0254-6124/2026/46(2)-0380-12 *Chin. J. Space Sci.* 空间科学学报

LUO Xi, ZHOU Qing, JIANG Yuanyuan. Single Event Upsets Fault Tolerance of Convolutional Neural Networks Based on Adaptive Boosting (in Chinese). *Chinese Journal of Space Science*, 2026, **46**(2): 380–391. CSTR:32142.14.cjss.2025–0025. DOI:10.11728/cjss2026.02.2025–0025

基于自适应增强算法的卷积神经网络单粒子翻转容错方法*

罗 熙[1,2] 周 晴[1] 江源源[1]

1(中国科学院国家空间科学中心 北京 100190)

2(中国科学院大学 北京 100049)

摘 要 空间辐射环境下的单粒子翻转效应严重威胁着星载智能系统的可靠性, 传统的三模冗余和周期性擦写等容错方法存在资源开销大、功耗高等问题. 提出一种基于自适应增强算法的轻量化容错方法 (AB-FTM), 通过该方法构建 ResNet20/32/44 异构弱模型集成架构, 结合动态权重调整机制, 不仅显著减少参数规模 (相比于原始 ResNet110 缩减 18.2%), 而且提升了分类精度与鲁棒性, 增强了容错能力. 在 CIFAR-10, MNIST, EuroSAT 和 Galaxy10 DECals 数据集上的实验验证表明, 当 0.032‰ 比例的参数发生单粒子翻转时, AB-FTM 较 ResNet110 三模冗余的准确率分别提升 53.25%, 63.49%, 57.67% 和 47.43%, 显著优于传统三模冗余方案. 该方法为未来空间科学卫星使用星载智能系统提供了兼顾可靠性、轻量化与计算效能的新型解决方案.

关键词 单粒子翻转, 自适应增强算法, 卷积神经网络, 容错, 航天器

中图分类号 P354

Single Event Upsets Fault Tolerance of Convolutional Neural Networks Based on Adaptive Boosting

LUO Xi[1,2] ZHOU Qing[1] JIANG Yuanyuan[1]

1(*National Space Science Center*, *Chinese Academy of Sciences*, *Beijing* 100190)

2(*University of Chinese Academy of Sciences*, *Beijing* 100049)

Abstract Single-Event Upsets (SEUs) in the space radiation environment pose a serious threat to the reliability of satellite-borne intelligent systems. Traditional fault-tolerance methods such as Triple Modular Redundancy (TMR) and periodic scrubbing face challenges including excessive resource overhead and high power consumption. This paper presents a lightweight fault-tolerance method based on Adaptive Boosting-based Fault-Tolerance Method (AB-FTM) to address SEU vulnerabilities in convolutional neural networks. The proposed approach constructs a heterogeneous ensemble architecture comprising three weak models (ResNet20, ResNet32, ResNet44) and integrated with a dynamic weight adjustment mecha-

* 中国科学院太空探源专项项目资助 (GJ110100)

2025-02-15 收到原稿, 2025-06-25 收到修定稿

©The Author(s) 2026. This is an open access article under the CC-BY 4.0 License (https://creativecommons.org/licenses/by/4.0/)

nism. By integrating a dynamic weight adjustment mechanism, the method not only significantly reduces the parameter scale (achieving an 18.2% reduction compared to ResNet110) but also enhances classification accuracy, robustness, and fault tolerance. Experimental validation on datasets including CIFAR-10, MNIST, EuroSAT, and Galaxy10 DECals demonstrates that when 0.032‰ of parameters are affected by single-event upsets, the proposed method improves classification accuracy by 53.25%, 63.49%, 57.67%, and 47.43% respectively compared to the TMR-based ResNet110, significantly outperforming traditional triple modular redundancy solutions. This approach provides a novel solution for future space science satellites employing satellite-borne intelligent systems, balancing reliability, lightweight design, and computational efficiency.

Key words Single event upset, Adaptive boosting, Convolutional neural network, Fault tolerance, Spacecraft

0 引言

随着航天器智能化、自主化水平大幅提升, 基于深度学习的在轨智能处理系统已成为空间任务的核心模块之一, 可集成于航天器的智能决策系统中, 实现自主导航、任务规划及空间非合作目标的意图识别[1] 等功能. 对于可能出现通信中断或高延迟的航天器, 在轨智能处理系统能够及时调整任务执行, 无需等待地面指令, 有效提升任务效率.

以祝融号火星车[2] 为例, 由于地火回路时延长达 6～44 min 且下行数据量有限, 地面无法实时遥控, 需借助智能处理系统在无测控支持下自主执行. 以中国首颗综合性太阳观测卫星[3] 夸父一号 (ASO-S) 为例, 未来可利用智能处理系统实时监测并解析筛选日冕物质抛射、耀斑加速高能粒子等空间天气事件. 自主优化观测策略, 例如调整 FMG 载荷指向太阳黑子磁流浮现区, 以捕获关键的瞬态数据, 更好地探测太阳爆发, 预报空间天气, 为中国空间环境的安全提供保障. 此外, 空间科学卫星鸿蒙计划[4] 未来也可以通过搭载智能处理系统, 实现实时分析海量射电频谱数据, 动态识别宇宙信号与噪声特征, 自主优化卫星编队观测模式、智能故障诊断与容错, 更好地实现高精度全天频谱测量、超长波巡天及行星射电活动监测的科学目标.

然而空间辐射环境对神经网络的可靠性构成了严峻威胁. 统计显示[5], 在地球低轨道 (LEO, 高度 100～1000 km) 上单粒子翻转 (Single Event Upset, SEU) 是卫星发生故障最主要的原因. 这种位翻转不仅会导致神经网络模型精度骤降, 更可能引发系统级决策功能失效, 严重威胁航天器任务安全.

传统的抗单粒子翻转方法主要采用三模冗余设计[6]、周期性擦写[7]、ECC 纠错等传统容错机制. 尽管这些方法在电子系统中已获验证, 但应用于神经网络时存在显著局限性. 由于神经网络的推理依赖于所有参数, 通常参数均是百万量级以上, 硬件冗余导致 3 倍以上资源开销, 计算复杂量增大, 难以适配星载平台严格的功耗约束; 周期性擦写虽能修复错误, 但是不能防止神经网络参数发生单粒子翻转时引发的错误传递. 同时, 频繁存取操作加剧存储单元磨损并产生额外能耗; ECC 纠错编解码过程中增加的延迟, 可能导致高速数据流因校验计算而使整体速度下降, 进而形成瓶颈. 因此, 研究开始探索针对神经网络的单粒子翻转评估与容错方法. 例如 Ruospo 等[8] 在 2021 年通过研究参数的位宽, 使用浮点和定点两种不同的数据类型来更高效地表示数据, 以提高卷积神经网络的效率与可靠性. Ibrahim 等[9] 在 2020 年测试了 GPU, FPGA, ASIC 三种类型的神经网络加速器的单粒子翻转情况, 发现 GPU 和 ASIC 在硬件资源、浮点数与定点数、错误传播行为和容错技术方面具有优势. Qian 等[10] 在 2021 年提出了一种基于 dropout 算法的单粒子容错方法, 可以在不增加额外资源消耗的情况下, 加固后的 LeNet-5 网络相对于未使用任何加固方法的 LeNet-5 网络准确率提升约 2%, 但该容错算法在更大规模的模型中效果并不明显.

