

# 一种基于Dropout约束深度极限学习机的雷达目标分类算法

赵飞翔 刘永祥\* 霍凯  
(国防科技大学电子科学学院 长沙 410073)

**摘要:** 雷达目标分类在军事和民用领域发挥着重要作用。极限学习机(Extreme Learning Machine, ELM)因其学习速度快、泛化能力强而被广泛应用于分类任务中。然而,由于其浅层结构,ELM无法有效地捕获数据深层抽象信息。虽然许多研究者已经提出了深度极限学习机,它可以用于自动学习目标高级特征表示,但是当训练样本有限时,模型容易陷入过拟合。为解决此问题,该文提出一种基于Dropout约束的深度极限学习机雷达目标分类算法,在雷达测量数据上的实验结果表明所提算法在分类准确率上达到93.37%,相较栈式自动编码器算法和传统深度极限学习机算法分别提高了5.25%和8.16%,验证了算法有效性。

**关键词:** 极限学习机; 深度学习; Dropout约束; 雷达目标分类; 栈式自动编码器

中图分类号: TN957.51

文献标识码: A

文章编号: 2095-283X(2018)05-0613-09

DOI: 10.12000/JR18048

**引用格式:** 赵飞翔, 刘永祥, 霍凯. 一种基于Dropout约束深度极限学习机的雷达目标分类算法[J]. 雷达学报, 2018, 7(5): 613-621. DOI: 10.12000/JR18048.

**Reference format:** Zhao Feixiang, Liu Yongxiang, and Huo Kai. A radar target classification algorithm based on Dropout constrained deep extreme learning machine[J]. *Journal of Radars*, 2018, 7(5): 613-621. DOI: 10.12000/JR18048.

## A Radar Target Classification Algorithm Based on Dropout Constrained Deep Extreme Learning Machine

Zhao Feixiang Liu Yongxiang Huo Kai

(College of Electronic Science, National University of Defense Technology, Changsha 410073, China)

**Abstract:** Radar target classification is very important in military and civilian fields. Extreme Learning Machines (ELMs) are widely used in classification because of their fast learning speed and good generalization performance. However, because of their shallow architecture, ELMs may not effectively capture the data high level abstractions. Although many researchers have proposed the Deep Extreme Learning Machine (DELME), which can be used to automatically learn high level feature representations, the model easily falls into overfitting when the training sample is limited. To address this issue, Dropout Constrained Deep Extreme Learning Machine (DCDELME) is proposed in this paper. The experimental results on the measured radar data show that the accuracy of the proposed algorithm can reach 93.37%, which is 5.25% higher than that of the stacked autoencoder algorithm, and 8.16% higher than that of the traditional DELME algorithm.

**Key words:** Extreme Learning Machine (ELM); Deep learning; Dropout constrained; Radar target classification; Stacked autoencoder

### 1 引言

雷达目标分类<sup>[1-3]</sup>在军事和民用领域都发挥着

重要的作用。传统的目标分类方法首先需要提取目标特征,然后通过分类器进行分类。其中特征提取是关键步骤,所提特征质量的好坏直接影响着分类性能的优劣。然而,传统的特征提取方法是浅层次的,且依赖于研究人员的专业知识和经验。因此如何自动地从数据中提取有利于目标分类的深层特征成了一个重要的问题。

目前,深度学习<sup>[4,5]</sup>在模式识别和机器学习中得到了广泛关注。深度学习的本质是构建一个包含

收稿日期: 2018-06-22; 改回日期: 2018-08-29; 网络出版: 2018-09-12

\*通信作者: 刘永祥 lyx\_bible@sina.com

基金项目: 国家自然科学基金(61422114, 61501481), 湖南省杰出青年科学基金(2015JJ1003)

Foundation Items: The National Natural Science Foundation of China (61422114, 61501481), The Natural Science Fund for Distinguished Young Scholars of Hunan Province (2015JJ1003)

多个隐含层的神经网络来映射数据以获得高层抽象信息。在训练深度学习模型时,首先利用无监督学习逐层训练网络,并将每层输出作为下一层的输入,然后利用监督学习从上到下对网络进行微调。目前广泛应用的深度学习模型有深度信念网络(Deep Belief Network, DBN)<sup>[4]</sup>,栈式自编码器(Stacked AutoEncoder, SAE)<sup>[6]</sup>和卷积神经网络(Convolutional Neural Networks, CNN)<sup>[7]</sup>。虽然这些模型在语音识别<sup>[7]</sup>、目标分类<sup>[8]</sup>和年龄估计<sup>[9]</sup>等领域有好的表现,但其微调过程需要耗费大量时间,从而降低网络训练速度。

为了解决这个问题,黄广斌等人<sup>[10]</sup>提出了极限学习机(Extreme Learning Machine, ELM)算法。该算法的思想是随机生成输入权重和隐层偏置,然后通过求解输出权重的最小二乘范数解来训练网络<sup>[11]</sup>。ELM不仅具有比传统学习方法更快的学习速度,而且具有良好的泛化性能。由于这些优点,ELM被应用在很多领域,如在线序列学习<sup>[12]</sup>、人脸识别<sup>[13]</sup>、图像分析<sup>[14]</sup>、聚类<sup>[15]</sup>以及雷达目标分类<sup>[16]</sup>等。然而,我们知道,由于其浅层架构,ELM可能无法有效地捕获数据的高层抽象信息。文献<sup>[17]</sup>提出一种深度极限学习机(Deep Extreme Learning Machine, DELM)算法,多层网络结构使其能够在数据中提取高层抽象信息,此外,与其他深度网络相比,DELM可以获得更好的性能。文献<sup>[18]</sup>提出一种多层模型DrELM,该算法依据栈式泛化理论通过堆叠ELM能获得目标深层特征表达。在文献<sup>[19]</sup>中,极限学习机-自动编码器(ELM-AE)被用作学习单元以学习网络中每层的局部感受野,同时为了保留更多信息,低层的输出被传送到最后一层以形成更完整的特征表示。文献<sup>[20]</sup>提出了一种新的基于ELM的多层感知学习框架,并在公共数据集上进行了实验,结果表明,该算法比现有的分层学习算法具有更好的收敛速度。

