to detect interference signals coexisting within the same frequency band, relying on the VAE decoders' output and the designed mask.
- Extensive comparisons with prevalent BSS techniques are performed across two distinct scenarios. The first scenario involves signals mixed in both the time and frequency domains, while the second scenario pertains to signals mixed solely in the time domain but separable in the frequency domain. A range of performance metrics including mean square error (MSE), SDR, SIR, SAR [21], and scale-invariant signal-to-distortion ratio (Si-SDR) [21] are employed for system evaluation.

The structure of this article is organised as follows. Section II describes the methodology adopted in this paper. Specifically, it starts by presenting the source separation problem and proceeds with a discussion on the generation of radar and data communication signals. The theoretical background of VAE is then explained, followed by the mask techniques that can be implemented with the proposed VAE model. Simulation and results discussion are provided in Section III for three different scenarios, and a conclusion is provided in Section IV.

II. METHODOLOGY

Fig. 1 illustrates an example of a joint communication and radar (JCR) system with a source separation base station for two possible application scenarios. In the first scenario, the JCR station can receive both the communication signal from the vehicle, and the sensing signal (echoes) reflected from the detected target, as shown on the right side of Fig. 1. In the second scenario, the base station has the capability to transmit and receive the sensing signal while simultaneously receiving the communication signal from a vehicle, as depicted on the left side of Fig. 1.

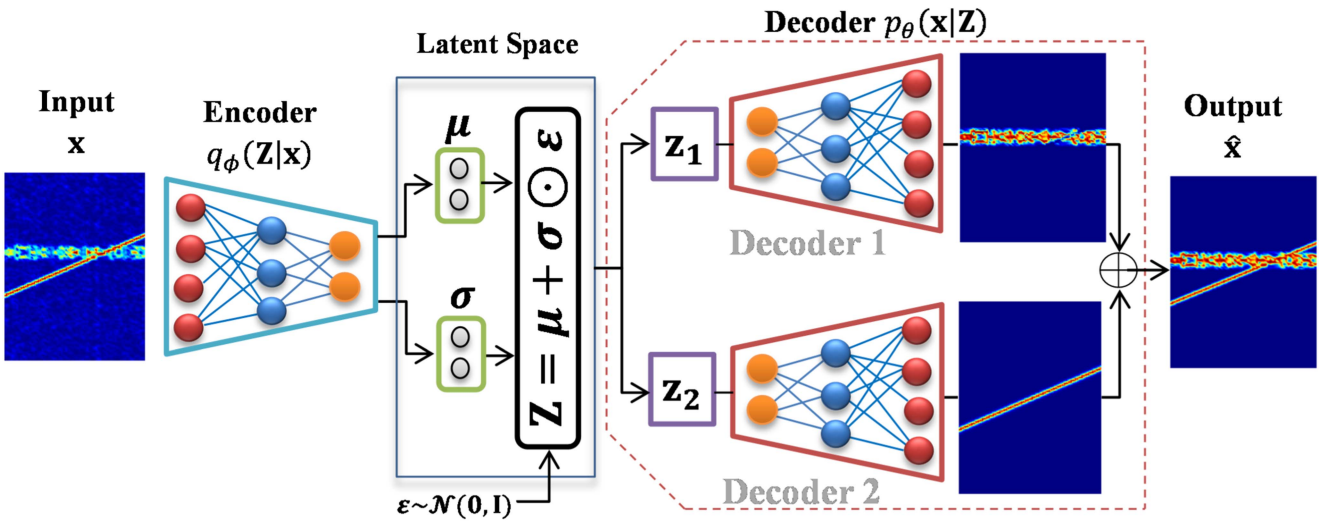

FIGURE 2. Structure of the variational autoencoder (VAE) used for the disentanglement of two signals.

A. SOURCE SEPARATION PROBLEM

We consider source separation of LFM and 4-ASK signals, where the received signal is given by:

$$x(t) = \sum_{k=1}^{K} s_k(t) + n(t) \tag{1}$$

where K is number of subsignals, and $s_k(t)$ is the kth subsignal. $n(t)$ is the additive white Gaussian noise (AWGN) with zero mean and variance σ^2. The channel fading effect is assumed to be estimated perfectly. The spectrogram of the mixed signals can be computed using short-time Fourier transform (STFT) such that:

$$X(t, f) = STFT(x(t)) \tag{2}$$

B. SIGNALS GENERATION

The linear frequency modulated (LFM) waveform is given by:

$$s_1(t) = A_1 \exp\left(j\left(\pi\ \mu\ (t - t_o)^2 + 2\pi f_{c_1} t\right)\right) \tag{3}$$

where A_1 is a constant representing the amplitude of LFM, μ is the frequency modulation slope, j is the imaginary unit, T_p is the time duration window, t_o is a constant, and f_{c_1} represents the carrier frequency.

The 4-ASK waveform is given by:

$$s_2(t) = a_2(t) \exp\left(\ j2\pi f_{c_2} t\right) \tag{4}$$

where $a_2(t) = A_2 a_{ASK}(t)\text{rect}(t/T_p)$, A_2 is a constant, $a_{ASK}(t)$ is the amplitude which contains four different levels, i.e., $\{-0.5, -1, 0.5, 1.5\}$, and f_{c_2} represents the carrier frequency of the 4-ASK waveform.

The Barker code (BC) waveform is given by:

$$s_3(t) = a_3(t) \exp\left(\ j2\pi f_{c_3} t\right) \tag{5}$$

where $a_3(t) = A_3 a_{BC}(t)\text{rect}(t/T_p)$, A_3 is a constant, $a_{BC}(t)$ is the Barker code with a length of 13 generated according to T_p, i.e., $\{+1, +1, +1, +1, +1, -1, -1, +1, +1, -1, +1, -1, +1\}$, and f_{c_3} represents the carrier frequency of the BC waveform. The BC waveform will be later employed as interference signal in the analysis section.

In the simulation, the time window T_p and sampling rate f_s are set to 75 μs and 30 MHz for all signals. For the LFM signal, $t_0 = T_p/2$, $\mu = BW/T_p$ and bandwidth $BW = 10$ MHz are used. The spectrum of the 4-ASK signal $a_{ASK}(t)$ is limited by a raised cosine filter with roll-off factor of 0.5, symbol span of 32 and upsampling factor of 15, while the BC signal $a_{BC}(t)$ is first filtered by a low-pass filter with a bandwidth of 2MHz.

C. VARIATIONAL AUTOENCODER (VAE)

AE is a type of artificial neural network that consists of three parts, namely encoder, latent space (bottleneck layer), and decoder; it is trained to learn the mapping of input data $\mathbf{x}$ to latent space $\mathbf{Z}$ and reproduce the input as output $\hat{\mathbf{x}}$ [16]. VAE is a probabilistic generative model and an extended version of AE, as shown in Fig. 2. VAE projects the data to a lower latent space and reformulates the input data's likelihood estimation as a variational inference problem [23]. The likelihood of data $\mathbf{x}$ given $\boldsymbol{Z}$ can be formulated based on the zero-mean Laplace distribution as [20]:

$$p_\theta(\mathbf{x}|\mathbf{Z}) = \prod_{i=1}^{D_\mathbf{x}} Lap(x_i|\hat{x}_i, b)$$

$$= \prod_{i=1}^{D_\mathbf{x}} \frac{1}{2b} \exp\left(-\frac{|x_i - \hat{x}_i|}{b}\right) \tag{6}$$

where $b = \sqrt{0.5}$ is the scaling factor for unit variance, $\mathbf{Z} = [\mathbf{z}_k]_{k=1}^{K}$ is the concatenation of the latent variables for sources $(k = 1, 2, \ldots, K)$ with the dimension of $D_\mathbf{Z} = D_\mathbf{z} K$, and $D_\mathbf{z} \ll D_\mathbf{x}$. $D_\mathbf{x}$ represents the dimensions (or the number of features) for the input data $\mathbf{x}$, and θ represents the parameters of the decoder (bias and weights) that are optimized during

the training stage. It is typical for VAEs to assume a Gaussian likelihood ($\ell 2$ loss), but small deviations around the mean and yields are allowed for blurry reconstructions. Meanwhile, the steep peak of the Laplace likelihood ($\ell 1$ loss) enables better reconstructions by equally penalising deviations around the mean, as suggested in [20].

The estimation of the mixed data is given by the sum of the output of the decoders as:

$$\hat{\mathbf{x}} = \sum_{k=1}^{K} \hat{\mathbf{s}}_k = \sum_{k=1}^{K} g_\theta\left(\mathbf{z}_k\right) \tag{7}$$

Over each source's latent variable, isotropic Gaussian priors are defined:

$$p\left(\mathbf{Z}\right) = \prod_{k=1}^{K} p\left(\mathbf{z}_k\right) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{z}_k|0, \mathbf{I}) \tag{8}$$

This prior assumes that each element varies independently and helps in the separation of variables in the data. Using estimates of latent sources $\mathbf{z}_k$ obtained by inferring $\mathbf{Z}$ from data $\mathbf{x}$, we can produce the source signal $\hat{\mathbf{s}}_k = g_\theta(\mathbf{z}_k)$ and thus achieve source separation.

In VAE, we cannot compute the true posterior distribution $p_\theta(\mathbf{Z}|\mathbf{x}) = p_\theta(\mathbf{x}|\mathbf{Z})p_\theta(\mathbf{Z})/p_\theta(\mathbf{x})$ directly because it is usually intractable or too complex to compute. Instead, we use the encoder network to learn the approximate posterior distribution $q_\phi(\mathbf{Z}|\mathbf{x})$ that is as close as possible to the true posterior distribution $p_\theta(\mathbf{Z}|\boldsymbol{x})$ with the help of the objective function. We can utilise variational inference to approximate the posterior distribution across the latent variables from the given data $\mathbf{x}$. The mean-field factorization of the approximate posterior q_ϕ ensures that the elements of $\mathbf{Z}$ are independent and Gaussian distributed:

$$q_\phi\left(\mathbf{Z}|\mathbf{x}\right) = \mathcal{N}(\mathbf{Z}|\mu_\phi\left(\mathbf{x}\right), \sigma_\phi^2\left(\mathbf{x}\right)\mathbf{I}) \tag{9}$$

where the mean $\boldsymbol{\mu} = \mu_\phi(\mathbf{x})$ and variance $\boldsymbol{\sigma}^2 = \sigma_\phi^2(\mathbf{x})$ are the outputs from the encoding neural network with parameters ϕ.

