A stacked discriminative auto-encoder based on center loss for radar target HRRP recognition

Rui Li*, Xiaodan Wang*, Wen Quan, Guoling Zhang, Qian Xiang
College of air and missile defense, air force engineering university
*
hejiuxing@stu.xidian.edu.cn  *Corresponding author’s e-mail: afeu_wang@163.com

Abstract. In the radar automatic target recognition filed, extracting representative features from the radar high resolution range profile (HRRP) is the key issues, which determines the accuracy and reliability of radar target recognition. In this paper, we propose a novel stacked discriminative auto-encoder(S-disAE), the center loss is integrated into auto-encoder, it can force the learned feature with large distance close to their class feature representation center, so as to reduce the intra-class variations while keeping the features of different classes separable. In pre-training stage of our discriminative auto-encoder(disAE), we combine the mean square error loss and center loss to learn the main factors of input raw data and minimize the intra-class variations of the feature, in fine-tuning stage, we also combine the softmax cross entropy loss and center loss to improve the classification accuracy. We conduct several experiments on the simulated HRRP data, the results demonstrate that the proposed algorithm can extract discriminant features and improve the HRRP recognition accuracy.

1. Introduction
A high-resolution range profile (HRRP) is amplitude of the coherent summations of the complex returns from target scattering centers in each range cell, which represents the projection of the complex returned echoes from the target scattering centers along the radar Line-Of-Sight (LOS) [1-3]. Due to the easy acquisition and processing of HRRP, HRRP plays an important role in the wide-band radar automatic target recognition (RATR). HRRP reflects abundant information of the scatters contained in the target, so radar automatic target recognition based on HRRP has received intensive attention [1-7].

Traditional HRRP RATR methods extract hand-crafted features and apply shallow architectures machine learning algorithm to recognize HRRP [4-7]. In recent years, deep neural networks have been developed very quickly, deep learning algorithms automatically learn high-level abstractions of data by using deep architectures composed of multiple nonlinear transformations. Deep learning techniques have demonstrated powerful representation learning capability and achieved remarkable successes in various applications, including computer vision [17], natural language processing [18], speech recognition [19], and object detection [20]. The large-scale deep neural networks have also been successfully applied for radar for the traditional RATR field in recent years, such as deep belief network (DBN) and stacked auto-encoder(SAE). Yan et al. [8] proposed an HRRP recognition method by sparse denoising autoencoder (SDAE) to extract the robust features of HRRP. Zhao et al. [9] used stacked autoencoder(SAE) to learn the characteristics of HRRP through unsupervised training, and used ELM as a classifier to achieve better results and improve the learning speed when the number of samples was very small. Pan et al. [10] used t-SNE and sampling method to provide well segmented and balanced HRRP data, and then explore discriminant deep belief network (DDBN) for HRRP recognition. Feng et al. [11] proposed a stacked corrective auto-encoders for HRRP recognition, they used the average profile of the
HRRP as a corrective output, and considered the covariance matrix of each HRRP frame for loss function establishing. Duet et al. [12] considered the labels information when building conditional variational auto-encoder to yield more discriminative and useful features for HRRP target recognition, at the same time, they factorized the parameters to effectively reduce model parameters.

We notice that many auto-encoder based methods are used for HRRP recognition, but auto-encoder is a typical unsupervised neural network, the representation is extracted in an unsupervised manner without the class information, we think that the learned features based on unsupervised auto-encoder are not always optimal for data separation. Generally, supervised learning method can learn more discriminative features that are useful for classification tasks. In this paper, we introduce the center loss into auto-encoder and propose a novel stacked discriminative autoencoder (S-disAE), the center loss will push the learned feature with large distance towards the center of its corresponding class, so as to reduce the intra-class variations while keeping the features of different classes separable. We combine the mean square error loss and center loss to minimize the intra-class variations when pre-training our discriminative auto-encoder, in fine-tuning stage, we also combine the softmax cross entropy loss and center loss to improve the classification accuracy. Experimental results on the HRRP simulation data demonstrate that the proposed algorithm can extract discriminant features and improve the HRRP recognition accuracy.