然而,在模型训练过程中,样本的数量和质量直接影响着网络权重参数的调整。当训练样本有限时,模型很容易陷入过拟合。因此有必要研究当训练样本有限时防止深度模型过拟合的方法。Dropout是Hinton<sup>[21]</sup>在2012提出的解决这一问题的新技术。该技术的核心是在网络训练时从网络中随机删除一些隐藏节点及其连接,从而防止由于特征间的共同作用而引起的过拟合<sup>[22]</sup>。研究表明,Dropout可以提高神经网络在计算机视觉、语音识别、文本分类等应用中的学习性能。文献<sup>[23]</sup>研究了Dropout的各种性质,包括收敛性质、优化性质和逼近性质。文献<sup>[24]</sup>从理论层面研究了Dropout能有效解

决过拟合问题的原因。文献<sup>[25]</sup>提出了在深度神经网络中自适应Dropout参数的方法。在文献<sup>[26]</sup>中,Dropout被引入到深度卷积神经网络的训练过程中,并在中文字符识别中得到应用。文献<sup>[27]</sup>提出了理解Dropout的通用框架。虽然深度极限学习机可以获得目标的深层抽象信息,但在训练样本有限的情况下,模型容易过拟合。为了解决这一问题,本文提出了一种基于Dropout约束的深度极限学习机(DCDELM)算法。

本文的主要内容如下:第2节介绍了深度极限学习机和Dropout的相关理论知识;在第3节中,提出基于Dropout约束的深度极限学习机。实验结果在第4部分进行了分析,第5部分对论文进行总结。

## 2 理论背景

### 2.1 极限学习机

基于梯度的学习算法存在训练速度慢、泛化性能差的缺点。为了解决这些问题,Huang等人<sup>[10]</sup>提出了极限学习机(ELM)算法。该算法模型由3层网络组成,分别为输入层、隐藏层和输出层。给定一个含有 $N$ 个训练样本的数据集 $(\mathbf{x}_i, \mathbf{t}_i)$ ,其中 $i = 1, 2, \dots, N$ ,  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbf{R}^n$ ,  $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbf{R}^m$ 。如果该模型结构有 $L$ 个隐藏层节点, $m$ 个输出节点,并且隐层激活函数为 $g(x)$ ,则隐层输出可用式(1)表示。如果这个含有 $L$ 个隐藏层节点的网络能够以0误差逼近输入的 $N$ 个训练样本,则网络输出可用式(2)表示。

$$\mathbf{h} = g(\mathbf{w}\mathbf{x} + \mathbf{b}) \quad (1)$$

$$\mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{t}_i^T, i = 1, 2, \dots, N \quad (2)$$

其中, $\boldsymbol{\beta}$ 是隐层节点和输出节点间的权重向量。

式(2)可以转化为:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (3)$$

其中

$$\begin{aligned} \mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_L, \mathbf{b}_1, \dots, \mathbf{b}_L, \mathbf{x}_1, \dots, \mathbf{x}_N) \\ = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + \mathbf{b}_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + \mathbf{b}_L) \\ \vdots & \dots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + \mathbf{b}_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_N + \mathbf{b}_L) \end{bmatrix}_{N \times L} \end{aligned} \quad (4)$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \quad (5)$$

因此训练ELM相当于求解 $\boldsymbol{\beta}$ 的最小二乘范数解,其表达式为:

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{T} \quad (6)$$

其中， $\mathbf{H}^+$ 为隐层输出矩阵 $\mathbf{H}$ 的Moore-Penrose广义逆。

式(6)可转化为：

$$\beta = \left( \frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T} \quad (7)$$

其中， $\mathbf{I}$ 为单位矩阵， $C$ 为岭参数(也称正则化系数)。

ELM也可以用优化的方法解释。ELM理论的目标是使得训练误差 $\|\mathbf{H}\beta - \mathbf{T}\|^2$ 和输出权值的范数解 $\|\beta\|$ 达到最小。则式(3)可通过式(8)求解：

$$\left. \begin{aligned} \text{Minimize: } & \frac{1}{2} \|\beta\|^2 + C \frac{1}{2} \sum_{i=1}^N \|\xi_i\|^2 \\ \text{Subject to: } & \mathbf{h}(\mathbf{x}_i)\beta = \mathbf{t}_i^T - \xi_i^T, \quad i = 1, \dots, N \end{aligned} \right\} \quad (8)$$

其中， $\xi_i$ 是 $m$ 个输出节点相对于训练样本 $\mathbf{x}_i$ 的训练误差向量。 $\mathbf{h}(\mathbf{x}_i)$ 是第 $i$ 个样本 $\mathbf{x}_i$ 的隐层输出向量。基于Karush-Kuhn-Tucker (KKT)理论，可以获得和式(7)相同的解。

### 2.2 极限学习机-自动编码器

自动编码器(AutoEncoder, AE)是一种无监督的神经网络。它通过对输入进行编码，然后再进行解码以重构该输入。极限学习机-自动编码器(ELM-AE)是一种新的神经网络方法，它可以像AE一样重构输入数据。ELM-AE也由3层网络组成，分别为输入层、隐藏层和输出层。ELM-AE的框架在图1中给出。ELM-AE和传统ELM的主要区别在于，ELM是一种监督学习算法，其输出是目标类别。而ELM-AE是一种无监督学习算法，它的输出即为其输入。给定一个含有 $N$ 个样本的数据集 $\mathbf{x}_i, i = 1, 2, \dots, N, \mathbf{x}_i \in \mathbf{R}^i$ ，则ELM-AE的隐层输出可用式(9)表示，其网络输出可用式(10)表示。

$$\mathbf{h} = g(\mathbf{w}\mathbf{x} + \mathbf{b}), \quad \mathbf{w}^T \mathbf{w} = \mathbf{I}, \quad \mathbf{b}^T \mathbf{b} = \mathbf{I} \quad (9)$$

$$\mathbf{h}(\mathbf{x}_i)\beta = \mathbf{x}_i^T, \quad i = 1, 2, \dots, N \quad (10)$$

ELM-AE隐层参数在随机生成后需要进行正交

化。这样可以有效地将输入数据映射到随机子空间。与ELM随机初始化输入权重和隐层偏置相比，正交化可以更好地捕获输入数据的各种边缘特征，从而使模型能够有效地学习数据的非线性结构。输出权重可通过式(11)进行计算：

$$\beta = \left( \frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{X} \quad (11)$$