Variational inference transforms the approximate inference into an optimisation problem with the goal of maximising the variational lower bound (also known as the evidence lower bound (ELBO) function) given by [16], [20]:

$$\mathcal{L}\left(\boldsymbol{\theta}, \phi; \mathbf{X}\right) = \sum_{n=1}^{N} \mathcal{L}\left(\theta, \phi; \mathbf{x}^{(\mathrm{n})}\right) \tag{10}$$

$$\mathcal{L}\left(\theta, \phi; \mathbf{x}^{(\mathrm{n})}\right) = \left\langle \mathrm{In} p_\theta\left(\mathbf{x}^{(n)}|\mathbf{Z}\right)\right\rangle_{q_\phi\left(\mathbf{Z}|\mathbf{x}^{(n)}\right)} - D_{KL}\left(q_\phi\left(\mathbf{Z}|\mathbf{x}^{(n)}\right)||p\left(\mathbf{Z}\right)\right) \tag{11}$$

where dataset $\mathbf{X} = \{\mathbf{x}^{(n)}\}_{n=1}^{N}$ consists of N samples. The first term on the right side of (11) is the expected log-likelihood under the approximate posterior that attempts to minimize the reconstruction error based on $\ell 1$, as given in (6). The second term is the negative Kullback–Leibler divergence (KLD) between the approximate posterior and the prior that minimises the difference between the two distributions to ensure that the approximate posterior distribution $q_\phi(\mathbf{Z}|\mathbf{x})$ is close to the true posterior distribution $p_\theta(\mathbf{Z}|\mathbf{x})$ in terms of KLD. KLD provides regularisation; coupled with the stochastic sampling of the latent space, it is critical in promoting disentanglement (separation). β-VAE is a modification of the VAE structure that introduces an adjustable parameter (β) to the KLD of the standard VAE in (11). This architecture is introduced to discover disentangled latent factors without supervision, usually with $\beta > 1$; however, a higher β may create a trade-off between the reconstruction quality and the extent of disentanglement. When $\beta = 1$, the β-VAE is equivalent to the standard VAE, and when $\beta = 0$, it becomes equivalent to the basic AE [15], [16]. It is typical for VAE to assume a Gaussian approximate posterior since it allows for a simple Monte Carlo simulation of the expected log-likelihood using the reparametrization trick:

$$\mathbf{Z}^{(n)} \sim q_\phi\left(\mathbf{Z}|\mathbf{x}^{(n)}\right), \mathbf{Z}^{(n)} = \boldsymbol{\mu}^{(n)} + \boldsymbol{\sigma}^{(n)} \odot \epsilon \tag{12}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ and $\odot$ represent the element-wise products.

D. VAE MASKS

Masking involves the manipulation of components within the mixture spectrogram to isolate individual sources. This process allows for precise control over the emphasis or suppression of specific parts of the mixture, thereby effectively disentangling the signals. The selection of specific masking techniques is guided by several factors, including the technique's robustness to noise, particularly valuable in handling noisy environments; its aptitude for managing sources that overlap in either time-frequency or time domain, a common occurrence in real-time JCR signals; and its computational complexity, as certain methods may require more resources and processing time. The significance of masking lies in its ability to not only enhance separation performance, but also to mitigate background noise, ultimately elevating the overall quality of the separated signals.

After training the VAE, an inferred source $\hat{\mathbf{s}}_k$ can be transformed into a mask-based source signal estimate $\breve{\mathbf{s}}_k$ that captures the fine features from the data signal. We test two existing masks [18], [20] and propose three new masks.

Mask 1: Based on [20], the estimated source signal is given by:

$$\breve{\mathbf{s}}_k\left(t\right) = STFT^{-1}\left(\frac{D_k\left(t, f\right)}{\sum_{k=1}^{K} D_k\left(t, f\right)} \odot |X\left(t, f\right)|_N \odot e^{i\Phi}\right) \tag{13}$$

where $D_k(t, f)$ is the kth decoder spectrogram output, Φ is the phase of the mixture spectrogram, $\odot$ is the element-wise multiplication operator, and $|\cdot|_N$ is the normalized modulus.

Mask 2: Based on [18], the separated sources in the time domain can be estimated using the soft time-frequency mask

(Wiener filter) and mixture phase as:

$$\breve{\mathbf{s}}_k(t) = STFT^{-1}\left(\frac{D_k^2(t,f)}{\sum_{k=1}^{K} D_k^2(t,f)} \odot |X(t,f)|_N \odot e^{i\Phi}\right) \tag{14}$$

Mask 3: We can multiply masks 1 and 2 with the threshold matrix M_k to improve the performance at low SNR, where M_k is a matrix obtained by equalizing the corresponding $D_k(t,f)$ to zeros and ones using a threshold value; here, we propose to use it for mask 3, which is given as:

$$\breve{\mathbf{s}}_k(t) = STFT^{-1}\left(M_k \odot \frac{D_k^4(t,f)}{\sum_{k=1}^{K} D_k^4(t,f)} \odot |X(t,f)|_N \odot e^{i\Phi}\right) \tag{15}$$

The difference between this mask and previous masks is the power of 4 used for the output of the decoders.

Mask 4: Using the original complex-valued mixture spectrogram $X(t,f)$, we can use the following equations:

$$\breve{\mathbf{s}}_k(t) = STFT^{-1}\left(G_k \odot X(t,f)\right),$$

$$G_k = \frac{D_k(t,f)}{\sum_{k=1}^{K} D_k(t,f)} \odot |X(t,f)|_N \tag{16}$$

where G_k is a matrix obtained based on Mask 1 by equalizing the result to zeros and ones using a threshold value. The threshold can be used to remove unwanted noise, which will be discussed later. This mask works well if the two signals are only mixed in the time domain and not mixed in the frequency domain.

Mask 5: Using the decoder output as the magnitude for the reconstructed signal and the phase taken from the original mixture spectrogram, we can obtain the following equation:

$$\breve{\mathbf{s}}_k(t) = STFT^{-1}\left(D_k(t,f) \odot e^{i\Phi}\right) \tag{17}$$

This technique is introduced to check VAE's capability to reconstruct the signals without masking. It does not perform well with data communication signals, but it performs well and outperforms some of the other masks with sensing signals. This technique has a lower computational complexity compared to previous masks.

III. RESULTS AND DISCUSSION

A. VAE ARCHITECTURE

In the simulation of the proposed VAE, we used Python with Keras and Tensorflow libraries. The encoder has an input layer with a dimension of $D_{\mathbf{x}}$ and five hidden layers with 700, 600, 500, 400, and 300 fully-connected neurons, respectively. Each linear layer is followed by BatchNormalization and Relu activation function layers. Then, two fully-connected layers with linear activation functions for the mean and log-variance were created, as depicted in Fig. 2. Finally, the $D_{\mathbf{Z}}$–dimensional approximate posterior mean and log-variance of $\mathbf{Z}$ were the outputs from the sampling layer that performed the reparametrization trick as given in (12). In the simulation, we set $D_{\mathbf{Z}} = 40$, where $D_{\mathbf{z}} = 20$ and $K = 2$. The sampling layer was then separated into two and given to each decoder. The decoders have a reverse structure of the encoder, with each decoder starting with a dimension of $D_{\mathbf{z}}$ for the sampled latent source $\mathbf{z}_k$ and progressing in the reverse order. After the final hidden layer, a dense layer with Sigmoid (or Softplus) activation functionto output $D_{\mathbf{x}}$-dimension source signal $\hat{\mathbf{s}}_k$ for each decoder, the outputs of the decoders were finally added to form the expected mixture signal $\hat{\mathbf{x}}$. The Adam optimizer with a learning rate of 0.001 was used to optimize the neural networks' parameters. The ELBO objective function discussed earlier in (11) was used, and the β value of VAE was set to 10, which linearly increased over the first ten epochs to avoid early posterior collapse in training.

B. TRAINING STAGE

The training dataset consists of spectrograms that result from mixing the LFM signal with the 4-ASK signal at different frequencies, with an SNR of 25 dB. For instance, LFM signals (with frequencies of 0, 6, and 10 MHz) and 4-ASK signals (with frequencies of 2, 8, and 12 MHz) were generated to form nine different mixed signals. We first scaled the power of the LFM and 4-ASK signals and then used STFT to transform the mixed time-domain signal to the time-frequency domain. The mixed spectrograms were transformed to magnitude and normalized to the range between 0 and 1. The dataset consists of 1080 balanced samples. Each sample contains 8576 features formed by reshaping 128 frequency bins and 67 time frames.

C. TESTING STAGE

To verify the capability of the proposed VAE to separate sources, three different cases were considered. The trained VAE model was used to test the three cases. To evaluate the performance of VAE, we used different performance metrics such as MSE, SDR, SIR, SAR [21], and Si-SDR [22]. The MSE measure is sensitive to the amplitude and proper scaling is required for fair comparison. SDR measures the degree of distortion between the separated sources. SIR shows how effectively the sources are separated based on the residual interference between the sources. SAR identifies the artefacts caused by the separation technique. SDR is usually considered as an overall performance measure for any source separation approach [24]. We also compared the performance of the proposed VAE with popular blind source separation techniques such as ExComplexFastICA [6], ACMNsym [7], CQAMsym [8], SOBI [9], and JADE [10].

