remote sensing

MDPI

Article

Unsupervised Radar Target Detection under Complex Clutter Background Based on Mixture Variational Autoencoder

Xueling Liang, Bo Chen, Wenchao Chen *, Penghui Wang and Hongwei Liu

National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China
* Correspondence: chenwenchao@xidian.edu.cn

Abstract: The clutter background in modern radar target detection is complex and changeable. The performance of classical detectors based on parametric statistical modeling methods is often degraded due to model mismatch. Existing data-driven deep learning methods require cumbersome and expensive annotations. Furthermore, the performance of the detection network is severely degraded when the detection scene changes, since the trained network with the data from one scene is not suitable for another scene with different data distribution. To this end, it is crucial to develop an unsupervised detection method that can finely model complex and changing clutter scenes. This problem is challenging yet rewarding because it completely eliminates the cost of obtaining cumbersome annotations. In this paper, we introduce GM-CVAE, a novel unsupervised Gaussian Mixture Variational Autoencoder with a one-dimensional Convolutional neural network approach to finely model complex and changing clutter. Furthermore, we develop an unsupervised narrow-band radar target detection strategy based on reconstructed likelihood. Comprehensive experiments are carried out to show that the proposed method realizes the refined modeling of clutter and guarantees superior detection performance in the simulated complex clutter environment. Compared with baselines, the proposed method shows better performance.

Keywords: complex clutter background; clutter refinement modeling; narrow-band low resolution radar; GMM; VAE

check for updates

Citation: Liang, X.; Chen, B.; Chen, W.; Wang, P.; Liu, H. Unsupervised Radar Target Detection under Complex Clutter Background Based on Mixture Variational Autoencoder. *Remote Sens.* **2022**, *14*, 4449. https://doi.org/10.3390/rs14184449

Academic Editors: Pawel Rotter, Wojciech Chmiel and Sławomir Mikrut

Received: 30 July 2022
Accepted: 2 September 2022
Published: 6 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Target detection technology, as the most basic and important function of radar [1,2], has always been paid attention by military and researchers. Adaptive threshold techniques are widely used in radar automatic detection to maintain a constant false alarm rate (CFAR) in an unknown clutter environment. Cell-averaging CFAR (CA-CFAR) [3,4] is the earliest proposed type of CFAR and is very popular in practical radar detection. The background level is estimated by averaging the amplitudes of the reference cells selected from the neighbor cells of the cell under test (CUT), which is optimal in a homogeneous Rayleigh distribution. Under the heterogeneous background environment, data in the reference cells will no longer satisfy the condition of independent and identically distributed (*i.i.d.*), and the detection performance of the CA-CFAR method will decline inevitably [5]. When the CUT is located in the noise region and the reference window contains clutter samples or other interfering targets, a masking effect will occur, which then will lead to serious degradation of the detection probability. Furthermore, at the clutter edges, the false alarm rate may increase intolerably. Therefore, the greatest of selection CFAR (GO-CFAR) method [6,7] is proposed, which is better than the CA-CFAR method in maintaining the CFAR property in the clutter edges case. After this, the order statistic CFAR (OS-CFAR) [8,9] has been proposed in a heterogeneous background environment. OS-CFAR ranks the amplitudes of the reference cells and selects the k-th sample as the average power level of the CUT, where the choice of the order k depends on the desired false alarm rate.

The CFAR detection algorithms aforementioned are mainly used to solve the target detection problem under the clutter with Gaussian statistical characteristics. Furthermore, aiming at the clutter background with non-Gaussian statistical characteristics, researchers have also proposed a series of models in recent years, such as the Weibull distribution model [10,11], K-distribution model [12,13], spherical invariant stochastic process model [14] and compound Gaussian model [15–19]. For heterogeneous clutter, some methods have been developed especially for the clutter edges case. For example, refs. [20,21] are put forward to solve the clutter edges problem in Gaussian background , while [10,22,23] studied this problem under non-Gaussian distribution. In addition, for the problem of inaccurate estimation of sample covariance matrix due to too little *i.i.d.* training data available in heterogeneous clutter environment, refs. [24,25] develop matrix information geometry (MIG) detectors and [26–28] studies adaptive subspace signal (ASD) detector to improve target detection performance.

The existing radar target detection methods consider different statistical distribution models, corresponding detectors have also been designed by researchers based on different models, and achieved the expected detection performance under the background of their respective assumptions. However, these methods have some common characteristics and limitations. Firstly, they are parametric-based statistical modeling methods. Existing detection methods are all statistical results of a large amount of data, which have been experimentally verified and have good universality. However, due to only utilizing empirical information, the statistical distribution form is fixed, and the pertinence is poor so that the model cannot be corrected with the change of data. In practical applications, there are problems such as model mismatch, which lead to a significant drop in performance. Secondly, the information dimension used is relatively simple, mainly the amplitude one-dimensional feature, and more high-dimensional features are not further extracted from Doppler and time-domain distributions, etc. The underlying assumption is that the target must have a strong amplitude relative to the clutter background. Therefore, underutilization of feature information often leads to its limited performance. Finally, the accuracy of clutter modeling is relatively coarse, and lacks refined modeling theories and methods. In practice, there may be a variety of clutter in the radar irradiation area, subjecting to different distribution and forming a complex clutter scene, which will also affect the final detection performance.

The ground-based warning radar is a typical application scenario of the CFAR detection method. As for the ground-based warning radar, it has an important responsibility to ensure homeland security, and needs to detect, track and even identify suspicious airborne targets (low altitude penetration, unmanned aerial vehicles and other low, small and slow targets.) and suspicious ground targets in advance. Radar is required to protect the large scope, meanwhile, the ground clutter is more complex (as shown in Figure 1, there may be various terrains in the range covered by the beam, such as mountains with trees, grass, towns and farmland, lakes, and bushes next to shrubs, etc.) and even include artificial active and passive interference. Such a complicated detection scene means that the above-mentioned detection methods for specific clutter backgrounds will no longer be applicable, so it is important to introduce a refined clutter modeling method.

Due to its tremendous capability in learning expressive representations of complex data, a recent trend is to utilize deep learning in a wide variety of tasks [29,30], including radar target detection [31–33]. The radar detection task in deep learning can be regarded as a binary classification problem [34,35]. The most typical approach is to learn a classifier on training samples with class labels and then apply it to classify test samples. However, the trained network with the data from one scene is not suitable for another scene with different data distribution. Oftentimes, we need to retrain the classifier model on the new dataset, which requires manually labeling many new training samples. To alleviate this problem and eliminate or reduce the need to label new datasets, domain adaptation methods [36–38] are proposed. This method utilizes annotated data from the source domain (there is a large amount of data in the source domain) to train the classifier to operate in the target domain, which is not friendly to unlabeled training datasets. For training

datasets lacking effective labels, unsupervised deep learning methods are more suitable. Unfortunately, to the best of our knowledge, there are almost no significant attempts to use unsupervised learning methods for target detection in a ground-based warning radar background. In addition, the military field is usually faced with non-cooperative targets, which are difficult to enumerate and obtain, leading to a serious shortage of target samples. In other words, clutter data and target data are extremely unbalanced, which inevitably leads to the problem of class distribution imbalance. Applying a classification model, in this case, is not a wise move. In the current work, we explore an unsupervised mode for radar target detection that is certainly more challenging than supervised methods. However, it is also more rewarding due to minimal assumptions and hence will encourage the development of novel and more practical algorithms. In approaching unsupervised target detection in ground-based warning radar, we exploit the simple fact that targets are less frequent than the clutter happenings, and attempt to leverage such domain knowledge in a structured manner.

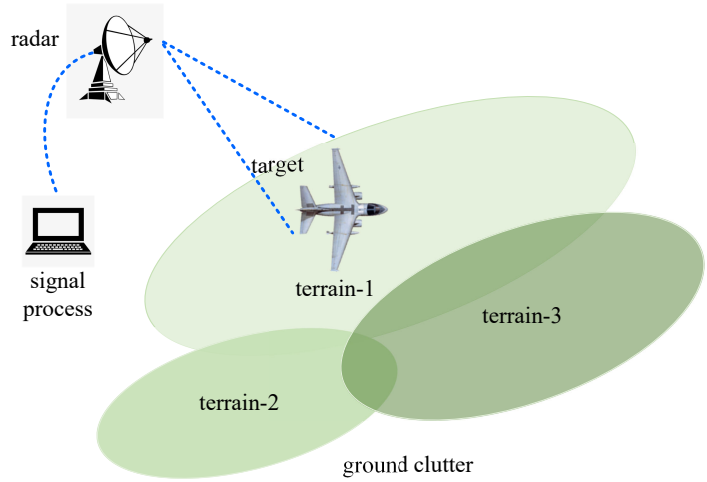


Figure 1. Schematic diagram of typical ground scenes. It depicts the terrain scene over which an aircraft target flies. This scene is a typical example of a complex ground scene, which is the original intention of this study.

Having said all of the above, in view of the differences in clutter distribution caused by environmental diversity, we aim to use an unsupervised deep generative model to learn robustly and model finely complex clutter distributions, find the specific distribution of interested target according to the characteristics of clutter and target difference, and then extract, amplify and detect the interested target in this paper. As a consequence, motivated by [39–42], we propose a Gaussian Mixture Variational Autoencoder with a one-dimensional convolutional neural network (GM-CVAE) method considering the attractive properties of the convolutional neural network (CNN). The model takes pure clutter data as input and learns the log-likelihood as output. Specifically, GM-CVAE learns a latent representation of clutter to capture its normal pattern, and then reconstructs the input data from the latent representation. Target and clutter distributions, meanwhile, are different, the reconstruction effect of the model on the target will be worse, so the target detection can be achieved by setting the reconstruction error threshold.

Our contributions are summarized as follows:

- In order to solve the problem of complex and changeable clutter in the radar scanning range in the actual environment, we propose a GM-CVAE framework to realize refined modeling of clutter.
- Considering that it is difficult to obtain the true labels of the data and the imbalance of clutter and target samples, we develop an unsupervised narrow-band radar target detection strategy based on reconstructed likelihood.

- Experiments are used to evaluate our approach on simulated complex clutter datasets. These demonstrate the superiority of our method compared to the baselines.

The remaining parts of this paper are organized as follows. Section 2 presents the relevant preliminary knowledge. Section 3 describes the proposed method and the target detection strategy is presented in Section 4. Section 5 summarizes the numerical results and a general conclusion is presented in Section 6.

2. Preliminaries

In this section, we first introduce the clutter and target characteristics of the ground-based warning radar, respectively. Then, we provide a brief overview of the variational autoencoder, which is the basis of our work.

2.1. Ground Clutter Characteristics

Ground clutter is the main factor affecting the detection and tracking performance of the ground-based warning radar for low-altitude targets and ground targets. Essentially, ground clutter is the vector sum of echoes from many scatterers within the radar resolution cell. The clutter scattering phenomenon can be understood as a random process related to the random shape of the ground, which is usually described by the probability distribution model of clutter amplitude and clutter correlation model [43]. Many experimental facts have verified that for a low-resolution radar with a beam incident angle greater than 5° (antenna beam width greater than 2°, pulse width greater than 1 μs), the clutter amplitude distribution of sea clutter, ground clutter as well as meteorological clutter obeys the Rayleigh distribution. Therefore, this paper assumes that the ground clutter amplitude follows a Rayleigh distribution, the ground clutter power spectrum is described by a Gaussian function in which spectral center shift and spectral broadening are different, and partial relevant parameters in [44] are referred to for ground clutter simulation. In particular, the Rayleigh distribution is the distribution of the envelope of a Gaussian random signal passing through a narrow band linear system and needs to satisfy the following conditions: (1) The scatters within the radar irradiation unit are statistically independent. (2) The distance difference between the two scatters and the radar in the irradiation unit is much smaller than the scale of the irradiation unit, and the gain of the antenna in the unit is constant. (3) The number of scatterers is large and the signal is normally distributed according to the central limit theorem.

2.2. Target Characteristics

The primary task of the ground-based warning radar is to detect unknown targets as early as possible, determine their positions within the detection range as far as possible, and closely monitor the flying targets in the airspace. In order to satisfy the needs of long detection distance, ground-based warning radars generally work in lower frequency bands, mainly S and L frequency bands, and some work in the UHF frequency band, of course, VHF, and even HF frequency bands. The detection distance of ground-based warning radar is more than 350 km, even up to 550 km, but it does not have high requirements on the accuracy and resolution of the detection target. Generally, in low-resolution ground-based warning radar, the target imaged on the range-Doppler spectrum is a point-like target with a small scale, and usually does not cross range cells. Therefore, the radar target model considered in this paper is a point-like target that occupies only one range cell.

2.3. Variational Autoencoder

GM-CVAE is a variant of Variational Autoencoder (VAE) [45–47], which uses the Gaussian mixture model (GMM) as the prior of VAE latent space, extracts and expresses features of data through probability modeling. The biggest difference between GM-CVAE and VAE is that the former can handle complex multi-modal data, while the latter may not be able to approximate the original multi-modal data distribution well due to a single

Gaussian distribution prior. Hence, a brief description of the variational autoencoder is necessary before introducing our methods.

