第59卷第4期 原 子 能 科 学 技 术 Vol. 59, No. 4
2025年4月 Atomic Energy Science and Technology Apr. 2025

基于交互式零知识协议的深度学习算法核查技术方案

吕彦锋，吕学升*，黄声慧，梁庆雷
（中国原子能科学研究院 放射化学研究所，北京 102413）

摘要：零知识协议的对敏感信息的保护及其可以接受的置信程度使其成为可能改变核军控核查的重要手段，但目前都是依靠模板进行统计学匹配核查，这会造成模板敏感信息泄露，且核查局限非常大，只能核查与模板同一设计型号的核弹头部件。为解决以上问题，本文提出了基于零知识协议的深度学习算法核查技术方案。利用蒙特卡罗方法建立深度学习样本库，经过深度神经网络学习后，算法可以在无需匹配模板的情况下按照零知识协议辨识出经过随机交互式核查的结果是否显示出待测核弹头为真实核弹头。本文所设计方案为核军控核查提供了一种新的不需要模板匹配的技术手段，有助于保障敏感信息不被泄露。本研究技术不仅可应用于核军控核查领域，也可同样应用于其他领域，具有重要意义和广阔的应用前景。

关键词：零知识协议；核保障；人工智能；深度学习；蒙特卡罗方法
中图分类号：TL99；E817；D815.2 **文献标志码：**A **文章编号：**1000-6931（2025）04-0946-11
doi: 10.7538/yzk.2024.youxian.0543

Verification Technical Scheme for Deep Learning Algorithm Based on Interactive Zero Knowledge Protocol

LYU Yanfeng, LYU Xuesheng*, HUANG Shenghui, LIANG Qinglei
(*Department of Radiochemistry*, *China Institute of Atomic Energy*, *Beijing* 102413, *China*)

Abstract: With the further development of nuclear disarmament and nuclear safeguards policies, future nuclear arms control agreements may limit all tactical nuclear warheads in the arsenal, as well as deployed and planned strategic nuclear warheads. Although there are corresponding military control verification methods now, the shielding methods for sensitive information are still not perfect. The existing methods include indirect verification of nuclear warhead vehicles, attribute measurement, and template matching. The zero knowledge protocol's protection of sensitive information and its acceptable level of confidence make it an important means of potentially changing nuclear arms control verification, belonging to a type of template matching method. However, because zero knowledge protocols currently rely on templates for statistical matching verification, relying on templates can lead to sensitive information leakage and verification limitations, only verifying nuclear weapons of the same design model as the template. Finding ways to replace templates has become an urgent problem to be solved. Innovative research on zero knowledge protocols based on deep learning algorithms was

收稿日期：2024-07-03；**修回日期：**2024-10-31
* **通信作者：**吕学升

proposed, which combines the discipline of artificial intelligence machine learning. The traditional method of template matching was replaced by the method of artificial intelligence deep learning. On the basis of relevant declaration data, a massive sample library was established through Monte Carlo method. After machine learning, artificial intelligence programs can identify whether the results of random interactive verification match their declaration situation according to the zero knowledge protocol without the need for matching templates. This new method innovatively combines artificial intelligence machine learning, zero knowledge protocols in cryptography, Monte Carlo simulation, and nuclear security verification. In order to demonstrate its protection against sensitive information, the Monte Carlo simulation of this study used multiple nuclear warhead design configurations, and even used nuclear material components of the same quality and abundance as the nuclear warhead. The results show that various design configurations, including nuclear material components with similar quality and abundance and sufficient to undergo chain reactions, can be identified as true warheads by deep learning programs, verifying their good sensitive information protection. In order to demonstrate the accuracy of its discrimination, the Monte Carlo simulation of this study used various missing abundance of nuclear warhead raw materials, the use of lead to replace fissionable materials, and even the use of low concentration nuclear material loose parts as non-nuclear warheads to reflect the situation of nuclear weapon nuclear material loss and replacement. The results show that various situations and even nuclear material components with similar quality and abundance and sufficient to undergo chain reactions can be identified as true warheads by deep learning programs, verifying its good discrimination accuracy. This research technology can not only be applied in the fields of nuclear arms control verification and nuclear safeguards supervision, but also in other fields, with important significance, practical application value, and broad application prospects.