2. Related works
Auto-encoder [13] aims at learning hidden representation of data in an unsupervised manner, it consists of an encoder module and decoder module, as shown in Figure 1.

![Fig.1. The generic flowchart of an auto-encoder.](image)

The encoder aims at mapping the input \( x \) to the latent representation space \( h \): \[
    h = f(x) = f(Wx + b)
\] (1)

Where \( f \) is non-linear function, such as the sigmoid function, and ReLu function, \( W \) is the weight matrix, and \( b \) is bias vector.

The decoder attempts to transform the latent representation \( h \) to output \( z \) for reconstruction:
\[
    z = g(h) = g(W'h + b')
\] (2)

Where \( g \) is activation function, \( W' \) is weight matrix, \( b' \) is bias vector.

Auto-encoder tries to reconstruct the input \( x \) with the output \( z \), so the objective function of AE is:
\[
    J_{AE}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \| x_i - g(f(x_i)) \|^2
\] (3)

Where \( \theta = (W, b, b') \) is the parameters set, \( N \) is the samples number of training dataset \( X \).

There are many other variants of auto-encoder has been investigated to learn underlying explanatory representation of data, such as stacked sparse auto-encoder(SSpAE), stacked denoising auto-encoder(SDAE), and stacked contractive auto-encoder (SCAE). The SSpAE [14] imposed the KL sparsity constraint on the hidden units to learn sparse representation. The SDAE [15] reconstructed the clean input with the corrupted data to capture the robust features of the data. The SCAE [16] forced the Frobenius norm of the Jacobean matrix of the hidden representation to extract robust features to slight
changes of input. Once the auto-encoder is trained, the encoder module is stacked greedily layer-by-
layer to construct the stacked auto-encoder(SAE), the SAE can learn high level abstract and complicated
representations of input, on the top of last layer, a softmax classifier is added for supervised fine-tuning
the whole deep network. In the fine-tuning stage, the cross entropy loss function is used for predict target:

$$J_{SAE}(\theta) = \frac{1}{N} \sum_{i=1}^{N} [-y_i \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i)]$$

Where $\hat{y}_i$ is the predict output of $i$-th sample, $y_i$ is the true label of $i$-th sample.

3. Proposed discriminative auto-encoder based on center loss
We usually train several auto-encoders layer-by-layer in an unsupervised manner and then stack them
to construct stacked auto-encoder. In the pre-training stage, the auto-encoder can capture the underlying
factors for the observed input, but we think the underlying features of the data are not always optimal
for classification. So, in this paper we investigate a stacked discriminative auto-encoder by introduce
the center loss to learn representation that are more discriminative for data separation.

For a classification problem with $N$ training samples and $m$ classes, suppose that the dataset is
\( \{X, Y\} = \{ (x_i, y_{ij}) | x_i \in \mathbb{R}^d, y_{ij} \in \mathbb{R}^m \}_{i=1}^{N} \), where $x_i$ is the $i$-th training data, $d$ is the dimension of data, $y_{ij}$ is
the corresponding one-hot label vector of $i$-th sample. The objective of the proposed discriminative auto-
encoder(dis-AE) method can be denoted as:

$$J_{AE}(\theta) = \frac{1}{2N} \sum_{x \in X} \| x - g(f(x)) \|^2 + \lambda \frac{1}{2N} \sum_{x \in X} \| f(x) - C_{y_{ij}} \|^2$$

The first term is the mean square error (MSE) reconstruction loss, the second term is the center loss,
$\lambda$ is a balance factor. $g(f(x))$ is the feature of hidden layer corresponding to sample $x_i$, $C_{y_{ij}}$ is
the feature class center corresponding to class of $i$-th sample. The center loss function [22] will punish the
feature with large distance to the center of its corresponding class, so as to reduce the intra-class
variations while keeping the features of different classes separable, so the center loss can be used to
extract the feature that are effective for the classification tasks. Thus, our method combines the
reconstruction loss and the center loss to both balance extracting underlying pattern, as well as providing
discriminative features.

