

2026 年度拟提名陕西省自然科学奖项目公示内容

一、项目名称

神经网络模型轻量化理论与参数高效微调关键技术研究

二、提名者及提名意见

提名者：

姓名	工作单位	职称	学科专业
王燕妮	西安建筑科技大学	教授	计算机科学与技术
张双喜	西北工业大学	教授	信息与通信工程

提名意见：

本人认真审阅了该成果提名书及附件材料，对该成果所取得的原创性科学发现、学术价值及学科贡献进行了全面评估。该成果围绕神经网络模型轻量化与参数高效微调中的关键科学问题，开展了系统深入的原创性研究，取得了以下重要科学发现：（1）揭示图拉普拉斯谱约束下关键拓扑支撑选择与多维表征保持机理，为图神经网络模型的轻量化结构设计夯实了理论基础；（2）揭示邻域协方差主子空间对随机噪声传播的抑制机理，成为图神经网络模型轻量化结构高效运作重要手段；（3）揭示预训练权重响应的稳定基底保持与低自由度重构机理，为视觉模型参数高效微调设计提供了统一的理论框架。相关研究成果发表于 CVPR、NeurIPS、TPAMI、TII 等本领域权威学术会议及期刊，得到国内外学术同行的广泛关注与正面引用，在推动神经网络模型高效部署发展方面发挥了重要作用。该成果研究思路新颖，方法科学严谨，数据翔实可靠，科学发现具有显著的原创性和重要的学术价值。项目提名材料齐全、规范，经完成单位公示，无知识产权纠纷，人员排序无争议，符合陕西省科学技术奖提名条件。

提名该项目为陕西省自然科学奖二等奖。

三、项目简介

立项背景与意义：以深度神经网络和大规模预训练模型为核心的人工智能正在由单一任务走向通用建模和规模化部署，随着模型大小、参数量与算子算法复杂度的持续增长，训练、迁移、存储和推理成本已成为制约人工智能广泛部署的关键瓶颈。2023—2026 年，国家战略部署由算力基础设施高质量建设逐步延伸至“人工智能+”规模化赋能，并提出算力、算法、数据协同发展。《“十五五”规划纲要》进一步提出统筹算力设施、模型算法和高质量数据资源供给，研究更加高效的模型训练与推理方法。由此，揭示神经网络模型轻量化机理，降低模型训练及任务间迁移的计算、存储和参数更新开销，是提高算力利用效率、推动人工智能在行业场景和资源受限终端高效部署的重要基础。

国内外高校与企业已从图采样、去噪聚合和参数高效适配等方向降低神经网络开销。图学习方法通过邻域节点采样提高大规模图训练效率，并关注深层图网络的表示特征对齐；视觉模型微调方法则通过冻结预训练主体、只更新少量参数来降低模型迁移成本。现有工作已提出多种方案，但训练阶段如何减少邻域采样与样本交互、聚合阶段如何兼顾去噪与高效计算存储、微调阶段如何提高可学习参数利用率，仍需建立理论可解释、机制可验证、并能够用工程指标量化的方法。

本成果以五篇代表性论文为基础，分别研究三类轻量化理论与参数高效微调关键技术问题：（1）图表示学习中如何减少邻域正样本特征的交互和对应负样本开销？（2）邻域聚合中如何抑制随机扰动并降低重复计算与中间存储？（3）预训练模型迁移中如何减少可训练参数？