针对上述问题, 提出一种基于自适应增强算法, 区别于传统单一模型加固思路, 通过构建 ResNet20/32/44 异构弱模型集成架构, 结合动态权重调整机制,

在显著减少参数规模 (相比原始 ResNet110, 缩减了 18.2%) 的同时, 提升了分类精度与鲁棒性, 具有更强的容错能力. 该方法创新性地将模型差异性转化为容错优势, 通过多层级错误屏蔽与补偿机制, 有效抑制单粒子翻转的传播影响. 实验在遥感数据集 EuroSAT[11] 与天文星系数据集 Galaxy10 DECals[12] 等多个数据集上验证表明, 新方法在 0.032‰比例的参数发生单粒子翻转时, 相比传统三模冗余方案, 准确率至少提高了 47.43%, 显著优于传统的三模冗余容错方案. 本文为星载智能系统提供了兼顾可靠性、轻量化与计算效能的新型解决方案.

1 神经网络参数单粒子翻转模拟方法

1.1 单粒子翻转机理

单粒子翻转是指高能带电粒子穿过半导体器件的敏感区域时, 电离作用导致存储单元中的数据发生翻转的现象. 以典型的 SRAM 存储单元为例, 其存储机制是通过两个反相器构成的双稳态电路来实现的. 当高能粒子击中 SRAM 的敏感区域时, 产生的瞬态电流可能会改变存储单元节点的电压状态, 使其从逻辑 1 翻转为逻辑 0, 或从逻辑 0 翻转为逻辑 1. 这种翻转效应不会对电子元器件造成永久性损伤, 但其引发的存储单元逻辑状态错误会持续存在, 直至对该存储单元执行下一次写入操作, 或系统复位后才会恢复正常. 对于神经网络而言, 这种翻转可能直接作用于权重矩阵或偏置参数的存储单元, 导致模型推理结果出现不可预知的偏差.

当前, 低轨大规模卫星星座 (例如 SpaceX 星链、OneWeb 及中国千帆计划) 呈现爆发式部署态势, 单星座规模可达数千至数万颗. 受商业卫星低成本设计约束, 此类平台普遍采用抗单粒子翻转阈值较低的商用级 SRAM 器件. 根据辐射环境模型分析, 在典型低轨高度 (1000～1200 km) 运行期间, SRAM 器件的 SEU 事件空间分布呈现显著地域相关性. 在轨单粒子翻转大部分发生在南大西洋异常区 (SAA), 对于抗单粒子翻转阈值为 0.7 MeV·cm^2·mg^{-1} 的低阈值 SRAM 器件, 发生概率约为 3.18×10^{-8} bit^{-1}·h^{-1}[13]. 以 ResNet110 网络为例, 其 1.7×10^6 参数规模在 SAA 区域暴露 1 h 可引发≥1 次 SEU 事件, 这对需要高可靠的航天任务来说是非常危险的, 因此有必要针对航天环境下的神经网络采取加固措施.

1.2 可控的单粒子翻转模拟方法

由于主流深度学习框架 (例如 PyTorch 和 TensorFlow) 默认使用 IEEE 754 标准的 32 位单精度浮点数 (Binary 32) 进行模型训练与推理, 且 32 位单精度浮点数更贴近航天器常用处理器的数据位宽, 因此, 这里采用 32 位单精度浮点数作为神经网络参数的存储格式. 根据 IEEE754 的规定, 32 位单精度浮点数由 1 位符号位, 8 位指数位, 23 位尾数位组成, 其表达式为

$$V=-1^S\times2^{(E-127)}\times(1+M). \tag{1}$$

其中, S 为符号位值 (0 表示正数, 1 表示负数), E 为指数位的无符号整数值, M 为尾数位代表小数.

在空间环境中单粒子翻转是随机的, 可能发生在任意浮点数的任意比特位, 对网络参数的影响差异显著, 高位翻转比低位翻转影响更大. 权重参数符号位翻转可能会导致数值正负号反转, 相当于将神经元连接从抑制变为激励 (或相反), 可能完全改变特征提取的方向. 指数位翻转也会导致权重参数的数量级的变化. 32 位单精度浮点数的模型权重参数尾数位发生翻转情况如图 1 所示.

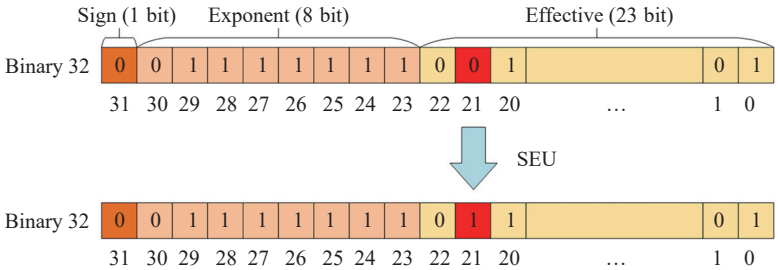


图 1 32 位单精度浮点数发生单粒子翻转

Fig. 1 Single Event Upset (SEU) in Binary 32

此外, 传统 SEU 仅限于单个比特位, 这种假设在早期半导体工艺中基本成立[14]. 随着半导体工艺的微缩化, 电荷共享导致的多比特翻转也变得日益显著[15]. 电荷共享 (Charge Sharing) 效应是指当高能粒子撞击 SRAM 存储单元时, 产生的电荷不仅会影响撞击点的存储单元, 还会通过扩散机制影响相邻的存储单元. 这将会导致连续多个比特位可能同时发生翻转, 造成多比特翻转 (Multiple Bit Upset, MBU). 根据 Bage 等[16] 于 2009 年的测试实验, 这种连续性翻转的概率分布符合几何分布模型, 即

$$P(X=k)=(1-r)^{k-1}r,\ k\geqslant 1. \tag{2}$$

其中, k 表示连续翻转的比特数; 参数 r 表示电荷扩散终止的概率, 即在翻转一个比特位后停止进一步翻转的概率, 基于 45 nm 工艺 SRAM 器件的辐照测试结果显示 $r=0.57$[16].

为了系统评估星载神经网络在空间环境中面临的单粒子翻转问题, 设计了一种可控的参数错误注入算法, 用于模拟高能粒子辐射导致的单粒子翻转现象. 使用的深度学习框架为 PyTorch, 模型参数存储于 pth 格式文件中, 参数以十进制数表示, 需将其转换为二进制表示的 32 位单精度浮点数, 再随机注入比特位翻转. 设计的神经网络单粒子翻转仿真注错算法核心步骤如下.

步骤 1 参数提取. 提取神经网络中所有的参数.

步骤 2 选择参数. 在所有参数中随机选择一个作为翻转的目标参数.

步骤 3 参数二进制转化. 获取目标参数的十进制浮点数值后, 将其转换为符合 IEEE754 格式的二进制表示形式, 该步骤是实现比特级操作的基础, 因为单粒子翻转作用的对象是二进制数据中的比特位.