### 2.3 深度极限学习机

在训练深度学习模型时，需要首先采用无监督学习算法对各层的参数进行训练，然后利用监督学习对网络进行微调。其中微调的过程需要耗费很多时间。2013年，黄广斌等人<sup>[17]</sup>提出了深度极限学习机(Deep Extreme Learning Machine, DELM)。与其他深度学习模型一样，DELM使用无监督学习方法ELM-AE来训练每层的参数，但不同的是DELM不需要对网络进行微调。这样，与其他深度学习算法相比，DELM不需要花费很长的时间来训练网络。图2显示了DELM的训练过程。

给定 $N$ 个训练样本集，网络隐层数为 $k$ ，网络需要学习的权重参数为 $\mathbf{W} = \{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_k, \mathbf{W}_{k+1}\}$ ， $\mathbf{X}_i$ 表示第 $i$ 个ELM-AE的输入。通过最小化重构误差，我们可以获得输入数据的特征表达，并将其作为下一层网络的输入。正如图2所示，网络的输入被看做是第1个ELM-AE的输入和输出，即 $\mathbf{T} = \mathbf{X}_1 = \mathbf{X}$ ，然后根据式(11)可以求得输出权重 $\beta_1$ 。将第1个ELM-AE的隐层输出 $\mathbf{H}_1$ (其中 $\mathbf{H}_k = g(\beta_k^T \mathbf{H}_{k-1})$ )作为第2个ELM-AE的输入和输出，即 $\mathbf{T} = \mathbf{X}_2 = \mathbf{H}_1$ 。用相同的方法，第 $k$ 个隐层之前的网络连接权重均可求得。根据自动编码器理论，其编码层和解码层的权重互为转置，因此，深度极限学习机第 $k$ 个隐层之前的网络连接权重为 $\mathbf{W}_\tau = \beta_\tau^T (\tau = 1, \dots, k)$ 。最后，将 $\mathbf{H}_k$ 视为监督层极限学习机的隐层输出，则其输出权重 $\mathbf{W}_{k+1}$ 可根据式(7)求得。至此，完成深度极限学习机的训练过程。

### 2.4 Dropout理论

Dropout是提高网络泛化能力的有效途径。Dropout的意思是指在网络训练过程中将神经元随机移除。移除一个神经元，是将它连同它的所有输入和输出连接从网络中暂时丢弃。正如图3所示。选择将哪一个神经元丢弃是随机的，通常每一个神经元以独立于其他神经元的固定概率被保留。由于神经元是随机丢弃的，故而对于每一次训练，网络都在发生变化，这样的设计可以避免网络对于某一局部特征的过拟合。