针对这三类问题，本成果提出三项重要科学发现：（1）以图拉普拉斯变换理论为基础，提出 Topology-Aware Positive Sampling (TAPS)/Node-to-Neighbourhood (N2N) 方法压缩邻域正样本采样并保持其表示特征对齐，且进一步提出了 Negative-Free Node-to-Neighbourhood (NF-N2N)/Graph Signal Decorrelation (GSD) 方法取消负样本依赖并保持表示特征的区分度；（2）以主成分分析 Principal Component Analysis (PCA) 理论为基础，提出 De-Noising AGgregation (DNAG) 方法提取邻域主成分方向，并通过 Aggregation-First (AF) 方法一次性在系统内存中提前计算邻域节点聚合，避免聚合计算过程对 GPU 资源的浪费；（3）以奇异值分解 Singular Value Decomposition (SVD) 理论为基础，分别提出 Adapter Re-Composing (ARC) 方法跨层共享和 Residual-based Low-Rank Rescaling (RLRR) 行列调制压缩可训练参数。三项重要科学发现分别利用正负样本数、每轮训练时间、GPU 显存、可运行性、可训练参数和任务精度等指标对方法进行验证，在神经网络模型轻量化理论与参数高效微调关键技术等研究领域取得显著突破。

成果主要创新性成果及科学价值如下：

揭示关键正样本对齐与无负样本学习的轻量化训练机理。针对图神经网络中全部邻域节点反复参与模型损失函数计算、负样本生成和随机筛选造成的时间与显存开销，本成果提出 TAPS/N2N 与 NF-N2N/GSD 等方法来减少这些资源耗费。TAPS 方法在图结构中按拓扑结构依赖程度从候选邻域节点中选取少量关键正样

本；特别是，TAPS 方法的变体 N2N-TAPS-1 仅使用 1 个关键正样本，即可使得模型达到良好的精度。相对无采样的 N2N 方法，N2N-TAPS-1 在 6 个数据集的每轮训练时间由 0.19—1.68 ms 降至 0.09—0.24 ms，降幅 36.8%—85.7%，准确度 Micro-F1 指标变化为-2.06 至+0.16 个百分点。此外，成果提出的 NF-N2N 方法采用邻域采样节点对齐、非邻域节点之间去相关等技术，能够在无需负样本的情况下保持正样本的表示特征区分度，且防止模型训练过程中的塌缩现象。本成果的轻量化图学习方法依赖提出的正样本采样对齐和取消负样本依赖等关键技术，能够较大幅度减少训练交互、时间耗费与显存开销，有利于图学习方法部署到资源受限的边缘设备上，为广泛推广图模型应用和节能减排做出重要贡献。

揭示邻域主方向去噪与聚合计算、存储轻量化机理。针对随机扰动随邻域传播、逐层 PCA 重复计算和中间表示占用过大的问题，DNAG 方法提取前 K 个邻域节点特征的主成分方向，并将这些特征转换为聚合权重，以保留主成分信息、抑制随机噪声扰动；算法中 K 为主成分方向数而非邻居数，且不显式删除候选邻居。方法采用的 AF 训练架构将 PCA 移到训练前一次计算并复用，减少模型训练中的节点特征重复聚合和中间存储。当 K=5 时，一次性的 PCA 预先聚合计算（系统内存中）用时在 2.098—91.398 s 之间，相比其他方法时间下降比例为 99.58%—99.98%。该科学发现通过一次性预计算聚合操作和聚合-编码运算次序的调整，同时降低了聚合计算与存储开销，为资源受限环境下的大规模图学习提供了可执行的低开销方案。

揭示跨层共享与结构化调制的参数高效微调机理。针对基础模型全参数微调需为每个任务更新与保存整套权重以及逐层适配器参数量大等问题，成果提出 ARC 方法以共享 SVD 参数空间的方式来减少可学习参数量。在 VTAB-1k 数据集上，ARC 方法仅更新 0.13M 可训练参数，相比全参数微调以及其他方法参数量减少 67%—99.85%，平均准确率提升至 73.4%。依据该研究思路，成果进一步提出的 RLRR 方法，该方法的主要贡献在于提出了一个基于 SVD 理论的高效参数微调统一框架，仅更新 0.33M 参数即可使模型的平均准确率提升至 74.5%。两种方法均可在部署前将可学习参数吸附到基础模型的主干中，实现推理成本零增量。该科学发现价值在于把基础模型适配下游任务从整套参数学习方法变更为少量共享系数和向量学习方法，并提出跨层共享可学习参数架构和 SVD 微调统一框架，为参数高效微调技术提供了新的理论和实践视角。