Case 1: Two signals are mixed in time and frequency domains. The time-domain waveform, frequency-domain waveform, and spectrogram of the mixed signal at an SNR of 10 dB are illustrated in Fig. 3. The bandwidth and centre frequency of the LFM signal are 10 MHz and 0 MHz, respectively. The centre frequency of the 4-ASK signal is 2 MHz.

Fig. 4 shows the normalized outputs of Decoder 1, Decoder 2, and VAE for the example shown in Fig. 3 at an SNR of

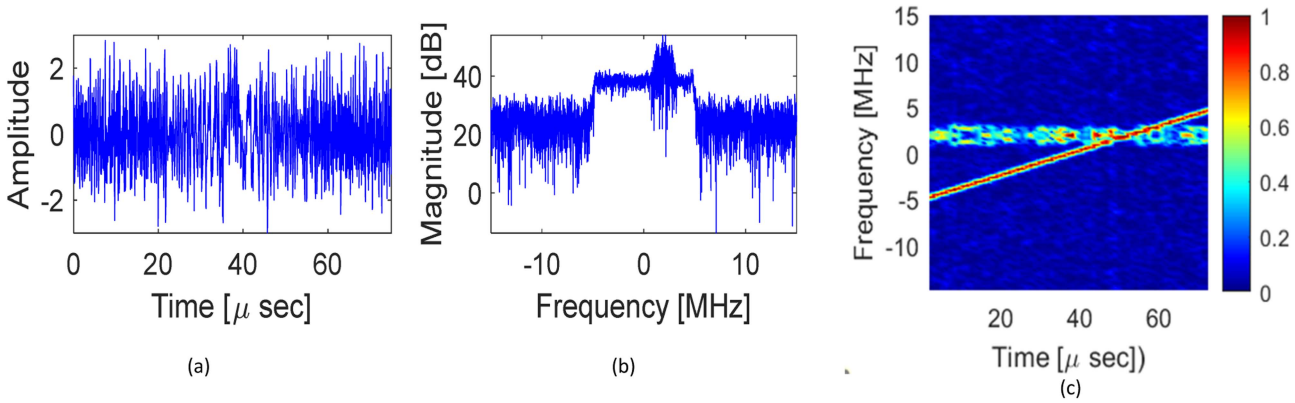

FIGURE 3. Received mixed signal for case 1: (a) time-domain waveform, (b) spectrum, and (c) corresponding spectrogram by STFT at an SNR of 10 dB.

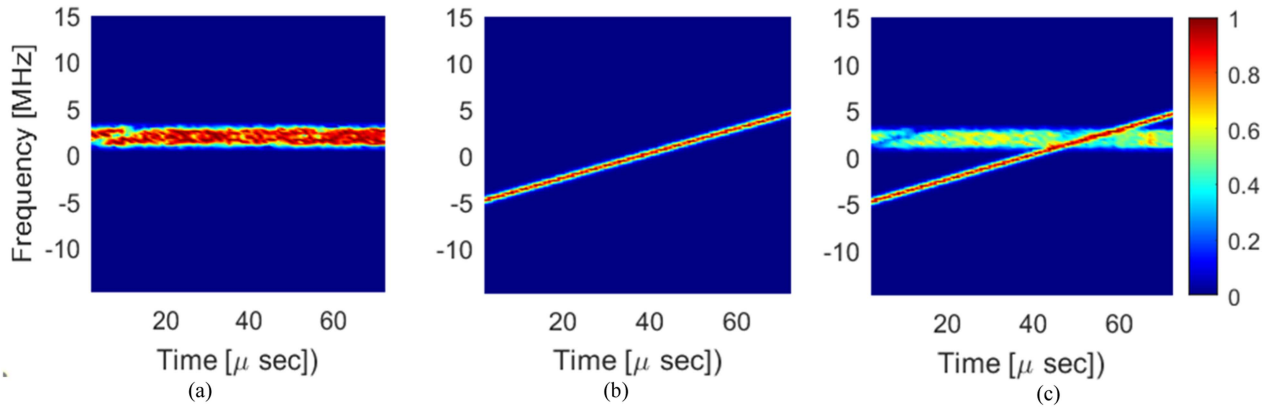

FIGURE 4. (a) Decoder 1 output, (b) decoder 2 output, and (c) VAE (decoder 1+decoder 2) output for the mixture input with 10 dB SNR for case 1.

10 dB. We can observe that the trained VAE model managed to separate the two signals and clean the noise effect. In other words, the VAE works well as a source separation and denoising model. The spectrogram of the 4-ASK signal shows deformation on the left side, though it should be on the right side. This error happened because the unsupervised VAE identified the magnitude spectrogram of another sample that was used during the training stage, such as the LFM and 4-ASK signals that had centre frequencies of 6 and 2 MHz, respectively. This issue can be solved by applying the mask technique as seen in Fig. 5. However, mask techniques may enhance the noise effects; thus, formulating a proper mask technique is important. The spectrogram of the LFM signal output by Decoder 2 shows a clean and smoothly separated source. VAE managed to separate the LFM signal more efficiently since it had a higher power, which is evident when combining the two spectrograms, as illustrated in Fig. 4(c).

Fig. 5 demonstrates the spectrograms of the recovered signal after applying Masks 1-5 for Decoder 1 (4-ASK signal) and Masks 2 and 5 for Decoder 2 (LFM signal). Table 1 shows the corresponding performance metrics. Masks 1-4 have correctly solved the issue at the output of Decoder 1 (deformation on the left side); however, the deformation on the right side has appeared, and all the masks have different performance metrics. By comparing all results, we can observe that Masks 1, 2, and 3 with no thresholds (Fig. 5(a)–(c)) are affected by noise; however, adding a threshold can significantly improve the spectrograms and the performance metrics, as revealed in Table 1. The best performance is achieved by the proposed Mask 3 with a threshold of 0.1, where it attained a 1 dB

TABLE 1 Performance Metrics for One Sample at an SNR of 10 dB in Figs. 3–5 (Case 1)

Signal	Mask	SDR [dB]	SIR [dB]	SAR [dB]	Si-SDR [dB]	MSE [dB]
4-ASK	Mask 1	8.5828	13.9990	10.2234	7.7911	-12.1833
	Mask 1 (0.1)	11.8981	16.4146	13.8903	10.4682	-14.8631
	Mask 2	8.1948	14.0499	9.6684	7.5652	-11.9124
	Mask 2 (0.1)	12.5690	17.3784	14.3896	11.2072	-15.5618
	Mask 3	7.5120	13.5922	8.9275	6.9836	-11.3172
	Mask 3 (0.1)	**12.8903**	**17.8198**	**14.6454**	**11.5542**	**-15.8971**
	Mask 4 (0.1)	6.7341	10.1377	9.7853	5.1732	-9.3096
	Mask 4 (0.4)	7.9183	11.9400	10.3781	6.0495	-10.7055
	Mask 5	7.1674	11.9675	9.1816	6.0723	-10.3559
LFM	Mask 5	13.7961	19.0927	15.3696	12.3907	-15.5340
	Mask 2	**16.1720**	**20.8699**	**18.0055**	**14.6303**	**-17.6445**

The bold values indicate the best performance.

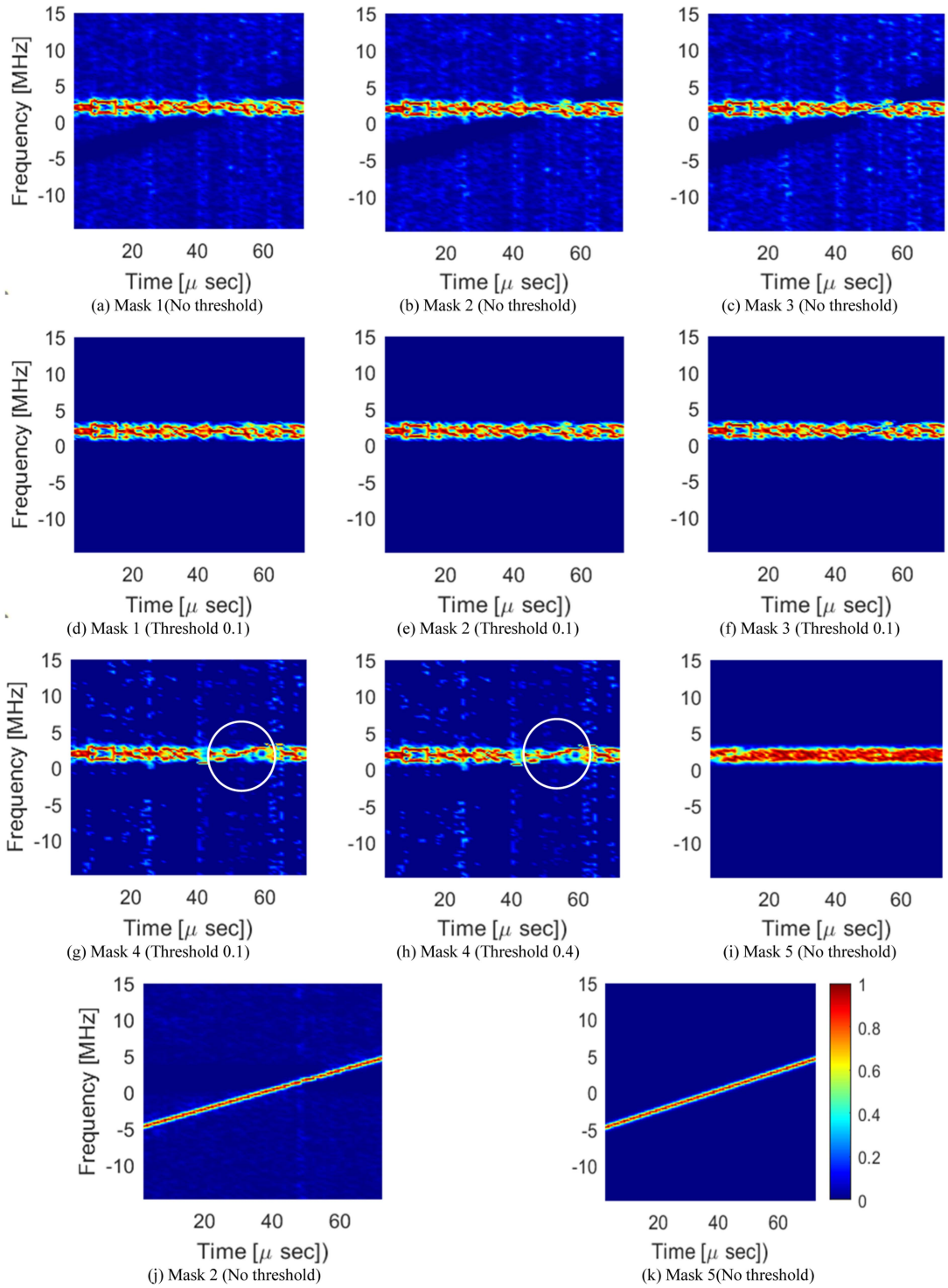

FIGURE 5. Spectrograms of the recovered signals after applying different masks, (a-i) 4-ASK signal, (j) (k) LFM signal.