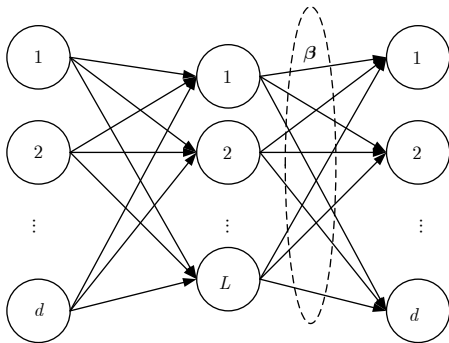


图1 ELM-AE框架

Fig. 1 The framework of ELM-AE

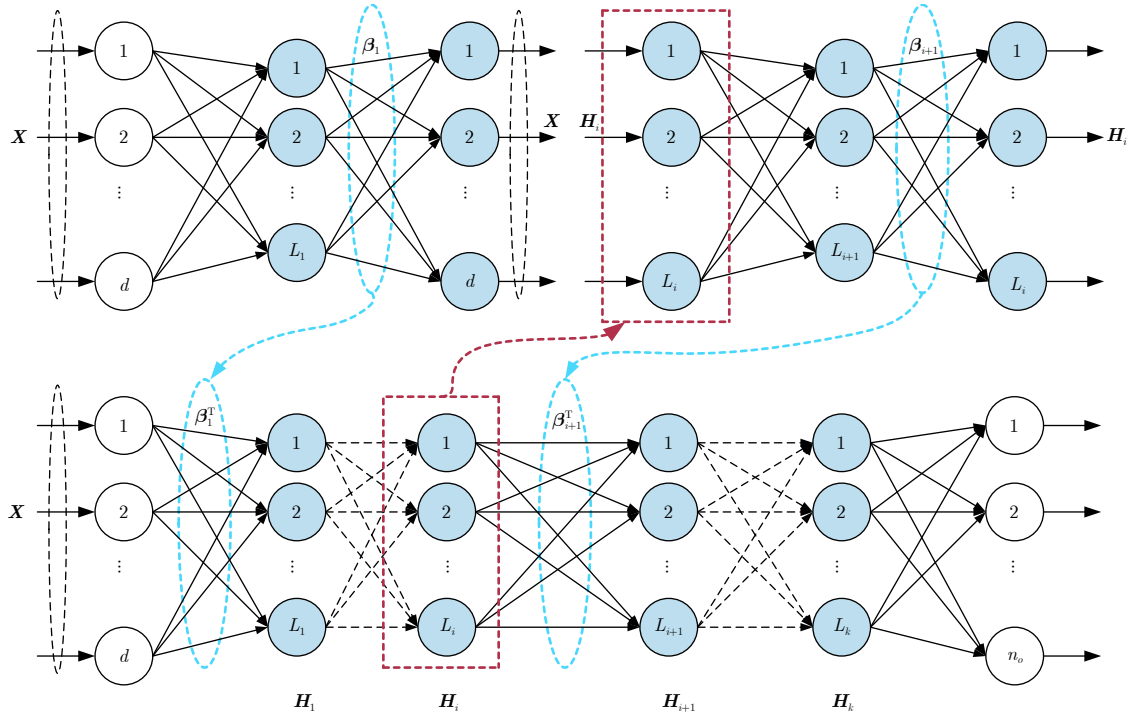


图 2 DELM训练过程

Fig. 2 The training process of DELM

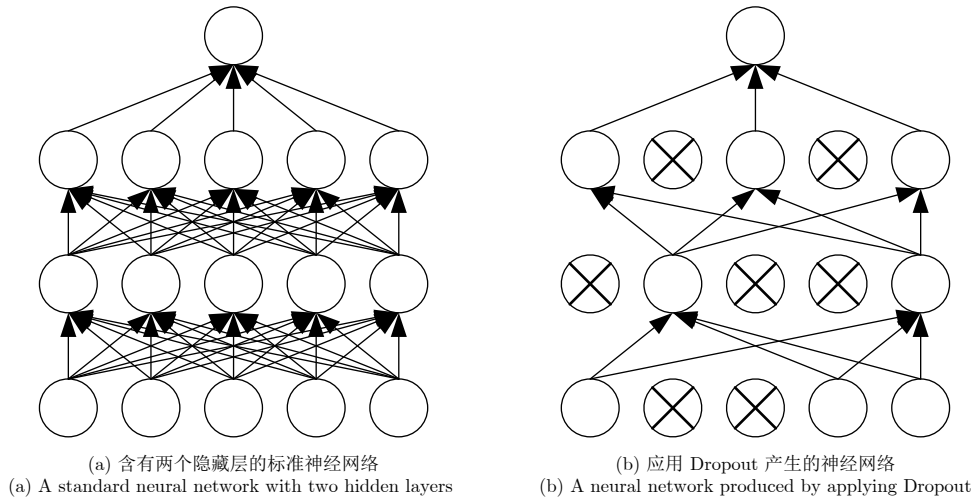


图 3 Dropout神经网络模型

Fig. 3 Dropout neural network model

### 3 基于Dropout约束的深度极限学习机

虽然深度极限学习机(DELM)能够捕获目标的抽象信息,但当训练样本有限时,模型容易陷入过拟合。为了解决这一问题,本文提出一种基于Dropout约束的深度极限学习机(DCDELIM)算法,将Dropout技术引入到DELM的训练过程中,以提高模型泛化能力。

通过引入Dropout,ELM-AE的优化过程可改写为:

$$\left. \begin{aligned} \text{Minimize: } & \frac{1}{2} \|\beta\|^2 + C \frac{1}{2} \sum_{i=1}^N \|\xi_i\|^2 \\ \text{Subject to: } & \mathbf{h}(\mathbf{x}_i)\beta = \mathbf{x}_i^T - \xi_i^T, \quad i = 1, \dots, N \\ & \mathbf{h}(\mathbf{x}_i)\beta = \mathbf{h}(\mathbf{x}_{i,t})\beta, \quad i = 1, \dots, N, \\ & t = 1, \dots, \hat{N} \end{aligned} \right\} (12)$$

其中,  $\mathbf{h}(\mathbf{x}_{i,t})$ 为引入Dropout后的隐层输出,  $\mathbf{h}(\mathbf{x}_{i,t}) = \gamma_{i,t} * \mathbf{h}(\mathbf{x}_i)$ 。  $\gamma_{i,t}$ 为服从Bernoulli分布的向量,其每一个元素以概率  $p$  等于1,以概率  $1 - p$  等于0。\*表示两个向量的哈达玛乘积。  $\hat{N}$ 为迭代次数。

定义 $\tilde{\mathbf{h}}(\mathbf{x}_{i,t}) = \mathbf{h}(\mathbf{x}_i) - \mathbf{h}(\mathbf{x}_{i,t})$ ，则

$$\mathbf{h}(\mathbf{x}_i)\beta - \mathbf{h}(\mathbf{x}_{i,t})\beta = \tilde{\mathbf{h}}(\mathbf{x}_{i,t})\beta = 0 \quad (13)$$

其中， $\tilde{\mathbf{h}}(\mathbf{x}_{i,t}) = \tilde{\gamma}_{i,t} * \mathbf{h}(\mathbf{x}_i)$ ， $\tilde{\gamma}_{i,t}$ 为服从Bernoulli分布的向量，其每一个元素以概率 $p$ 等于0，以概率 $1-p$ 等于1。

通过定义Lagrange函数来解决式(12)的优化问题：

$$\begin{aligned} \min: J_{\text{DCDELM}} = & \frac{1}{2} \|\beta\|^2 + C \frac{1}{2} \sum_{i=1}^N \|\xi_i\|^2 \\ & - \sum_{i=1}^N \alpha_i (\mathbf{h}(\mathbf{x}_i)\beta - \mathbf{x}_i^T + \xi_i^T) \\ & + \frac{\lambda}{2\hat{N}} \sum_{i=1}^N \sum_{t=1}^{\hat{N}} \|\tilde{\mathbf{h}}(\mathbf{x}_{i,t})\beta\|^2 \end{aligned} \quad (14)$$

则

$$\left. \begin{aligned} \frac{\partial J_{\text{DCDELM}}}{\partial \beta} = 0 & \rightarrow \beta - \mathbf{H}^T \alpha + \lambda \mathbf{R} \beta = 0 \\ \frac{\partial J_{\text{DCDELM}}}{\partial \xi_i} = 0 & \rightarrow \alpha_i = C \xi_i, \quad i = 1, \dots, N \\ \frac{\partial J_{\text{DCDELM}}}{\partial \alpha_i} = 0 & \rightarrow \mathbf{h}(\mathbf{x}_i)\beta - \mathbf{x}_i^T + \xi_i^T = 0, \\ & \quad i = 1, \dots, N \end{aligned} \right\} \quad (15)$$

$$\text{其中, } \mathbf{R} = \frac{1}{\hat{N}} \sum_{t=1}^{\hat{N}} \tilde{\mathbf{H}}_t^T \tilde{\mathbf{H}}_t.$$

则式(12)的解为：

$$\beta = \left( \mathbf{H}^T \mathbf{H} + \frac{\mathbf{I}}{C} + \frac{\lambda}{C} \mathbf{R} \right)^{-1} \mathbf{H}^T \mathbf{X} \quad (16)$$

需要注意的是，当 $\hat{N}$ 为有限值时，在求解输出

权重 $\beta$ 时，还需要计算 $\tilde{\mathbf{H}}_t$ 和 $\mathbf{R}$ ，这样势必会增加网络计算复杂度，从而增加网络训练时间。而当 $\hat{N}$ 趋于无穷大时，根据弱大数定律，计算 $\mathbf{R}$ 的公式可转化为：

$$\mathbf{R} = \mathbf{E} \left[ \frac{1}{\hat{N}} \sum_{t=1}^{\hat{N}} \tilde{\mathbf{H}}_t^T \tilde{\mathbf{H}}_t \right] \quad (17)$$

因此 $\mathbf{R}$ 可表示为：

$$\mathbf{R} = (1-p)^2 \mathbf{H}^T \mathbf{H} + (1-p)^2 \mathbf{D} \quad (18)$$

其中， $\mathbf{D} = (\mathbf{H}^T \mathbf{H}) * \mathbf{I}$ 。

通过转化，在计算 $\mathbf{R}$ 时，我们只需要计算 $\mathbf{H}$ 即可。当使用式(16)求解输出权重 $\beta$ 时，所提算法的计算复杂度和DELM是一样的。因此所提算法能在不增加网络计算复杂度的条件下提高模型泛化能力。