Key words: zero knowledge protocol; nuclear safeguard; artificial intelligence; deep learning; Monte Carlo method

随着核裁军政策进一步发展，未来的核军备控制协议可能会限制武库中全部的战术核弹头以及已部署和将部署的战略核弹头[1]。国内外对核弹头进行核军控核查的方法研究有很多，主要分为模板匹配法（template approach）和属性测量法（attributes approach）两种方法[2]。模板匹配法是指将弹头与模板的某组特征进行比较，以确定核弹头真伪，而属性测量法是指直接获取某项性质特征的测量值来判别真伪。传统的模板法和属性法都需要信息屏障来屏蔽敏感信息[3]，但是对信息屏障方法的认证和证实目前仍然有许多顾虑与争论，虽然文献 [4-5] 对信息屏障的基本功能要求和设计基础已经进行了非常深入的讨论，但到目前为止对如何构建信息屏障系统仍未达成一致。

目前最新的研究采用密码学的零知识协议（zero knowledge protocol, ZKP）概念，将零知识证明与模板匹配方法相结合。文献 [6]提出了一种采用零知识协议的核军控核查方法，预设中子非电子学探测器为辐射图像的精确负片对待测弹头进行核查。文献 [7] 提出了一种不同的交互式零知识协议核军控核查方法，采用高密度聚乙烯墙对特定中子源诱核弹头裂变信号进行干扰保护敏感信息，然后与模板进行匹配的方法。

深度学习就是很深层的神经网络学习，是一种发源于连接主义学习的机器学习方法[8]。神经网络模型可以完成复杂的机器学习任务，这也意味着训练低效且容易过拟合，但随着云计算、大数据时代的到来，由于计算能力的快速发展，一直以来阻碍深度学习发展的训练低效缺点得到大幅缓解，算力的大幅提升降低了过拟合的风险[9]。本文提出一种基于深度学习算法和零知识协议的军控核查方法。利用蒙特卡罗方法建立机器学习样本库，经过训练的神经网络模型可以在无需匹配核武模板的情况下按照零知识协议辨识出探测

结果对应的核弹头真假及其有无核材料遗失。

1 零知识协议理论背景

零知识协议是证明者(P)可以向验证者(V)证明某事是真实的,同时除了这个特定声明是真实的事实之外不会透露任何信息的一种方法。如图 1 所示,一个简单的环形隧道,C 与 D 之间有一道门需要密码口令才能将其打开,P 想要向 V 证明自己可以打开这道门,但又不愿意向 V 泄露密码口令。故采取以下步骤:1)V 在协议开始时停留在位置 A;2)P 一直走到迷宫深处,随机选择位置 C 或 D;3)P 消失后,V 走到位置 B,然后命令 P 从某个出口返回位置 B;4)P 服从 V 的命令,必要时利用秘密口令打开 C 与 D 之间的门;5)P 和 V 重复以上过程 n 次。

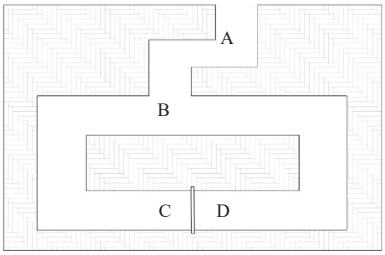


图 1 零知识协议环形隧道示意图

Fig. 1 Schematic diagram of circular tunnel zero knowledge protocol

协议中,如果 P 不知道秘密口令,就只能从来路返回 B,而不能走另一条路。此外,P 每次猜中 V 要求走哪一条路的概率为 1/2,因此每一轮中 P 能欺骗 V 的概率为 1/2,假定 n 取 16,则执行 16 轮后,P 能成功欺骗 V 的概率为 $1/2^{16}$=1/65 536。于是,如果 P 在这 16 轮中都能按 V 的要求返回,V 即可证明 P 确实知道秘密口令。此外还可看出,V 无法从上述证明过程中获取丝毫关于 P 的秘密口令的信息,所以这是一个交互式的零知识协议。