improvement over Mask 1 (0.1) and about 0.3 dB over Mask 2 (0.1) in almost all metrics. Mask 4 and 5 do not seem to suit the data communication signal. Mask 4 is significantly affected by the overlapping in the frequency domain, and it introduced some energy leakage from the LFM source as highlighted in the circles for Fig. 5(g) and (h). Mask 5 used the direct output of Decoder 1 as the magnitude (in other words, no masking was applied). However, Mask 5 performs very well with the LFM sensing signal as it outperforms Mask 2 and provides ∼3 dB improvement in terms of SDR and SAR.

The recovered signals based on Mask 2, Mask 3 (0.1 threshold), and ExComplexFastICA are depicted in Fig. 6. It is observed that the signals recovered by VAE are affected in amplitude at the time duration between 45–60 microseconds

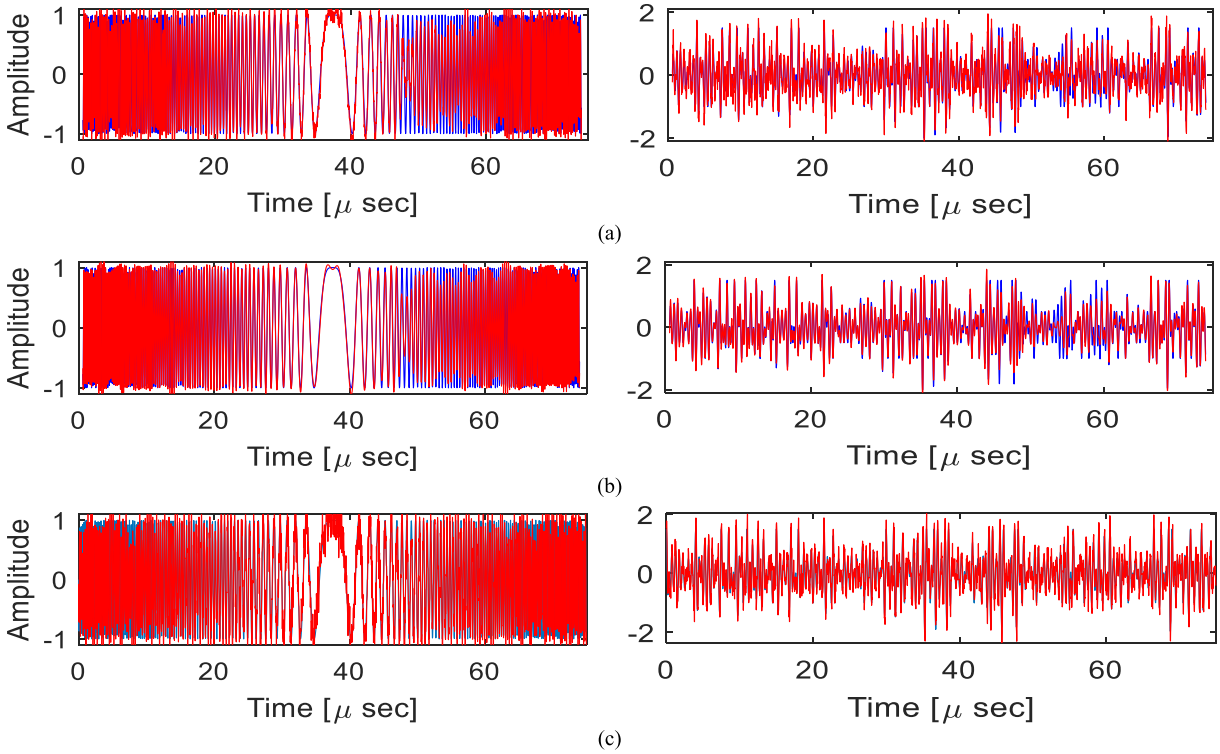

FIGURE 6. True LFM and 4-ASK signals (in blue colour) and recovered signals (in red colour) based on (a) mask 2, (b) mask 3 (0.1), and (c) ExComplexFastICA.

due to the signals overlapping in the time-frequency domain; these overlapping causes significant impacts on the performance metrics, as indicated in Table 1. The recovered signals by Mask 2 have slight noise effect, while Mask 3 (0.1) managed to clean the noise, as evidenced by the clarity of the LFM signal's midsection in both Mask 2 and Mask 3 (0.1) plots shown in Fig. 6. On the other hand, BSS recovered the LFM and 4-ASK signals but with a high noise effect that will slightly affect the performance metrics. The SDR, SIR, and SAR obtained by the ExComplexFastICA algorithm are 9.8512, 17.2250, and 10.8109 dB for the LFM signal and 11.2354, 21.7094, and 11.6731 dB for the 4-ASK signal, respectively. Comparing the performance of the two techniques, VAE achieves excellent performance for the LFM signal with a 6 dB gain in SDR, while ExComplexFastICA provides slightly better performance with a 1.6 dB gain in SDR for the 4-ASK signal.

Table 2 shows the averaged performance metrics of 100 samples for all discussed masks in recovering the LFM and 4-ASK signals. The results confirm that the proposed Mask 3 (0.1) outperforms the other masks for both 4-ASK and LFM signals. For the LFM signal, Mask 5's performance is also very close to Mask 3 (0.1) while requiring less computation complexity. Hence, Mask 3 (0.1) can be selected for a data communication signal and Mask 5 for a sensing signal.

We also evaluated the performance of the VAE with different masks at higher SNR, such as 25 dB; the results are given in Table 3. From Table 3, we can observe that Mask 3 (without threshold), followed by Mask 3 (with a threshold of 0.1) and Mask 5, outperform the other masks for the LFM signal, with only a slight difference between them. For the 4-ASK signal, Mask 3 (0.1), followed by Masks 3 and 2 outperform the other masks, where the difference between them is just fractions of dBs. The superiority of Mask 3 with a threshold is clearly observable at low SNR.

In Fig. 7, we compared the performance of the proposed VAE model to BSS techniques. Fig. 7(a) shows that the source

TABLE 2 Performance Metrics Averaged Over 100 Test Samples at SNR = 10 dB for Case 1

Signal	Mask	SDR [dB]	SIR [dB]	SAR [dB]	Si-SDR [dB]	MSE [dB]
LFM	Mask 1	12.1233	18.1619	13.4660	10.8847	-14.2298
	Mask 2	13.1289	19.1945	14.4699	11.8341	-15.0802
	Mask 3	13.6007	19.6591	14.9587	12.2636	-15.4252
	Mask 3 (0.1)	**16.5304**	**21.5115**	18.2585	**15.0390**	**-18.2054**
	Mask 4 (0.1)	12.4488	16.8997	14.5014	10.7941	-13.7740
	Mask 4 (0.4)	14.1716	18.8843	16.0532	12.5901	-15.7636
	Mask 5	16.4483	21.2385	**18.3121**	14.8358	-18.0861
4-ASK	Mask 1	8.5795	14.5656	10.0442	7.6325	-11.9202
	Mask 2	8.4650	14.8903	9.7991	7.5596	-11.8107
	Mask 3	8.2354	14.8736	9.5217	7.3467	-11.5969
	Mask 3 (0.1)	**12.1754**	**17.0848**	**13.9979**	**11.0041**	**-15.1586**
	Mask 4 (0.1)	6.7407	10.4531	9.5437	5.4017	-9.2469
	Mask 4 (0.4)	8.0913	12.0890	10.5809	6.6247	-10.9428
	Mask 5	7.4329	12.4321	9.3587	6.0461	-10.1972

The bold values indicate the best performance.