VAE is a popular reconstruction probabilistic model which combines Bayesian inference with the autoencoder framework. The idea of VAE is that the distribution of complex data can be modeled by lower-dimensional latent variables or representations. In a regular VAE, ref. [47] first presents a computationally tractable method for training this model. Specifically, as shown in Figure 2, the prior distribution of the latent variables z, denotes $p_\theta(z)$, are often assumed to obey the unimodal Gaussian distribution, and generally for simplicity, $p_\theta(z) \sim \mathcal{N}(0,1)$, where $p(\cdot)$ is the probability distribution function and θ is the parameters of decoder. Through the encoder network with parameters ϕ, VAE learn posterior distribution $q_\phi(z|x)$ so as to approximate the true posterior distribution $p_\theta(z|x)$. Then, it learns likelihood distribution $p_\theta(x|z)$ so as to reconstruct the original input x through the decoder network with parameter θ. The better the z learned by the network, the more accurate the reconstruction will be. Given that the input data x is known and the latent variable z is unknown, we hope that two distributions $q_\phi(z|x)$ and $p_\theta(z|x)$ are as close as possible. Therefore, the following objective function [47] is obtained

$$\min_{\phi,\theta} D_{\mathrm{KL}}(q_\phi(z|x)||p_\theta(z|x)), \tag{1}$$

where D_{KL} stands for Kullback–Leibler divergence, which definition equation is $D_{\mathrm{KL}}(q(x)||p(x)) = \int q(x)\log\frac{q(x)}{p(x)}dx$. It is mainly used to measure the similarity between two probability distributions. Next, through statistical derivation, the marginal log-likelihood of the input data x is obtained

$$\log p(x) = D_{\mathrm{KL}}(q_\phi(z|x)||p_\theta(z|x)) + \mathcal{L}_{\mathrm{VAE}}(\phi,\theta;x), \tag{2}$$

where

$$\mathcal{L}_{\mathrm{VAE}}(\phi,\theta;x) = \mathbb{E}_{q_\phi(z|x)}[\log p(x,z)] - \mathbb{E}_{q_\phi(z|x)}[\log q_\phi(z|x)], \tag{3}$$

and $\mathcal{L}_{\mathrm{VAE}}(\phi,\theta;x)$ is called the variational lower bound.

Since x is deterministic, $\log p(x)$ is a constant. Consequently, minimizing D_{KL} term is equivalent to maximizing $\mathcal{L}_{\mathrm{VAE}}$ term according to Equation (2). Furthermore, the objective function of VAE can be rewritten as follows

$$\max_{\phi,\theta} \mathcal{L}_{\mathrm{VAE}}(\phi,\theta;x) \tag{4}$$

Further derivation, Equation (4) can be rewritten as

$$\mathcal{L}_{\mathrm{VAE}}(\phi,\theta;x) = -D_{\mathrm{KL}}(q_\phi(z|x)||p_\theta(z)) + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] \tag{5}$$

The first term on the right side of equal sign in Equation (5) is the regularization term, which is used to minimize the difference between the posterior distribution $q_\phi(z|x)$ and prior distribution $p_\theta(z)$; the second term is the reconstruction term. The solution of $\mathcal{L}_{\mathrm{VAE}}(\phi,\theta;x)$ is a maximum likelihood estimation process of input data x and can be solved by stochastic gradient descent (SGD) methods [47]. In addition, reconstruction loss is related to the input data form. If the input data is binary, binary cross entropy (BCE) can be used to approximate the loss; if the input data is continuous, the loss can be replaced by the mean square error (MSE).

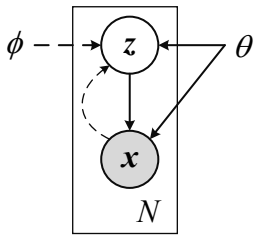


Figure 2. Graphical model of VAE. Without loss of generality, $\boldsymbol{x} \in \mathbb{R}^D$ stands for input data, $\boldsymbol{z} \in \mathcal{R}^K$ is the latent variable, where D and K are the dimension of the input and that of the latent variable, respectively. The solid line denotes generative process (decoder) and the dash lines denote variational approximation (encoder).

However, regular VAEs often fail to approximate hidden layer distributions of complex data due to too simple prior assumptions. The idea behind our GM-CVAE is to use a Gaussian mixture distribution as a prior for the stochastic variable $\boldsymbol{z}$, which depends on the categorical variable $\boldsymbol{c}$. The prior of this variant is multi-modal and can learn more complex distributions than the unimodal prior. In this paper, we study a high-dimensional Gaussian mixture model to estimate the sample densities for target detection, which will be clear in next section.

3. Method

This section mainly discusses the Gaussian Mixture Variational Autoencoder with a one-dimensional Convolutional neural network (GM-CVAE), which is proposed to refine clutter modeling in an unsupervised manner to optimize target detection.

3.1. Overview of the Framework

The overall framework shown in Figure 3 comprises data pre-processing, the representation model and detection three modules.

The data pre-processing module firstly performs pulse compression on the radar echo data; then selects Moving Target Indication (MTI) or Adaptive Moving Target Indication (AMTI) for filtering processing according to static or moving clutter; after that, it utilizes Moving Target Detection (MTD) for coherent integration to obtain the range-Doppler spectrum (also called the R-D spectrum) and referring to [33,48], we take the modulus of the R-D spectrum; lastly, it performs a normalization operation to avoid numerical problems and speed up network convergence. we define the normalized R-D spectrum as $\mathbf{X}$, $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N] \in \mathbb{R}^{M \times N}$, where M and N represent the number of Doppler channel and range cell, respectively, and $\mathbb{R}$ denotes the set of real number. The clutter data varies drastically in the spatial domain, and the assumption that data in the CUT and the reference cells are independent and identically distributed cannot be guaranteed. Therefore, $\boldsymbol{x}_i \in \mathbb{R}^{M \times 1}, i = 1, 2, \ldots, N$ denotes the Doppler feature vector of the i-th range cell and makes it a sample in this paper. For the sake of description, we will ignore the subscript i for the rest of the discussion.

The difference between the data pre-processing process in this paper and the traditional radar signal processing process is the normalization step. Furthermore, we do not discuss a novel method of clutter suppression, which is beyond the scope of this paper. In the representation module, we propose a GM-CVAE to learn the complex structures and Doppler characteristics of clutter data. Its details will be covered in Section 3.2. Subsequently, target detection is performed based on the reconstruction results inferred from the model. The specific target detection method will be introduced in Section 4.

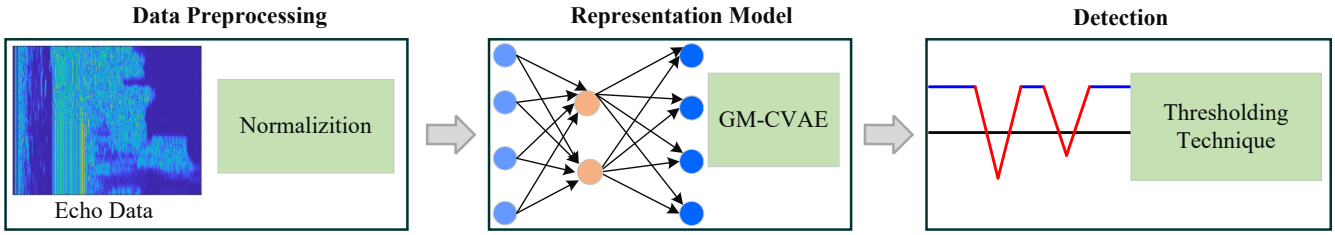


Figure 3. An overall framework for unsupervised target detection based on the GM-CVAE model. The framework mainly consists of three parts: data pre-processing, the representation model and detection.

3.2. Network Architecture

As shown in Figure 4, GM-CVAE is a reconstruction-based model, including an inference network and a generative network. Next, we will introduce the two sub-networks, respectively.

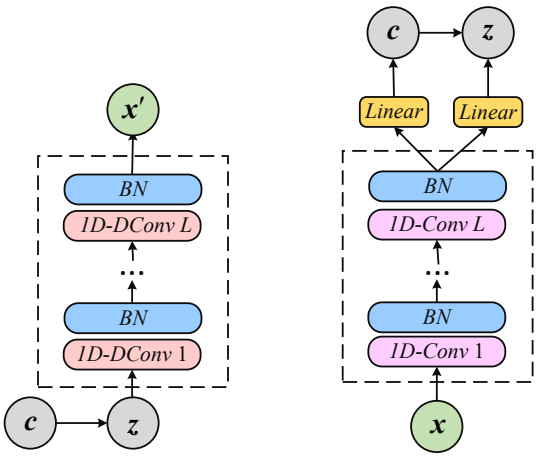


Figure 4. Network architecture of GM-CVAE, which is composed of two parts: (**a**) generative network and (**b**) inference network. $\boldsymbol{x}$ is the input data, and L is the layers of 1D-convolution and 1D-deconvolution. $\boldsymbol{x}'$ is the reconstruction output, $\boldsymbol{z}$ and $\boldsymbol{c}$ are stochastic and categorical latent variables, respectively.

Generative Network. It measures the likelihood of generating data instances given the latent variables. The generative model (as shown in Figure 4a) can factorize as

$$p_\theta(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{c}) = p_\theta(\boldsymbol{c}, \boldsymbol{z}) p_\theta(\boldsymbol{x}|\boldsymbol{z}) = p_\theta(\boldsymbol{c}) p_\theta(\boldsymbol{z}|\boldsymbol{c}) p_\theta(\boldsymbol{x}|\boldsymbol{z}), \tag{6}$$

and generate an observed sample $\boldsymbol{x}$ from the latent random variable $\boldsymbol{z}$ and latent discrete variable $\boldsymbol{c}$, where θ is the network parameter. Specifically, we assign the form of the prior of the latent variable $\boldsymbol{z}$ as

$$\boldsymbol{z}|\boldsymbol{c} \sim \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_k, \mathrm{diag}(\boldsymbol{\sigma}_k))^{c_k} \tag{7}$$

where $\{\boldsymbol{\mu}, \boldsymbol{\sigma}\} = \{\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k\}_{k=1}^{K}$ represent the mean and variance parameters of multiple Gaussian distribution, which are obtained through the fully connected networks, respectively. K denotes the predefined number of components in the Gaussian mixture and $\mathrm{diag}(\cdot)$ function constructs a diagonal matrix. $\boldsymbol{c} = (c_1, \ldots, c_K)^{\mathrm{T}}$ is an explicit latent variable associated with each $\boldsymbol{x}$ and is a one-hot vector subject to categorical prior distribution $\boldsymbol{c} \sim \mathrm{Cat}(\boldsymbol{\pi})$ with parameter $\boldsymbol{\pi} \in \mathbb{R}_{+}^{K \times 1}$. Thus, the latent variable $\boldsymbol{z}$ conditions on a different Gaussian

distributed prior $\boldsymbol{c}$, which indicates the diverse distribution characteristics of the input. Furthermore, by marginalizing $\boldsymbol{z}$, the following equation can achieve

$$\boldsymbol{z} \sim \sum_{\boldsymbol{c}} p_\theta(\boldsymbol{c}) p_\theta(\boldsymbol{z}|\boldsymbol{c}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{\mu}_k, \mathrm{diag}(\boldsymbol{\sigma}_k)) \tag{8}$$

Apparently, it is a mixture Gaussian distribution, which has higher representation power than a Gaussian distribution, so it is an ideal choice for characterizing the R-D spectrum characteristics of complex clutter.

Similar to [49], the generated distribution $p_\theta(\hat{\boldsymbol{x}}|\boldsymbol{z})$ is assigned as a Gaussian distribution and conditioned on both $\boldsymbol{z}$, which can be expressed as

$$\begin{aligned} \hat{\boldsymbol{x}}|\boldsymbol{z} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\sigma}_{\boldsymbol{x}}) \\ \boldsymbol{\mu}_{\boldsymbol{x}} &= f(\boldsymbol{w}^{\mu_x}\boldsymbol{z} + \boldsymbol{b}^{\mu_x}) \\ \boldsymbol{\sigma}_{\boldsymbol{x}} &= f(\boldsymbol{w}^{\sigma_x}\boldsymbol{z} + \boldsymbol{b}^{\sigma_x}) \end{aligned} \tag{9}$$

where $\boldsymbol{\mu}_{\boldsymbol{x}}$ and $\boldsymbol{\sigma}_{\boldsymbol{x}}$ denotes mean and variance parameters of the generated distribution. All $\boldsymbol{w}^*$-s, $\boldsymbol{b}^*$-s represent the weight parameters and bias parameters of the corresponding fully connected networks, respectively. $f(\cdot)$ is a nonlinear activation function, here we use LeakyReLU activation function and its expression is

$$\mathrm{LeakyReLU}(\boldsymbol{x}) = \begin{cases} \boldsymbol{x}, & if\ \boldsymbol{x} \geq 0, \\ \alpha\boldsymbol{x}, & if\ \boldsymbol{x} < 0, \end{cases} \tag{10}$$

The value of α is 0.1 in our model.

Finally, to better generate the structures of data, the decoder performs the one-dimensional deconvolutional network as $\boldsymbol{x}' = \mathrm{DCNN}(\hat{\boldsymbol{x}})$, where DCNN denotes the one-dimensional deconvolutional operation with trainable parameters $\boldsymbol{D}$. As the deconvolution operation proceeds, the spatial resolution gradually increases, where the output of the final deconvolution layer is aimed at reconstructing the input $\boldsymbol{x}$. Note that the constant change of model parameters during deep network training will cause the internal covariate shift [50], resulting in slow and inefficient network training. Hence, a BN layer is connected after the one-dimensional deconvolution layer in the deconvolution network to solve this problem.