1.1 从山洞模型到核军控零知识协议

零知识协议在 20 世纪 80 年代由 Goldwasser 等[10] 提出,是现代密码学的一种重要理论,可以实现在不泄露敏感信息的同时证明某个特定申明是正确的。目前核军控核查技术的主要矛盾是敏感信息的保护性和核查的有效性之间的矛盾。敏感信息的保护对于核军控核查非常重要,且往往与核查的有效性不可兼得。如在山洞模型中,不能得知钥匙的存在,也不能得知口令的类型或其他信息,但与此同时还要确认被核查方必须具有打开山洞门的口令。在核弹头部件的军控核查中,必须注意的是不能被核查方获取武器设计的敏感信息,同时也要保证核查的有效性。目前解决这个问题的方法主要是由英国、美国和俄国的国家实验室开发的信息屏障法[4-5],信息屏障是由复杂的自动化系统组成的。通过这些系统处理在核查时测量的高度机密信息,但最终只以是或否的方式显示核查结果。这样的系统存在很多缺点,且系统既复杂,又要求核查与被核查的双方都相信对方没有隐藏的后门,过度依赖电子学系统使得这种方法存在通过电子学和信息技术的手段设置后门以获取敏感信息的可能。

信息屏障法对敏感信息的保护并不到位。2014 年 Glaser 等[6] 利用气泡探测器的特性,即每个气泡探测器都有最大的气泡探测数量,通过外置的中子源诱发裂变核弹头内易裂变物质,通过气泡探测器阵列测量出模板弹头即黄金弹头的中子通量与位置的关系,反向制作底片,即用每个探测器最大的气泡探测数量减去黄金弹头实际探测到的气泡数量组成底片,当底片制作完成后,核查方可以使用此底片相同条件下对待测弹头进行测量,并通过统计学去验证,是否与模板弹头一致。由于气泡探测器是非电子学探测器,拥有卓越的敏感信息保护性能,利用零知识协议取代了信息屏障的方法,并且和模板匹配法相结合,既保证了敏感信息的保护同时又保证了测量的有效性。

但 Glaser 等[6] 和言杰等[7] 的方法具有很大的局限性,即需要模板。这带来了以下问题:核查方既无法确定模板弹头就是真实有效的核弹头,也无法保证在核查过程中模板不会泄露相关的敏感信息。因此取消模板的存在成为新的改进方向。

1.2 机器学习训练库的建立与模型可行性分析

为实现取代使用模板进行匹配对比验证核弹头真伪的方法,核查方可通过测量裂变材料中 ^{235}U 的丰度与质量两个参数实现对核弹头的核查,并在测量丰度和质量的同时避免敏感信息的泄露。

假设核查方与被核查方设置的随机编码孔数据库中共有 n 个随机编码,对于同一种核弹头,经过 n 个随机编码,高密度聚乙烯板会出现 m 种

16×16 探测器阵列中子气泡结果, 每个结果记为 $X_i=\{x_1, x_2, \cdots, x_{256}\}$。只要被核查方用任意的随机编码孔进行诱发裂变实验, 结果是某个 X_i, 则在多次实验后因为与山洞模型相同的原理, 实现核查目标。所以只要确定 X_i 的集合 $\boldsymbol{X}$ 即可实现对核弹头真伪的核查, 对于这样的大数据, 由于数量太大, 采用枚举法基本不可能实现, 所以只能通过蒙特卡罗模拟仿真与机器学习相结合, 通过蒙特卡罗创建一系列 X_i, 再通过机器学习的方法实现压缩解 256 维的探测器阵列气泡数解空间以寻找集合 $\boldsymbol{X}$ 的边界。

本文采用 MCNP5 程序作为蒙特卡罗仿真计算的程序, 通过 matlab 程序语言对中子源、高密度聚乙烯(HDPE)板的随机编码孔图形、真伪核弹头、气泡探测器、空间位置构建数据库, 并生成可供 MCNP5 程序读取的.i 文件。经过 MCNP 计算, 模型构建的中子源打出一定数量的中子后, 统计穿过气泡探测器阵列并超过一定能量的中子数量, 再通过 matlab 程序将 MCNP 得到的结果读取到 excel 文件中。每完成 1 次这样的操作就为之后的机器学习提供 1 次样本, 多次重复这样的操作, 即可得到机器学习所需要的样本库。使用样本库训练神经网络即可得到最终的用于核查的深度学习程序。