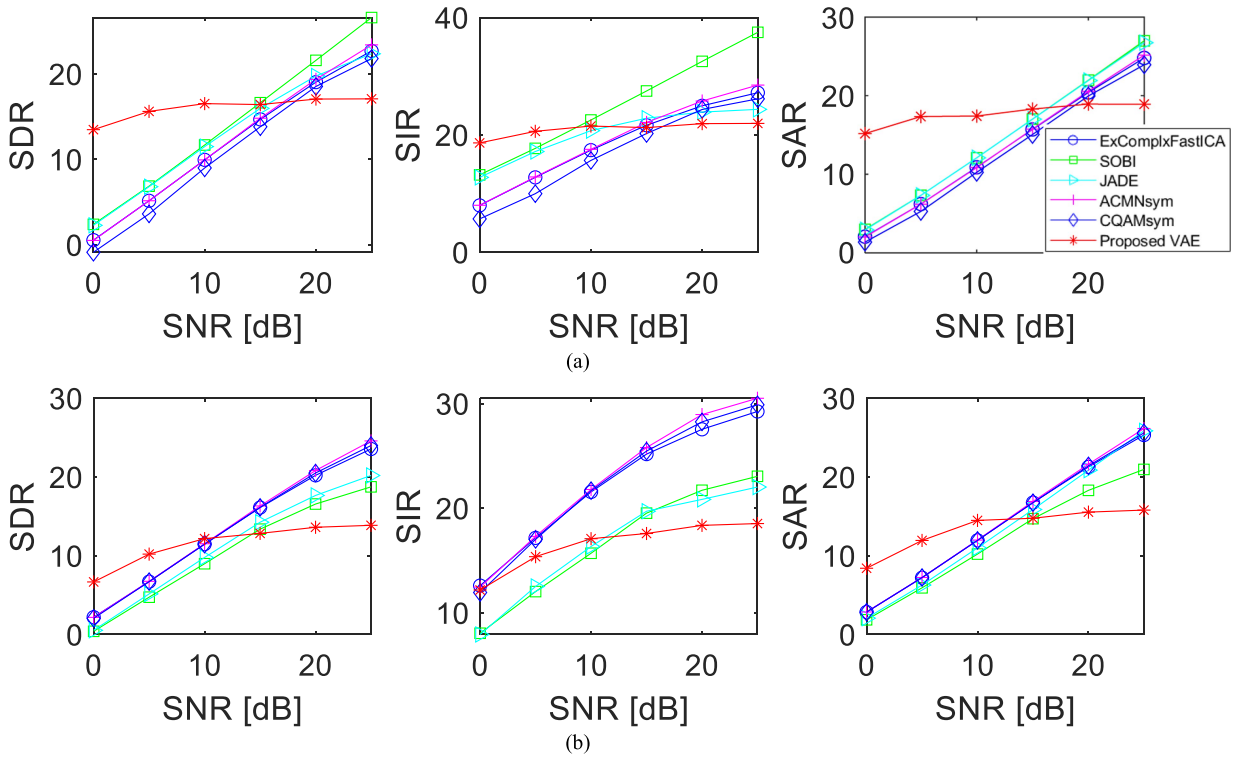

FIGURE 7. Performance metrics averaged over 100 test samples based on case 1 for the (a) LFM signal and (b) 4-ASK signal.

TABLE 3 Performance Metrics Averaged Over 100 Test Samples at SNR = 25 dB for Case 1

Signal	Mask	SDR [dB]	SIR [dB]	SAR [dB]	Si-SDR [dB]	MSE [dB]
LFM	Mask 1	14.7485	19.6911	16.4943	13.4393	-16.6153
	Mask 2	16.2366	21.2298	17.9471	14.8787	-18.0344
	Mask 3	**17.1368**	**22.1340**	18.8426	**15.7398**	-18.8605
	Mask 3 (0.1)	17.0884	21.9628	18.9195	15.6064	**-18.9263**
	Mask 4 (0.1)	12.8805	17.1787	15.0161	11.3604	-14.3661
	Mask 4 (0.4)	14.5820	19.2628	16.4696	13.1500	-16.2599
	Mask 5	17.0910	21.6979	**19.0605**	15.5050	-18.7677
4-ASK	Mask 1	12.6227	16.9123	14.7819	11.4917	-15.5923
	Mask 2	13.3775	17.9955	15.3418	12.2757	-16.3270
	Mask 3	13.6378	18.4270	15.5101	12.5511	-16.5893
	Mask 3 (0.1)	**13.8626**	**18.5334**	**15.7958**	**12.7223**	**-16.7584**
	Mask 4 (0.1)	7.4652	10.6690	10.6602	6.1029	-9.9484
	Mask 4 (0.4)	8.3415	12.1616	10.9511	6.8893	-11.1618
	Mask 5	7.8265	12.8164	9.7472	6.4984	-10.4818

The bold values indicate the best performance.

separation of the LFM signal based on the proposed VAE outperforms the BSS techniques in the noise-limited situation. At low SNR (<15 dB), the proposed VAE outperforms the BSS techniques; for instance, at SNR = 0 dB, the SDR and SAR achieved by VAE are 13.5 and 15 dB, which are significantly higher than the SDR and SAR of 2.4 and 3dB obtained by the SOBI method. This result represents a difference of 11.1 and 12 dB in SDR and SAR between the two methods. However, at higher SNR, the VAE's performance became saturated while the BSS techniques performed better. The reason for this saturation is because at high SNR, the interaction of the two signals increased, and the overlapping in the frequency domain became critical due to the received mixed signal containing both strong and weak signals, consequently causing power leakage. Fig. 7(b) demonstrates the source separation of the 4-ASK signal. We can observe that the source separation based on VAE outperforms the BSS techniques in terms of SDR and SAR only at an SNR of less than 10 dB; however, for a higher SNR, the BSS techniques perform much better. It is worth noting that VAE can extract the signal component at an acceptable quality even when the SNR is low. For instance, at SNR = 0 dB, the SDR and SAR for VAE are 6.5 and 8.4 dB compared to 0.4 and 1.9 dB for the SOBI method, indicating that VAE outperforms the SOBI method in this scenario by a gain of more than 6 dB.

Case 2: The two subsignals are mixed in the time domain but are separable in the frequency domain. The time-domain waveform, frequency-domain waveform, and spectrogram of the mixed signal at an SNR of 10 dB are illustrated in Fig. 8. The bandwidth and centre frequency of the LFM signal are 10MHz and 0MHz, respectively. The centre frequency of the 4-ASK signal is 12 MHz.

The output of Decoder 1 when the input to VAE was the mixed signal provided in Fig. 8(c) is shown in Fig. 9(a). It can be observed from the output of Decoder 1 that it is able to separate and denoise the 4-ASK signal. Examples of the spectrograms obtained after applying Mask 3 without a threshold and with a threshold of 0.1 are shown in Fig. 9(b)–(c).

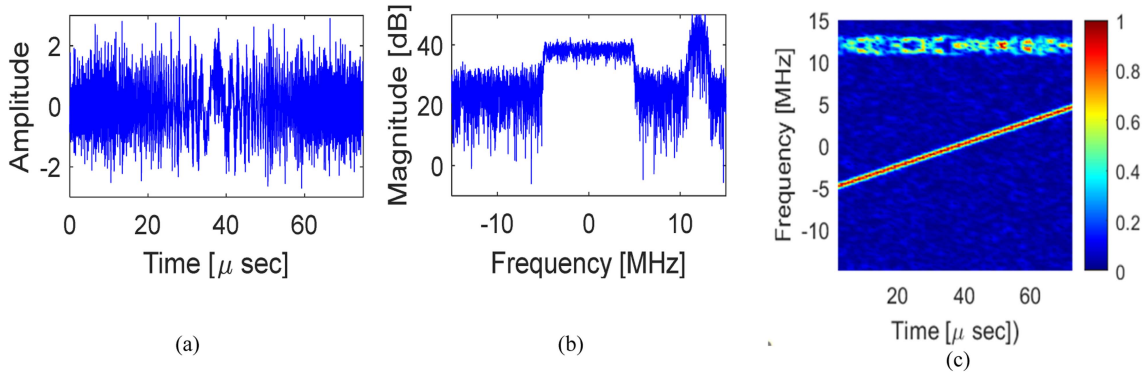

FIGURE 8. Received mixed signal for case 2: (a) time-domain waveform, (b) spectrum, and (c) corresponding spectrogram by STFT at an SNR of 10 dB.

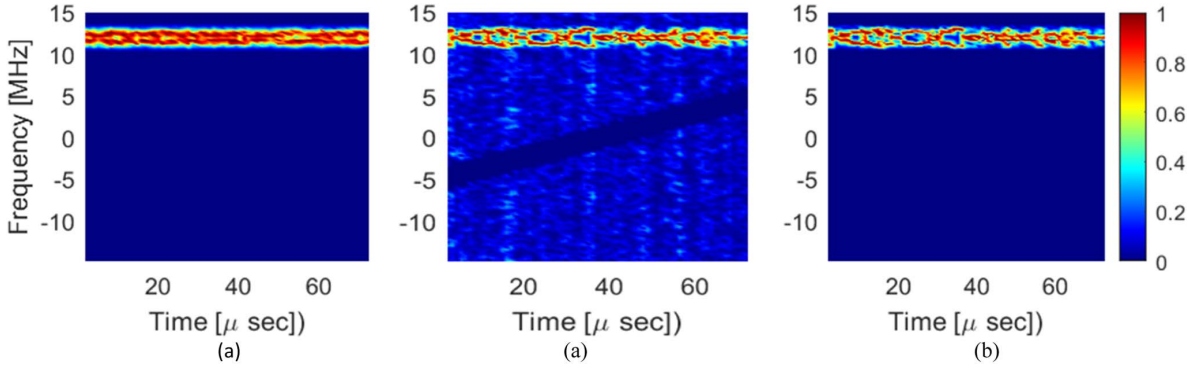

FIGURE 9. Spectrograms for (a) decoder 1 output and the recovered signals after applying, (b) mask 3, and (c) mask 3 (0.1).