步骤 4 翻转注入. 根据实验需求, 随机翻转目标参数中某一比特位, 并根据几何分布, 翻转该比特位附近的比特位.

步骤 5 数值重构与保存模型. 将翻转后的二进制串转换为十进制浮点数.

步骤 6 根据预设的单粒子翻转次数 N_{error}, 执行多次参数修改操作, 每次修改操作均是独立的单粒子翻转过程. 重复步骤 2～5.

步骤 7 保存注入比特位错误后的模型参数.

该算法可以灵活地设置翻转次数, 设置不同类型比特位 (符号位、指数位、尾数位) 翻转, 在神经网络模型的权重参数中, 能够按照翻转次数注入任意比特位翻转错误, 且支持多比特翻转, 同时模拟了电荷共享效应, 较真实地仿真了空间单粒子翻转机制, 为后续容错方法的有效性验证提供实验基础.

2 基于 ResNet 架构的自适应增强容错方法

整体三模冗余 (Triple Modular Redundancy, TMR) 通过部署三个相同的神经网络模型, 结合多数投票机制实现容错, 虽然可以有效应对单粒子翻转问题, 但存在资源开销大、功耗高等缺陷. 虽然已有研究[17,18] 通过对神经网络的不同模块进行敏感性分析, 并对其中关键的模块进行三模冗余, 从而达到减少资源开销的目的, 但由于神经网络的决策能力分布在所有参数中, 难以仅通过保护少量关键参数获得接近整体三模冗余的容错能力[17]. 为了克服这些限制, 提出了一种基于自适应增强 (Adaptive Boosting, AdaBoost) 算法[19] 的轻量级容错方法. 该方法通过集成多个更小的弱神经网络模型, 并根据弱模型的表现实时调整权重, 在大幅降低资源消耗的同时保持较强的容错性能.

2.1 ResNet 架构

ResNet (残差神经网络) 由 He 等[20] 提出, 其核心创新是通过图 2 的残差块 (Residual block) 结构解决深度神经网络中的梯度消失问题. 每个残差块引入跳跃连接 (Skip connection), 将输入信号直接传递到

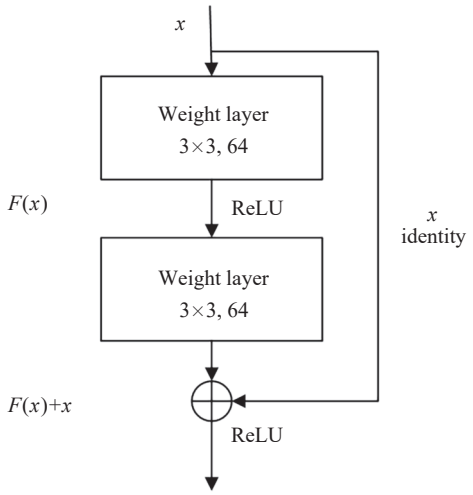


图 2 残差学习的一个基础模块

Fig. 2 Building block of residual learning

后续层, 使网络能够学习输入与输出之间的残差映射. 该设计使 ResNet 可扩展至数百层, 而无需牺牲训练稳定性, 从而在图像分类、目标检测等任务中取得突破性性能. 相比于 VGG 等其他网络, ResNet 复杂度较低, 并且可以达到更深的层次, 其结构特点适合使用不同深度和参数量的变体来研究加固方法. 表 1 是基于不同数据集任务的 ResNet 架构对比. 其中 ResNet18, ResNet34, ResNet50 和 ResNet101 使用的参数量均较多. 考虑到航天器资源较为紧张, 这里选用 ResNet 的轻量化变体 (ResNet20, ResNet32, ResNet44) 作为自适应增强算法的基础模型, 参数量更多的 ResNet110 作为对比模型. 这些变体通过调整残差块数量和通道维度控制模型复杂度.

2.2 基于自适应增强算法的轻量化容错方法

提出一种基于自适应增强算法的卷积神经网络抗单粒子翻转容错方法 (AB-FTM). AB-FTM 算法将 Adaboost 集成学习框架与 ResNet 深度残差网络的优势相结合, 形成一种既能保持高分类性能, 又具备强大单粒子翻转容错能力的星载神经网络解决方案.

AB-FTM 算法的设计核心在于使用多个结构差异化的较浅 ResNet 变体作为基础模型, 在训练阶段, 通过 Adaboost 机制动态调整其权重系数, 在推理阶段, 根据基础模型的表现实时调整权重系数, 构建一个能够减轻单粒子翻转影响的鲁棒模型系统. 相比于使用单一线性层集成几个模型, 若单粒子翻转影响了某个关键权重, 在单一模型中可能导致严重错误, 而在异构集成的 AB-FTM 中, 其他模型不会同时失效.

表 1 基于不同数据集任务的 ResNet 架构对比

Table 1 Comparison of ResNet architectures based on different dataset tasks

架构	数据集	数据集尺寸/pixel	参数量 ($\times 10^6$)	浮点运算次数
ResNet18	ImageNet	256×256	11.69	1.8×10^9
ResNet34	ImageNet	256×256	21.80	3.6×10^9
ResNet50	ImageNet	256×256	25.56	3.8×10^9
ResNet101	ImageNet	256×256	44.55	7.6×10^9
ResNet20	CIFAR-10	32×32	0.27	4.0×10^7
ResNet32	CIFAR-10	32×32	0.46	6.9×10^7
ResNet44	CIFAR-10	32×32	0.66	9.7×10^7
ResNet110	CIFAR-10	32×32	1.72	2.5×10^8

此外, 与传统三模冗余方法不同, AB-FTM 充分考虑了神经网络参数的多样性与互补性, 显著降低了硬件资源占用, 同时保持甚至提升了容错性能. 根据上一节对于 ResNet 模型选择的分析, 将结构相似但深度不同的 ResNet20, ResNet32, ResNet44 作为基础模型, 其模型集合为 $\{f_1, f_2, f_3\}$, f_1对应 ResNet20, f_2对应 ResNet32, f_3对应 ResNet44. AB-FTM 的训练过程基于 Adaboost 算法, 对于给定的训练集 $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$, 其中 x_i表示输入图像, $y_i \in \{1, 2, \cdots, C\}$表示对应的类别标签, 对于 $C = 10$的 CIFAR-10 数据集, 算法的训练过程如下.

(1) 初始化样本权重分布. 样本的权重分布表达式为

$$D_1(i) = \frac{1}{N}, \quad i = 1, 2, \cdots, N. \tag{3}$$

(2) 基础模型的迭代步骤. 对每个基础模型 $\{f_1, f_2, f_3\}$, $t = 1, 2, 3$ 执行以下迭代步骤.

步骤 1 使用当前权重分布 D_t对 ResNet 模型 f_t进行训练, 训练时采用加权采样策略确保样本按照 D_t分布被使用.

步骤 2 计算 f_t在训练集上的加权错误率δ_{ER}, 其表达式为

$$\delta_{\mathrm{ER}} = \sum_{i=1}^{N} D_t(i) \cdot I\left[f_t(x_i) \neq y_i\right]. \tag{4}$$

其中, $I[\cdot]$ 为指示函数, 当括号内条件成立时取 1, 否则取 0.