Inference Network. It is a difficult problem to directly solve the generative model, i.e., finding maximum a posterior (MAP) of latent variables and maximum likelihood estimation (MLE) of parameters. In response to this issue, a new distribution $q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})$ is employed to approximate the true posterior distribution $p_\theta(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})$, which is drawn from specific class and parameterized by trainable parameter ϕ. Specifically, according to the mean-field theory approximation [51], the inference model (as shown in Figure 4b) can be factorized as

$$q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x}) = q_\phi(\boldsymbol{c}|\boldsymbol{x}) q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c}) \tag{11}$$

In the inference network, for capturing the structural characteristics of input, a one-dimensional convolution network is conducted, i.e., $\bar{\boldsymbol{x}} = \mathrm{CNN}(\boldsymbol{x})$, where CNN denotes the one-dimensional convolutional operation with parameters $\bar{\boldsymbol{D}}$. Similarly, a BN layer is connected after the one-dimensional convolution layer in Figure 4b. To increase the content covered by a convolutional kernel and the sparsity of the hidden units, a convolutional layer is usually followed by a pooling layer. However, the pooling layer that may reduce the resolution of the feature map is removed from the convolutional encoder, because the radar cross-sectional area of the target itself in the detection scene is small, the reflected energy is weak, and the scale of the target is small on the R-D spectrum.

After that, the parameter of the category distribution, that is, the indicator variable $\boldsymbol{\pi}$, can be expressed as

$$\boldsymbol{\pi} = \mathrm{softmax}(\boldsymbol{w}^{\pi}\bar{\boldsymbol{x}} + \boldsymbol{b}^{\pi}) \tag{12}$$

where $\boldsymbol{w}^{\pi}$ and $\boldsymbol{b}^{\pi}$ are the weight and bias parameters learned by the network, respectively. The softmax function guarantees the non-negativity of probabilities. Then, we can obtain $\boldsymbol{c}$ by sampling $\boldsymbol{\pi}$ through Equation (14).

Finally, We parametrize variational factors with networks that output mean and diagonal covariance of variational distributions as well as assign their form as Gaussian posteriors

$$\begin{aligned} \boldsymbol{z}|\bar{\boldsymbol{x}}, \boldsymbol{c} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{z}}, \mathrm{diag}(\boldsymbol{\sigma}_{\boldsymbol{z}})) \\ \boldsymbol{\mu}_{\boldsymbol{z}} &= f\left(\prod_{k=1}^{K}(\boldsymbol{w}^{\mu_z}\bar{\boldsymbol{x}} + \boldsymbol{b}^{\mu_z})^{\boldsymbol{c}_k}\right) \\ \boldsymbol{\sigma}_{\boldsymbol{z}} &= \mathrm{softplus}\left(\prod_{k=1}^{K}(\boldsymbol{w}^{\sigma_z}\bar{\boldsymbol{x}} + \boldsymbol{b}^{\sigma_z})^{\boldsymbol{c}_k}\right) \end{aligned} \tag{13}$$

where $f(\cdot)$ denotes the LeakyReLU activation function, $\boldsymbol{\mu}_{\boldsymbol{z}}$ and $\boldsymbol{\sigma}_{\boldsymbol{z}}$ are the mean and variance parameters of the latent variable $\boldsymbol{z}$ inferred by networks. Likewise, all $\boldsymbol{w}^{*}$-s, $\boldsymbol{b}^{*}$-s represent the weight parameters and bias parameters of the corresponding fully connected networks, respectively. $\mathrm{softplus}(\cdot) = \log(1 + \exp(\cdot))$ is to ensure original data non-negative after nonlinear transformation.

Note that, categorical variable $\boldsymbol{c}$ controls the structural prior of the latent variable $\boldsymbol{z}$. Motivated by [47,52], we employ variational inference to learn $\boldsymbol{c}$ in GM-CVAE. However, $\boldsymbol{c}$ is a discrete variable, and the backpropagation algorithm is only applicable to differentiable layers, so it is not practical to directly optimize $\boldsymbol{c}$. Hence, inspired by [53], we introduce the Gumbel-softmax trick to approximate the original categorical distribution $\boldsymbol{c}$ so that $\boldsymbol{c}$ is differentiable and learnable. Specifically, let the variational distribution $q_{\phi}(\boldsymbol{c}) = \mathrm{Gumbel} - \mathrm{softmax}(\boldsymbol{\pi})$, $\boldsymbol{\pi} \in \mathbb{R}^{K\times 1}$, and

$$\begin{aligned} \boldsymbol{c}_k &= \frac{\exp((\log\pi_k + g_k)/\lambda)}{\sum_{l=1}^{K}\exp((\log\pi_l + g_l)/\lambda)}, k = 1, \ldots, K, \\ g_k &\sim \mathrm{Gumbel}(0, 1) = -\log(-\log(\varepsilon_k)), \end{aligned} \tag{14}$$

where $\lambda > 0$ indicates the softmax temperature and controls the softness of softmax. ε_k represents a standard uniform variable. When λ tends to 0, the samples from the Gumbel-softmax distribution become one-hot distribution, and the Gumbel-softmax distribution $q_{\phi}(\boldsymbol{c})$ is consistent with the categorical distribution prior $p_{\theta}(\boldsymbol{c})$.

Furthermore, the same as regular VAE algorithms, we also use a Gaussian variational distribution $q_{\phi}(\boldsymbol{z})$ to infer the posterior distribution of $\boldsymbol{z}$ in GM-CVAE, i.e.,

$$q_{\phi}(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2), \tag{15}$$

where the two parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are transformed by the inference network, and sampling from the latent variables can be achieved by [47]

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \tag{16}$$

thus ensuring normal execution of the back-propagation algorithm. Herein, $\boldsymbol{\epsilon}$ is a vector with all its entries drawn independently from the standard Gaussian distribution, and "$\odot$" denotes element-wise product.

The goal of GM-CVAE is to find a better stochastic latent representation $\boldsymbol{z}$ for data $\boldsymbol{x}$. For large and complex terrain scenes, it is difficult to accurately describe the entire ground clutter area with one distribution model, and even the same terrain scene obeys different distributions under the influence of different time and climatic conditions. Different clutter backgrounds obey different probability distributions, resulting in the existence of multiple normal modes in the data center. Therefore, it is an excellent choice for the prior of $\boldsymbol{z}$ to follow the mixture Gaussian distribution.

The generative network and inference network are trained together to approximate the true posterior distribution. Details are described below.

3.3. Offline Training Process

GM-CVAE model can be trained by maximizing the evidence of lower bound (ELBO) [54] on the log-likelihood of input data, $\mathcal{L}_{\mathrm{GM-CVAE}}(\phi, \theta; \boldsymbol{x})$,

$$\mathcal{L}_{\mathrm{GM-CVAE}}(\phi, \theta; \boldsymbol{x}) = \mathbb{E}_{q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{c})}{q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})}\right] \tag{17}$$

Combine Equations (6) and (11), and rewrite Equation (17)

$$\begin{aligned}
&\mathcal{L}_{\mathrm{GM-CVAE}}(\phi, \theta; \boldsymbol{x}) \\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{c})}{q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})}\right] \\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})}\left[\log \frac{p_\theta(\boldsymbol{c})}{q_\phi(\boldsymbol{c}|\boldsymbol{x})} + \log \frac{p_\theta(\boldsymbol{z}|\boldsymbol{c})}{q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c})} + \log\, p_\theta(\boldsymbol{x}|\boldsymbol{z})\right] \\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}, \boldsymbol{c}|\boldsymbol{x})}[\log\, p_\theta(\boldsymbol{x}|\boldsymbol{z})] - D_{\mathrm{KL}}\left[q_\phi(\boldsymbol{c}|\boldsymbol{x})||p_\theta(\boldsymbol{c})\right] - D_{\mathrm{KL}}\left[q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c})||p_\theta(\boldsymbol{z}|\boldsymbol{c})\right]
\end{aligned} \tag{18}$$

In Equation (18), the first term is the expected log-likelihood of the data $\boldsymbol{x}$, which aims to guarantee the reconstruction capacity of the generative model. The second term is the regularization between the prior $p_\theta(\boldsymbol{c})$ and posterior $q_\phi(\boldsymbol{c}|\boldsymbol{x})$ of the categorical variable $\boldsymbol{c}$ that constrains the two distributions as close as possible. The third term represents the regularization between the actual posterior $p_\theta(\boldsymbol{z}|\boldsymbol{c})$ and approximate posterior $q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c})$ of the latent variable $\boldsymbol{z}$. The closer the prior and posterior distributions are, the smaller the regularization value. Specifically, since $\boldsymbol{c}$ is a discrete variable, the second term in Equation (18) can be written as

$$\begin{aligned}
&D_{\mathrm{KL}}\left[q_\phi(\boldsymbol{c}|\boldsymbol{x})||p_\theta(\boldsymbol{c})\right] \\
&= \sum_{k=1}^{K} q_\phi(\boldsymbol{c}_k = 1|\boldsymbol{x}) \log \frac{q_\phi(\boldsymbol{c}_k = 1|\boldsymbol{x})}{p_\theta(\boldsymbol{c}_k = 1)} \\
&= \sum_{k=1}^{K} q_\phi(\boldsymbol{c}_k = 1|\boldsymbol{x}) \left[\log(q_\phi(\boldsymbol{c}_k = 1|\boldsymbol{x})) - \log\left(\frac{1}{K}\right)\right]
\end{aligned} \tag{19}$$

As we assign the prior of latent variables as mixture Gaussian distribution, the KL divergence of the third term in Equation (18) can be calculated as

$$\begin{aligned}
&D_{\mathrm{KL}}\left[q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c})||p_\theta(\boldsymbol{z}|\boldsymbol{c})\right] \\
&= \int \left(\sum_{k=1}^{K} \boldsymbol{c}_k q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c}) \log \frac{q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c})}{\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)}\right) \\
&= \sum_{k=1}^{K} \boldsymbol{c}_k D_{\mathrm{KL}}\left[q_\phi(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{c})||\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k)\right]
\end{aligned} \tag{20}$$

Moreover, the analytical expression of KL divergence between two Gaussian distributions is

$$\begin{aligned}
&D_{\mathrm{KL}}[\mathcal{N}(\boldsymbol{\mu}_1,\ \mathrm{diag}(\boldsymbol{\sigma}_1))||\mathcal{N}(\boldsymbol{\mu}_2,\ \mathrm{diag}(\boldsymbol{\sigma}_2))] = \\
&\frac{1}{2}\left[\log \frac{\boldsymbol{\sigma}_1}{\boldsymbol{\sigma}_2} - d + \boldsymbol{\sigma}_2^{-1}\boldsymbol{\sigma}_1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}\mathrm{diag}(\boldsymbol{\sigma}_2^{-1})(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right],
\end{aligned} \tag{21}$$

where d is the dimension of variables; $\mathrm{diag}(\cdot)$ function constructs a diagonal matrix and $(\cdot)^{\mathrm{T}}$ denotes a transpose operation. The analytical KL divergence expressions in Equations (19)–(21), and the reparameterization tricks of Gumbel-softmax in Equation (14)

and Gaussian distribution in Equation (16) allow the gradients of ELBO with respect to parameters in the inference network to be accurately evaluated.

Algorithm 1 summarizes the GM-CVAE training algorithm under complex clutter background in an unsupervised learning fashion.

Algorithm 1 : GM-CVAE Training Algorithm

input: The pre-processed R-D spectrum training set $\mathbf{X}_{\text{train}} = \{\boldsymbol{x}_i\}_{i=1}^{N_{\text{train}}}$; number of components K; batch-size M.
output: The encoder parameters ϕ and the decoder parameters θ; reconstruction probability vector for all samples $\boldsymbol{s}_{\text{train}}$ (see Equation (22));
$\theta, \phi \leftarrow$ Initialize parameters.
repeat
  Select a mini-batch training subset $\{\boldsymbol{x}_n\}_{n=1}^{M}$ randomly;
  Draw random noise $\{\varepsilon\}_{n=1}^{M}$ from uniform distribution for generating samples $\{\boldsymbol{c}\}_{n=1}^{M}$ according to Equation (14);
  Draw random noise $\{\boldsymbol{\epsilon}\}_{n=1}^{M}$ from normal distribution for generating latent variable $\{\boldsymbol{z}\}_{n=1}^{M}$ according Equation (16);
  Calculate $\bigtriangledown \mathcal{L}_{\text{GM-CVAE}}(\phi, \theta; \boldsymbol{x})$ according to Equations (18)–(21), and update parameters of the two sub-nets jointly;
until convergence
return model parameters ϕ and θ, as well as $\boldsymbol{s}_{\text{train}}$.