本成果将模型轻量化贯穿训练、聚合和部署过程，并为低资源条件下的参数高效微调方法提供了可解释、可验证的实现路径。这些成果发表在 CVPR (2 篇)、IEEE TPAMI、IEEE TII 和 NeurIPS 等 5 篇顶级会议及期刊。本成果部分方法应用于厦门律点科技有限公司的便携脉诊仪设备、陕西切念科技有限公司的医疗仪器设备、西安超验科技有限公司和深圳市金析得科技服务有限公司的边缘工业检测设备。这些应用体现了成果面向仪器设备和边缘端场景的工程适配能力与进一步转化潜力。

四、客观评价

本项目 5 篇代表作中，三篇图神经网络成果分别发表于 CVPR 2022、IEEE TPAMI 2024 和 IEEE TII 2023，两篇视觉 Transformer 高效微调成果分别发表于 CVPR 2024 和 NeurIPS 2023。以下仅选取 8 篇与成果完成人无重合的独立评价论文，来源覆盖 IJCV、IEEE TPAMI、NeurIPS、ICCV、ACM MM、ICML 和 IEEE TII 等顶级会议及期刊；权威评价者包括多位国内外高层次人才。典型引用及评价如下：

1. **IEEE Fellow、ICCV 2019 程序主席 Ming-Hsuan Yang** 作为作者之一，在视觉领域顶刊 **IJCV 综述**[1，作者公开版 PDF 第 5 页，§3.1.1]中指出，ARC (NeurIPS 2023) 方法首创跨层共享瓶颈下投影与上投影，并利用低维重组系数生成层自适应适配器。该综述以“introduces the idea”明确将上述构思归属于 ARC (NeurIPS 2023) 方法，肯定了项目在跨层参数复用和层自适应调制方面的原创贡献。

2. **NeurIPS 2025 程序主席 Piotr Koniusz** 署名的 **NeurIPS 2024 论文**[2，官方 PDF 第 2 页，§2]指出，ARC (NeurIPS 2023) 方法利用跨层 ViT 相似性、参数共享适配器和缩放因子降低微调成本；RLRR (CVPR 2024) 方法则通过残差项提高适配灵活性，同时保持预训练表征。该评价准确识别了两项成果“低成本共享”和“残差保真”的技术作用。

3. **IEEE TPAMI 副编辑、CVPR 2024 程序主席吴建鑫教授团队**在 **TPAMI 论文**[3，作者公开 PDF 第 7 页，§7.3]中写道，DTLv1 和 DTLv2 相较“previous state-of-the-art method RLRR (CVPR 2024)”平均准确率提高 0.4 个百分点；同页 Table 3 报告 RLRR (CVPR 2024) 方法在五个细粒度数据集上的平均准确率为 90.4%。这表明 RLRR (CVPR 2024) 方法在该公开比较时已被权威同行明确认定为此前的最先进方法，并成为后续方法必须超越的强基准。

4. **中国工程院院士王耀南**署名的 **ICCV 2025 论文**[4，印刷页 853，§5.2 及 Tables 1–2]将 RLRR (CVPR 2024) 方法纳入主流 PEFT 方法比较：在 VTAB-1K 数据集上平均准确率为 76.8%，仅比 CVPT 方法低 0.4 个百分点；在 FGVC 数据集上平均准确率为 90.4%，仅比 CVPT 方法低 0.1 个百分点。该量化证据说明 RLRR (CVPR 2024) 方法在新一代顶会方法出现后仍保持接近最新方法的基准竞争力。

5. **IEEE Fellow、国家杰出青年基金获得者卢湖川教授**署名的 **ACM MM 2025 论文**[5，作者公开 PDF 第 2 页，§2.1]将 ARC (NeurIPS 2023)、RLRR (CVPR 2024) 等方法与另一类适配方法并列，明确评价其“achieve strong performance”，并分别概括为多适配器重组与残差低秩调节。该项会评价从性能和参数效率两方面对两项成果给出直接正面评价。

