

冯乐, 汤华椿, 高亮, 等. 基于粒子群优化的自适应微调算法[J]. 智能计算机与应用, 2024, 14(4): 232-237. DOI: 10.20169/j.issn.2095-2163.240438

## 基于粒子群优化的自适应微调算法

冯乐<sup>1,2</sup>, 汤华椿<sup>1,2</sup>, 高亮<sup>3</sup>, 邹红梅<sup>1,2</sup>, 王林<sup>1,2</sup>, 谭棉<sup>1,2</sup>

(1 贵州民族大学 数据科学与信息工程学院, 贵阳 550025; 2 贵州民族大学 贵州省模式识别与智能系统重点实验室, 贵阳 550025; 3 集美大学 轮机工程学院, 福建 厦门 361000)

**摘要:** 模型微调是一种重要的迁移学习技术, 在各种缺乏训练数据的任务中取得了显著成功, 并且只需少量训练时间就能获得满意效果。然而, 随着网络模型规模和结构复杂性增加, 为特定目标任务设计合适的微调方案变得越来越困难。为了解决这个问题, 本文提出了一种基于粒子群优化的自适应微调算法。该算法将预训练模型中的微调层选择问题建模为参数优化问题, 并基于种群进化思想, 自适应地为目标任务找到合适的层微调方案。实验结果表明, 该算法能够精确地求解适合目标任务的层微调方案, 并且能够显著提升微调性能。与标准微调基线方法相比, 本文算法只需要几次迭代即可实现具有竞争力的性能, 从而大幅度减少了训练时间。

**关键词:** 模型微调; 微调层选择; 粒子群优化; 迁移学习

**中图分类号:** TP183 **文献标志码:** A **文章编号:** 2095-2163(2024)04-0232-06

### Adaptive fine-tuning algorithm based on particle swarm optimization

FENG Le<sup>1,2</sup>, TANG Huachun<sup>1,2</sup>, GAO Liang<sup>3</sup>, ZOU Hongmei<sup>1,2</sup>, WANG Lin<sup>1,2</sup>, TAN Mian<sup>1,2</sup>

(1 School of Data Sciences and Information Engineering, Guizhou Minzu University, Guiyang 550025, China;

2 Guizhou Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University, Guiyang 550025, China;

3 School of Marine Engineering, Jimei University, Xiamen 361000, Fujian, China)

**Abstract:** Model fine-tuning is an important transfer learning technique that can achieve significant success in tasks with limited training data. Satisfactory results can be obtained with minimal training time. However, as the scale and complexity of network models increase, it becomes increasingly difficult to design appropriate fine-tuning strategies for specific target tasks. To address this issue, this paper proposes an adaptive fine-tuning algorithm based on particle swarm optimization. The algorithm models the problem of selecting fine-tuning layers in pre-trained models as a parameter optimization problem. Based on the concept of population evolution, it adaptively identifies appropriate layer fine-tuning schemes for the target task. Experimental results show that the algorithm can accurately solve layer fine-tuning schemes that are suitable for the target task and significantly improve the performance of fine-tuning. Compared to the standard fine-tuning baseline, our method achieves competitive performance with only a few iterations, thereby significantly reducing training time.

**Key words:** model fine-tuning; fine-tuning layer selection; particle swarm optimization; transfer learning

## 0 引言

随着人工智能的快速发展, 以机器学习为核心的技术逐渐改变了人们的生活方式和工业生产模式。深度神经网络等机器学习方法通过学习大量标

注数据中的特征, 在计算机视觉<sup>[1]</sup>、自然语言处理<sup>[2]</sup>、推荐系统<sup>[3]</sup>及医学影像分析<sup>[4]</sup>等领域取得了重大突破。这些进展主要归因于大数据时代对标注数据集的广泛利用, 因为机器学习模型的性能通常与标注数据集的质量和规模相关。然而, 在实际应

**基金项目:** 国家自然科学基金(62162012); 贵州省科技计划项目(QKHJZK2022YB195, QKHJZK2023YB143, QKHPTRCZCKJ2021007); 贵州省教育厅自然科学研究项目(QJJ2023061, QJJ2023012, QJJ2022015)。