步骤 3 若$\delta_{\mathrm{ER}} \geqslant 0.5$, 则调整模型参数, 并且重新训练, 直至$\delta_{\mathrm{ER}} < 0.5$.

步骤 4 计算模型权重系数a_t, 其表达式为

$$a_t = \frac{1}{2} \ln\left(\frac{1 - \delta_{\mathrm{ER}}}{\delta_{\mathrm{ER}}}\right) + \ln(C - 1). \tag{5}$$

步骤 5 更新样本权重分布D_{t+1}, 其表达式为

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \cdot \begin{cases} \mathrm{e}^{-a_t}, & f_t(x_i) = y_i; \\ \mathrm{e}^{\frac{a_t}{(C-1)}}, & f_t(x_i) \neq y_i. \end{cases} \tag{6}$$

若样本被基础模型 f_t正确分类, 其权重缩小e^{-a_t}倍; 若分类错误, 则权重放大$\mathrm{e}^{\frac{a_t}{(C-1)}}$倍. Z_t是规范化因子, 确保更新后的权重分布D_{t+1}总和为 1, 其表达式为

$$Z_t = \sum_{i=1}^{N} D_t(i)\{\mathrm{e}^{-a_t} I\left[f_t(x_i) = y_i\right] + \mathrm{e}^{a_t/(C-1)} I\left[f_t(x_i) \neq y_i\right]\}. \tag{7}$$

(3) 定义最终集成模型. 最终集成模型定义为

$$F(x) = \underset{y\in\{1,2,\cdots,C\}}{\operatorname{argmax}} \sum_{t=1}^{3} a_t\, I\left[f_t(x) = y\right]. \tag{8}$$

其中, argmax[·] 为使目标函数取得最大值的自变量的函数.

在推理阶段, AB-FTM 引入了动态权重调整机制, 以减少某个基础模型性能下降或失效时对整个模型的影响, 并将该机制表示为一个权重调整函数 AdjustWeight[·], 用于根据模型输出的一致性动态调整各基础模型的权重, 其表达式为

$$\mathrm{AdjustWeight}\left[f_1(x), f_2(x), f_3(x), \alpha_1, \alpha_2, \alpha_3\right] = [\beta_1, \beta_2, \beta_3]. \tag{9}$$

该函数的输入为三个基础模型的输出预测类别$f_1(x)$, $f_2(x)$, $f_3(x)$以及对应的初始权重系数α_1, α_2, α_3, 输出为调整后的权重系数β_1, β_2, β_3. 其具体实现逻辑如下.

如果三个基础模型的预测结果是一致的, 即 $f_1(x) = f_2(x) = f_3(x)$, 则保持原有权重不变, 有

$$\beta_t = \alpha_t, t \in \{1, 2, 3\}. \tag{10}$$

如果两个基础模型的预测结果一致, 另一个不同, 即 $f_i(x) \neq f_j(x) = f_k(x)$, 其中 $i, j, k \in \{1, 2, 3\}$, 且 $i \neq j \neq k$, 则按以下步骤调整权重.

步骤 1 降低不一致模型的权重, 即

$$\beta_i = \alpha_i\ \gamma. \tag{11}$$

步骤 2 计算权重降低导致的总权重减少量, 即

$$\varDelta = \alpha_i\ (1 - \gamma). \tag{12}$$

步骤 3 将减少的权重按比例分配给其他两个模型, 保证总权重和不变, 即

$$\beta_j = \alpha_j + \varDelta \frac{\alpha_j}{\alpha_j + \alpha_k}, \tag{13}$$

$$\beta_k = \alpha_k + \varDelta \frac{\alpha_k}{\alpha_j + \alpha_k}, \tag{14}$$

$$\beta_i + \beta_j + \beta_k = \alpha_i + \alpha_j + \alpha_k. \tag{15}$$

其中, γ为调节系数, 建议取值范围为$[0.95, 1)$, 以平缓降低可能受到干扰的模型的影响. 较高的γ值确保权重变化不会过于剧烈, 保留了基础模型的部分贡献, 同时仍能有效降低异常输出的影响.

如果三个基学习器的预测结果各不相同, 即 $f_i(x) \neq f_j(x) \neq f_k(x)$, 则保持原有权重不变, 有

$$\beta_t = \alpha_t,\ \ t \in \{1, 2, 3\}; \tag{16}$$

$$F(x) = \underset{y\in\{1,2,\cdots,C\}}{\operatorname{argmax}} \sum_{t=1}^{3} \beta_t I\left[f_t(x) = y\right]. \tag{17}$$

考虑到在未发生单粒子翻转时的权重系数β_1, β_2, β_3的变化会导致精度受影响, 设置基学习器$f_i(x)$及另外两个基学习器在最近m次预测中预测结果相同的次数为n_1和n_2. 若n_1或n_2大于 cm, 则将$f_i(x)$的权重系数由β_i恢复为初始的权重系数a_i. 这里m设为 100, c 为恢复阈值系数, 用于控制权重恢复的灵敏度, 其取值范围为 $0<c<1$.

3 实验

3.1 数据集

3.1.1 CIFAR-10 数据集

CIFAR-10[21] 由 10 类彩色图像组成, 涵盖飞机、汽车、鸟类等常见物体类别, 每类包含 6000 张 32×32 pixel 图像, 训练集与测试集分别分配 50000 张和 10000 张. 该数据集因类别多样且图像分辨率较低, 常被用于评估轻量化模型的泛化能力. 选用 CIFAR-10 验证自适应增强算法在中等复杂度任务中对单粒子翻转的容错效果, 并分析不同弱模型集成策略的精度损失与容错性能的权衡关系.

3.1.2 MNIST 数据集

MNIST[22] (Mixed National Institute of Standards and Technology) 是计算机视觉领域应用广泛的基准数据集, 包含 0～9 共 10 类手写数字的灰度图像. 每张图像分辨率为 28×28 pixel, 训练集包含 60000 张样本, 测试集包含 10000 张样本. 该数据集因数据规模适中、标注准确且分布均匀, 常被用于验证模型的基础分类性能. 通过 MNIST 测试不同 ResNet 变体在单粒子翻转干扰下的容错能力, 以评估算法在低复杂度场景下的鲁棒性.

3.1.3 EuroSAT 数据集

EuroSAT[11] 是基于 Sentinel-2 卫星多光谱遥感影像构建的高分辨率数据集, 包含 10 类土地覆盖类

型 (例如森林、农田、道路等), 每类提供 27000 张 64×64 pixel 图像, 涵盖 13 个光谱波段. 该数据集因贴近实际空间任务需求且包含复杂地表特征, 被广泛用于遥感图像分类研究. 采用 EuroSAT 测试所提方法在高价值空间场景下的实用性, 通过模拟单粒子翻转对模型参数的扰动, 验证集成模型在单粒子翻转干扰与保持分类精度方面的综合性能优势.