4. Target Detection Strategy

The GM-CVAE model learns the network parameters to capture the normal patterns of clutter samples after offline training. Now based on the difference between clutter and target echo in Doppler feature space, we can ensure whether the input sample $\boldsymbol{x}$ is a target by utilizing the trained model. The reconstruction probability in the GM-CVAE model indicates its reconstruction difficulty [51]: the lower the reconstruction probability value, the more difficult it is to be accurately reconstructed. For a sample with a target, it will be difficult to be reconstructed accurately, and its reconstruction probability is often lower than that of a sample only with clutter [55]. Inspired by [56], we decided to take the reconstruction probability of $\boldsymbol{x}$ as the final score, denoted as s, and compare it with threshold t to determine whether the observed sample is a target. The calculation formula is as follows

$$s = \mathbb{E}_{q_\phi(\boldsymbol{z},\boldsymbol{c}|\boldsymbol{x})} \log p_\theta(\boldsymbol{x}|\boldsymbol{z}) \tag{22}$$

If the final score s is less than threshold t, observed sample $\boldsymbol{x}$ is determined as a target; otherwise, it will be judged as a pure clutter sample. The threshold t is learned automatically in the GM-CVAE model. Extreme Value Theory (EVT) takes extreme data as the research object, and infers the distribution of extreme data that may be observed without any distribution based on the original data. The second theorem in EVT, the Peaks-Over-Threshold (POT) [57] model, is based on the generalized Pareto distribution to model the distribution of all observed data that exceeds a sufficiently large threshold to study the tail characteristics of the distribution. POT demonstrates excellent performance in threshold tuning, so we apply the adjusted POT method to help choose the threshold t for GM-CVAE. Algorithm 2 describes the process of threshold detection.

Algorithm 2 : GM-CVAE Detection Algorithm

input: The pre-processed R-D spectrum test set $\mathbf{X}_{\text{test}} = \{\boldsymbol{x}_i\}_{i=1}^{N_{\text{test}}}$;
output: reconstruction probability vector for all samples $\boldsymbol{s}_{\text{test}}$ (see Equation (22)) and labels $\boldsymbol{y}$.
ϕ, θ, $\boldsymbol{s}_{\text{train}} \leftarrow$ Trained GM-CVAE network.
Similar to the loop body of Algorithm 1, reconstruction probability vector $\boldsymbol{s}_{\text{test}}$ is obtained;

Feed $\boldsymbol{s}_{\text{train}}$ as the initial threshold to the POT algorithm to generate the final threshold vector $\boldsymbol{t}$;
Element-wise comparison of $\boldsymbol{s}_{\text{test}}$ and $\boldsymbol{t}$:
if $\boldsymbol{s}_{\text{test}}(i) < \boldsymbol{t}(i)$, then **then**
  $\boldsymbol{x}_i$ is a target, label $\boldsymbol{y}_i = 1$
else
  $\boldsymbol{x}_i$ is a clutter, label $\boldsymbol{y}_i = 0$
end if

5. Numerical Examples

In this section, we first introduce our experiment setup in Section 5.1, including experiment data, evaluation metrics, compared baselines, model hyper-parameters and hardware platform. We then evaluate and compare the performances of GM-CVAE and baselines in different clutter scenes in Section 5.2. Finally, we qualitatively analyze the proposed method in Section 5.3.

5.1. Experiment Setup

Experiment Data. In the simulation, the linear frequency modulated signal is chosen for the waveform of the ground-based warning radar, which operates at a wavelength of 0.23 m, a bandwidth of 2 MHz, an azimuth beam width of 2°, a pulse width of 0.15 ms, a pulse repetition period of 2.37 ms, a sampling frequency of 4 MHz, and 23 coherent pulses in one CPI. In order to fit different terrains with different data distributions, three types of clutter are designed for simulating complex clutter scenes, and Figure 5 shows their R-D spectrum. The x-axis represents 50 range cells of radar, and the y-axis represents the normalized Doppler frequency, which can be converted to the corresponding radial velocity. The red solid box in the figure represents the primary clutter area, and targets will locate in it if data is not dealt with MTI or AMTI in the data pre-processing stage. The green dotted box represents the secondary clutter area and targets will fall in it if MTI or AMTI is conducted in the data pre-processing stage. Figure 6a demonstrates the Doppler characteristics of three types of clutter. The center frequency and clutter spectrum width of three types of clutter are different, Type-2 is static clutter, while Type-1 and Type-3 are moving clutter with different speeds. Figure 6b exhibits the power intensity of three types of clutter partial range cells. It can be seen that the three types of clutter fluctuate violently in space, and it is difficult for the samples to satisfy the conditions of independent and identical distribution. Therefore, combined with the characteristics of the target data, we use each Doppler vector in the R-D spectrum as sample data, and 1D-CNN is used to extract sample features in our GM-CVAE. In the experiment, we have 2400 clutter instances in total and divide the dataset into two parts: the former 2000 for training and the latter 400 for testing. The ratios of target instances in the testing set is 0.05.

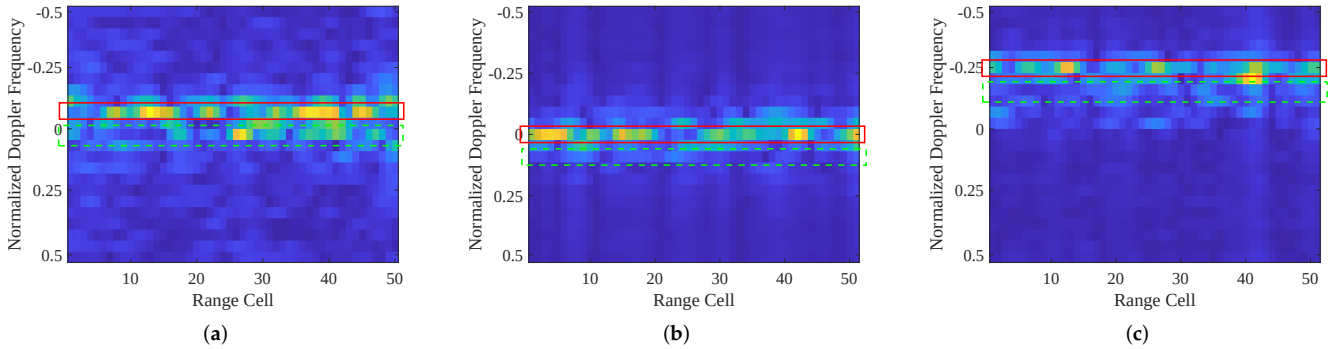


Figure 5. R-D spectrum of the three types of clutter. In the figure, the red solid box and the green dotted box, respectively, represent the area where the target is located: if the data is dealt with MTI or AMTI in the pre-processing stage, targets fall in the green dotted box; otherwise, targets fall in the red solid box. (**a**) Type-1; (**b**) Type-2; (**c**) Type-3.

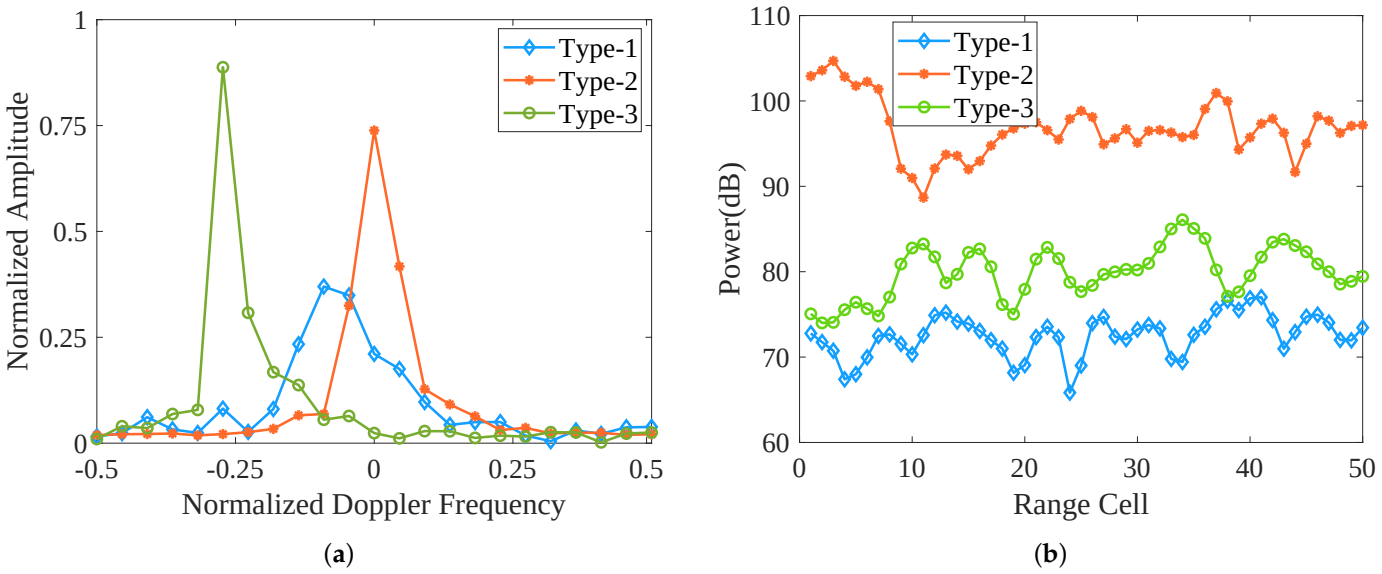


Figure 6. Characteristics of the three types of clutter. (**a**) Doppler characteristics; (**b**) Power comparison of three types of clutter part range cell.

Evaluation Metrics. We mainly employ two performance metrics to evaluate GM-CVAE and other baseline methods: detection probability (denoted as P_d) and the false alarm rate (denoted as P_{fa}). They can be shown as following [35]

$$\begin{aligned} P_d &= \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \\ P_{fa} &= 1 - \text{Precision} = \frac{\text{FP}}{\text{TP} + \text{FP}} \end{aligned} \tag{23}$$

Compared Baselines. To demonstrate the effectiveness of GM-CVAE, one-dimensional convolutional VAE (CVAE), classical CA-CFAR [3], GO-CFAR [7] and OS-CFAR [9] detectors as well as adaptive detectors, i.e., adaptive normalized matched filter (ANMF) [19] and ASD [27] are chosen as baselines. For ASD detector, we consider the rank-1 case (that is, the subspace signal model is represented as the product of the target amplitude and the steering vector). In other words, the performance of GM-CVAE is compared with the five classical detection methods and CVAE uses a simple Gaussian prior to learn the latent spatial distribution of the data.

Hyper-parameters. In our experiments, GM-CVAE is implemented based on Pytorch. The hyper-parameters of GM-CVAE are set empirically in our experiments as follows. Both the convolutional and deconvolutional layer L = 2, the kernel sizes and strides of which are $\{3 \times 1, 3 \times 1\}$ and $\{2 \times 1, 2 \times 1\}$, respectively. The dimension of z-space is 5 empirically and the components of categorical-space are related to the number of clutter types, and the value set is from $\{1, 2, 3\}$. For the adjusted POT method, the low quantile $= 0.2$ and $q = 10^{-4}$ [57]. GM-CVAE is trained with Adam optimizer [58], with a learning rate of 0.0005. We run 50 epochs for training, and the batch size is 64. More studies about the choice of hyper-parameters will be discussed later.

Hardware Platform Our experiments are conducted on a host with Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz accelerated by NVIDIA RTX 3080, with 10 GB VRAM of the graphic card.

5.2. Performance Comparison

In the range-Doppler domain, clutter or targets with different velocities occupy different Doppler cells, which is an important feature to effectively distinguish signals of different velocities. The Doppler frequency of the slow target is extremely low, and it often overlaps with the Doppler of low-frequency clutter. At this time, the target is submerged in the strong clutter, and feature extraction is difficult, thereby reducing the detection performance. During data pre-processing, the MTI filter can suppress the static zero-frequency ground clutter well, and has better detection performance for higher-speed moving targets, such as the echo of target 2 in Figure 7. However, when detecting a slow-moving target with a lower Doppler frequency, the MTI filter also greatly suppresses the moving target near the zero frequency, such as the attenuated target 1 echo in Figure 7. Likewise, the AMTI filter can better suppress motion clutter, but it can also suppress targets with similar clutter Doppler frequencies. To this end, two scenarios in target detection are considered, one is to employ the MTI or AMTI filter during data pre-processing, and the position of the target avoids the filter notch in the Doppler dimension (see the green dotted box in Figure 5) and is random in the range dimension; another, is to not utilize MTI or AMTI filter during data pre-processing, and the target is added to primary clutter area (See the red solid box in Figure 5).

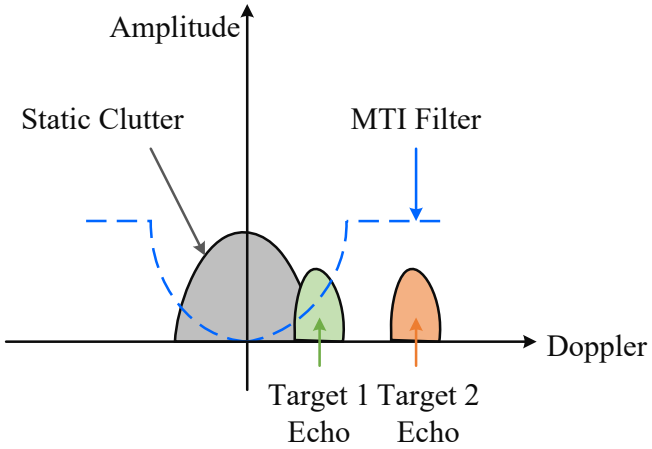


Figure 7. Schematic diagram of MTI filtering. It shows the relationship between filter notch and static clutter, and two targets in the Doppler domain.