#### 4 实验结果与讨论

在这一部分，我们使用雷达测量的5类飞机数据验证所提方法的有效性。其中飞机类型分别为轰炸机、战斗机、教练机、无人机以及运输机。雷达测量频段为34.7~35.7 GHz，频率间隔为2 MHz，方位角为0°~30°，步长为1°。在我们的实验中，每一个飞机目标共有31个HRRP样本，并且每一个HRRP样本为500维。图4和图5分别显示的是每一个飞机目标的1维距离像序列和距离像的时域特征。我们从每一个目标样本中取奇数位的16个样本组成训练样本，其余偶数位的样本组成测试样本。接下来，我们将分析网络隐层节点数、岭参数以及Dropout参数对模型分类效果的影响。实验运行软件环境为Matlab R2013a，硬件环境为Intel(R) Core(TM), 3.60 GHz CPU, 8 GB内存的PC机。

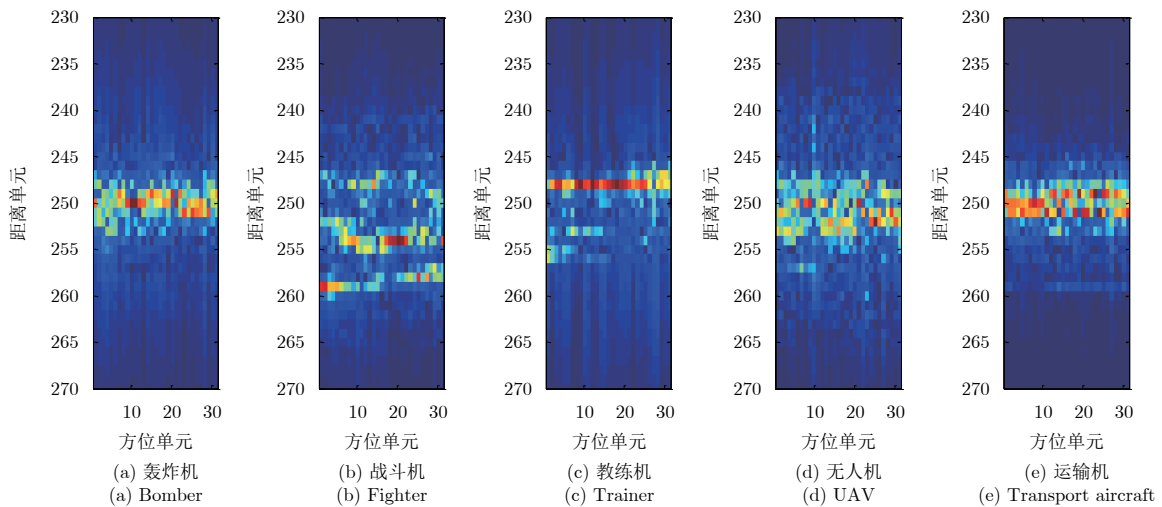


图4 每一个飞机目标HRRP序列

Fig. 4 The HRRP sequence for each aircraft target

### 4.1 隐层节点数对分类效果的影响

由于HRRP样本维数为500，所以将深度网络中可见层的节点数设置为500。众所周知，随着网络深度的增加，可以获得更多抽象的特征表示。然而，太多的隐藏层会使网络难以有效训练，并带来更多的参数需要学习。因此，我们不能盲目追求网络深度，而应该根据样本大小和实验需求来设置

网络隐藏层的数量。通过对实验数据和任务要求的分析，发现具有两个隐含层的网络能够满足实验需要。因此，我们将网络隐藏层的数量设为2。图6显示了第1和第2隐藏层不同节点数对分类效果的影响。从图6(a)和图6(c)可以看出，在网络训练阶段，随着节点数的增加，两种算法(DELM和DCDELM)的分类准确率也在增加。而从图6(b)可