**作者简介:** 冯乐(1998-), 男, 硕士研究生, 主要研究方向: 迁移学习; 汤华椿(1998-), 男, 硕士研究生, 主要研究方向: 计算机视觉; 高亮(2001-), 男, 本科生, 主要研究方向: 深度学习; 邹红梅(2000-), 女, 硕士研究生, 主要研究方向: 迁移学习; 王林(1965-), 男, 博士, 教授, 主要研究方向: 数字图像处理。

**通讯作者:** 谭棉(1984-), 女, 硕士, 副教授, 主要研究方向: 数字图像处理。Email: tanmian@gzmu.edu.cn

收稿日期: 2023-11-27

用中,获取大量标注数据可能存在成本和可行性方面的挑战。如:生物信息学数据集中的每个样本通常代表一个临床试验或病人的痛苦经历。此外,人们付出巨大代价获得的数据集,很容易过时,无法有效地应用于新任务,这限制了机器学习模型的应用范围和性能提升<sup>[5]</sup>。

为了解决数据不足的问题,迁移学习成为重要的机器学习技术。通过利用已有知识来辅助相关但不同目标任务的学习,取得了显著成功<sup>[6-7]</sup>。其中,模型微调是重要的迁移学习技术之一。其通过将预训练好的神经网络模型的参数从源域迁移到目标任务中,并通过少量数据对其进行再训练,以适应特定目标任务的需求<sup>[8-9]</sup>。通过在现有模型的基础上进行微调,可以提高模型适应特定任务的能力,同时节省训练时间和计算资源。模型微调的研究具有重要意义,可帮助深入理解深度学习模型的泛化能力,推动机器学习技术在各个领域的应用和发展。

在微调过程中,源域的预训练模型参数被迁移到目标模型作为“知识”。为提高模型性能,需要微调目标模型的一些参数,模型微调过程如图 1 所示。一般来说,与目标任务密切相关的参数应该被冻结,以便在目标任务中重复使用,而与目标任务无关的参数应该进行微调<sup>[10]</sup>。然而,由于参数与目标任务之间的相关性不明确,参数的选择变得困难。在传统微调方法中,参数的选择通常是基于专家经验手动完成的<sup>[11]</sup>。一种常见做法是进行全微调或标准微调,即在目标任务上微调所有参数<sup>[12-13]</sup>。这种方法相对于从头训练模型,能显著提升模型性能,但在目标数据稀缺或模型庞大时容易出现过拟合问题<sup>[14]</sup>。为了缓解过拟合问题,一些学者发现只微调模型的一个子集参数,冻结其他参数,可以有效减轻过拟合问题并显著提高模型性能<sup>[15]</sup>,因为这样允许目标模型学习源任务和目标任务之间的共同特征。因此,微调参数的选择对于提高模型性能至关重要。然而,随着深度神经网络规模的增加,基于专家经验的手动试错方法很难准确确定需要微调的参数<sup>[16]</sup>。因此,迫切需要一种不需要人工输入的自动微调策略,以确定迁移学习中需要微调的参数。

为了解决预训练模型中微调层选择的困难,本文将微调层的选择问题建模为一种参数优化问题,并提出了一种基于粒子群优化的自适应微调算法。该算法可以根据预训练模型与目标任务之间的关系,自适应地选择合适的层进行微调,从而显著提升模型的微调性能。

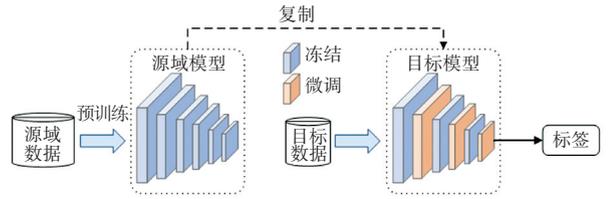


图 1 模型微调过程

Fig. 1 The process of the model fine-tuning

## 1 基于粒子群优化的自适应微调算法

在微调过程中,预训练的源域模型中一些层可以提取源域与目标域间通用的特征,这些层对于目标任务具有帮助,因此应该被冻结,而只适用于特定任务的层应该进行微调,以帮助模型适应目标任务。然而,源域模型可能有数十层甚至上百层,因此如何有效选择需要微调的层成为迁移学习中亟待解决的关键问题。针对该问题,本文提出了一种基于粒子群优化的自适应微调算法,其框架如图 2 所示。