In the fine-tuning stage, traditional stacked auto-encoder use the softmax cross entropy loss, it is
prone to cause the phenomenon that the intra-class distance even larger than the inter-class distance
between features, which will lead to the unsatisfactory recognition result. In order to overcome this
phenomenon and reduce the intra-class variations, we also integrate the center loss into stacked auto-
encoder to build our stacked discriminative auto-encoder (S-disAE). So in the fine-tuning stage, the
objective function of S-disAE can be denoted as:
\[ J_{\text{sub}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left[ y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \right] + \lambda \frac{1}{N} \sum_{i=1}^{N} \| F(x_i) - C_i \|^2 \]  

(6)

Where, \( F(x_i) \) is the feature of last hidden layer corresponding to sample \( x_i \), \( C_i \) is the feature class center corresponding to class of \( x_i \) belongs to, \( \lambda \) is a balance factor. In short, both in the pre-training stage and the fine-tuning stage, we add the center loss to reduce the intra-class variations and learn discriminative features. Therefore, our method will improve the discriminative feature extraction power and improve the classification performance. The whole framework of our method is shown in figure 2.

4. Experiments and results

4.1. Experimental setting

In this paper, we focus on ballistic missiles (BM) HRRP target recognition. In order to simulate the BM target HRRP data, we use FEKO and MATLAB joint simulation to get five types of target simulation radar returned echo data. The simulation targets are shown in Figure 3. After completed the simulation, we use Inverse Fourier transform to convert the FEKO simulated complex returned echo data into 256 dimension HRRP. At last, we can get 3601 HRRP samples of different degrees for each target, the examples HRRP at 40 degree of five different targets are given in figure 4. The total samples of the dataset are 18005, we random construct two HRRP data sets A and B. Data set A contains 12005 training data and 6000 test samples, and data set B contains 6005 training data and 12000 test samples.

In this paper, we train three discriminative auto-encoders and stacked them to construct the stacked discriminative auto-encoder for HRRP recognition. The final network is set as 256-600-300-150-5, the activation function is ReLu. We employ the Adam algorithm to optimize the model. Each layer is trained for 100 epochs, the mini-batch size is set as 256, learning rate is 0.001. We implement model using TensorFlow framework. All experiments are performed on a desktop computer equipped with an Intel Core i7 4770 CPU, 64 GB RAM, and single NVIDIA GeForce GTX 1080Ti GPU.
4.2. Experiments on the parameter influence of $\lambda$

For our method, there is a parameter $\lambda$. In order to analyze the impact of $\lambda$ on HRRP recognition rate, we chosen $\lambda$ from the range $[0.0001, 0.001, 0.01, 0.1, 1]$. For simplicity, in the pre-training and fine-tuning stage, $\lambda$ is set as same. The change curves of the recognition rate on the dataset B of our S-disAE under different $\lambda$ rates are compared in in Figure 5.

As we can see from the figure, the parameter $\lambda$ have relative large impact on HRRP dataset, the recognition accuracy of our method increases firstly and then decreases when $\lambda$ change from 0.0001 to 1. When $\lambda$ =0.001, the recognition accuracy reaches the optimal value. We also notice that when $\lambda$ =1 and $\lambda$ =0.1 , the recognition results are lower, so large $\lambda$ will influence the results of our method, parameter selection is important to achieve better performance.