5.2.1. Data Pre-Processing with MTI or AMTI

The size of the reference window with the aforementioned traditional detectors is 24, and the length of the guard cell is 2. A total of 10^4 Monte Carlo trials are performed. The selection of order k in OS-CFAR is 14 and the target obeys the Swerling type-0/5 distribution. Note that the comparison between network methods and other detectors: the actual false alarm rate of the test set is used as the constant false alarm rate of CA-CFAR,

GO-CFAR, OS-CFAR, ANMF and ASD methods after the network training is completed, and then detection probability of the two methods is compared. If the actual false alarm rate in network methods is zero, the constant false alarm rate in the traditional methods is set to 10^{-3}.

Performances in Single-clutter Scene. Taking Type-2 as an example, the detection probabilities of GM-CVAE, CVAE and the five traditional methods versus Signal to Noise Ratio (SNR) are presented in Figure 8a,b, respectively, where the noise power is about 59.3 dB. The actual false alarm rate of GM-CVAE, CVAE and the five traditional detection methods versus SNR are presented in Figure 8d,e, respectively. Figure 8c,f compare the detection and false alarm performance of GM-CVAE and CVAE, respectively. Specifically, GM-CVAE is far superior to CA-CFAR, GO-CFAR and OS-CFAR methods in terms of detection performance under low SNR conditions. The detection performance of ANMF and ASD in the single clutter scene is similar, and compared with the two detectors, our proposed GM-CVAE has a performance improvement of about 3 dB. Since ANMF is a coherent accumulator, it can improve the accumulation gain by accumulating multiple samples, and thus has better detection performance than the three CFAR detectors. From the perspective of false alarm, the five traditional detection methods are slightly higher than the GM-CVAE method, in which the false alarm value fluctuates reasonably. Similar conclusions occur for CVAE and CA-CFAR, GO-CFAR, OS-CFAR, ANMF and ASD methods. Furthermore, compared with Figure 8c,f, GM-CVAE performs slightly better than CVAE under single clutter background, indicating that GMM prior extracts more information than single Gaussian prior.

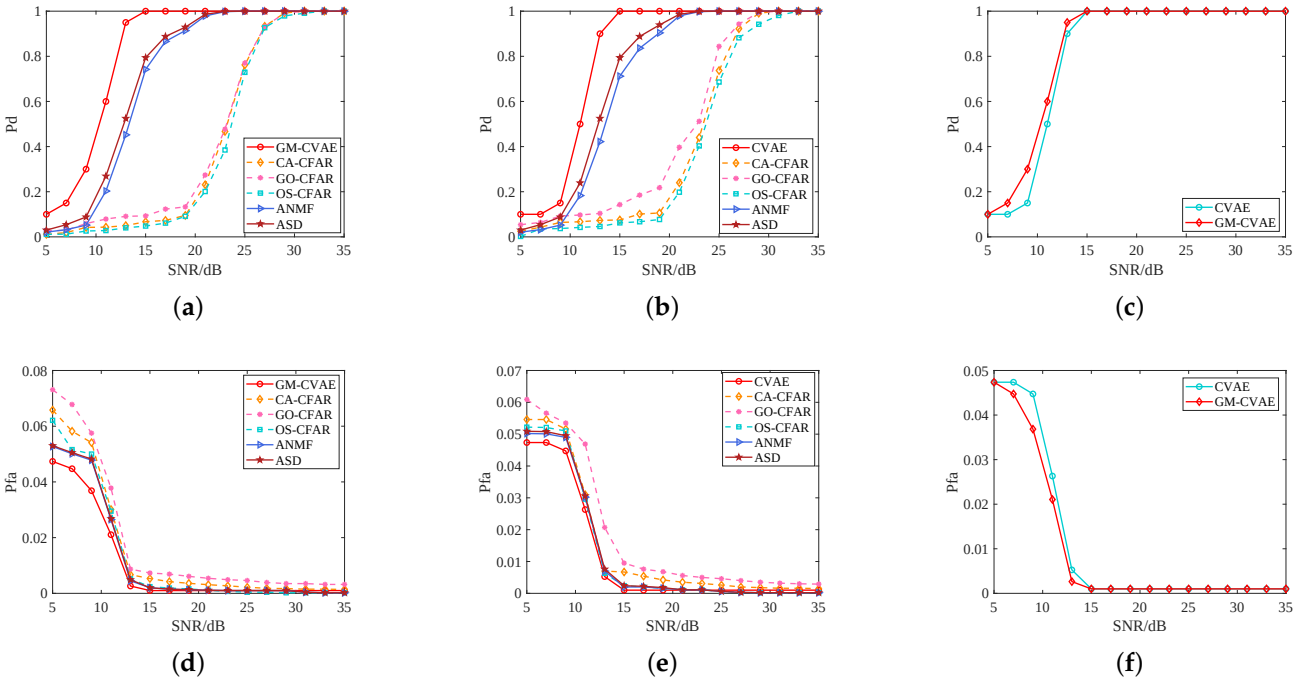


Figure 8. Single-scene performance for Type-2 clutter: Comparison of detection probability (**a**) between GM-CVAE and CA-CFAR, GO-CFAR, OS-CFAR, ANMF and ASD methods; (**b**) between CVAE and the five traditional methods; (**c**) between GM-CVAE and CVAE; Comparison of actual false alarm rate (**d**) between GM-CVAE and the five traditional methods; (**e**) between CVAE and the five traditional methods; and (**f**) between GM-CVAE and CVAE.

Performances in Complex-clutter Scene. To simulate the complexity of multiple clutter in radar irradiation area, different types of clutter regions are spliced according to the range dimension. The complex scene is shown in Figure 9, where the three types of clutter are concatenated in equal proportions. The noise powers of Type-1, Type-2 and

Type-3 clutter are 56.0 dB, 59.3 dB, and 64.7 dB, respectively. Figure 10 exhibits the detection performances and actual false alarm rates under complex clutter scene.

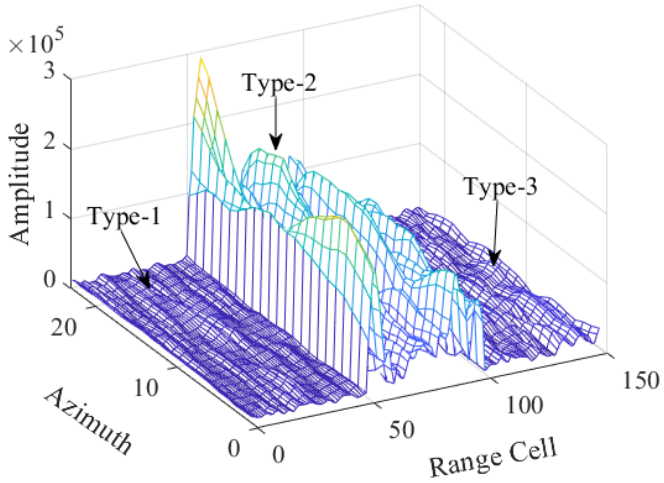


Figure 9. The three types of clutter splicing scene after pulse compression. In the schematic diagram, each clutter occupies 50 range cells.

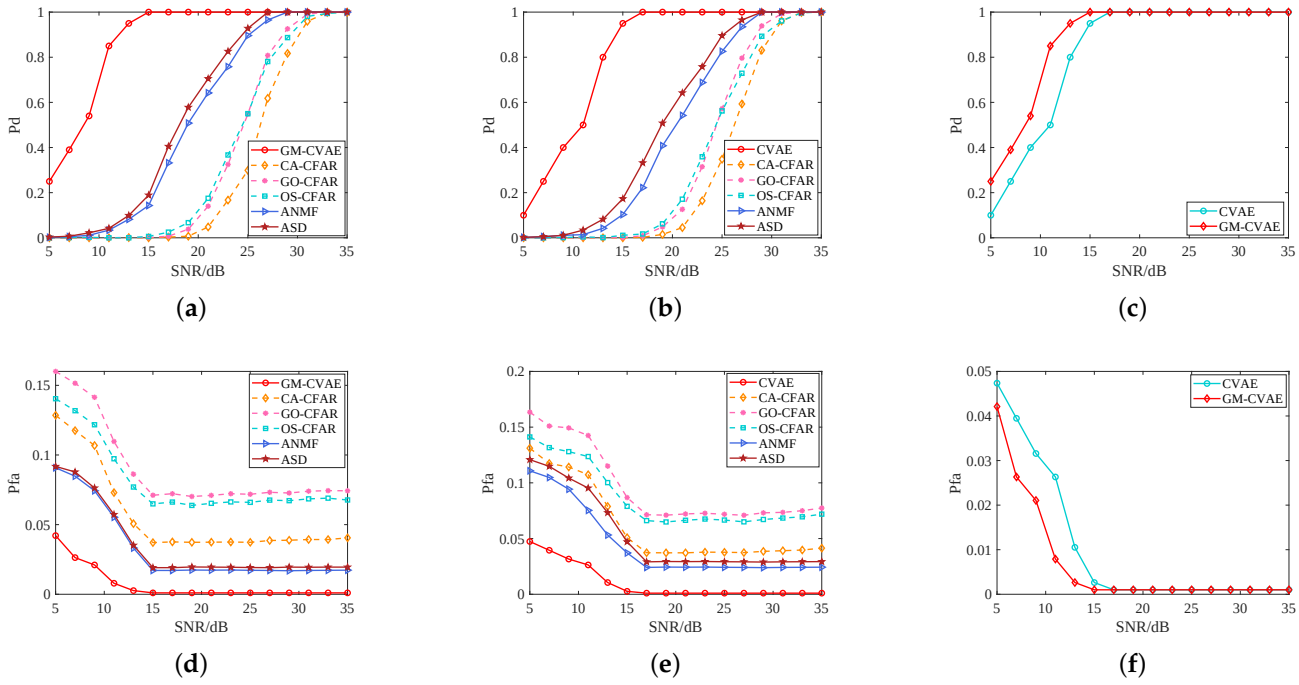


Figure 10. Complex-scene performance composed of clutter Type-1, Type-2 and Type-3: Comparison of detection probability (**a**) between GM-CVAE and CA-CFAR, GO-CFAR, OS-CFAR, ANMF and ASD methods; (**b**) between CVAE and the five traditional detection methods; (**c**) between GM-CVAE and CVAE; Comparison of actual false alarm rate (**d**) between GM-CVAE and the five traditional detection methods; (**e**) between CVAE and the five traditional detection methods; and (**f**) between GM-CVAE and CVAE.

GM-CVAE and CVAE in Figure 10a,b provide excellent detection performances while CA-CFAR, GO-CFAR and OS-CFAR detection methods suffer a performance degradation owing to the clutter edge effect. Compared with ANMF (or ASD) and GM-CVAE, the latter is improved by about 9 dB. This is because ANMF and ASD detectors rely on clutter covariance matrix estimation. When the clutter is heterogeneously distributed, it is difficult

to have a sufficient number of *i.i.d.* training samples. Furthermore, P_{fa} of CA-CFAR, GO-CFAR and OS-CFAR increases so significantly that it can no longer maintain the constant false alarm rate, our GM-CVAE continues to work robustly yet. Compared with the detection performances of GM-CVAE and CVAE methods in Figures 8c and 10c, it is evident from the result of comparison that the more complex the clutter distribution is, the more significant the P_d improvement of the proposed method is, i.e., the performance of the two methods in a single-clutter scene is close, while P_d is improved by about 2.5 dB under the complex clutter background. This phenomenon indicates that for complex data distribution, GM-CVAE can learn more abundant representational features than CVAE, realizes refined modeling of clutter, and reflects the superiority of GMM prior.

The SNR improvements of the proposed method under different clutter scenes when compared with the CA-CFAR method are summarized in Table 1. CVAE, OS-CFAR, GO-CFAR, ANMF and ASD methods are also considered for better comparison. Since the constant false alarm parameter in the five traditional methods come from the actual false alarm rate of the network method, CVAE and GM-CVAE are compared with CA-CFAR, respectively. Herein, we set the required SNR of the CA-CFAR method to satisfy the reference. In addition to the two clutter scenarios given above, Table 1 also quantitatively analyzes the SNR enhancement of the other two complex scenarios. The statistic in Table 1 demonstrates that our GM-CVAE provides excellent detection performances compared to traditional CA-CFAR, GO-CFAR, OS-CFAR, ANMF and ASD detection methods, both in single-clutter and complex-clutter scenes. Moreover, with the diversification of clutter distribution, the improvement of GM-CVAE detection performance is more evident.

Table 1. SNR Improvement of the Referred Methods with Respect to CA-CFAR.

	Method	**Type-2 (dB)**	**Type-1 $\rightarrow$[1] 2 (dB)**	**Type-2 $\rightarrow$ 3 (dB)**	**Type-1 $\rightarrow$ 2 $\rightarrow$ 3 (dB)**
case A [2]	GO-CFAR	+0.8	-1.7	-2.4	+2.4
	OS-CFAR	−0.7	−2.5	−3.0	0.0
	ANMF	+11	+8	+6.4	+5.1
	ASD	+12	+9	+7.5	+6.3
	CVAE	+13.7	+14	+13.4	+14.3
	GO-CFAR	0.0	−1.8	−2.0	+1.8
	OS-CFAR	-0.1	−2.6	−1.4	+1.0
	ANMF	+10.5	+9	+7.7	+8
	ASD	+11.5	+10	+8.8	+9
	GM-CVAE	**+14.0**	**+16.0**	**+15.3**	**+18.0**

[1] '$\rightarrow$' represents a splicing operation along the distance dimension; [2] case A represents 'Data pre-processing with MTI or AMTI'.