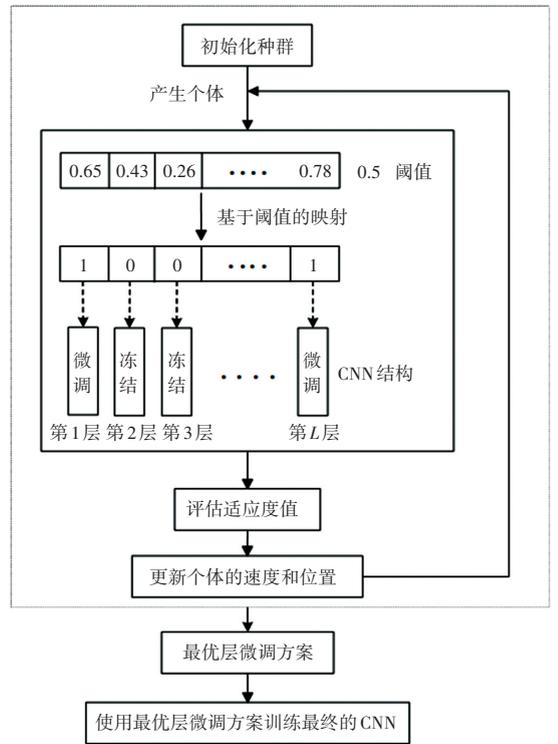


图 2 基于粒子群优化的自适应微调框架图

Fig. 2 Adaptive fine-tuning framework diagram based on particle swarm optimization

该算法使用粒子群优化来自动搜索适合微调的层,从而提高参数微调的准确性。

本文利用粒子群优化算法 (Particle Swarm Optimization, PSO) 对源域模型的微调层进行搜索。PSO 是一种流行的基于元启发式的算法,操作简单易实现且具有强大的全局搜索能力<sup>[17]</sup>。在 PSO

中,每个粒子代表搜索空间中的一个候选解,通过个体历史最优解和群体全局最优解的指导,调整自身的位置和速度。通过迭代不断更新粒子的位置和速度,直到找到满足停止条件的最优解。

在微调问题中,假设预训练模型  $W$  共有  $L$  层,针对每一层微调还是冻结的问题,可以将其对应的个体(候选解)定义为如下形式:

$$x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,L}) \quad (1)$$

其中,  $x_i$  表示种群中第  $i$  个个体,  $x_{i,j}$  表示预训练模型的第  $j$  层参数微调方案,并且  $x_{i,j} \in [0,1]$ 。当  $x_{i,j} \leq 0.5$  时,表示冻结预训练模型的第  $j$  层参数,当  $x_{i,j} > 0.5$  时表示微调预训练模型的第  $j$  层参数。

为了评估每个个体的质量,本文定义个体的适应度函数  $f(*)$  如下:

$$f(x_i) = \text{Acc}(\hat{W}; D_i) \quad (2)$$

其中:

$$\hat{W} = W - \eta \frac{\partial \mathcal{L}(W)}{\partial W} \odot M \quad (3)$$

$$M_j = \begin{cases} 0, & x_{i,j} \leq 0.5 \\ 1, & x_{i,j} > 0.5, j = 1, 2, \dots, L \end{cases} \quad (4)$$

公式(3)中的  $M = (M_1, M_2, \dots, M_L)$  表示与模型  $W$  一一对应的掩码矩阵。其中,  $M_j$  表示与模型  $W$  的第  $j$  层参数意义对应的掩码值,主要用于实现参数的冻结和微调;  $L$  表示模型的总层数;  $\hat{W}$  表示源域模型  $W$  根据掩码矩阵  $M$  微调后的模型,当  $M_j = 0$  时,则表明冻结模型  $W$  的第  $j$  层参数,而当  $M_j = 1$  时,则表明微调第  $j$  层参数以适应目标任务。掩码矩阵  $M$  的值由个体  $x_i$  决定,  $\mathcal{L}$  表示交叉熵损失函数,  $\eta$  表示学习率。  $\text{Acc}(\{\hat{W}; D_i)$  表示基于个体  $x_i$  对相应层参数进行微调少量周期后,在目标任务  $D_i$  上的分类精度。为了节约计算成本,将每个个体的微调周期数设置为  $\text{maxEpoch}$  (本文中为 5)。