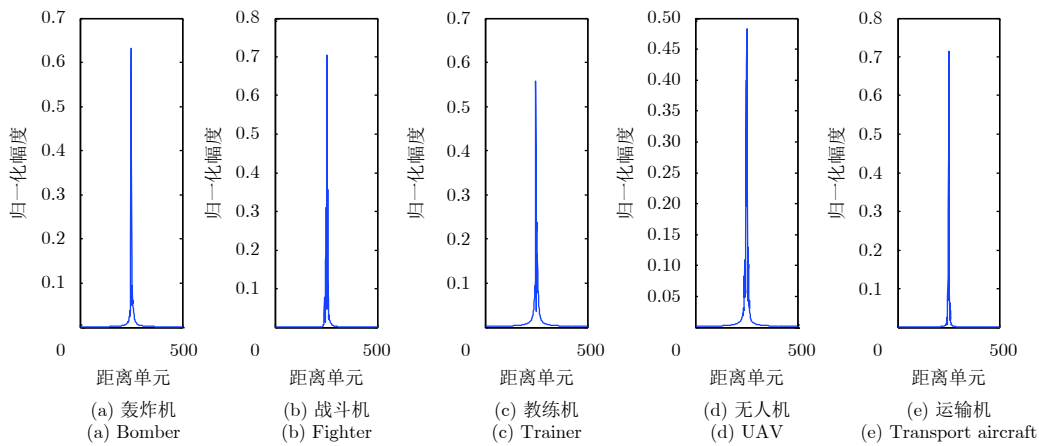


图 5 每一个飞机目标距离像

Fig. 5 Range profiles of each aircraft target

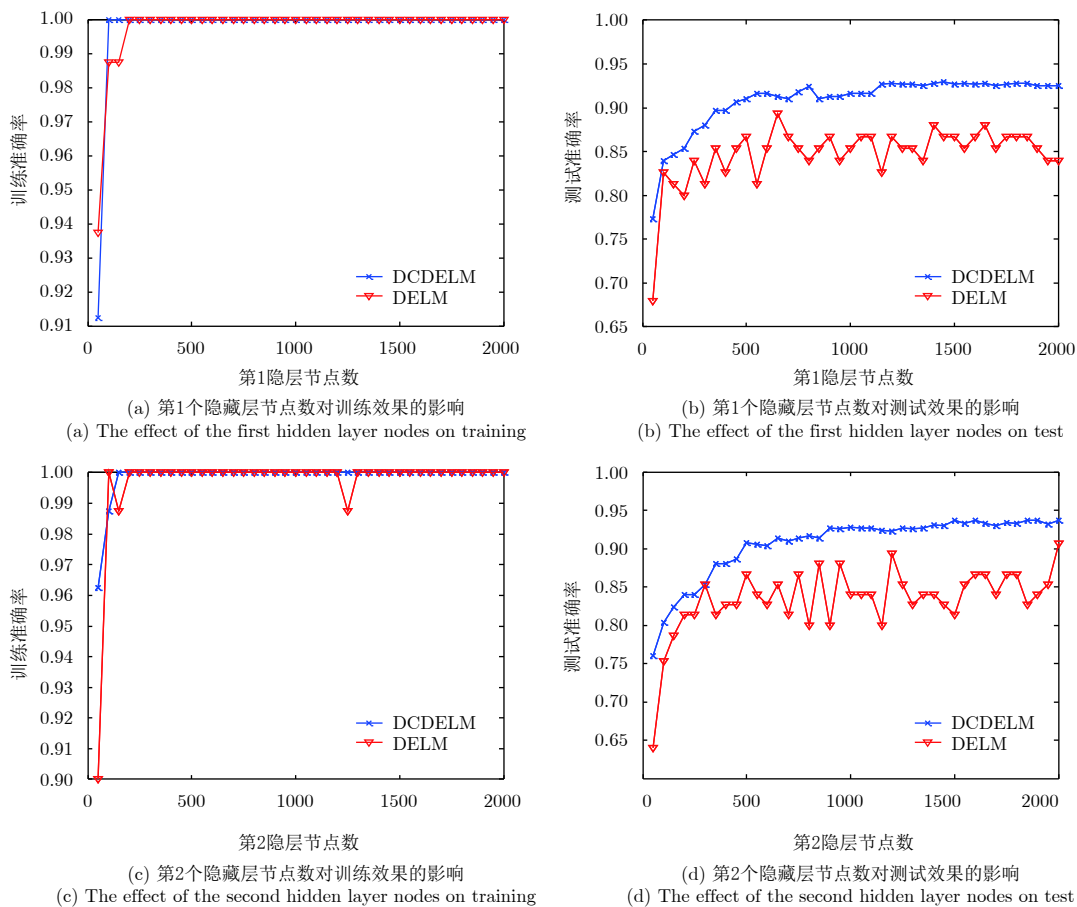


图 6 第1和第2隐藏层不同节点数对分类效果的影响

Fig. 6 The effect of different number of nodes in the first and second hidden layers on the classification performance

知在测试阶段，随着节点数的增加，DELM算法的分类准确率出现波动，其变化趋势和训练效果不一致，表明该算法陷入了过拟合。由于引入了Dropout约束，所提算法(DCDELM)可以有效地解决训练样本有限时DELM的过拟合问题。在测试阶段，算法的分类精度随着隐藏节点数目的增加而增加，其结果如图6(d)所示。从图7可知，随着隐层节点数的增加，网络训练时间和测试时间也在增加。因此，在保证分类精度的前提下，有必要选择合适的隐含层节点。经过分析，我们选择隐藏层节点的数量分别为850和1000。

4.2 岭参数对分类效果的影响

通过上面的实验，我们确定了网络隐藏层数和隐层节点数。接下来，我们需要确定岭参数  $C = [C_1, C_2, C_3]$  的取值。我们在  $\{10^{-10}, 10^{-9}, \dots, 10^9, 10^{10}\}$  范围内进行网格搜索以确定合适的参数值。我们遵循如下规则：首先固定其他两个参数  $C_2$  和  $C_3$ ，然后在  $\{10^{-10}, 10^{10}\}$  范围内选择可取得最好分类效果的  $C_1$  的值。照此方法可确定每一个合

适的岭参数。图8展示了不同的岭参数  $C = [C_1, C_2, C_3]$  对分类效果的影响。从图中可以看出，当  $C_1 = 0.1$ ,  $C_2 = 10^{10}$ ,  $C_3 = 10^8$  时，可以获得最好的分类结果。

4.3 Dropout参数对分类效果的影响

所提算法模型中引入了Dropout约束，能够在训练样本较少的情况下解决模型过拟合的问题。其中约束参数  $\lambda$  和Dropout概率  $p$  影响着Dropout的效果。在这一部分，我们在  $\{10^{-10}, 10^{-9}, \dots, 10^9, 10^{10}\}$  范围内进行网格搜索以确定合适的  $\lambda$  值，并在  $\{0.05, 0.10, \dots, 0.95, 1.00\}$  范围内确定合适的  $p$  值。这些参数对分类结果的影响在图9(a)中显示。从图中可以明显地看出，模型敏感于概率  $p$  和参数  $\lambda$ 。由图9(b)可知，当  $\lambda$  取值小于  $10^6$  时， $p = 0.50$  时模型分类效果要优于  $p$  等于其他值时的模型分类效果，这是由于  $p = 0.50$  时Dropout随机生成的网络结构最多，能够使模型具有最高的泛化水平<sup>[21,22]</sup>。当  $\lambda$  取值大于  $10^6$  时，模型分类效果下降较为明显，这是因为若  $\lambda$  过大，将造成网络中其他参数变得很小，而使网络陷入欠拟合。因此可以得出结论，所