TABLE 4 Performance Metrics Averaged Over 100 Test Samples at SNR = 10 dB for Case 2

Signal	Mask	SDR [dB]	SIR [dB]	SAR [dB]	Si-SDR [dB]	MSE [dB]
LFM	Mask 1	19.8855	28.5031	20.6733	18.3509	-21.0520
	Mask 2	20.1508	28.7585	20.9507	18.5327	-21.2175
	Mask 3	20.1266	28.7282	20.9272	18.5018	-21.1876
	Mask 3 (0.1)	24.3020	32.9963	25.2239	22.8150	-25.4657
	Mask 4 (0.1)	22.8554	31.9198	23.7291	21.2502	-24.2597
	Mask 4 (0.4)	21.1297	29.5595	22.1364	19.8390	-22.0170
	Mask 5	**25.4559**	**34.2663**	**26.3139**	**23.8212**	**-26.4960**
4-ASK	Mask 1	8.8874	16.8261	9.7646	8.1891	-11.9496
	Mask 2	8.3004	16.2526	9.1873	7.5989	-11.3594
	Mask 3	8.1703	16.1090	9.0627	7.4674	-11.2283
	Mask 3 (0.1)	**17.3452**	**24.4324**	**18.4402**	**15.6348**	**-20.1304**
	Mask 4 (0.1)	11.8438	19.6808	12.7016	11.1325	-15.0466
	Mask 4 (0.4)	15.0607	21.5265	16.2755	13.4800	-17.5045
	Mask 5	9.0584	15.5838	10.4376	7.7249	-11.6165

The bold values indicate the best performance.

Mask 3 without a threshold enhances the noise effect, while Mask 3 with a threshold managed to remove the noise effect efficiently.

Table 4 presents the averaged performance metrics of 100 samples for different masks at an SNR of 10 dB. It can be seen that the performance in this scenario is much better than the results obtained for Case 1, which is presented earlier in Table 2, especially for the LFM signal. Masks 1, 2, and 3 did not achieve a good improvement for the 4-ASK signal compared to the results obtained in Table 2. However, the proposed Mask 5 and Mask 3 (0.1) outperforms the other masks for the LFM and 4-ASK signals, respectively. Moreover, Mask 4 (0.4) achieves good performance metrics compared to the results obtained in Case 1 (Table 2).

Fig. 10 displays the plots comparing the average performance metrics between the proposed VAE and BSS algorithms. The performance metrics for the LFM signal are better than the 4-ASK signal because the instantaneous amplitude of the 4-ASK signal varied rapidly. The results reveal significant improvement in the performance of the proposed VAE compared to the BSS techniques in the scenario where the two signals were only mixed in the time domain. For instance, at SNR of 0 and 10 dB, the VAE provides SDR gains of 13.6 and 12.6 dB, respectively, over the SOBI method for the LFM signal and gains of 6.4 and 7.3 dB, respectively, for the 4-ASK signal.

Examples of recovered waveforms based on Mask 3 (0.1) and ExComplexFastICA at an SNR of 10 dB are depicted in Fig. 11. The two algorithms successfully reconstructed the input signals; however, the VAE with Mask 3 (0.1) shows excellent performance compared to the BSS technique due to its capability to remove the noise effects and produce a smooth

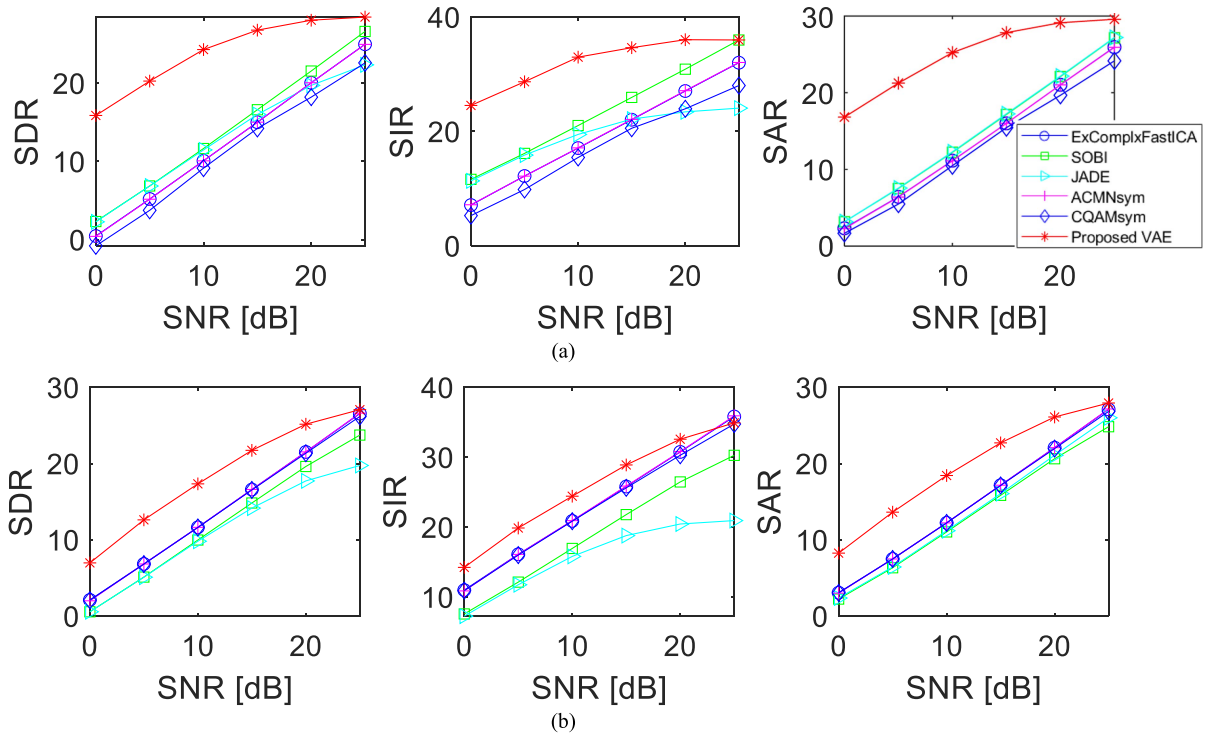

FIGURE 10. Performance metrics averaged over 100 test samples based on case 2 for the (a) LFM signal and (b) 4-ASK signal.

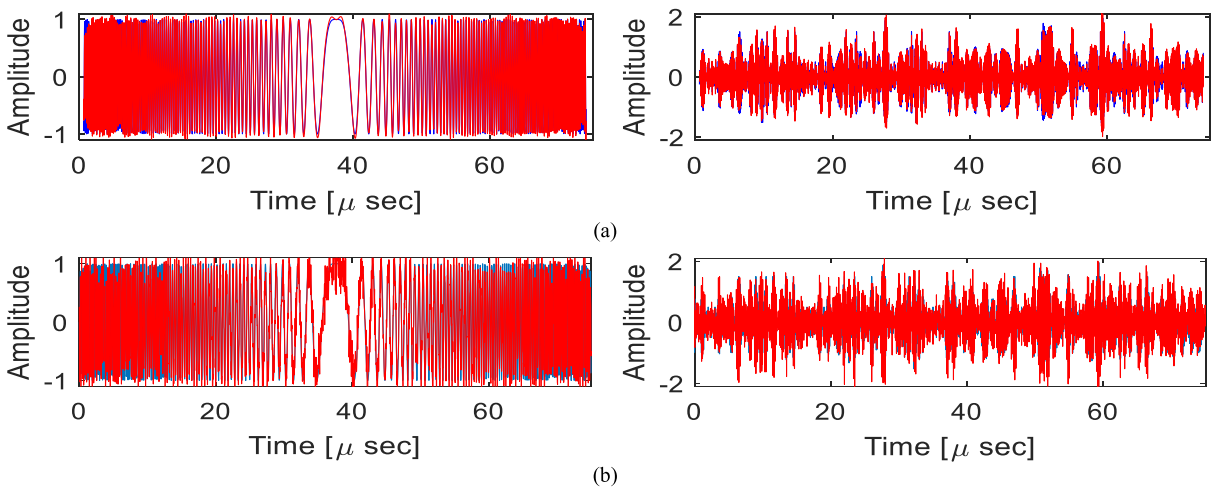

FIGURE 11. The true LFM and 4-ASK signals (in blue colour) and recovered signals (in red colour) based on (a) mask 3 (0.1) and (b) ExComplexFastICA.

signal, as illustrated in Fig. 11(a). The SDRs for the recovered LFM and 4-ASK signals are 24.6 dB and 16.7 dB, respectively, based on the proposed VAE (Mask 3 (0.1)) and 10.5 dB and 11.6 dB, respectively, based on the ExComplexFastICA technique. In other words, in this scenario, the proposed VAE achieved gains of 14 dB and 5 dB in recovering the sensing and data communication signals, respectively.

Case 3: The two subsignals are mixed in both time and frequency domains, in addition to the presence of an interference signal represented by the Barker code (BC) waveform operating in the same frequency band. The time-domain waveform, frequency-domain waveform, and spectrogram of the mixed signal at an SNR of 10 dB are illustrated in Fig. 12. The bandwidth and centre frequency of the LFM signal are 10 MHz and 0 MHz, respectively. The centre frequency of the 4-ASK and BC signals are 2 MHz and 12 MHz, respectively.

To estimate the interference signal, the following steps are proposed:

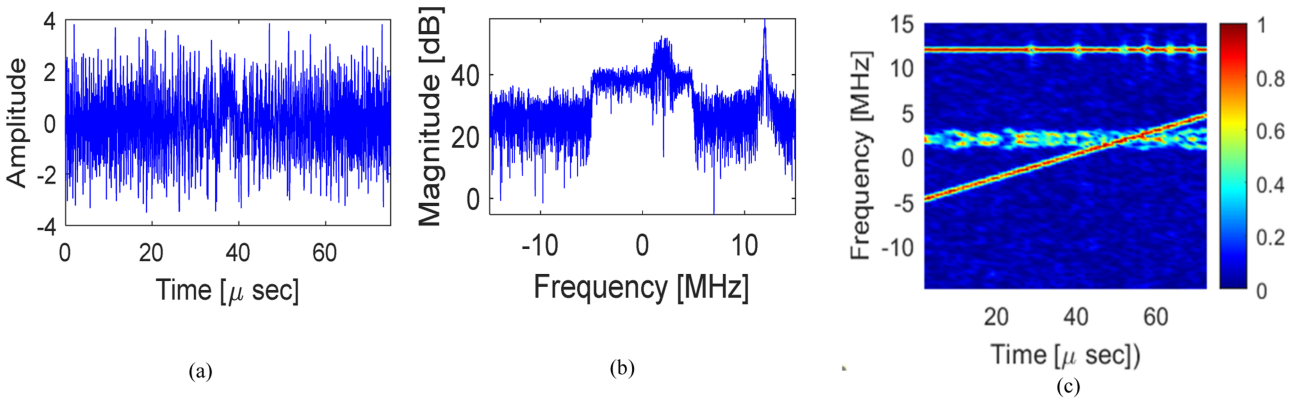

FIGURE 12. Received mixed signal for case 3: (a) time-domain waveform, (b) spectrum, and (c) corresponding spectrogram by STFT at an SNR of 10 dB.