4.3. Classification performance evaluation

In this section, we compare the classification performance of our method with SAE, SSsAE, SDAE and DBN, they have same network structure as our S-disAE, we also compare with support vector machine (SVM) directly applying on the original HRRP data. The penalty coefficient $C$ is set as 10 for SVM. We also add 30db, 20db and 10db Gaussian noise to test the robust performance. The experiment was conducted on two data sets A and B with different degrees of signal to noise ratio(SNR) conditions, we perform 10 rounds tests for each dataset, and report the average test classification error. Table 1 shows the test classification performance, the bold values in the text indicate the best results.

| Method   | HRRP dataset A | HRRP dataset B |
|----------|----------------|----------------|
|          | Original data  | 30db | 20db | 10db | Original data  | 30db | 20db | 10db |
| SVM      | 92.31%         | 90.52% | 88.42% | 86.37% | 91.70% | 89.12% | 87.11% | 85.02% |
| SAE      | 93.58%         | 92.83% | 91.23% | 86.59% | 92.97% | 91.72% | 89.78% | 84.98% |
| SSsAE    | 94.05%         | 92.85% | 91.75% | 86.22% | 93.23% | 92.08% | 90.38% | 85.13% |
| SDA      | 93.83%         | 92.90% | 91.91% | 87.06% | 93.33% | 91.92% | 90.45% | 85.61% |
| DBN      | 93.69%         | 92.53% | 91.13% | 86.76% | 93.07% | 91.69% | 89.99% | 85.25% |
| S-disAE  | **95.17%**     | **93.78%** | **92.95%** | **88.17%** | **94.28%** | **92.88%** | **91.39%** | **86.08%** |

From the table we can easily observe that all the deep learning algorithm achieve remarkable improvement results compared with SVM. So the deep neural networks have stronger ability to extract deep essential features and improve the target recognition rate. DBN achieve comparable results compared with SAE, SSsAE encourage sparsity on hidden layer activation, so it can avoid over-fitting and achieve higher accuracy than SAE. SDAE can learn more robust feature and achieve slightly improvement of the test accuracy compared with SAE. It can be observed that our proposed method outperforms all the other algorithms, there are about more than 1-1.5% improvements under different
SNR conditions. This is mainly because the introduction of the center loss in the leaning process, center loss reduces the intra-class distance while increases the inter-class distance, which greatly improves the separability of the features.

### 4.4. Visualization experiments

In this section, we conduct experiment to visualize the learned representation of our method on the HRRP dataset B. After trained the whole network, we can project the feature into a 2D space to visualize the embedding latent representations. Figure 6(a) shows the visualization of on the original HRRP, figure 6(b) shows the third layer features extracted by traditional SAE, figure 4 (c) shows the third layer features extracted by our S-disAE, both of them are reduce to 2 dimension via using t-Distributed Stochastic Neighbor Embedding (t-SNE) [21].

![Fig. 6. Representations learned by different methods on HRRP dataset B, reduced to 2 dimensions using t-SNE. (a) Original input data; (b) 3-th layer features extracted by SAE; (c) 3-th layer features extracted by S-disAE.](image)

In the figures, each dot indicates a sample, the colors of the points correspond to the class labels. 0-4 represent warhead, decoy1, decoy2, decoy3 and mother-cabin. As we can see, the data points of different classes overlap seriously derived from the original HRRP, at the same time, the warhead, decoy1, decoy2 and mother-cabin have very large scatter. The features extracted by SAE still have large distance between samples. Compare with traditional SAE, the features of each class extracted by our S-disAE are more close to class centers, so our method can reduce the intra-class variations while keeping the features of different classes separable. Therefore, the features learned by the proposed method is undoubtedly useful for recognition.

### 5. Conclusion

In order to extract discriminative features and improve HRRP recognition accuracy, we propose a stacked discriminative auto-encoder based on center loss for radar target HRRP recognition. Our S-disAE introduce the center loss in the pre-training stage and fine-tuning stage, so as to learn discriminative feature that have small intra-class variations while keeping the features of different classes separable. We conduct several experiments on the simulated HRRP data, the results demonstrate that the proposed algorithm can extract discriminant features and improve the HRRP recognition accuracy.

### Acknowledgements

This work was supported by National Natural Science Foundation of China under Grants 61806219, 61876189, 61503407, 61703426, 61273275. This work was also supported by Young Talent fund of University Association for Science and Technology in Shaanxi, China, NO.20190108.