6. **NeurIPS 2024 领域主席 Lin Ma**署名的 **TPAMI 论文**[6，正式期刊页 8398，§II-B]指出，N2N (CVPR 2022) 互信息最大化方法能够“effectively obtain informative node representations”——通过邻域上下文有效获得信息丰富的节点表

示。其中“effectively”构成明确正面学术判断，肯定了 N2N (CVPR 2022) 方法利用邻域上下文提升节点表征信息量的机制价值。

7. ICME 和 PRCV 领域主席梁建青、CCF 人工智能与模式识别专委会执行委员王智强共同署名的 ICML 2025 论文[7, 官方 PDF 第 4 页, §2]将 NF-N2N (IEEE TPAMI 2024) 方法概括为利用互信息衡量锚节点与邻居的拓扑相似性以选择正样本，准确识别了“互信息—拓扑相似性—正样本采样”机制。该文下一句同时指出相关方法可能面临随机性和复杂度问题，故本材料仅以此说明 NF-N2N (IEEE TPAMI 2024) 方法已进入顶会后续方法比较和改进基准。

8. IEEE Fellow、欧洲科学院院士 Zidong Wang 署名的 IEEE TII 论文[8, 作者公开版 PDF 第 2 页, Introduction]在讨论小样本高速列车故障诊断时，将 DNAG (IEEE TII 2023) 作为参考文献[6]纳入“numerous lightweight models ... with limited samples”的技术链条。该引用体现了 DNAG (IEEE TII 2023) 轻量化思想在后续工业智能研究中的场景化传播。

上述 8 篇均为独立公开引文，且未使用项目成员参与的团队自引论文。论文 [1][3][5][6]属于明确正面评价或先进地位判断；论文[2]属于正向机制识别；论文[4]属于独立量化基准采用；论文[7]属于带局限讨论的机制引用；论文[8]属于场景价值定位。

代表性引文

- [1] Xin Y, Yang J, Luo S, Du Y, et al. Parameter-Efficient Fine-Tuning for Pre-Trained Vision Models: A Survey and Benchmark[J]. IJCV, 2026, 134: Article 304.
- [2] Ni Y, Zhang S, Koniusz P. PACE: Marrying Generalization in PArAmeter-efficient Fine-tuning with Consistency rEgularization[C]//NeurIPS 37. 2024.
- [3] Fu M, Zhu K, Ding Z, Wu J. DTL: Parameter- and Memory-Efficient Disentangled Vision Learning[J]. TPAMI, 2026, 48(2): 1736–1749.
- [4] Huang L, Mao J, Yi J, Tao Z, Wang Y. CVPT: Cross Visual Prompt Tuning[C]//ICCV. 2025: 848–858.
- [5] Zhu Y, Diao H, Gao S, et al. Regularizing Subspace Redundancy of Low-Rank Adaptation[C]//ACM MM. 2025: 1666–1675.
- [6] Zhang L, Song R, Tan W, Ma L, Zhang W. IGCN: A Provably Informative GCN Embedding for Semi-Supervised Learning With Extremely Limited Labels[J]. TPAMI, 2024, 46(12): 8396–8409.
- [7] Liang J, Li Z, Wei X, Liu Y, Wang Z. ML²-GCL: Manifold Learning Inspired Lightweight Graph Contrastive Learning[C]//ICML. PMLR, 2025, 267: 37222–37236.
- [8] Xue Y, Yang R, Chen X, Song B, Wang Z. Separable Convolutional Network-Based Fault Diagnosis for High-Speed Train: A Gossip Strategy-Based Optimization Approach[J]. TII, 2025, 21(1): 307–316.