2　核弹头蒙特卡罗仿真建模

本文机器学习所需要的核弹头部件样本有多种, 所以需要构建多种类型的核弹头部件模型, 其中一个公开的核弹头简化模型如图 2 所示。该简化模型是由马里兰大学的 Fetter 等[11] 根据核弹头结构的一般特点进行简化后所得, 模型中, 由一系列同心球壳组成的裂变材料为武器级铀, 中子反射层为铍, 惰层为贫化铀, 外壳为铝, 外壳内填充有机炸药。该模型的参数如表 1 所列。

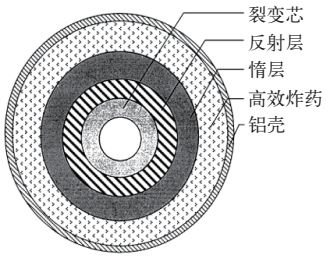


图 2　核弹头简化模型[11]

Fig. 2　Simplified model of nuclear warhead[11]

表 1　核弹头简化模型的参数

Table 1　Parameters of nuclear warhead simplified model

结构名称	组成	质量/kg	密度/(g/cm^3)	同心球壳内外径/cm
空腔	真空			外径 5.8
裂变芯	96.3% ^{235}U+1% ^{234}U+2.5% ^{238}U+0.2% 氧	12.0	15.6	内径 5.8, 外径 7.0
反射层	铍	3.0	1.8	内径 7.0, 外径 9.0
惰层	0.3% ^{235}U+99.7% ^{238}U	79.0	19.7	内径 9.0, 外径 12.0
炸药层	2.44% 氢+14.63% 碳+39.03% 氮+43.9% 氧	71.0	2.8	内径 12.0, 外径 22.0
铝壳	铝	17.0	2.7	内径 22.0, 外径 23.0

同时为了满足机器学习的需要, 在核弹头样本建模数据库中同样添加伪核弹头部件的数据库作为神经网络训练的伪核弹头识别输入, 与真核弹头部件数据相比, 改变了铀的丰度与质量参数, 使伪核弹头部件不再能发生链式反应。在具体的实际应用场景中表现为武器部件中^{235}U 丰度没有降低, 武器级的核材料没有丢失或被盗, 以及没有用其他丰度的铀部件来伪造核弹头部件达到隐瞒数量的目的。

2.1　中子源蒙特卡罗仿真建模

本研究需要选择一个释放中子能量在 1 MeV 以下的源, 以最大限度地提高对^{235}U 的探测效率, 同时避免^{238}U 干扰。为了确保实验的可行性和数据的可靠性, 选择实验室现有的有稳定中子产量和可预测能谱分布的 Am-Li 中子源作为蒙特卡罗仿真模型。

在 MCNP 程序中, 根据 Am-Li 中子源的特性设置了中子发射的参数, 包括中子的能量分布、空间分布和时间分布。通过仿真模拟了中子与核弹头材料相互作用的全过程, 包括中子的传播、裂变反应的发生以及裂变中子的探测。

2.2　气泡探测器蒙特卡罗仿真建模

本文所用中子气泡探测器如图 3[12] 所示, 该探测器由固化体、过热液滴、添加剂和施压装置

组成。因为其可以重复使用[13]和非电子学的特性[14]，不会轻易被电子学手段窃取敏感信息，是核军控核查零知识协议的首选探测器。气泡探测器除了非电子学可反复使用的优点外，还有阈能低、灵敏度高、直接可视测量、对伽马射线不灵敏的优点[15]。

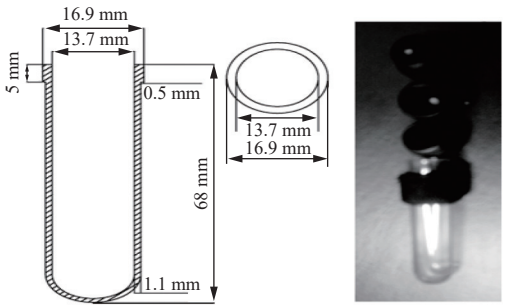


图 3 气泡探测器设计图和实物照片[12]

Fig. 3 Design drawing and physical image of bubble detector[12]