3.1.4 Galaxy10 DECals 数据集

Galaxy10 DECals[12] 是一个包含 17736 张彩色星系图像的星系分类数据集, 由 DESI[23] 提供的高质量图像与 Galaxy Zoo[24] 分类标签组成. 每张图像分辨率为 256×256 pixel, 被分为 10 个类别, 如扰动星系、合并星系、圆形光滑星系等. Galaxy10 DECals 数据集具有非平衡分布的特点, 其中, 各类样本数量差异较大, 这种不平衡性更真实地反映了天文观测中的实际情况, 为算法在空间任务中的应用提供了更贴近现实的验证环境. 能够反映模型在处理高维度、多特征、类别不平衡数据时的容错性能与算法适应性, 为星系分类等空间应用场景下的辐射加固神经网络提供参考.

3.2 实验设置

所有实验均在 Linux 环境下基于 PyTorch 框架实现, 硬件平台采用 AMD Ryzen 5500 CPU 和 NVIDIA RTX 4090D GPU. 实验设计了四种方法进行对比: 单一 ResNet110、传统三模冗余方法 (TMR, 使用三个相同的 ResNet110)、静态集成方法 (使用 ResNet20/32/44 但无动态权重调整) 及本文提出的 AB-FTM 方法. 评估指标主要包括基准准确率 (无错误注入时的分类准确率) 和容错后准确率 (单粒子翻转后的分类准确率).

为了确保实验结果的可靠性, 针对每个参数翻转比例进行 1000 次独立的随机错误注入, 取平均准确率作为实验结果数据. 实验的整体流程如图 3 所示.

为了验证 AB-FTM 方法的有效性, 开展以下实验.

(1) 无参数翻转实验. 测试在神经网络参数未发生单粒子翻转时各方法的准确率, 开展基准性能评估.

(2) 相同比例参数翻转实验. 由于模型的参数量与模型在 SRAM 中所占用的面积成正比, 参数量越多的模型在单位时间内发生单粒子翻转的次数越多. 参数翻转比例从 0.001‰按倍数递增至 0.128‰, 共设八个梯度. 表 2 列举了不同方法在各翻转比例下实际发生单粒子翻转的参数数量.

(3) 相同数量参数翻转实验. 在 CIFAR-10 数据

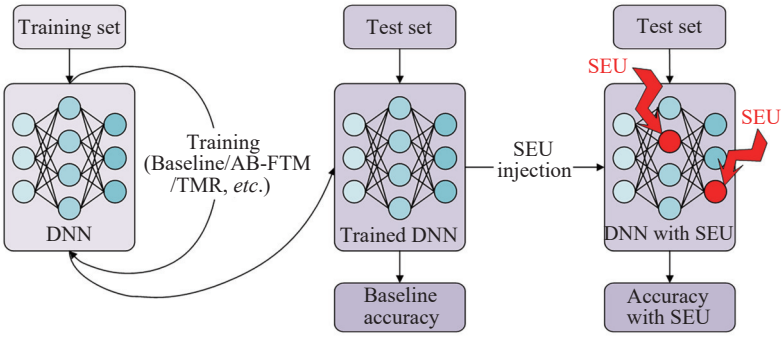


图 3 实验整体流程

Fig. 3 Overall flow chart of the experiment

表 2 不同错误注入比例下发生单粒子翻转的参数数量

Table 2 Number of parameters with SEU under different error injection rates

方法	0.001‰	0.002‰	0.004‰	0.008‰	0.016‰	0.032‰	0.064‰	0.128‰
单一模型ResNet110	2	3	7	14	27	54	109	218
三模冗余ResNet110 (TMR)	5	10	20	41	82	163	326	653
静态集成ResNet20/32/44 (SE)	1	3	6	11	22	45	90	179
AB-FTM	1	3	6	11	22	45	90	179

集上, 分别测试 1～200 个参数发生翻转时的情况, 验证在中等复杂度任务中不同方法在相同故障数量下的表现.

(4) 参数按不同比特位翻转影响实验. 在 EuroSAT 数据集上, 分别测试在参数的符号位、指数位、尾数位发生翻转时的情况, 参数翻转比例从 0.001‰按倍数递增至 0.128‰, 共设 8 个梯度. 评估不同比特位翻转对模型准确性的影响.

3.3　实验结果

3.3.1　无参数翻转实验

对各方法在无错误注入情况下的性能进行评估. 表 3 给出了四种方法在 4 个数据集上的分类准确率. 可见在无错误注入的条件下, AB-FTM 方法在分类准确率上明显提升, 在 4 个数据集上, 相比 ResNet110 模型准确率提升 0.03%～0.59%, 在正常工作条件下也能获得更好的性能收益.

3.3.2　相同比例参数翻转实验

图 4 为不同方法在各数据集上随着注错比例变化的性能变化趋势. 随着参数翻转比例的增加, 所有方法的准确率均呈现下降趋势, 但下降速率存在显著差异. 单一 ResNet110 模型对单粒子翻转最为敏感, 即使参数翻转比例仅为 0.001‰时, 平均准确率已

表 3　各方法在无错误注入情况下的分类准确率

Table 3　Classification accuracy of each method without error injection

方法	MNIST/(%)	CIFAR-10/(%)	EuroSAT/(%)	Galaxy10 DECals/(%)
单一模型ResNet110	99.65	93.68	97.37	84.33
三模冗余ResNet110 (TMR)	99.65	93.68	97.37	84.33
静态集成ResNet20/32/44 (SE)	99.68	93.83	97.56	84.49
AB-FTM	99.68	94.27	97.81	84.50

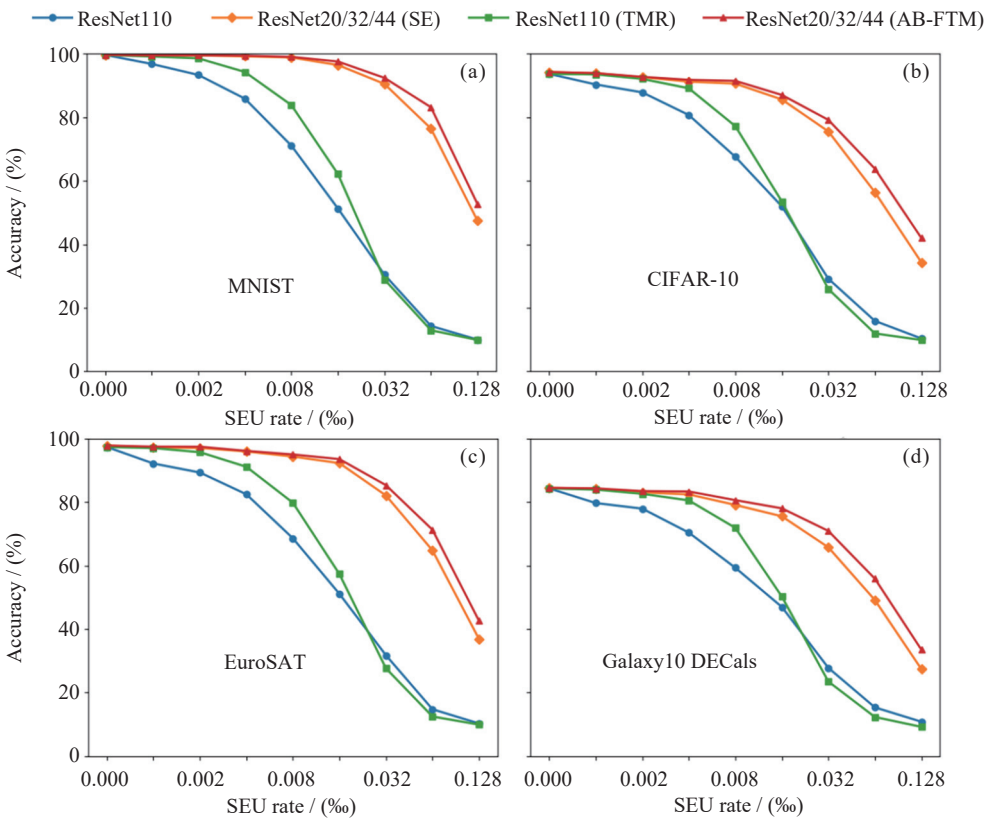


图 4　单粒子翻转容错性能评估 (SEU 比例)