5.2.2. Data Pre-Processing without MTI or AMTI

Performances in Single-clutter Scene. In order to contrast with the previous experiments, we still choose Type-2 clutter as an example, the size of the reference window with the three CFAR methods is 24, and the length of the guard cell is 2. The selection of order k in OS-CFAR is 15 and the noise power is 52.4 dB. The comparison results of different methods are shown in Figure 11. The five traditional detectors suffer a performance degradation owing to the absence of clutter suppression compared to GM-CVAE (Figure 11a,d) and CVAE (Figure 11b,e). It also can be seen that our GM-CVAE increases by about 2 dB compared with that of CVAE under the same P_d (Figure 11c), the P_{fa} drops significantly at lower SNRs concurrently. This further verifies that the proposed GM-CVAE method can more effectively model the distribution of data in harsh clutter environments. In addition, regardless of the network methods or the traditional CFAR methods, the detection performance drops significantly compared to Figure 8, and the traditional CFAR methods almost completely fail at this time. This indicates that when the target velocity is

extremely low, it will be seriously affected by clutter, and it is extremely difficult to extract Doppler information.

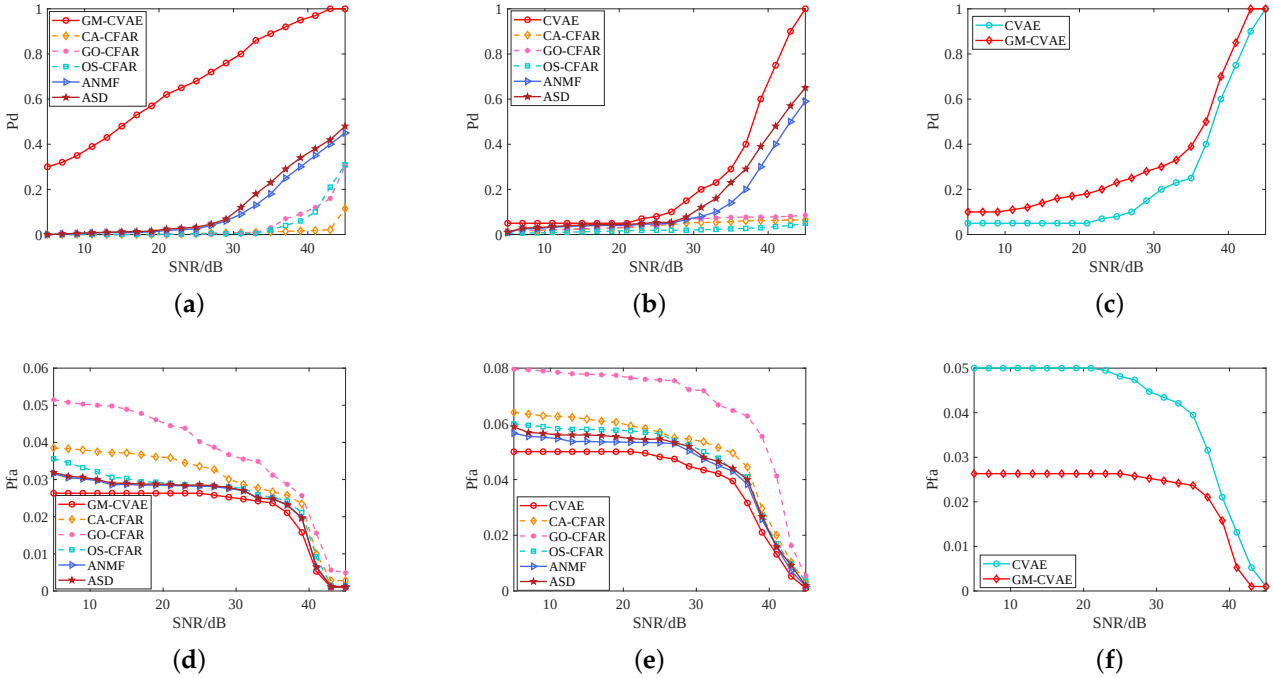


Figure 11. Single-scene performance for Type-2 clutter without MTI filter: Comparison of detection probability (**a**) between GM-CVAE and CA-CFAR, GO-CFAR, OS-CFAR, ANMF and ASD methods; (**b**) between CVAE and the five traditional detection methods; (**c**) between GM-CVAE and CVAE; Comparison of actual false alarm rate (**d**) between GM-CVAE and the five traditional detection methods; (**e**) between CVAE and the five traditional detection methods; and (**f**) between GM-CVAE and CVAE.

Performances in Complex-clutter Scene. The selection of order k in OS-CFAR is 16 and the noise power of the three regions is 53.8 dB, 52.4 dB, and 59.4 dB, respectively. Likewise, Figure 12 exhibits the performance curves under the background of complex clutter. With the dual effects of clutter filtering-free and clutter edge, CA-CFAR, GO-CFAR, OS-CFAR, ANMF and ASD detection methods work poorly compared with CVAE and GM-CVAE methods, and CA-CFAR has the worst effect. Compared to the three CFAR detectors, ANMF and ASD achieve slightly higher detection gains due to the coherent accumulation of samples. Moreover, the five traditional detectors need to sacrifice high SNR to detect the target. P_{fa} of the three CFAR methods multiplies approximately compared with CVAE and GM-CVAE methods, and the order k needs to be chosen carefully. In addition, our proposed GM-CVAE method in this scenario continues to work robustly, with higher detection probability and lower false alarm rate than CVAE. This further explains that the isotropic distribution of simple Gaussian in VAE fails to model the intrinsic multimodality of high dimensional data. In conclusion, our GM-CVAE method does have advantages for modeling complex clutter, and it can improve the performance of radar target detection.

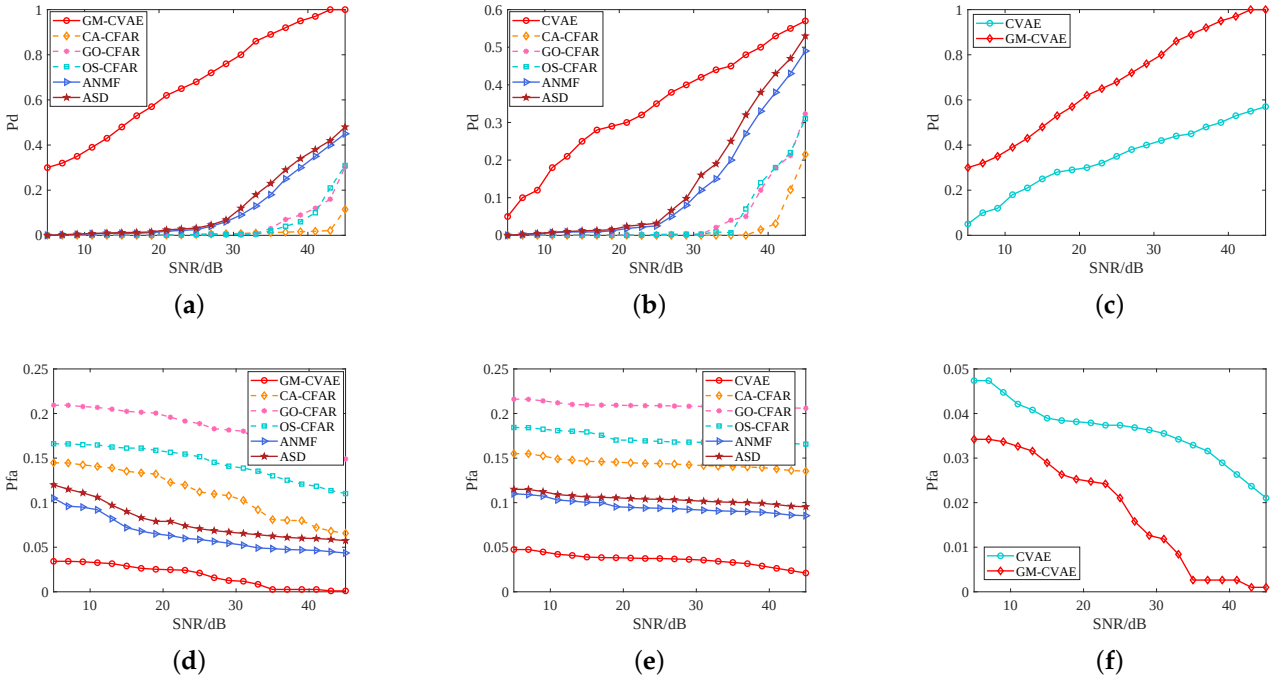


Figure 12. Complex clutter scene: (**a**) Comparison of detection probability between GM-CVAE and the five traditional detectors; (**b**) Comparison of detection probability between CVAE and the five traditional detectors; (**c**) Comparison of detection probability between GM-CVAE and CVAE; (**d**) Comparison of actual false alarm rate between GM-CVAE and the five traditional detectors; (**e**) Comparison of actual false alarm rate between CVAE and the five traditional detectors; and (**f**) Comparison of actual false alarm rate between GM-CVAE and CVAE.

5.3. Qualitative Analysis

The Influence of Model Parameters. In this part, we discuss the sensitivity of model parameters, namely the dimension of hidden stochastic variables (i.e., z). The reason we do not talk about the number of components of Gaussian mixture c is because the number of clutter types is assumed to be exactly known in each experiment.

GM-CVAE is a reconstruction-based model. For an input observation, GM-CVAE compresses it to a low dimension z-space representation and then uses the representation to reconstruct it. That is to say, in this process, the key features in the original data are retained and finally used for reconstruction; some unimportant features are eliminated. Therefore, if the dimension of the latent space z is too low, it means that there are too few key features, and it is easy to lose mode which will lead to the under-fitting of the model; if the dimension of z is too high, it means that more rare features will contribute to the model, making the search space larger and more saddle points present. This makes the learning rate of the model more slowly, and in addition, it will take up too much video memory. Based on this, when SNR = 11 dB, we set the dimension of z ranges from 2 to 8 in the experiment, observe its impact on the detection performance, and choose the most suitable dimension. It can be seen from Figure 13 that the detection probability P_d increases significantly when the dimension of z increases from 2 to 5. This means that adding key features improves the representation ability of the model. When the dimension of z is 5, the detection probability is the highest, which means that the key features retained at this time are enough to represent the original data. When we continue to increase the dimension of z, it will not bring significant improvement, but may reduce the performance of the model due to the increase in the search space and the increase in saddle points. Hence, the dimension of z is set to 5 in this paper.

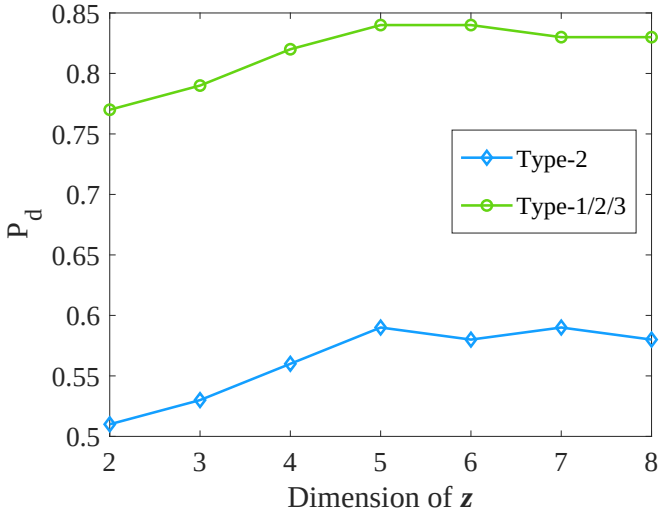


Figure 13. The P_d results of GM-CVAE by varying hyper-parameters z.

Visualization of z-space Representations. To illustrate that our model can fine model complex clutter distributions, here we visualize the 2D embedding in the original data domain and z-latent space on the GM-CVAE model for complex clutter scene using t-distributed stochastic neighborhood embedding (t-SNE) [59] method in Figure 14. Each dot in the figure indicates the latent variable of the observation vector, and each color represents a cluster group. The class boundaries of the training samples are blurred in the original data domain according to Figure 14a,c, while samples are separable in z-latent space and work well according to Figure 14b,d. This result indicates that our model can represent different clutter features with the mixture distributed latent variables, thus verifying that GM-CVAE can refine modeling.

The number of components of the Gaussian mixture $\boldsymbol{c}$ in the experiments is consistent with the set number of clutter types. If the number of components of Gaussian mixture $\boldsymbol{c}$ is changed in a multiple, i.e., more than one Gaussian component is used for each cluster to approximate the distribution, and the detection performance in single clutter and complex clutter scenes is considered, the results are shown in Figure 15a. The horizontal axis in Figure 15a represents the multiple expansion of the original Gaussian mixture component number. It can be seen that the detection performance has improved in both scenes, but the improvement effect is minimal. The analysis is as follows: objects in a cluster are similar or related to each other, and objects in different clusters are dissimilar or irrelevant. The more similar within clusters and the greater the gap between clusters, the better the effect. However, as shown in Figure 14b,d, our three kinds of data have a large distance in the $\boldsymbol{z}$ latent space itself. Even if it is divided into more clusters (as shown in Figure 15b, 6 clusters), the distance among the clusters does not widen each other. So the performance gain is quite limited.