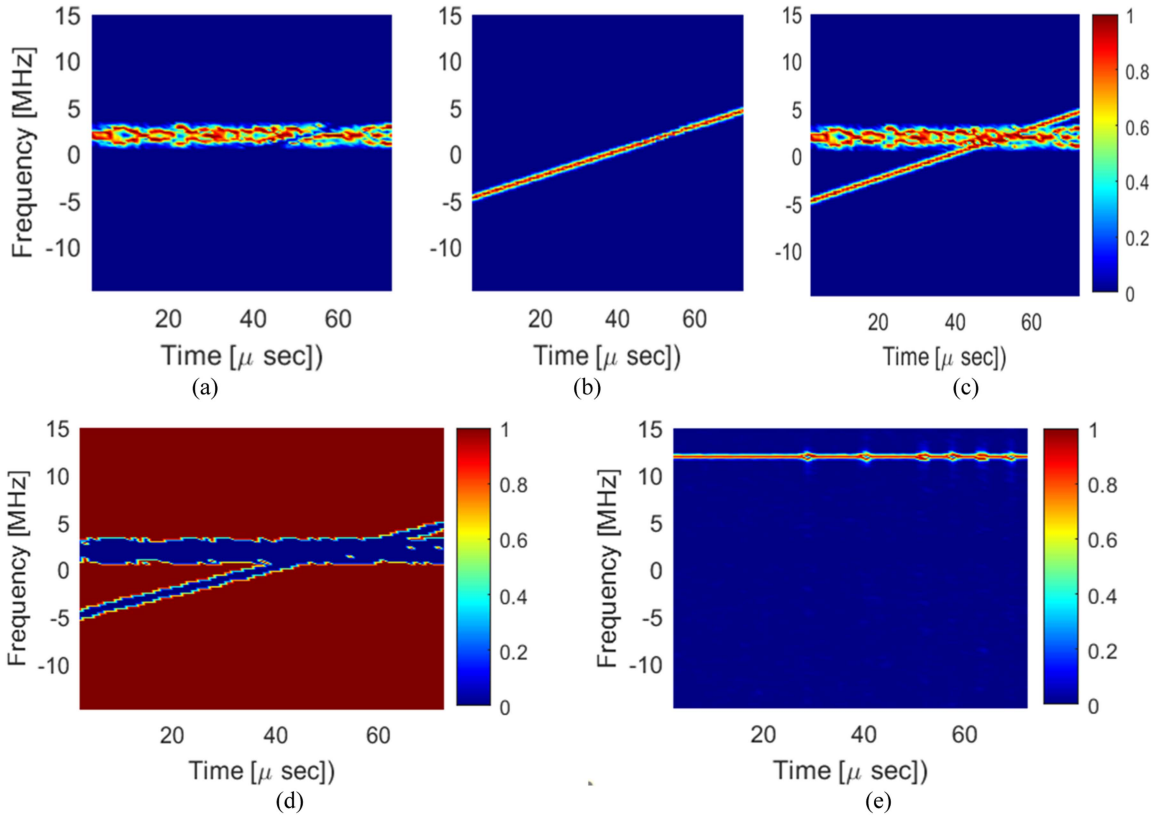

FIGURE 13. Steps used to estimate the interference signal: (a) decoder 1 output based on mask 3 (0.1), (b) decoder 2 output based on Mask 3 (0.1), (c) the normalized VAE output (decoder 1 + decoder 2 outputs), (d) the created mask, and (e) the estimated spectrogram of the interference signal.

Step 1: Use the trained VAE model to separate the 4-ASK and LFM signals (as shown in Fig. 13*(a), (b)).*

Step 2: Combine the outputs of Decoder 1 and Decoder 2, and normalise them to obtain the spectrogram magnitude as shown in Fig. 13*(c).*

Step 3: Create a mask I_k by assigning ones to the spectrogram magnitude (obtained in step 2) for values that are less than a certain threshold value (i.e., 0.1) and zeros elsewhere, as shown in Fig. 13*(d).*

Step 4: To estimate the interference signal, apply the following equation:

$$\breve{s}_k(t) = STFT^{-1}\left(I_k \odot |X(t,f)|_N \odot X(t,f)\right)$$

Fig. 13 illustrates the spectrograms obtained by following the proposed steps to detect the interference signal. Fig. 13(e) displays the spectrogram of the estimated interference signal, where it is clearly shown that the designed mask in Fig. 13(d) has removed the desired signals (LFM and 4-ASK signals),

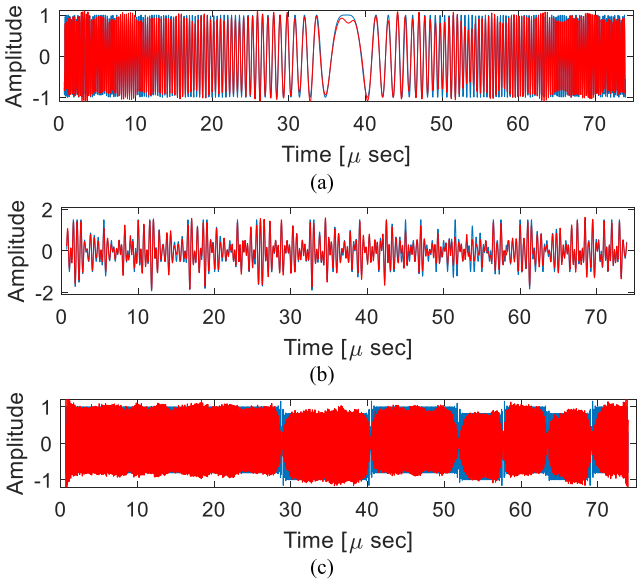

FIGURE 14. Separated signals in the time domain by VAE: (a) LFM, (b) 4-ASK, and (c) interference signal (BC).

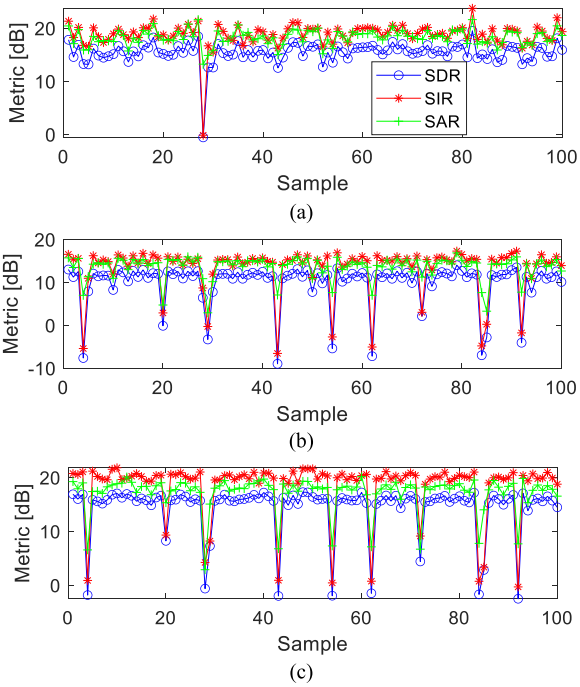

FIGURE 15. Performance metrics (in dB) of the proposed VAE in estimating the (a) LFM, (b) 4-ASK, and (c) interference (BC) signals.

TABLE 5 Average Performance Metrics for the 100 Samples Given in Fig. 15 at an SNR of 10 dB for Case 3

Signal	SDR [dB]	SIR [dB]	SAR [dB]
LFM	15.4212	18.7798	18.2931
4-ASK	9.9205	13.2969	13.5357
BC	14.3743	18.3516	17.1788

leaving only the interference signal represented by the BC signal.

Fig. 14 depicts the separated signals obtained by the proposed VAE model for Case 3 at an SNR of 10 dB. It is seen that VAE can successfully separate and reconstruct the input signals with a sufficiently low noise level. The SDRs for the LFM, 4-ASK, and BC signals are 18, 13, and 16.8 dB, respectively, indicating the excellent performance of the proposed VAE in detecting the interference signal at 10 dB SNR.

Fig. 15 shows the ability of the trained VAE model to recover the LFM and 4-ASK signals and detect the BC interference signal for 100 test samples. Table 5 outlines the averaged performance metrics for each signal. The proposed VAE achieves an accuracy of 99% and ∼91% in recovering the LFM and 4-ASK signals, respectively, and ∼91% in detecting the interference signal. The 9% drop occurred because the unsupervised VAE confused between the 4-ASK signal and the BC signal. This incident happened due to the BC signal having a centre frequency that was similar to the 4-ASK signal used during the training of the VAE model.

The signal separation speeds for various methods, averaged over 100 runs, are outlined in Table 6. Notably, JADE demonstrates a relatively rapid separation, requiring just 0.0007 seconds, closely followed by SOBI. In contrast, CQAMsym

TABLE 6 Speed in Separation of the Signals

Method	Time (second)
ExComplxFastICA	0.0016
SOBI	0.0007
JADE	**0.0005**
ACMNsym	0.0049
CQAMsym	0.1270
Proposed VAE	0.0631

exhibits a notably longer time of 0.1270 seconds, indicating a comparatively slower process of signal separation. The proposed VAE method offers a moderate speed, accomplishing separation in 0.0631 seconds. This recorded time includes both the prediction time for both decoders and the implementation time for the masking technique and inverse STFT.