Fig. 4　Single Event Upset (SEU) fault tolerance performance evaluation (SEU rate)

明显下降. 传统三模冗余 (TMR) 方法表现出较强的抗干扰能力, 但在四个数据集上均出现了随着翻转比例的增加, 三模冗余模型准确率反而不如单一 ResNet110 模型的情况, 且此时单一 ResNet110 模型的平均准确率为 30%～40%, 出现此种情况的原因是, 当三模冗余中两个模型失效时, 即使其中一个模型仍然有效, 但系统整体由于三取二的投票机制, 导致系统无法得到正确的结果. 而单一 ResNet110 模型仍然有 30%～40% 的概率能输出正确的结果.

值得注意的是, 在参数翻转比例较低时 (例如 0.001‰～0.004‰), 各方法的性能差异已经开始显现, 但差距相对较小. 随着翻转比例的增加, 性能差异逐渐扩大, 特别是在 0.032‰以上的高翻转比例区间, AB-FTM 方法的优势变得尤为显著. 在 CIFAR-10, MNIST, EuroSAT 和 Galaxy10 DECals 数据集上较 ResNet110 (TMR) 的准确率分别提升 53.25%, 63.49%, 57.67%, 47.43%.

进一步分析不同数据集上的实验结果发现, 基准准确率越低的数据集, 降幅越显著. 例如 CIFAR-10 的准确率降幅普遍大于 MNIST, 在参数翻转比例为 0.032‰时, AB-FTM 方法在 MNIST 上的准确率为 92.43%, 降幅仅为 7.25%; 而在 CIFAR-10 上的准确率为 79.19%, 降幅为 15.08%. 这一结果表明, 随着任务复杂度的增加, 单粒子翻转的影响更为显著, 而 AB-FTM 方法的容错优势也相应增大.

3.3.3 相同数量参数翻转实验

在 CIFAR-10 数据集上, 不同方法在相同单粒子翻转数量下的抗单粒子翻转容错表现如图 5 所示. 结果显示, 即使在相同的单粒子翻转数量时, AB-FTM 在绝大多数情况下的准确率仍然高于 ResNet110 基准模型, 并且容错能力也优于 ResNet110 (TMR). 当单粒子翻转次数大于 25 个, 容错优势更为明显. 实验结果表明, 在长时间运行中, 虽然单粒子翻转数量不断增加, AB-FTM 方法容错能力始终优于其他方法.

3.3.4 不同类型比特位翻转影响实验

图 6 为在 EuroSAT 数据集上, 四种方法分别在符号位 (S)、指数位 (E)、尾数位 (M) 发生单粒子翻转后的错误率. 实验表明, 指数位翻转对模型性能的影响最为严重, 这是因为指数位翻转会导致权重参数的数值量级发生巨大变化; 符号位翻转次之, 会导致权重参数的正负号反转, 改变神经元连接的抑制/激励特性; 尾数位翻转的影响相对较小, 主要影响权重参数的精度, 而不改变数量级.

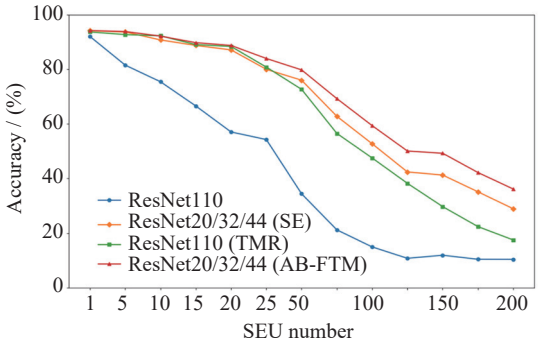


图 5 单粒子翻转容错性能评估 (SEU 数量)

Fig. 5 Single Event Upset (SEU) tolerance performance evaluation (SEU number)

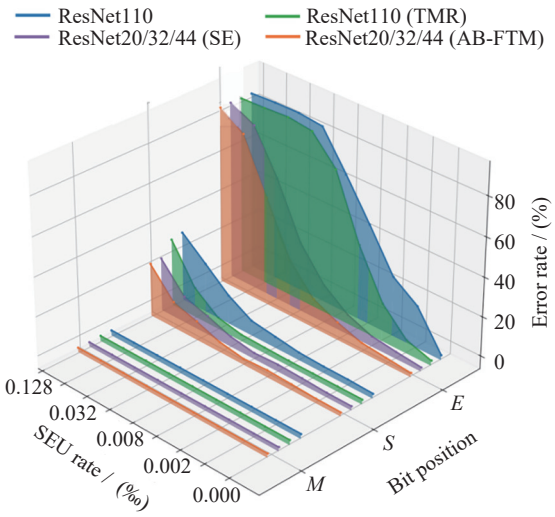


图 6 不同位发生单粒子翻转容错性能评估 (SEU 比例)

Fig. 6 Evaluation of SEU tolerance performance at different bit positions (SEU injection rate)

在同类型比特位发生翻转时, AB-FTM 方法容错能力仍优于其他方法.

4 分析与评估

从资源开销、动态权重调整机制重要性和容错效果三个维度, 对提出的 AB-FTM 方法与现有方案进行全面对比分析, 评估其在空间应用场景下的综合性能优势.

4.1 计算资源开销分析

四种方法的资源开销对比见表 4. 根据表 4 中各

模型在关键资源维度上的量化数据, 从系统资源利用效率角度分析, 本文提出的 AB-FTM 方法比传统方案具有更显著优势. 在参数规模方面, 该模型总参数量为 1.39×10^6, 较 ResNet110 的 1.70×10^6 降低了 18.2%, 这种轻量化设计源于采用多个较浅层次网络替代单一深层网络的策略. 更重要的是, 与 ResNet110 (TMR) 方法所需的 5.10×10^6 参数相比, 本方案仅需约 27.3% 的参数量, 这种数量级的降低对于受资源约束的航天器具有重要意义.

存储效率方面, AB-FTM 方法需 5.56 MByte 存储空间, 较 ResNet110 方法减少 18.2%, 较 ResNet110 (TMR) 方法降低 72.7%. 这种存储开销的显著降低不仅有助于减少存储器件的使用数量, 还能降低整体系统的单粒子翻转敏感区域, 从而在硬件层面上提升可靠性并降低了成本.