在粒子群优化过程中,每个粒子都有自己的位置(候选解)  $x_i$  和速度  $v_i$ ,其根据个体历史最优和群体全局最优的引导,不断更新自己的位置和速度,速度和位置的更新公式如下:

$$v_{i,d}^{(t+1)} = \alpha v_{i,d}^{(t)} + c_1 r_1 (p_{i,d} - x_{i,d}^{(t)}) + c_2 r_2 (p_{g,d} - x_{i,d}^{(t)}) \quad (5)$$

$$x_{i,d}^{(t+1)} = x_{i,d}^{(t)} + v_{i,d}^{(t+1)} \quad (6)$$

其中,  $x_{i,d}^{(t)}$  和  $v_{i,d}^{(t)}$  分别表示第  $t$  代种群第  $i$  个个体在第  $d$  维上的位置和速度;  $p_{i,d}$  和  $p_{g,d}$  分别表示第  $i$  个个体在第  $d$  维的历史最优位置和种群的全局最优位置;  $r_1$  和  $r_2$  均为取值范围在 0~1 之间的随机

数;  $\alpha$  表示惯性权重;  $c_1$  和  $c_2$  表示个体学习因子。

基于粒子群优化的自适应微调算法的伪代码如算法 1 所示。在算法 1 中,首先随机产生一个拥有  $n$  个个体的种群,并且初始化速度  $v = \{v_i, i = 1, \dots, n\}$ ,初始化位置  $x = \{x_i, i = 1, \dots, n\}$  (每个个体位置  $x_i$  表示预训练模型中  $L$  层的微调方案)。在第 2 行中,分别对  $n$  个个体的历史最优解  $p_i, i = 1, \dots, n$  和群体最优解  $p_g$  进行初始化。然后,基于每个个体对应的微调方案,使用公式(3)和公式(4)对模型  $W$  进行微调。接着,根据公式(2)计算出每个个体的适应度值,更新个体最优解  $p_i, i = 1, 2, \dots, n$  以及群体最优解  $p_g$ ,并使用公式(5)和公式(6)更新个体的速度和位置。经过  $G$  次更新迭代后,最终可以得到一个最佳的微调方案  $x^*$ ,最后根据  $x^*$  微调模型 110 个周期数得到最终的模型  $\hat{W}$ 。

**算法 1** 基于粒子群优化的自适应微调算法

**输入** 预训练模型  $W$ , 种群大小  $n$ , 种群迭代次数  $G$

- 1 随机产生拥有  $n$  个个体的种群:  $(v_i, x_i), i = 1, \dots, n$
- 2 初始化个体历史最优解  $p_i = x_i, i = 1, \dots, n$ ; 种群最优解:  $p_g = x_1$
- 3 for  $i = 1:G$  do
- 4     for  $j = 1:n$  do
- 5         基于个体位置  $x_j$ , 使用公式(4)计算掩码矩阵  $M$
- 6         使用公式(3)微调模型  $W_{\text{maxEpoch}}$  个周期
- 7         使用公式(2)计算第  $j$  个个体的适应度值  $\varphi_j$
- 8         if  $\varphi_j$  优于  $p_j$  then
- 9              $p_j = x_j$
- 10         if  $\varphi_j$  优于  $p_g$  then
- 11              $p_g = x_j$
- 12         end if
- 13     end if
- 14     end for
- 15      $x^* = p_g$
- 16     使用公式(5)更新所有个体速度  $v$
- 17     使用公式(6)更新所有个体的位置  $x$
- 18     end for
- 19     根据  $x^*$  微调模型 110 个周期

**输出** 微调后的最终模型  $\hat{W}$

## 2 实验结果与分析

本节中,报告并讨论了实验结果,以验证所提出算法对解决迁移学习中微调层选择问题的有效性。

### 2.1 数据集和预训练模型

通过在 4 个常用的公开图像分类数据集上的实验,评估了基于粒子群优化的自适应微调算法和现有微调方法的性能。数据集包括: MIT Indoors<sup>[18]</sup>、Stanford Dogs<sup>[19]</sup>、Caltech256-30 和 Caltech256-60<sup>[20]</sup>,数据集的统计信息见表 1。考虑到大多数方法都在 ResNet50 上进行实验,本文选择在 ImageNet<sup>[21]</sup> 数据集上预训练的 ResNet50 作为预训练模型。