2.3 蒙特卡罗仿真建模布局

采用 MCNP5 软件可视化 Vised 软件绘制的蒙特卡罗布局可视化图形如图 4 所示，在 z=0 的空间位置剖视图空间结构布局如下：中子源位于原点(0, 0)处，核弹头部件质心位于中子源 y 轴正方向 30 cm 处，高密度聚乙烯(HDPE)板质心位于中子源 y 轴正方向 65 cm 处，HDPE 板厚度 10 cm，有 20×20 共 400 个随机编码孔，编码孔有 60% 的孔隙被 HDPE 棒填充，探测器阵列有 32×32 共 1 024 个中子气泡探测器，探测器质心位于中子源 y 轴正方向 85 cm 处。

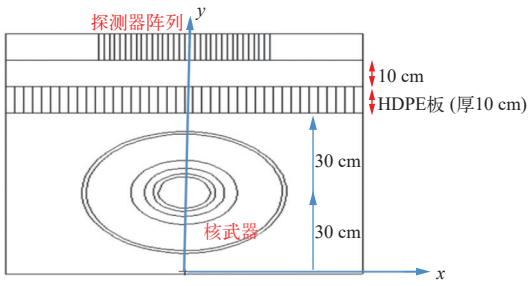


图 4 Vised 软件绘制的蒙特卡罗仿真建模可视化图形

Fig. 4 Visual graphics for Monte Carlo simulation modeling drawn by Vised

2.4 模型可行性分析

在核物理领域，^{235}U 和^{238}U 是两种主要的铀同位素。由于它们的对中子的核裂变反应截面会随中子能量的变化而变化，且两者的中子诱发裂变反应截面完全不同，所以它们对核军控核查具有不同的意义。^{235}U 和^{238}U 的反应截面与中子能量的关系如图 5 所示(图中数据来源于文献 [16])。^{235}U 在热中子区吸收截面较大，能引发裂变反应，而^{238}U 则需要更高能量的中子才能发生足量裂变。理想的核查中子源应该能提供足够的中子通量，同时其能量分布应足以诱发^{235}U 的裂变，但不会使^{238}U 裂变规模影响到探测结果。这样，通过测量裂变产生的中子，可以特异性地检测^{235}U 的存在，从而为核军控核查提供关键信息。

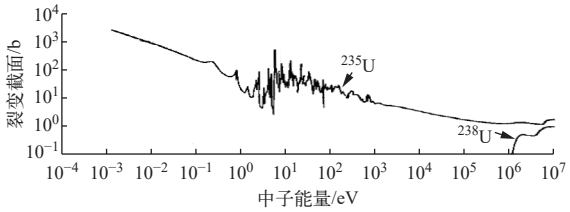


图 5 ^{235}U 和^{238}U 的反应截面与中子能量的关系(数据来源于文献 [16])

Fig. 5 Relationship between reaction cross section and neutron energy of ^{235}U and ^{238}U (data from reference [16])

MCNP 软件可以实现蒙特卡罗模拟仿真，可以对仿真中诱发裂变中子信号进行统计，通过 ELPT: n 0.5 的编程语句限制 500 keV 以下能量的中子不能通过气泡探测器层(气泡探测器对中子能量存在探测下限，本文定为 500 keV)，屏蔽源中子的干扰。虽然 Vised 软件可以直接识别并读取 MCNP 程序的建模并提供模型的可视化，但其图形界面并不直观，因此进一步用 SolidWorks 软件重新绘制本文所需搭建的仿真模型，结果如图 6 所示。由图 6 可见，由 SolidWorks 绘制的三维视