### REFERENCES

[1] Xu B., Chen B, Wang J.W., et al. (2019) Target-aware recurrent attentional network for radar HRRP target recognition. Signal Processing, 155: 268–280.

[2] Zhou D.Y. (2016) Radar target HRRP recognition based on reconstructive and discriminative dictionary learning. Signal Processing, 126: 52-64.
[3] Feng B., Chen B., Liu H.W. (2017) Radar HRRP target recognition with deep networks, Pattern Recognition, 61: 379–393

[4] Du L., Wang P., Liu H., et al. (2011) Bayesian spatiotemporal multitask learning for radar HRRP target recognition. IEEE Transaction on Signal Processing, 59(7): 3182–3196.

[5] Feng B., Du L., Liu H.W., et al. (2012) Radar HRRP target recognition based on K-SVD algorithm, in: Proceedings of the IEEE CIE International Conference on Radar, pp. 642–645.

[6] Liu J., Fang N., Xie Y.J., et al. (2016) Scale-space theory-based multi-scale features for aircraft classification using HRRP. Electronic Letter, 52 (6): 475–477.

[7] Li R., Wang X.D., Lei, L., et al. (2017) Ballistic Target HRRP Fusion Recognition Combining Multi-class Relevance Vector Machine and DS. Information and Control, 46: 0065-0071.

[8] Zhao F., Liu Y., Huo K., et al. (2018) Radar HRRP target recognition based on stacked autoencoder and extreme learning machine. Sensors, 18(1): 173-188.

[9] Zhao, F.X., Liu, Y.X., Huo, K., et al. (2018) Radar HRRP Target Recognition Based on Stacked Autoencoder and Extreme Learning Machine. Sensors, 18: 173-187.

[10] Pan M., Jiang J., Kong Q.P., et al. (2017) Radar HRRP target recognition based on t-SNE segmentation and discriminant deep belief network, IEEE Geoscience and Remote Sensing Letters, 14(9):1609-1613.

[11] Feng B., Chen B., Liu H.W. (2017) Radar HRRP target recognition with deep networks, Pattern Recognition, 61:379–393.

[12] Du C., Chen B., Xua B., et al. (2019) Factorized discriminative conditional variational auto-encoder for radar HRRP target recognition. Signal Processing, 158: 176–189

[13] Bengio Y., Lamblin P., Dan P, et al. (2006) Greedy layer-wise training of deep networks, in: Proceedings of the Advances in Neural Information Processing Systems 19, pp.153–160.

[14] Ngiam Q. V. Le, Coates I. A., Lahiri A., et al. (2011) On optimization methods for deep learning, in: Proceedings of the 28th International Conference on Machine Learning, pp. 265-272.

[15] Vincent P., Larochelle H., Lajoie I., et al. (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research, 11(12): 3371-3408.

[16] Rifai S., Vincent P., Muller X., et al. (2011) Contractive auto-encoders: Explicit invariance during feature extraction, in Proc. the 28th International Conference on Machine Learning, Bellevue, Washington, USA, pp. 833–840.

[17] Szegedy C., Vanhoucke V., Ioffe S., et al. (2016) Rethinking the inception architecture for computer vision, in Proc. the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 2818–2826.

[18] Ren S., He K., Girshick R., et al. (2017) Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence. 39 (6): 1137–1149.

[19] Young T., Hazarika D., Poria S. et al. (2017) Recent trends in deep learning based natural language processing, IEEE Computational Intelligence Magazine, 13(3): 55-75.

[20] Khalil P. A., Jones E., Babar M. I., et al. (2019) Speech Emotion Recognition Using Deep Learning Techniques: A Review. IEEE Access, 7: 117327-117345.

[21] Maaten L.V.D., and Hinton G.E. (2008) Visualizing data using t-SNE, Journal of machine learning research, 9: 2579-2605.

[22] Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016). A Discriminative Feature Learning Approach for Deep Face Recognition. Lecture Notes in Computer Science, 499–515.