表 1 用于评估微调性能的数据集

目标数据集	训练样本数	测试样本数	类别数
Stanford Dogs	12 000	8 580	120
MIT Indoor	5 360	1 340	67
Caltech256-30	7 680	5 120	256
Caltech256-60	15 360	5 120	256

### 2.2 实现细节和对比方法

所有实验均使用 Pytorch 框架在 NVIDIA 3090 GPU 上进行,利用 SGD 作为训练过程中的优化器,权重衰减和动量分别设置为 0.000 5 和 0.9,初始学习率和批量大小分别为 0.01 和 64。在用本文方法评估每个个体时微调周期数量设置为 5,最终微调时周期数量被设置为 110,并且每 30 个周期进行一次学习率衰减。每个实验均重复执行 5 次,以获得平均分类精度。

### 2.3 层微调方案的有效性分析

为了验证所提基于粒子群优化的自适应微调算法获得的层微调方案的有效性,进行了与随机微调方案的性能比较分析。

在所提算法的实验中,种群大小  $n$  的取值为 10,种群最大迭代次数  $G$  为 12,每个个体在评估适应度时的周期数  $\max Epoch$  设置为 5。在个体速度更新过程中,学习因子  $c_1$  和  $c_2$  均设置为 1.5,惯性权重  $\alpha$  采用线性递减的方式,且  $\alpha \in [0.7, 1.4]$ 。在随机微调方案的实验中,是否对源域模型的各层进行微调,取决于一个随机生成的 0~1 之间的数值。具体而言,若该随机数小于 0.5,则相应层被冻结;反之,则进行微调。因此,源域模型中微调的层是随机选定的。根据所提算法和随机选择获得的层微调方案,对源域模型进行了 110 个周期的微调,并在 4 个图像分类数据集上分别评估了其分类精度。

表 2 展示了两种微调方案在所有数据集上的测试集分类精度。由此可见,所提算法在所有数据集上获得了最好的微调结果,而随机选择的层微调后的结果比所提算法平均低了 3.8%,这表明精确选择微调层的重要性,一个不适合的选择会导致模型微调性能大幅度下降,所提算法可以在源域模型中精确的找到适合的层进行微调,从而有效地提高了模型在目标任务上的性能。此外,图 3 展示了所提算法在源域模型上获得的微调方案可视化结果。从图 3 可以看出,不同的目标任务在源域模型上的最佳微调方案具有显著差异。

表 2 不同微调方案的比较

数据集	MIT Indoors	Stanford-Dogs	Caltech256-30	Caltech256-60
随机微调方案	77.38	81.32	77.59	79.02
所提算法获得的微调方案	<b>78.28</b>	<b>86.45</b>	<b>80.97</b>	<b>84.84</b>

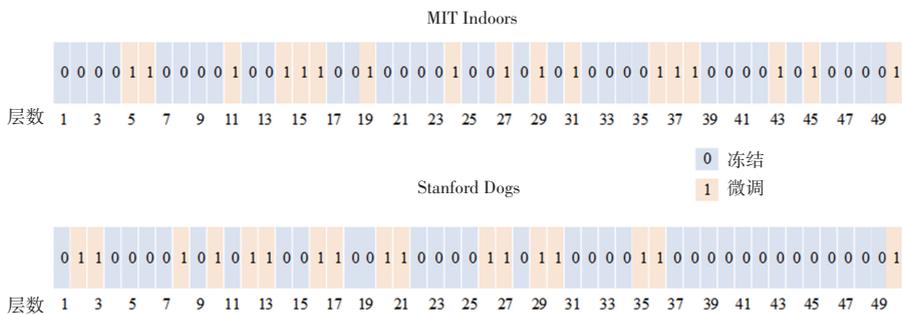


图 3 所提算法获得的微调方案可视化结果

Fig. 3 Visualization results of the fine-tuning scheme obtained by the proposed algorithm

综上所述,表 2 和图 3 的实验结果表明,在微调的过程中,源域模型中微调层的选择对模型的微调性能至关重要,所提算法可以准确的寻找到适合特定目标任务的微调层,从而提高了模型在目标任务上的性能。