Furthermore, to show intuitively the efficiency of our mixture distributed latent variable, four case studies of GM-CVAE are shown in Figure 16. Figure 16a shows the sequence of test samples in the background of single-clutter filtered by MTI or AMTI and Figure 16c displays the corresponding log-likelihood score. It can be seen from Figure 16a,c that even if some targets are similar to the clutter distribution, our model is still able to correctly learn the distribution of the clutter and gives the target points a lower log-likelihood. A similar conclusion can be observed in Figure 16b,d. In addition, the target in the red circle in Figure 16b is submerged in the clutter, but the corresponding score exhibits a clear peak, which further validates the superiority of our model in capturing complex clutter features. In Figure 16e,f, the phenomenon of some targets submerged in clutter is more prominent, but the GM-CVAE score remains stable. For the complex-clutter scenes in Figure 16b,f, GM-CVAE can model the non-stationary input assigned into different cluster

groups with different parameters, thus leading to a more stable score, while it exhibits considerable spikes in the regions of targets. Consequently, GM-CVE is robust and effective for modeling complex clutter.

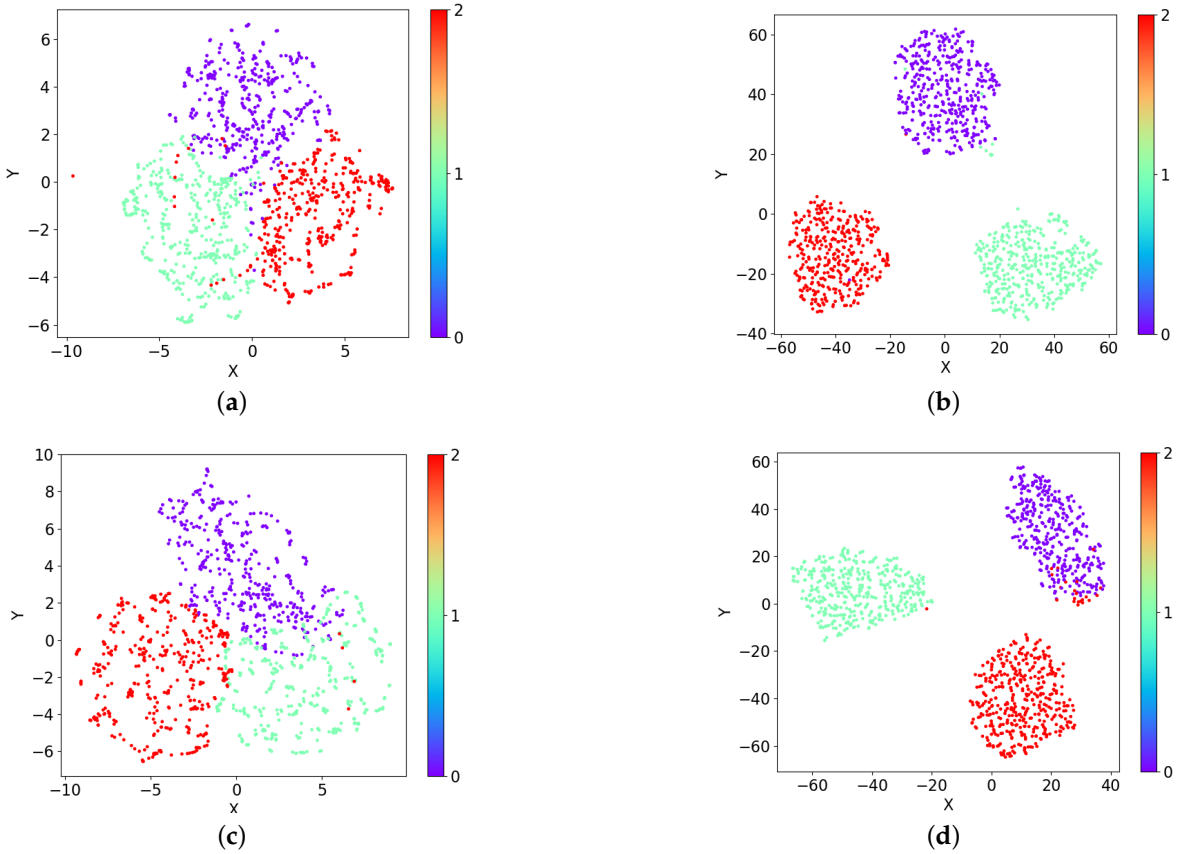


Figure 14. Representations of complex clutter scenes in different spaces. Case A: (**a**) original space representations, (**b**) *z*-space representations of GM-CVAE; case B: (**c**) original space representations, (**d**) *z*-space representations of GM-CVAE. Herein, case A represents data pre-processing with MTI or AMTI and case B indicates data pre-processing without MTI or AMTI.

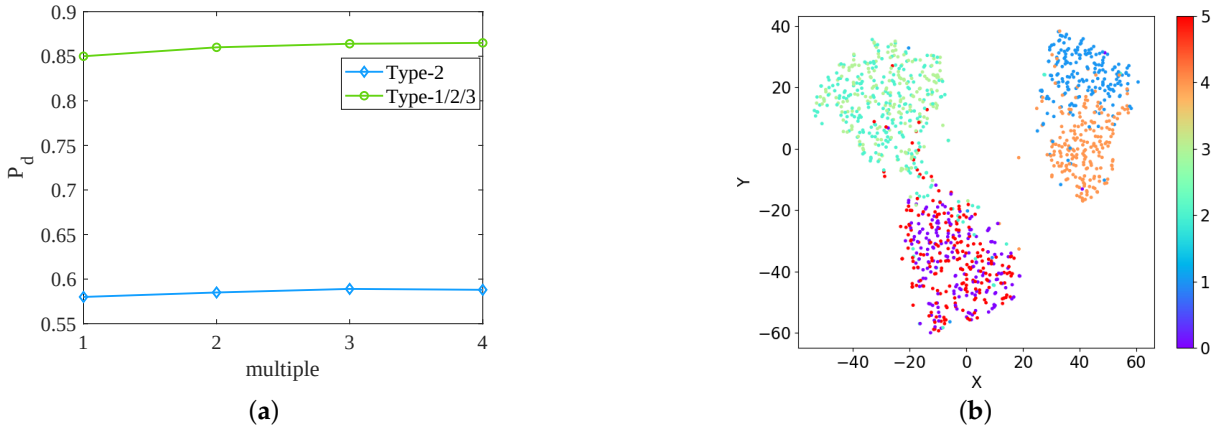
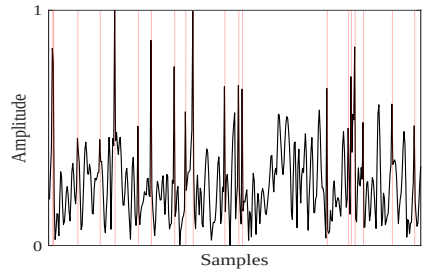
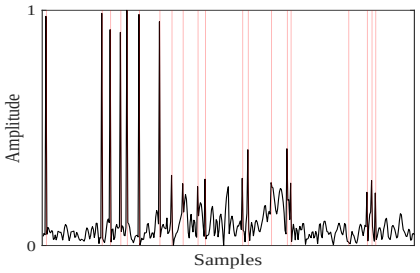


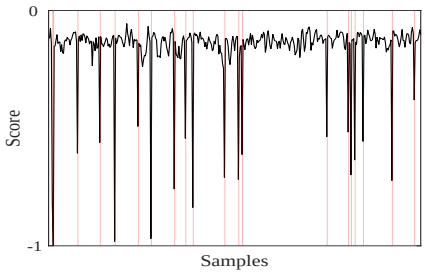
Figure 15. When the number of components of Gaussian mixture c changes in a multiple: (**a**) when SNR = 11 dB, the detection performance curves in single clutter and complex clutter scenes; (**b**) when SNR = 11 dB, considering complex clutter scene, the number of components of c is 6 (multiple is 2), the visualization results of the z-space at this time.



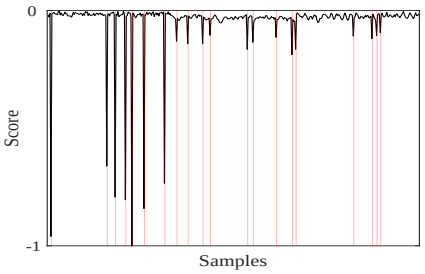
(**a**) Case A: single-clutter scene with SNR = 15 dB



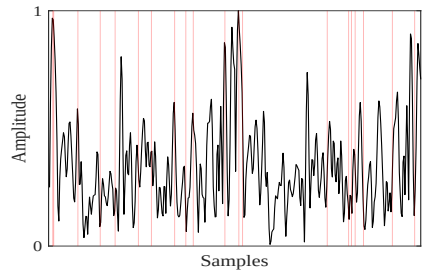
(**b**) Case A: complex-clutter scene with SNR = 15 dB



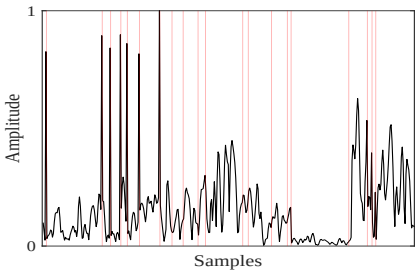
(**c**) Case A: log-likelihood score



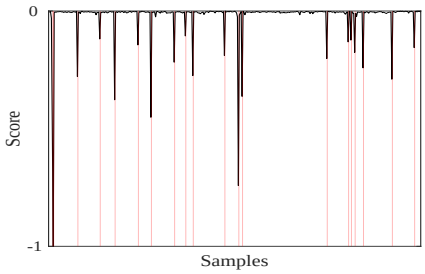
(**d**) Case A: log-likelihood score



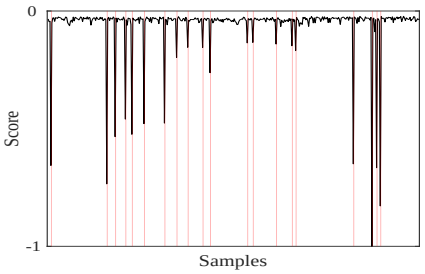
(**e**) Case B: single-clutter scene with SNR = 43 dB



(**f**) Case B: complex-clutter scene with SNR = 43 dB



(**g**) Case B: log-likelihood score



(**h**) Case B: log-likelihood score

Figure 16. (**a**–**h**) Case study of reconstruction score on test datasets with different SNR. Regions highlighted in red represent the ground-truth target. Case A denotes data pre-processing with MTI or AMTI and case B indicates data pre-processing without MTI or AMTI.

6. Conclusions

Traditional detection methods often face the problem of model mismatch under complex clutter backgrounds. Moreover, it is difficult to obtain target data and labels, which makes supervised deep learning methods difficult to implement. To solve this problem, We propose an unsupervised target detection approach (GM-CVAE). GM-CVAE assigns its latent variables into a mixture Gaussian distribution and thus it can model the normal patterns of clutter samples finely by learning their low-dimensional representations, fully

capturing Doppler shape features and various data distributions. Moreover, GM-CVAE provides an intuitive and effective way to detect targets based on reconstruction probability. The extensive experimental results demonstrate that GM-CVAE outperforms the traditional detection approaches and CVAE model with single Gaussian as prior in terms of detection probability and the false alarm rate, and show the superiority of one model that fits a great variety of clutter distributions. In addition, our model also achieves effective detection of slow targets in complex clutter backgrounds without sample labels. Although this work adopts simulated data for training, we believe that it has promising potential to apply unsupervised deep learning to radar target detection on complex clutter data. In addition, we expect to obtain sufficient training data with multiple distributions from the radar system which will definitely prompt the detection model to be more convincing. In the future, we will try to extend and evaluate our model with real-world radar data. Other signal features will be analyzed to improve the model performance to adapt complex clutter environment.

Author Contributions: Conceptualization, X.L. and B.C.; methodology, X.L., B.C. and W.C.; data curation, X.L. and B.C.; formal analysis, X.L. and W.C.; writing—original draft preparation, X.L.; visualization, X.L. and W.C.; supervision, B.C., P.W. and H.L.; funding acquisition, B.C. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant U21B2006 and Grant 61771361; in part by Shaanxi Youth Innovation Team Project; in part by the 111 Project under Grant B18039; and in part by the Thousand Young Talent Program of China.

Data Availability Statement: Not applicable.