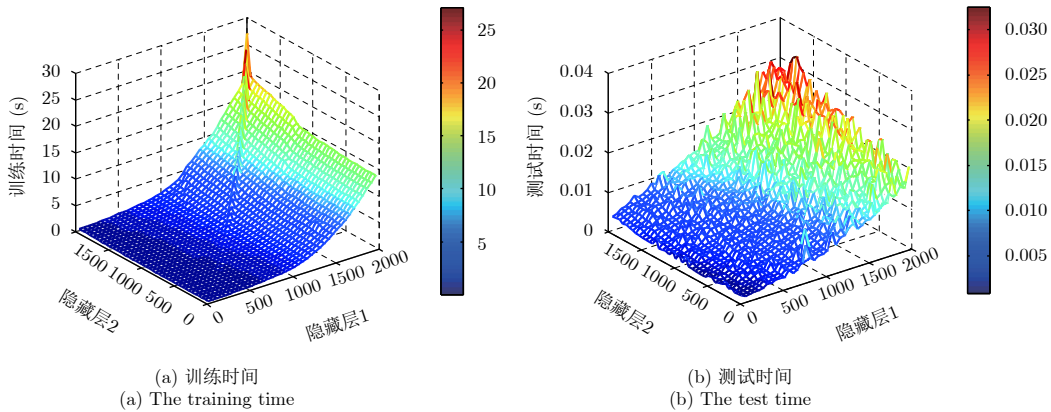


图7 不同隐层节点对网络训练和测试时间的影响

Fig. 7 The effect of different hidden nodes on training and test time of network

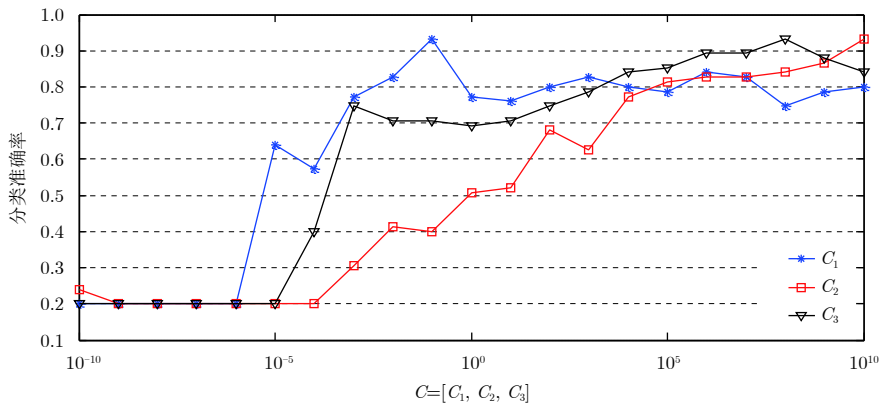


图8 不同岭参数  $C = [C_1, C_2, C_3]$  对网络分类效果的影响

Fig. 8 The effect of different ridge parameters  $C = [C_1, C_2, C_3]$  on classification performance

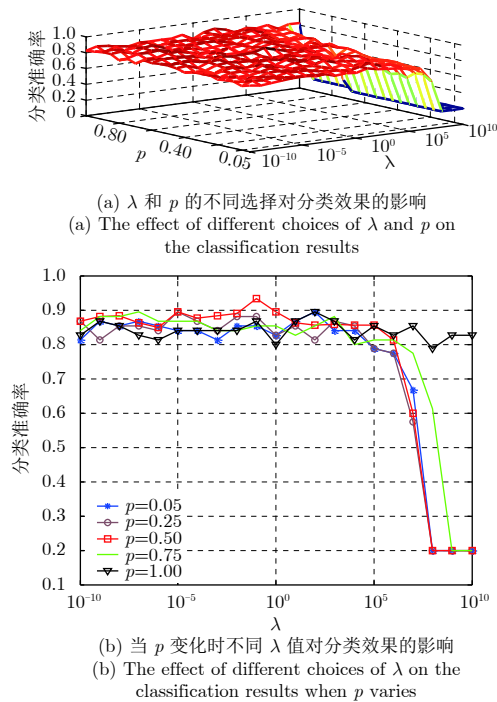


图 9 不同Dropout参数对分类效果的影响

Fig. 9 The effect of Dropout parameters on the classification performance

提算法敏感于概率 $p$ 和参数 $\lambda$ ，只有选择合适的 $\lambda$ 和 $p$ 的值，模型才能取得更好的分类效果。需要说明的是，当 $p=1$ ，DELM是所提算法的一种特殊形式。

#### 4.4 所提算法和其他算法分类效果比较

在之前的工作<sup>[17-20]</sup>中，研究者已经验证了DELM相比传统常用的深度学习模型的优越性。在本文中，我们将重点考虑在训练样本有限时，所提算法相较SAE和DELM在分类效果上的提高。由于样本维数为500，我们设置3种算法结构的可见层节点数为500，隐藏层数为2。其中DELM和所提算法隐层激活函数为‘sigmoid’。SAE的隐层节点数分别为850和300， $\rho$ 取0.66，训练迭代次数为30。将每一个算法均运行10次，然后求其平均值做为最终的分类结果。实验结果在表1中给出。

从表1中可以看出，所提方法的分类性能要优于DELM和SAE。这是因为所提出的算法在训练过程中引入了Dropout约束，可以有效地提高模型的泛化能力。SAE的分类准确率虽然高于DELM，但

表 1 所提方法和其他算法分类准确率比较

Tab. 1 Comparison of classification accuracy between the proposed method and other algorithms

方法	分类准确率(%)	训练时间(s)
SAE	88.12	7.8746
DELM	85.21	0.4121
所提方法	93.37	0.4205

其训练时间更长，这是由于SAE模型在训练过程中需要进行迭代微调，而DELM模型在训练过程中无需微调操作，从而极大地减少了训练时间。另外，所提算法和DELM的训练时间差别不大，验证了所提算法在计算复杂度上与DELM相同。

## 5 结束语

本文提出了一种基于Dropout约束的深度极限学习机算法。深度极限学习机能够有效地捕获目标的深度抽象特征表示，与其他深度学习模型相比，学习速度更快。然而，当训练样本有限时，模型容易陷入过拟合。Dropout是解决神经网络过拟合问题的一种有效方法。因此，将其应用到深度极限学习机的训练过程中，可以有效地提高模型的泛化能力。实测雷达数据验证了该算法的有效性。