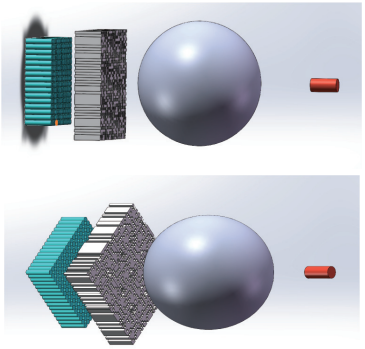


图 6 SolidWorks 软件绘制的蒙特卡罗仿真建模可视化图形

Fig. 6 Visual graphics for Monte Carlo simulation modeling drawn by SolidWorks

图能更清晰地展示核弹头模型的复杂结构。通过三维视图可以清晰地看到 HDPE 板随机编码孔的构形、中子气泡探测器的排布和整体模型的布局。

2.5　HDPE 板随机码孔的敏感信息屏蔽性

HDPE 板的随机编码孔在核军控核查中的作用是使中子通过时产生偏折，从而有效屏蔽中子中携带的敏感信息。无 HDPE 板干扰下，核弹头诱发裂变中子通量随空间位置的分布如图 7 所示，可以明显看出待测弹头呈球状。可见，无 HDPE 板干扰时，探测器阵列会将核弹头的形态设计信息泄露，甚至如果再有中子源的相关信息，核查方可以反推出核弹头的设计构型。有 HDPE 板干扰下，核弹头诱发裂变中子通量随空间位置的分布如图 8 所示。由图 8 可见，HDPE 板干扰了诱发裂变的中子，改变了中子通量的空间分布，使其不再与弹头外形相关，从而在不牺牲核查有效性的前提下，保护了核弹头设计等关键信息不泄露。

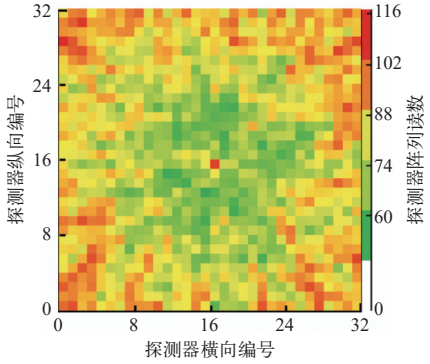


图 7　无 HDPE 板干扰下的探测器阵列读数

Fig. 7　Detector array reading without HDPE board interference

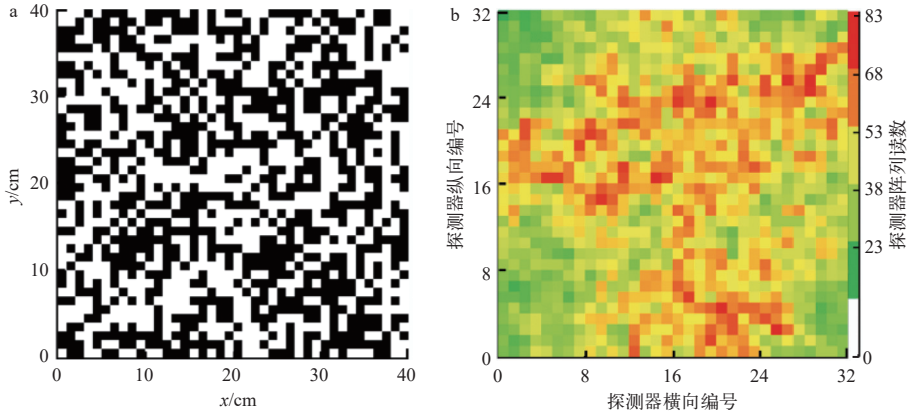


图 8　第 1 种随机编码孔下 HDPE 板干扰时的探测器阵列读数

Fig. 8　Detector array reading with HDPE board interference under the first type of random coding hole

图 8a 和图 9a 为诱发裂变的两种随机编码孔（黑色代表此格装载了 HDPE 棒，白色代表无 HEPE 棒，HDPE 板厚度为 10 cm，尺寸为 20 cm×20 cm，有 20×20 共 400 个随机编码孔，编码孔有 60% 的孔隙被 HDPE 棒填充，每个孔隙大小为 1 cm×1 cm）HDPE 板的诱发裂变信号受到干扰后的探测结果，诱发裂变中子穿过 HDPE 板受到干扰，改变了原有的通量随空间分布（图 7），由气泡探测器阵列记录下的读数，由于 HDPE 板的编码孔不同，图 8b 和图 9b 的探测器阵列的读数结果不同。可以看出不同随机编码孔的 HDPE 板干扰会使相同的诱发裂变中子信号产生偏折，从而使得探测器阵列探测出不同的中子通量空间分布。由于随机编码孔的设计不被核查方所知，受到其干扰后的诱发裂变信号也不再能被反推出敏感的核弹头设计信息。

由于不同的随机编码孔对诱发裂变信号的干扰效果使得中子通量随空间位置存在差异，这些差异为生成多样化的深度学习样本提供了可能，从而增强了模型的泛化能力和鲁棒性。通过使用这些样本对神经网络进行训练，即可构建出一个既能有效识别核材料，又能屏蔽敏感信息的核查模型。