Acknowledgments: B. Chen acknowledges the support of NSFC (U21B2006 and 61771361), Shaanxi Youth Innovation Team Project, the 111 Project (No. B18039) and the Program for Oversea Talent by Chinese Central Government. W. Chen acknowledges the support of National Key Lab Fund project.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANMF	Adaptive normalized matched filter
ASD	Adaptive subspace detector
CFAR	Constant false alarm rate
CA-CFAR	Cell-averaging CFAR
GO-CFAR	The greatest of selection CFAR
OS-CFAR	The order statistic CFAR
CUT	Cell under test
CNN	Convolutional neural network
GMM	Gaussian mixture model
VAE	Variational autoencoder
CVAE	Variational autoencoder with 1D-CNN
GM-CVAE	Gaussian mixture variational autoencoder with 1D-CNN
EVT	Extreme value theory
POT	Peaks over threshold

References

1. Yan, J.; Liu, H.; Pu, W.; Liu, H.; Liu, Z.; Bao, Z. Joint Threshold Adjustment and Power Allocation for Cognitive Target Tracking in Asynchronous Radar Network. *IEEE Trans. Signal Process.* **2017**, *65*, 3094–3106. [CrossRef]
2. Zhang, H.; Liu, W.; Shi, J.; Fei, T.; Zong, B. Joint Detection Threshold Optimization and Illumination Time Allocation Strategy for Cognitive Tracking in a Networked Radar System. *IEEE Trans. Signal Proc.* **2022**, *126*, 1–15. [CrossRef]
3. Finn, H.M. Adaptive Detection Mode with Threshold Control as A Function of Spatially Sampled Clutter Level Estimates. *RCA Rev.* **1968**, *29*, 414–465. [CrossRef]
4. Weiss, M. Analysis of Some Modified Cell-Averaging CFAR Processors in Multiple-target Situations. *IEEE Trans. Aeros. Electron. Syst.* **1982**, *18*, 102–114. [CrossRef]

5. Gandhi, P.; Kassam, S. Analysis of CFAR Processors in Nonhomogeneous Background. *IEEE Trans. Aerosp. Electron. Syst.* **1988**, *24*, 427–445. [CrossRef]
6. Hansen, V.G. Constant False Alarm Rate Processing in Search Radars. In Proceedings of the IEE Conference on Radar-Present and Future, London, UK, 23–25 October 1973; pp. 325–332.
7. Hansen, V.G.; Sawyers, J.H. Detectability Loss Due to "Greatest Of" Selection in a Cell-Averaging CFAR. *IEEE Trans. Aeros. Electron. Syst.* **1980**, *16*, 115–118. [CrossRef]
8. Rohling, H. Radar CFAR Thresholding in Clutter and Multiple Target Situations. *IEEE Trans. Aerosp. Electron. Syst.* **1983**, *19*, 608–621. [CrossRef]
9. Elias Fusté, A.; de Mercado, G.G.; de los Reyes, E. Analysis of Some Modified Ordered Statistic CFAR: OSGO and OSSO CFAR. *IEEE Trans. Aerosp. Electron. Syst.* **1990**, *26*, 197–202. [CrossRef]
10. Pourmottaghi, A.; Taban, M.; Gazor, S. A CFAR Detector in A Nonhomogenous Weibull Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 1747–1758. [CrossRef]
11. Zhang, X.; Zhang, R.; Sheng, W.; Ma, X.; Han, Y.; Cui, J.; Kong, F. Intelligent CFAR Detector for Non-homogeneous Weibull Clutter Environment Based on Skewness. In Proceedings of the 2018 IEEE Radar Conference (RadarConf18), Oklahoma City, OK, USA, 23–27 April 2018; [CrossRef]
12. Roy, L.; Kumar, R. Accurate K-distributed Clutter Model for Scanning Radar Application. *IET Radar Sonar Navig.* **2010**, *4*, 158–167. [CrossRef]
13. Yang, Y.; Xiao, S.p.; Feng, D.j.; Zhang, W.m. Modelling and Simulation of Spatial-temporal Correlated K Distributed Clutter for Coherent Radar Seeker. *IET Radar Sonar Navig.* **2014**, *8*, 1–8. [CrossRef]
14. Conte, E.; Longo, M. Characterisation of Radar Clutter as A Spherically Invariant Random Process. *IEE Proc. Part F* **1987**, *134*, 191–197. [CrossRef]
15. Sangston, K.J.; Gini, F.; Greco, M.V.; Farina, A. Structures for Radar Detection in Compound Gaussian Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1999**, *35*, 445–458. [CrossRef]
16. Gini, F.; Farina, A. Vector Subspace Detection in Compound-Gaussian Clutter. Part I: Survey and New Results. *IEEE Trans. Aerosp. Electron. Syst.* **2002**, *38*, 1295–1311. [CrossRef]
17. Gini, F.; Farina, A.; Montanari, M. Vector Subspace Detection in Compound-Gaussian Clutter. Part II: Performance Analysis. *IEEE Trans. Aerosp. Electron. Syst.* **2002**, *38*, 1312–1323. [CrossRef]
18. Xu, S.W.; Shui, P.L.; Cao, Y.H. Adaptive Range-spread Maneuvering Target Detection in Compound-Gaussian Clutter. *Digit. Signal Process.* **2015**, *36*, 46–56. [CrossRef]
19. Conte, E.; Lops, M.; Ricci, G. Adaptive Detection Schemes in Compound-Gaussian Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1998**, *34*, 1058–1069. [CrossRef]
20. Chen, B.; Varshney, P.K.; Michels, J.H. Adaptive CFAR Detection for Clutter-edge Heterogeneity using Bayesian Inference. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 1462–1470. [CrossRef]
21. Zaimbashi, A. An Adaptive Cell Averaging-based CFAR Detector for Interfering Targets and Clutter-edge Situations. *Digit. Signal Process.* **2014**, *31*, 59–68. [CrossRef]
22. Doyuran, U.C.; Tanik, Y. Expectation Maximization-based Detection in Range-heterogeneous Weibull Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 3156–3166. [CrossRef]
23. Meng, X. Rank Sum Nonparametric CFAR Detector in Nonhomogeneous Background. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *57*, 397–403. [CrossRef]
24. Hua, X.; Ono, Y.; Peng, L.; Xu, Y. Unsupervised Learning Discriminative MIG Detectors in Nonhomogeneous Clutter. *IEEE Trans. Commun.* **2022**, *70*, 4107–4120. [CrossRef]
25. Aubry, A.; De Maio, A.; Pallotta, L.; Farina, A. Covariance Matrix Estimation via Geometric Barycenters and Its Application to Radar Training Data Selection. *IET Radar Sonar Navig.* **2013**, *7*, 600–614. [CrossRef]
26. Wang, Z.; Li, G.; Chen, H. Adaptive Persymmetric Subspace Detectors in the Partially Homogeneous Environment. *IEEE Trans. Signal Process.* **2020**, *68*, 5178–5187. [CrossRef]
27. Kraut, S.; Scharf, L.; McWhorter, L. Adaptive Subspace Detectors. *IEEE Trans. Signal Process.* **2001**, *49*, 1–16. [CrossRef]
28. Liu, J.; Sun, S.; Liu, W. One-Step Persymmetric GLRT for Subspace Signals. *IEEE Trans. Signal Process.* **2019**, *67*, 3639–3648. [CrossRef]
29. Bidart, R.; Wong, A. Affine Variational Autoencoders. In Proceedings of the International Conference on Image Analysis and Recognition, Waterloo, ON, Canada, 27–29 August 2019; pp. 461–472.
30. Cai, F.; Ozdagli, A.I.; Koutsoukos, X. Detection of Dataset Shifts in Learning-Enabled Cyber-Physical Systems using Variational Autoencoder for Regression. In Proceedings of the 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), Victoria, BC, Canada, 10–12 May 2021; pp. 104–111. [CrossRef]
31. Liu, N.; Xu, Y.; Ding, H.; Xue, Y.; Guan, J. High-dimensional Feature Extraction of Sea Clutter and Target Signal for Intelligent Maritime Monitoring Network. *Comput. Commun.* **2019**, *147*, 76–84. [CrossRef]
32. Lopez-Risueno, G.; Grajal, J.; Diaz-Oliver, R. Target Detection in Sea Clutter using Convolutional Neural Networks. In Proceedings of the 2003 IEEE Radar Conference (Cat. No. 03CH37474), Huntsville, AL, USA, 5–8 May 2003; pp. 321–328. [CrossRef]

33. Jing, H.; Cheng, Y.; Wu, H.; Wang, H. Adaptive Network Detector for Radar Target in Changing Scenes. *Remote Sens.* **2021**, *13*, 3743. [CrossRef]
34. Wang, L.; Tang, J.; Liao, Q. A Study on Radar Target Detection Based on Deep Neural Networks. *IEEE Sens. Lett.* **2019**, *3*, 1–4. [CrossRef]
35. Xie, Y.; Tang, J.; Wang, L. Radar Target Detection using Convolutional Neutral Network in Clutter. In Proceedings of the 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 11–13 December 2019; pp. 1–6. [CrossRef]
36. Kan, M.; Shan, S.; Chen, X. Bi-shifting Auto-encoder for Unsupervised Domain Adaptation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3846–3854. [CrossRef]
37. Rostami, M. Lifelong Domain Adaptation via Consolidated Internal Distribution. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 11172–11183. [CrossRef]
38. Lasloum, T.; Alhichri, H.; Bazi, Y.; Alajlan, N. SSDAN: Multi-Source Semi-Supervised Domain Adaptation Network for Remote Sensing Scene Classification. *Remote Sens.* **2021**, *13*, 3861. [CrossRef]
39. Zhang, Y.; Shen, D.; Wang, G.; Gan, Z.; Henao, R.; Carin, L. Deconvolutional paragraph representation learning. *arXiv* **2017**, arXiv:1708.04729.
40. Xie, R.; Sun, Z.; Wang, H.; Li, P.; Rui, Y.; Wang, L.; Bian, C. Low-Resolution Ground Surveillance Radar Target Classification Based on 1D-CNN. In Proceedings of the Eleventh International Conference on Signal Processing Systems, Chengdu, China, 15–17 November 2019; Volume 11384, pp. 199–204. [CrossRef]
41. Xie, R.; Dong, B.; Li, P.; Rui, Y.; Wang, X.; Wei, J. Automatic Target Recognition Method For Low-Resolution Ground Surveillance Radar Based on 1D-CNN. In Proceedings of the Twelfth International Conference on Signal Processing Systems, Xi'an, China, 23–26 July 2021; Volume 11719, pp. 48–55. [CrossRef]
42. Su, Y.; Zhao, Y.; Sun, M.; Zhang, S.; Wen, X.; Zhang, Y.; Liu, X.; Liu, X.; Tang, J.; Wu, W.; et al. Detecting Outlier Machine Instances Through Gaussian Mixture Variational Autoencoder with One Dimensional CNN. *IEEE Trans. Comput.* **2022**, *71*, 892–905. [CrossRef]
43. Greco, M.S.; Watts, S. Radar Clutter Modeling and Analysis. In *Academic Press Library in Signal Processing*; Elsevier: Amsterdam, The Netherlands, 2014; Volume 2, pp. 513–594. [CrossRef]
44. Billingsley, J.B. *Low-Angle Radar Land Clutter: Measurements and Empirical Models*; United States of America by William Andrew Publishing: Norwich, NY, USA, 2002. [CrossRef]
45. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–58. [CrossRef]
46. Chalapathy, R.; Chawla, S. Deep Learning for Anomaly Detection: A Survey. *arXiv* **2019**, arXiv:1901.03407.
47. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2014**, arXiv:1312.6114.
48. Jing, H.; Cheng, Y.; Wu, H.; Wang, H. Radar Target Detection with Multi-Task Learning in Heterogeneous Environment. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]
49. Chung, J.; Kastner, K.; Dinh, L.; Goel, K.; Courville, A.C.; Bengio, Y. A Recurrent Latent Variable Model for Sequential Data. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015.
50. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
51. Guo, Y.; Liao, W.; Wang, Q.; Yu, L.; Ji, T.; Li, P. Multidimensional Time Series Anomaly Detection: A GRU-based Gaussian Mixture Variational Autoencoder Approach. In Proceedings of the ACML 2018: The 10th Asian Conference on Machine Learning, Beijing, China, 14–16 November 2018. [CrossRef]
52. Blei, D.M.; Jordan, M.I. Variational Inference for Dirichlet Process Mixtures. *Bayesian Anal.* **2006**, *1*, 121–143. [CrossRef]
53. Jang, E.; Gu, S.; Poole, B. Categorical Reparameterization with Gumbel-softmax. *arXiv* **2016**, arXiv:1611.01144.
54. Shi, W.; Zhou, H.; Miao, N.; Zhao, S.; Li, L. Fixing Gaussian Mixture VAEs for Interpretable Text Generation. *arXiv* **2019**, arXiv:1906.06719.
55. Su, Y.; Zhao, Y.; Niu, C.; Liu, R.; Sun, W.; Pei, D. Robust Anomaly Detection for Multivariate Time Series Through Stochastic Recurrent Neural Network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2828–2837. [CrossRef]
56. Xu, H.; Chen, W.; Zhao, N.; Li, Z.; Bu, J.; Li, Z.; Liu, Y.; Zhao, Y.; Pei, D.; Feng, Y.; et al. Unsupervised Anomaly Detection via Variational Auto-encoder for Seasonal KPIs in Web Applications. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 187–196. [CrossRef]
57. Siffer, A.; Fouque, P.A.; Termier, A.; Largouet, C. Anomaly Detection in Streams with Extreme Value Theory. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1067–1075. [CrossRef]
58. Basu, S. Analyzing Alzheimer's Disease Progression from Sequential Magnetic Resonance Imaging Scans using Deep Convolutional Neural Networks. Master's Thesis, McGill University, Montreal, QC, Canada, 2019. [CrossRef]
59. Van Der Maaten, L. Accelerating t-SNE using Tree-based Algorithms. *J. Mach. Learn. Res.* **2014**, *15*, 3221–3245. [CrossRef]