五、代表性论文专著目录（不超过 8 条，其中代表性论文不超过 5 篇，代表性专著不超过 3 部）

序号	论文专著 名称	刊名	作者	年卷页码 (xx 年 xx 卷 xx 页)	发表 时间	通讯 作者	第一 作者	国内作 者	他 引 总 次 数	检索数据 库	知识产权 是否归国内所有
1	Self-supervised node representation learning via node-to-neighbourhood alignment	IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)	Wei Dong, Dawei Yan, Peng Wang	2024 年 46 卷 4218-4233 页	2024 年 1 月 25 日	Peng Wang	Wei Dong	董玮 闫大伟	8	Web of science	是
2	Node representation learning in graph via node-to-neighbourhood mutual information maximization	2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	Wei Dong, Junsheng Wu, Yi Luo, Zongyuan Ge, Peng Wang	2022 年 16599-16608 页	2022 年 6 月 18 日	Junsheng Wu, Peng Wang	Wei Dong	董玮 武君胜 罗毅	39	Web of science	是
3	Efficient adaptation of large vision transformer	Advances in Neural	Wei Dong, Dawei Yan,	2023 年 36 卷	2023	Peng Wang	Wei Dong	董玮 闫大伟	42	Web of science	是

	via adapter re-composing	Information Processing Systems (NeurIPS)	Zhijun Lin, Peng Wang	52548-52567 页	年 12 月 15 日	Wan g	Do ng	林智君			
4	Low-rank rescaled vision transformer fine-tuning: A residual design approach	2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	Wei Dong, Xing Zhang, Bihui Chen, Dawei Yan, Zhijun Lin, Qingsen Yan, Peng Wang, Yang Yang	2024 年 16101-16110 页	20 24 年 6 月 16 日	Pen g Wan g	We i Do ng	董玮 张兴 陈碧辉 闫大伟 林智君 闫庆森 杨阳	33	Web of science	是
5	Denoising aggregation of graph neural networks by using principal component analysis	IEEE Transactions on Industrial Informatics (TII)	Wei Dong, Marcin Woźniak, Junsheng Wu, Weigang Li, Zongwen Bai	2023 年 19 卷 2385-2394 页	20 23 年 3 月 7 日	Mar cin Woź nia k	We i Do ng	董玮 武君胜 李伟刚 白宗文	77	Web of science	是
合 计									199		

六、主要完成人情况（不超过 6 人）

姓名	排名	行政职务	技术职称	工作单位	完成单位	对本项目贡献
董玮	1	西安建筑科技大学计算机与信息工程学院计算机系支部书记	教授	西安建筑科技大学	西安建筑科技大学	承担了图表示学习模型的轻量化方法、模型参数高效微调方法的设计工作，主要包括：轻量化的自监督图对比学习方法；模型参数高效微调 LoRA 方法。该完成人为代表性论文 1、2、3、4 和 5 的唯一第一作者。
林智君	2	无	无	西北工业大学	西北工业大学	参与了模型参数高效微调方法的设计工作，主要方向为模型参数高效微调 LoRA 方法研究。该完成人为代表性论文 3 和 4 的作者。
闫大伟	3	无	无	西北工业大学	西北工业大学	参与了图表示学习模型的轻量化方法、模型参数高效微调方法的研究工作，主要包括：轻量化的自监督图对比学习方法；模型参数高效微调 LoRA 方法。该完成人为代表性论文 1、3 和 4 的作者。
闫庆森	4	无	教授	西北工业大学	西北工业大学	参与了模型参数高效微调方法的设计工作，主要研究方向为 LoRA 方法的持续学习特性研究。该完成人为代表性论文 4 的作者。

白宗文	5	延安大学 科技处处长	教授	延安大学	延安大学	参与了图表示学习模型的轻量化方法的研究工作, 主要研究方向为轻量化的自监督图对比学习方法的主成分聚合设计。该完成人为代表性论文 5 的作者。
-----	---	---------------	----	------	------	--

七、主要完成单位情况（不超过 3 个）

完成单位	排名	对本项目主要贡献（限 600 字）
西安建筑科技大学	1	第一完成人以西安建筑科技大学为第一完成单位，一直以来从事图表示学习模型的轻量化方法、模型参数高效微调方法等研究工作。该完成单位为成果的测试与应用提供了良好的硬件软件平台，支撑完成人在轻量化图表示学习、模型参数高效微调领域进一步探索与研究。
西北工业大学	2	第二完成单位主要承担的是视觉模型参数高效微调方法与持续学习等研究工作。西北工业大学在视觉模型深度学习技术方面拥有丰富的研究与应用经验，为本成果打下了坚实的科研基础。
延安大学	3	第三完成单位主要承担的是图表示学习方法的轻量化研究工作。延安大学在图数据挖掘技术方面拥有丰富的研究积累，为本成果打下了坚实的科研基础。