在计算复杂度方面, 基于 CIFAR-10 数据集的测试结果显示, ResNet110 单次前向传播需 253.15×10^6 FLOPs, 而 AB-FTM 得益于其异构并行架构, 总计算量降至 207.37×10^6 FLOPs. 相比之下, ResNet110 (TMR) 方案由于采用三模冗余设计, 其计算量升至 759.45×10^6 FLOPs. AB-FTM 方法计算效率的提升直接转化为更低的功耗和更快的推理速度, 对于航天器的实时任务处理具有重要价值.

4.2 动态权重调整机制重要性分析

为评估动态权重调整机制在 AB-FTM 方法中的重要性, 对比了带权重调整的 AB-FTM 与不带权重调整的静态集成 (Static Ensemble, SE) 方法在 CIFAR-10 数据集上参数翻转比例为 0.064‰时的性能, 如表 5 所示. 在 4 个数据集上, 准确率的下降率均降低 6% 左右, 静态集成依赖于简单多数投票, 当两个或以上模型发生错误时会完全失效; 而 AB-FTM 通过权重动态调整, 能够在部分型受损的情况下仍保持较高性能.

4.3 容错效果评估

图 7 和图 8 分别给出了 ResNet110 单模型和 AB-FTM 方法在 CIFAR-10 数据集上, 不同翻转比例条件下的各 1000 次独立实验的准确率分布, 每个紫点代表每次实验的准确率. 由图 7 和图 8 可见, 随着翻转比例的增加, ResNet110 单模型的准确率不仅整体下降, 更呈现出明显的双峰分布特征. 尤其在 0.016‰～0.064‰的翻转比例范围内, 准确率分布明显分化为两个集群. 一个集群维持在较高水平 (70%～90%), 另一个集群则急剧下降至较低水平 (10%～30%).

这引出了神经网络应对单粒子翻转时的一个关键问题——准确率的断崖式下降. 当特定区域发生单粒子翻转时, 模型性能并非渐进式退化, 而是呈现出灾难性失效的特性. 相比之下, AB-FTM 方法的准确率分布呈现出差异化显著的特征. 在各翻转比例下, 其准确率分布始终保持单峰特性, 且分布范围相对

表 4 四种方法的资源开销对比

Table 4 Comparison of resource cost among four methods

方法	模型	参数量($\times10^6$)	存储空间/MByte	浮点运算次数($\times10^8$)
单一模型ResNet110	1×ResNet110	1.70	6.78	2.5
三模冗余ResNet110 (TMR)	3×ResNet110	5.10	20.34	7.5
静态集成ResNet20/32/44 (SE)	ResNet20+32+44	1.39	5.56	2.1
AB-FTM	ResNet20+32+44	1.39	5.56	2.1

表 5 动态权重调整机制贡献分析

Table 5 Contribution analysis of dynamic weight adjustment mechanism

数据集	静态集成		AB-FTM		动态权重调整机制提高的准确率/(%)
	准确率/(%)	准确率下降率/(%)	准确率/(%)	准确率下降率/(%)	
MNIST	76.47	23.21	83.17	16.51	6.70
CIFAR-10	56.43	37.40	63.82	30.45	7.39
EuroSAT	64.80	32.76	71.30	26.51	6.50
Galaxy10 DECals	49.17	35.32	55.90	28.60	6.73

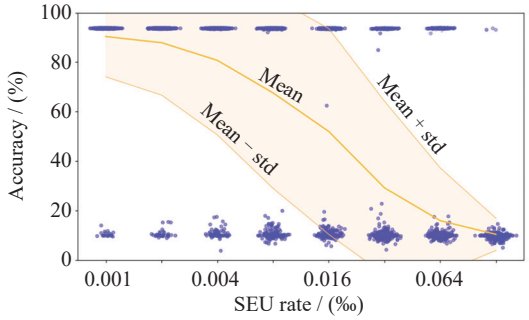


图 7 ResNet110 在 CIFAR-10 上不同翻转比例的准确率分布

Fig. 7 Accuracy distribution of ResNet110 at different SEU ratios on CIFAR-10

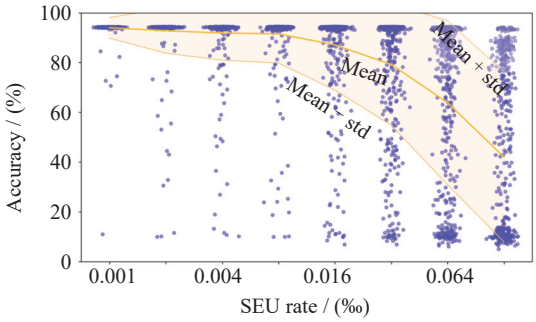


图 8 AB-FTM 在 CIFAR-10 上不同翻转比例的准确率分布

Fig. 8 Accuracy distribution of AB-FTM at different SEU ratios on CIFAR-10

分散, 这表明 AB-FTM 方法虽然也存在准确率断崖式下降的现象, 但性能退化曲线更加平缓, 一定程度上有效抑制了单粒子翻转导致的准确率断崖式下降现象.

5 结论

针对空间辐射环境下神经网络单粒子翻转的可靠性问题, 提出了一种基于自适应增强算法的轻量化容错方法 AB-FTM. 通过构建异构弱模型集成架构与动态权重调整机制, 在仅占基准模型 81.8% 参数量的条件下, 显著提升了星载智能系统在单粒子翻转干扰下的分类精度与鲁棒性, 有效降低了功耗和硬件成本.

实验表明, 该方法在 CIFAR-10, MNIST, EuroSAT 和 Galaxy10 DECals 数据集上均优于传统三模冗余方案, 在 CIFAR-10 数据集上, 当单粒子翻转比例为 0.008‰时, AB-FTM 保持 91.48% 的分类准确率, 而 ResNet110 (TMR) 仅为 77.23%,在遥感数据集 EuroSAT 和天文星系数据集 Galaxy10 DECals 上, 当 0.032‰比例的参数发生单粒子翻转时, 准确率分别提升 57.67% 和 47.43%. 这种高效的容错机制特别适合资源受限的航天器, 为星载智能系统的可靠性保障提供了兼顾轻量化、低功耗与高可靠性的新型解决方案.

未来将探索神经网络内部结构的加固方法, 实现系统架构层面与网络内部结构相结合的协同加固方法. 进一步优化异构模型架构的硬件适配性, 探索动态容错机制与辐射加固电路的协同设计, 并验证该方法在深空探测极端环境下的长期稳定性, 以推动空间智能系统自主化与抗辐射能力的协同发展, 为空间科学探测与航天器可靠性设计提供技术支撑.

利益冲突 所有作者均声明不存在利益冲突.