Fig. 16 shows a comparison of computational complexity in terms of the time required for signal separation. To expand the time duration of the signals, we extended them to 75, 150, 450, 700 μs. Consequently, this resulted in an increased number of signal samples and input features for the proposed VAE, totalling 8576, 17536, 53504, and 89600, respectively. The figure showcases that the JADE method exhibits the lowest computational complexity, followed by SOBI, ExComplexFastICA and ACMNsym. The CQAMsym method demonstrates the highest computational complexity. Notably, the proposed VAE attains a lower complexity than CQAMsym while surpassing the other four BSS methods in terms of complexity.

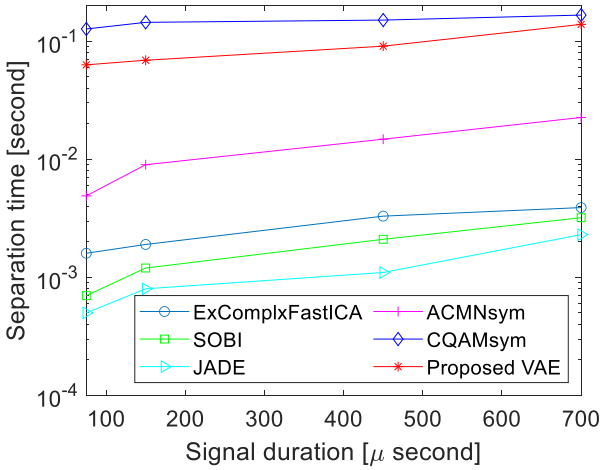

FIGURE 16 **Complexity comparison in terms of separation time (log scale).**

IV. CONCLUSION

In this article, unsupervised VAE was used for the first time in the domain of JCR systems to blindly separate radar and communication signals, as well as to mitigate interference in a spectral coexistence scenario. Different masking techniques were evaluated for data communication and sensing signals. We found that the proposed mask with a power of 4 and a threshold value was very suitable for the data communication signal, whereas the simple mask technique with less computation complexity can be used for recovering the sensing signal. Simulation results demonstrated that our proposed VAE model could separate and reconstruct the data communication signal with rapidly varying instantaneous amplitude. In addition, the proposed model was able to separate complex mixed signals where the subsignals were overlapping in the time-frequency domain and in the time domain only. To conclude our findings, the effect of the overlapping in the time-frequency domain was critical for the VAE at a high SNR only; at a low SNR, the VAE was superior to the BSS methods in terms of SDR and SAR. On the other hand, a low SNR was very critical for the BSS algorithms in both cases when the subsignals were overlapping in the time-frequency domain and in the time domain only. Recognizing the respective strengths of both the VAE and BSS methods, there exists an opportunity for future work to design a hybrid model that effectively addresses the challenge of mutual interference in JCR system.

REFERENCES

[1] C. Masouros, R. W. Heath, J. A. Zhang, Z. Feng, L. Zheng, and A. P. Petropulu, "Editorial: Introduction to the issue on joint communication and radar sensing for emerging applications," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 6, pp. 1290–1294, Nov. 2021, doi: 10.1109/JSTSP.2021.3119395.

[2] J. A. Zhang et al., "Enabling joint communication and radar sensing in mobile networks—A survey," *IEEE Commun. Surv. Tuts.*, vol. 24, no. 1, pp. 306–345, Firstquarter 2022, doi: 10.1109/COMST.2021.3122519.

[3] B. Jin et al., "Data-driven sparsity-based source separation of the aliasing signal for joint communication and radar systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2161–2174, Feb. 2023, doi: 10.1109/TVT.2022.3212408.

[4] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020, doi: 10.1109/TCOMM.2020.2973976.

[5] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999, doi: 10.1109/72.761722.

[6] M. Novey and T. Adali, "On extending the complex FastICA algorithm to noncircular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2148–2154, May 2008, doi: 10.1109/TSP.2007.911278.

[7] M. Novey and T. Adali, "Complex ICA by negentropy maximization," *IEEE Trans. Neural Netw.*, vol. 19, no. 4, pp. 596–609, Apr. 2008, doi: 10.1109/TNN.2007.911747.

[8] M. Novey and T. Adali, "Complex fixed-point ICA algorithm for separation of QAM sources using Gaussian mixture model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2007, pp. II-445–II-448, doi: 10.1109/ICASSP.2007.366268.

[9] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997, doi: 10.1109/78.554307.

[10] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. F (Radar Signal Process.)*, vol. 140, no. 6, pp. 362–370, 1993, doi: 10.1049/ip-f-2.1993.0054.

[11] K. A. Alaghbari, L. H. Siong, and A. W. C. Tan, "Robust correntropy ICA based blind channel estimation for MIMO-OFDM systems," *COMPEL- Int. J. Computation Math. Elect. Electron. Eng.*, vol. 34, no. 3, pp. 962–978, 2015, doi: 10.1108/COMPEL-08-2014-0199.

[12] Z. Luo, C. Li, and L. Zhu, "A comprehensive survey on blind source separation for wireless adaptive processing: Principles, perspectives, challenges and new research directions," *IEEE Access*, vol. 6, pp. 66685–66708, 2018, doi: 10.1109/ACCESS.2018.2879380.

[13] A. Naeem and H. Arslan, "Joint radar and communication based blind signal separation using a new non-linear function for fast-ICA," in *Proc. IEEE 94th Veh. Technol. Conf.*, 2021, pp. 1–5, doi: 10.1109/VTC2021-Fall52928.2021.9625477.

[14] K. A. Alaghbari, M. H. Md. Saad, A. Hussain, and M. R. Alam, "Activities recognition, anomaly detection and next activity prediction based on neural networks in smart homes," *IEEE Access*, vol. 10, pp. 28219–28232, 2022, doi: 10.1109/ACCESS.2022.3157726.

[15] I. Higgins et al., "Beta-VAE: Learning basic visual concepts with a constrained variational framework," in *Int. Conf. Learn. Representations* Toulon, France, 2017.

[16] C. P. Burgess et al., "Understanding disentangling in β-VAE," in *Proc. NIPS Workshop Learn. Disentangled Representations*, Long Beach Convention Center, CA, USA, Dec. 9, 2017. [Online]. Available: https://doi.org/10.48550/arXiv:1804.03599

[17] L. Pandey, A. Kumar, and V. Namboodiri, "Monoaural audio source separation using variational autoencoders," in *Proc. Interspeech*, 2018, pp. 3489–3493, doi: 10.21437/Interspeech.2018-1140.

[18] E. Karamatlı, A. T. Cemgil, and S. Kırbız, "Audio source separation using variational autoencoders and weak class supervision," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1349–1353, Sep. 2019, doi: 10.1109/LSP.2019.2929440.

[19] H. D. Do, S. T. Tran, and D. T. Chau, "Speech source separation using variational autoencoder and bandpass filter," *IEEE Access*, vol. 8, pp. 156219–156231, 2020, doi: 10.1109/ACCESS.2020.3019495.

[20] J. Neri, R. Badeau, and P. Depalle, "Unsupervised blind source separation with variational auto-encoders," in *Proc. 29th Eur. Signal Process. Conf.*, 2021, pp. 311–315, doi: 10.23919/EUSIPCO54536.2021.9616154.

[21] C. Fevotte, R. Gribonval, and E. Vincent, "BSS eval toolbox user guide–revision 2.0," INRIA, Rennes Cedex, France, Tech. Rep. 00564760, 2005.

[22] J. L. Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "SDR–half-baked or well done?," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 626–630, doi: 10.1109/ICASSP.2019.8683855.

[23] B. Vanessa and U. Seljak, "Probabilistic autoencoder," *Trans. Mach. Learn. Res.*, vol. 9, pp. 1–25, 2022, doi: 10.48550/ARXIV.2006.05479.
[24] E. M. Grais and M. D. Plumbley, "Single channel audio source separation using convolutional denoising autoencoders," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2017, pp. 1265–1269, doi: 10.1109/GlobalSIP.2017.8309164.

KHALED A. ALAGHBARI received the B.Eng. degree (Hons.) in electronics engineering majoring in telecommunication and the M.Eng.Sc. degree in wireless communications and the Ph.D. degree in optical communications from Multimedia University (MMU), Malacca, Malaysia, in 2011, 2014, and 2020, respectively. From 2021 to 2022, he was a Postdoctoral Researcher with the Institute of IR 4.0, National University of Malaysia (UKM), Bangi, Malaysia. He is currently a Postdoctoral Research Fellow with the Center for Sustainable Communications and IoT, MMU. His research interests include signal processing, communication systems, and artificial intelligence (AI).

HENG SIONG LIM (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from Universiti Teknologi Malaysia, Malaysia, in 1999, and the M.Eng.Sc. and Ph.D. degrees in wireless communications from Multimedia University (MMU), Malacca, Malaysia, in 2002 and 2008, respectively. He is currently a Professor with the Faculty of Engineering and Technology, MMU. His research interests include the areas of signal processing for wireless communications and advanced digital communication receivers design.

BENZHOU JIN (Member, IEEE) received the B.E. degree in instrument science and technology from Xidian University, Xi'an, China, in 2008, and the Ph.D. degree in information and communication engineering from Tsinghua University, Beijing, China, in 2013. He is currently a Professor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. From 2013 to 2018, he was a Senior Engineer with the Nanjing Research Institute of Electronics Technology, Nanjing. From 2018 to 2022, he was an Associate Professor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. His research interests include radar system engineering, nonlinear signal processing, and array signal processing.

YUTONG SHEN received the B.E. degree in electronic science and technology from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2021. She is currently working toward the M.E. degree in electronic information with the College of Electronic and Information Engineering, NUAA, Nanjing. Her research interests mainly include aliased signal separation and main lobe interference suppression of radar system.