八、完成人合作关系说明

西北工业大学博士生林智君，在本项目中主要承担视觉模型参数高效微调框架设计，并做出重要贡献。与董玮教授合作共同完成《Efficient adaptation of large vision transformer via adapter re-composing》(NeurIPS, 2023), 《Low-rank rescaled vision transformer fine-tuning: A residual design approach》(CVPR, 2024) 等多篇论文的撰写与发表。

西北工业大学博士生闫大伟，在本项目中主要承担“高效微调+持续学习”联合训练策略的构建与优化等研究工作，并做出重要贡献。与董玮教授合作共同完成《Self-supervised node representation learning via node-to-neighbourhood alignment》(TPAMI, 2024), 《Efficient adaptation of large vision transformer via adapter re-composing》(NeurIPS, 2023), 《Low-rank rescaled vision transformer fine-tuning: A residual design approach》(CVPR, 2024) 等多篇论文的撰写与发表。

西北工业大学闫庆森教授，在本项目中主要承担视觉模型参数高效微调方法与持续学习的实验验证等研究工作，并做出重要贡献。与董玮教授合作共同完成《Low-rank rescaled vision transformer fine-tuning: A residual design approach》(CVPR, 2024) 等多篇论文的撰写与发表。

延安大学白宗文教授，在本项目中主要承担的是图表示学习方法的轻量化研究工作，并做出了重要贡献。与董玮教授合作共同完成《Denoising aggregation of graph neural networks by using principal component analysis》(TII, 2023) 等多篇论文的撰写与发表。

完成人合作关系情况汇总表

序号	合作方式	合作关系人及排名	合作时间	合作成果	证明材料
1	论文合著	董玮（1），林智君（2）	2023 年 12 月 15 日	NeurIPS 2023 会议论文一篇（论文 3，CCF-A 类）	《Efficient adaptation of large vision transformer via adapter re-composing》(NeurIPS, 2023)
2	论文合著	董玮（1），林智君（2）	2024 年 6 月 16 日	CVPR 2024 会议论文一篇（论文 4，CCF-A 类）	《Low-rank rescaled vision transformer fine-tuning: A residual design approach》(CVPR ,2024)
3	论文合著	董玮（1），闫大伟（3）	2024 年 1 月 25 日	IEEE TPAMI 2024 期刊论文一篇（论文 1，CCF-A 类，SCI 一区）	《Self-supervised node representation learning via node-to-neighbourhood alignment》(TPAMI , 2024)
4	论文合著	董玮（1），闫大伟（3）	2023 年 12 月 15 日	NeurIPS 2023 会议论文一篇（论文 3，CCF-A 类）	《Efficient adaptation of large vision transformer via adapter re-composing》(NeurIPS, 2023)
5	论文合著	董玮（1），闫大伟（3）	2024 年 6 月 16 日	CVPR 2024 会议论文一篇（论文 4，CCF-A 类）	《Low-rank rescaled vision transformer fine-tuning: A residual design approach》(CVPR ,2024)
6	论文合著	董玮（1），闫庆森（4）	2024 年 6 月 16 日	CVPR 2024 会议论文一篇（论文 4，CCF-A 类）	《Low-rank rescaled vision transformer fine-tuning: A residual design approach》(CVPR ,2024)
7	论文合著	董玮（1），白	2023 年 3 月	IEEE TII 2023	《Denoising

	著	宗文（5）	7 日	期刊论文一篇 （论文 5，CAA-A 类，SCI 一区）	aggregation of graph neural networks by using principal component analysis》 （TII, 2023）
--	---	-------	-----	------------------------------------	---