### 2.4 与现有方法对比分析

本节中,将所提微调方法与 Train-From-Scrach

(不使用任何微调方法,直接使用随机初始化的权重从头训练)、Standard Fine-Tuning<sup>[12]</sup>(标准微调)、L2-SP<sup>[14]</sup>、Child-Tuning<sup>[15]</sup>等目前典型的微调方法分别应用在 4 个公开数据集上,以验证所提方法的有效性。所有的方法均进行 110 次迭代。为了消除随机性的影响,每个方法均重复 5 次实验以获得平均分类准确率。实验结果见表 3。

表 3 所提算法与现有方法的对比结果

Table 3 Comparison of the proposed algorithm with existing methods

%

Method	MIT Indoors	Stanford-Dogs	Caltech256-30	Caltech256-60
Train-From-Scrach	40.82	42.45	25.41	47.55
Standard Fine-Tuning <sup>[12]</sup>	76.64	79.02	77.53	82.57
L2-SP <sup>[14]</sup>	76.41	79.69	79.33	82.89
Child-Tuning <sup>[15]</sup>	77.83	81.13	80.19	83.63
Our Method	<b>78.28</b>	<b>86.45</b>	<b>80.97</b>	<b>84.84</b>

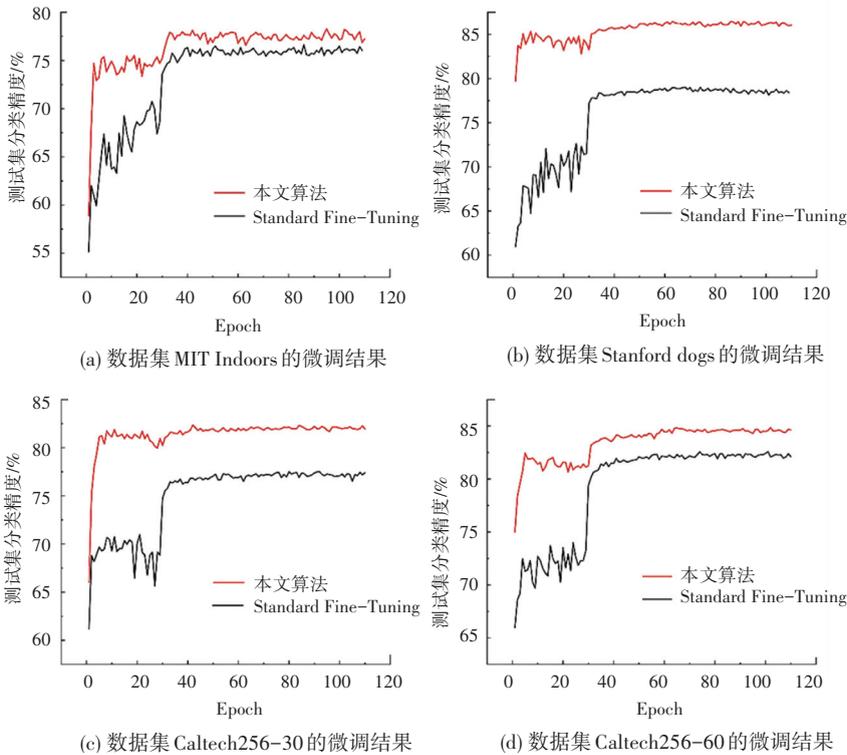


图 4 所提算法与标准微调基线方法的可视化结果

Fig. 4 Visualization of the proposed algorithm and the standard fine-tuning baseline method

由表 3 中数据可知,所提算法在所有数据集上均获得了最高的精度。其中,Train-From-Scrach 方法采用随机初始化的权重参数在目标任务上进行训练,没有使用任何微调技术,因此该方法的精度最低。

标准微调方法是迁移学习中常用的一种微调基

线,本文所提的算法相比标准微调基线平均增加了 3.7%,其原因在于本文算法可以提供准确的层微调方案,极大化的利用了源域知识,因此获得了更好的结果。图 4 中展示了所提算法与标准微调基线在 4 个数据集上的分类精度曲线。由图 4 可知,所提算法相比标准微调基线在所有数据集上大幅度提高了