## 参考文献

- [1] Liu Y X, Zhu D K, Li X, *et al.* Micromotion characteristic acquisition based on wideband radar phase[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(6): 3650-3657. DOI: [10.1109/TGRS.2013.2274478](https://doi.org/10.1109/TGRS.2013.2274478).
- [2] Bigdeli B and Pahlavani P. Quad-polarized synthetic aperture radar and multispectral data classification using classification and regression tree and support vector machine-based data fusion system[J]. *Journal of Applied Remote Sensing*, 2017, 11(1): 016007. DOI: [10.1117/1.JRS.11.016007](https://doi.org/10.1117/1.JRS.11.016007).
- [3] Shi J F, Li L L, Liu F, *et al.* Unsupervised polarimetric synthetic aperture radar image classification based on sketch map and adaptive Markov random field[J]. *Journal of Applied Remote Sensing*, 2016, 10(2): 025008. DOI: [10.1117/1.JRS.10.025008](https://doi.org/10.1117/1.JRS.10.025008).
- [4] Hinton G E and Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507. DOI: [10.1126/science.1127647](https://doi.org/10.1126/science.1127647).
- [5] Lecun Y, Bengio Y, and Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [6] Zhao F X, Liu Y X, Huo K, *et al.* Radar HRRP target recognition based on stacked autoencoder and extreme learning machine[J]. *Sensors*, 2018, 18(1): 173. DOI: [10.3390/s18010173](https://doi.org/10.3390/s18010173).
- [7] Abdel-Hamid O, Mohamed A R, Jiang H, *et al.* Convolutional neural networks for speech recognition[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2014, 22(10): 1533-1545. DOI: [10.1109/TASLP.2014.2339736](https://doi.org/10.1109/TASLP.2014.2339736).
- [8] Ding J, Chen B, Liu H W, *et al.* Convolutional neural network with data augmentation for SAR target recognition[J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(3): 364-368. DOI: [10.1109/LGRS.2016.2511109](https://doi.org/10.1109/LGRS.2016.2511109).

- 2015.2513754.
- [9] Wang X L, Guo R, and Kambhampettu C. Deeply-learned feature for age estimation[C]. Proceedings of 2015 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2015: 534–541. DOI: [10.1109/WACV.2015.77](https://doi.org/10.1109/WACV.2015.77).
- [10] Huang G B, Zhu Q Y, and Siew C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1/3): 489–501. DOI: [10.1016/j.neucom.2005.12.126](https://doi.org/10.1016/j.neucom.2005.12.126).
- [11] Huang G, Huang G B, Song S J, *et al.* Trends in extreme learning machines: A review[J]. *Neural Networks*, 2015, 61: 32–48. DOI: [10.1016/j.neucom.2014.10.001](https://doi.org/10.1016/j.neucom.2014.10.001).
- [12] Liang N Y, Huang G B, Saratchandran P, *et al.* A fast and accurate online sequential learning algorithm for feedforward networks[J]. *IEEE Transactions on Neural Networks*, 2006, 17(6): 1411–1423. DOI: [10.1109/TNN.2006.880583](https://doi.org/10.1109/TNN.2006.880583).
- [13] Zong W W and Huang G B. Face recognition based on extreme learning machine[J]. *Neurocomputing*, 2011, 74(16): 2541–2551. DOI: [10.1016/j.neucom.2010.12.041](https://doi.org/10.1016/j.neucom.2010.12.041).
- [14] Liu N and Wang H. Evolutionary extreme learning machine and its application to image analysis[J]. *Journal of Signal Processing Systems*, 2013, 73(1): 73–81. DOI: [10.1007/s11265-013-0730-x](https://doi.org/10.1007/s11265-013-0730-x).
- [15] Ding S F, Zhang N, Zhang J, *et al.* Unsupervised extreme learning machine with representational features[J]. *International Journal of Machine Learning and Cybernetics*, 2017, 8(2): 587–595. DOI: [10.1007/s13042-015-0351-8](https://doi.org/10.1007/s13042-015-0351-8).
- [16] Zhao F X, Liu Y X, Huo K, *et al.* Radar target classification using an evolutionary extreme learning machine based on improved quantum-behaved particle swarm optimization[J]. *Mathematical Problems in Engineering*, 2017: 7273061. DOI: [10.1155/2017/7273061](https://doi.org/10.1155/2017/7273061).
- [17] Kasun L L C, Zhou H M, Huang G B, *et al.* Representational learning with ELMs for big data[J]. *IEEE Intelligent Systems*, 2013, 28(6): 31–34.
- [18] Yu W C, Zhuang F Z, He Q, *et al.* Learning deep representations via extreme learning machines[J]. *Neurocomputing*, 2015, 149: 308–315. DOI: [10.1016/j.neucom.2014.03.077](https://doi.org/10.1016/j.neucom.2014.03.077).
- [19] Zhu W T, Miao J, Qing L Y, *et al.* Hierarchical extreme learning machine for unsupervised representation learning[C]. Proceedings of 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015: 1–8. DOI: [10.1109/IJCNN.2015.7280669](https://doi.org/10.1109/IJCNN.2015.7280669).
- [20] Tang J X, Deng C W, and Huang G B. Extreme learning machine for multilayer perceptron[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(4): 809–821. DOI: [10.1109/TNNLS.2015.2424995](https://doi.org/10.1109/TNNLS.2015.2424995).
- [21] Hinton G E, Srivastava N, Krizhevsky A, *et al.* Improving neural networks by preventing co-adaptation of feature detectors[OL]. <https://arxiv.org/abs/1207.0580>.2012.07.
- [22] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: A simple way to prevent neural networks from overfitting[J]. *The Journal of Machine Learning Research*, 2014, 15(1): 1929–1958.
- [23] Iosifidis A, Tefas A, and Pitas I. DropELM: Fast neural network regularization with dropout and dropconnect[J]. *Neurocomputing*, 2015, 162: 57–66. DOI: [10.1016/j.neucom.2015.04.006](https://doi.org/10.1016/j.neucom.2015.04.006).
- [24] Baldi P and Sadowski P. The dropout learning algorithm[J]. *Artificial Intelligence*, 2014, 201: 78–122.
- [25] Wager S, Wang S D, and Liang P. Dropout training as adaptive regularization[C]. Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, Nevada, 2013: 351–359.
- [26] Yang W X, Jin L W, Tao D C, *et al.* DropSample: A new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition[J]. *Pattern Recognition*, 2016, 58: 190–203. DOI: [10.1016/j.patcog.2016.04.007](https://doi.org/10.1016/j.patcog.2016.04.007).
- [27] Baldi P and Sadowski P. Understanding dropout[C]. Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, Nevada, 2013: 2814–2822.

### 作者简介



赵飞翔(1989–), 男, 河南洛阳人, 国防科技大学电子科学学院在读博士生, 研究方向为雷达目标识别。

E-mail: zfxkj123@sina.cn



刘永祥(1976–), 男, 河北唐山人, 博士, 国防科技大学电子科学学院智能感知系主任, 教授, 博士生导师, 主要研究方向为目标微动特性分析与识别。

E-mail: lyx\_bible@sina.com



霍凯(1983–), 男, 湖北黄冈人, 博士, 国防科技大学电子科学学院讲师, 主要研究方向为雷达信号处理与目标识别。

E-mail: huokai2001@163.com