参考文献

[1] YU Jing, PENG Xiaodong, XIE Wenming, *et al.* Spatial non-cooperative target behavior intent recognition based on data generation and deep neural networks[J]. *Chinese Journal of Space Science*, 2024, **44**(6): 1134-1146 (余静, 彭晓东, 谢文明, 等. 基于数据生成和深度神经网络的空间非合作目标行为意图识别 [J]. 空间科学学报, 2024, **44**(6): 1134-1146)

[2] ZHANG Hui, LU Hao, YU Tianyi, *et al.* Teleoperation technology of Zhurong Mars Rover[J]. *Journal of Deep Space Exploration*, 2021, **8**(6): 582-591 (张辉, 卢皓, 于天一, 等. “祝融号”火星车遥操作技术 [J]. 深空探测学报 (中英文), 2021, **8**(6): 582-591)

[3] GAN W Q, ZHU C, DENG Y Y, *et al.* The *Advanced Space-based Solar Observatory* (ASO-S)[J]. *Solar Physics*, 2023, **298**(5): 68

[4] CHEN Xuelei, YAN Jingye, XU Yidong, *et al.* Discovering the sky at the longest wavelength mission—a pathfinder for exploring the cosmic dark ages[J]. *Chinese Journal of Space Science*, 2023, **43**(1): 43-59 (陈学雷, 阎敬业, 徐怡冬, 等. 宇宙黑暗时代探路者—鸿蒙计划 [J]. 空间科学学报, 2023, **43**(1): 43-59)

[5] ZHANG Xiaofang, LIU Songtao, WU Yaoping. Statistical analysis of space weather effects on satellite anomalies[J]. *Chinese Journal of Space Science*, 2015, **35**(4): 461-472 (张晓芳, 刘松涛, 吴耀平. 影响卫星故障的空间天气分析 [J]. 空间科学学报, 2015, **35**(4): 461-472)

[6] LI Yan, HU Yueming, ZENG Xiaoyang. Cost-effective TMR soft error tolerance technique for commercial aerospace: utilization of approximate computing[J]. *Jour-*

nal of Electronics & Information Technology, 2024, **46**(5): 1604-1612 (李炎, 胡岳鸣, 曾晓洋. 面向商业航天卫星成本效益的三模冗余软错误防护技术: 近似计算的实践 [J]. 电子与信息学报, 2024, **46**(5): 1604-1612)

[7] XIA Jun, ZHANG Jiawei, SUN Chen, *et al.* System design of on-orbit anti-SEU based on XC7V690T[J]. *Computer Measurement & Control*, 2024, **32**(3): 267-272,279 (夏俊, 张嘉伟, 孙晨, 等. 基于 XC7V690T 的在轨抗单粒子翻转系统设计 [J]. 计算机测量与控制, 2024, **32**(3): 267-272,279)

[8] RUOSPO A, SANCHEZ E, TRAIOLA M, *et al.* Investigating data representation for efficient and reliable convolutional neural networks[J]. *Microprocessors and Microsystems*, 2021, **86**: 104318

[9] IBRAHIM Y, WANG H B, LIU J Y, *et al.* Soft errors in DNN accelerators: a comprehensive review[J]. *Microelectronics Reliability*, 2020, **115**: 113969

[10] QIAN Huan, XIE Zhuochen, LIANG Xuwen. Single-event upsets fault tolerance of convolutional neural networks based on dropout algorithm[J]. *Journal of University of Chinese Academy of Sciences*, 2021, **38**(5): 712-719 (钱欢, 谢卓辰, 梁旭文. 基于 dropout 算法的卷积神经网络单粒子翻转容错方法 [J]. 中国科学院大学学报, 2021, **38**(5): 712-719)

[11] HELBER P, BISCHKE B, DENGEL A, *et al.* EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, **12**(7): 2217-2226

[12] LEUNG H W, BOVY J. Deep learning of multi-element abundances from high-resolution spectroscopic data[J]. *Monthly Notices of the Royal Astronomical Society*, 2019, **483**(3): 3255-3277

[13] SHANG Lin, LIU Xiaona, CAO Caixia, *et al.* Analysis of in-orbit single event upset of low-Earth-orbit internet satellite and protection measures[J]. *Spacecraft Environment Engineering*, 2021, **38**(5): 503-507 (尚琳, 刘晓娜, 曹彩霞, 等. 低轨互联网卫星在轨单粒子翻转分析及防护措施 [J]. 航天器环境工程, 2021, **38**(5): 503-507)

[14] REVIRIEGO P, MAESTRO J A, CERVANTES C. Reliability analysis of memories suffering multiple bit upsets[J]. *IEEE Transactions on Device and Materials Reliability*, 2007, **7**(4): 592-601

[15] BLACK J D, DODD P E, WARREN K M. Physics of multiple-node charge collection and impacts on single-event characterization and soft error rate prediction[J]. *IEEE Transactions on Nuclear Science*, 2013, **60**(3): 1836-1851

[16] BAEG S, WEN S J, WONG R. SRAM interleaving distance selection with a soft error failure model[J]. *IEEE Transactions on Nuclear Science*, 2009, **56**(4): 2111-2118

[17] RUOSPO A, GAVARINI G, BRAGAGLIA I, *et al.* Selective hardening of critical neurons in deep neural networks[C]//Proceedings of 2022 25th International Symposium on Design and Diagnostics of Electronic Circuits and Systems (DDECS). Prague: IEEE, 2022: 136-141

[18] CHEN Ziyang, ZHANG Meng, ZHANG Jiliang. A fault-tolerant design of spaceborne onboard neural network[J]. *Journal of Electronics & Information Technology*, 2023, **45**(9): 3234-3243 (陈子洋, 张萌, 张吉良. 一种星载在轨神经网络的容错设计方法 [J]. 电子与信息学报, 2023, **45**(9): 3234-3243)

[19] TAHERKHANI A, COSMA G, MCGINNITY T M. AdaBoost-CNN: an adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning[J]. *Neurocomputing*, 2020, **404**: 351-366

[20] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778

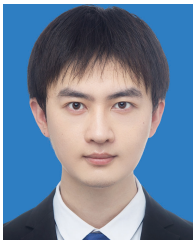
[21] KRIZHEVSKY A. Learning Multiple Layers of Features from Tiny Images[R]. Toronto: University of Toronto, 2009

[22] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, **86**(11): 2278-2324

[23] DEY A, SCHLEGEL D J, LANG D, *et al.* Overview of the DESI legacy imaging surveys[J]. *The Astronomical Journal*, 2019, **157**(5): 168

[24] WALMSLEY M, LINTOTT C, GÉRON T, *et al.* Galaxy Zoo DECaLS: detailed visual morphology measurements from volunteers and deep learning for 314 000 galaxies[J]. *Monthly Notices of the Royal Astronomical Society*, 2022, **509**(3): 3966-3988

作者简介



罗 熙 男, 2000 年 5 月出生于湖南省永州市. 现为中国科学院国家空间科学中心硕士研究生, 专业为计算机技术, 主要研究方向为星载人工智能.
E-mail: chellyluoxi@foxmail.com

周 晴（通信作者） 女, 1972 年 2 月出生于湖南省醴陵市, 现为中国科学院国家空间科学中心正研级高级工程师, 硕士生导师, 主要研究方向为空间信息与数据处理技术, 高可信嵌入式软件技术等.
E-mail: zhouqing@nssc.ac.cn