微调性能。例如,经过5个 epochs 后,本文方法得到的分类准确率为 73.21%、83.87%、81.11% 和 82.44%,而基线方法仅有 62.61%、67.73%、69.66% 和 71.28%,表明本文方法可以用更少的训练周期达到甚至超过基线的精度水平,极大的减少了新任务的训练时间,在实际应用中具有重要作用。

### 3 结束语

本文提出了一种基于粒子群优化的自适应微调算法,通过将预训练模型中微调层选择问题转化为参数优化问题,并通过粒子群优化算法求解合适的层微调方案,克服了传统微调过程中存在的微调层难以选择的问题。实验结果表明,所提算法可以为目标任务精确的搜索出合适的层微调方案。与传统的微调方法相比,所提方法取得了更加优异的性能;与标准微调基线方法相比,本文方法只需几个训练周期便能达到与之相当甚至更高的性能。

### 参考文献

[1] DAI Z, CAI B, LIN Y, et al. Unsupervised pre-training for detection transformers [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (11): 12772-12782.

[2] CHI Zewen, HUANG Heyan, LIU Luyang, et al. Can pretrained english language models benefit non-english NLP systems in low-resource scenarios? [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023(32): 1061-1074.

[3] LIAO W, ZHANG Q, YUAN B, et al. Heterogeneous multidomain recommender system through adversarial learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34 (11): 8965-8977.

[4] LI Honglin, ZHU Chenglu, ZHANG Yunlong, et al. Task-specific fine-tuning via variational information bottleneck for weakly-supervised pathology whole slide image classification [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023; 7454-7463.

[5] CALDERARO J, GHAFARI L N, ZENG Q, et al. Deep learning-based phenotyping reclassifies combined hepatocellular-cholangiocarcinoma[J]. Nature Communications, 2023, 14(1): 8290.

[6] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 22 (10): 1345-1359.

[7] ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning [J]. Proceedings of the IEEE, 2021, 109 (1):

43-76.

[8] ZHU Z, MAO K. Knowledge-based bert word embedding fine-tuning for emotion recognition [J]. Neurocomputing, 2023, 552: 126488.

[9] LI H, FOWLKES C, YANG H, et al. Guided recommendation for model fine-tuning [C]// Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2023; 3633-3642.

[10] MENSINK T, UIJLINGS J, KUZNETSOVA A, et al. Factors of influence for transfer learning across diverse appearance domains and task types [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44 (12): 9298-9314.

[11] TAJBAKHS N, SHIN J Y, GURUDU S R, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? [J]. IEEE Transactions on Medical Imaging, 2016, 35 (5): 1299-1312.

[12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. arXiv preprint arXiv:2010.11929, 2020.

[13] RAGHU S, SRIRAAM N, TEMEL Y, et al. Eeg based multi-class seizure type classification using convolutional neural network and transfer learning [J]. Neural Networks, 2020, 124: 202-212.

[14] LI X, GRANDVALET Y, DAVOINE F. Explicit inductive bias for transfer learning with convolutional networks [C]// Proceedings of the 35th International Conference on Machine Learning. 2018; 2825-2834.

[15] XU R, LUO F, ZHANG Z, et al. Raise a child in large language model: Towards effective and generalizable fine-tuning [C]// Proceedings of Association for Computational Linguistics (ACL). 2021; 9514-9528.

[16] BASHA S H S, VINAKOTA S K, PULABAIGARI V, et al. AutoTune: Automatically tuning convolutional neural networks for improved transfer learning [J]. Neural Networks, 2021, 133: 112-122.

[17] ZENG N, WANG Z, LIU W, et al. Adynamic neighborhood-based switching particle swarm optimization algorithm [J]. IEEE Transactions on Cybernetics, 2022, 52 (9): 9290-9301.

[18] QUATTONI A, TORRALBA A. Recognizing indoor scenes [C]// Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009; 413-420.

[19] KHOSLA A, JAYADEVAPRAKASH N, YAO B, et al. Novel dataset for fine-grained image categorization: Stanford dogs [C]// Proceedings of CVPR (FGVC). 2011; 1577-1584.

[20] GRIFFIN G, HOLUB A, PERONA P. Caltech-256 object category dataset [DS]. Los Angeles: California Institute of Technology, 2007.

[21] YAMADA Y, OTANI M. Does robustness on imageNet transfer to downstream tasks? [C]// Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022; 9205-9214.