2.6　真伪核弹头对蒙特卡罗仿真结果的影响

本研究的算法核心在于通过分析诱发裂变信号来鉴别核弹头的真伪，该信号源自核弹头中的易裂变材料。当易裂变材料的丰度和质量相同时，相同的中子源使易裂变材料产生的诱发裂变

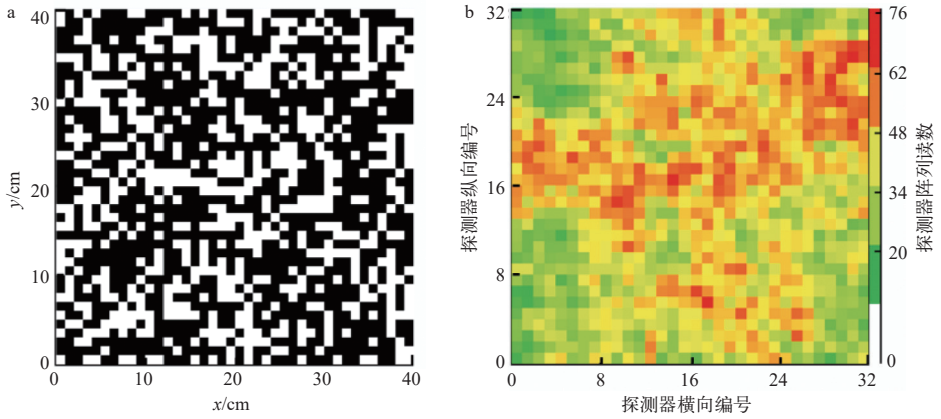


图 9 第 2 种随机编码孔下 HDPE 板干扰时的探测器阵列读数

Fig. 9 Detector array reading with HDPE board interference under the second type of random coding hole

信号也大致相同。即便在引入不同编码孔的 HDPE 板进行干扰的情况下，只要这些 HDPE 板上随机编码孔的占空比保持一致，裂变信号受到的干扰强度也将保持一致的特性。基于此致性，可以利用深度学习神经网络训练学习经过 HDPE 板干扰后的诱发裂变信号特征。

在深度学习模型的训练过程中，采用了大量经过 HDPE 板干扰的诱发裂变信号样本。这些样本的多样性和复杂性使得模型能够学习到不同干扰条件下的信号特征，从而增强了模型的泛化能力。用于深度学习的样本库数据如图 10、11 所示。图 10a、图 11a 是仿真实验的中子源诱发待测真弹头易裂变物质裂变，诱发裂变中子穿过如图 10a、图 11a 所示的编码孔（与图 8、9 尺寸一致）HDPE 板的诱发裂变信号受到干扰后的探测结果。图 10b 和图 11b 是仿真实验的中子源诱发待测伪核弹头易裂变物质裂变，诱发裂变中子穿过如图 10c 和图 11c 所示的编码孔（与图 8、9 尺寸一致）HDPE 板的诱发裂变信号受到干扰后的探测结果。其中图 10b 所示伪核弹头是将表 1 和图 2 所示的弹头裂变层^{235}U 丰度降低为 49%，而图 11b 所示的伪核弹头是移除了核弹头裂变层 4 kg 丰度为

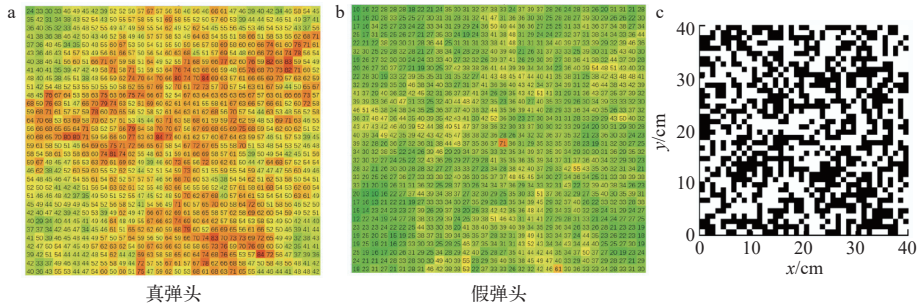


图 10 第 1 种 HDPE 干扰下真伪核弹头诱发裂变的探测器阵列读数

Fig. 10 Reading of detector array induced fission by real and fake nuclear warheads under the first HDPE board interference

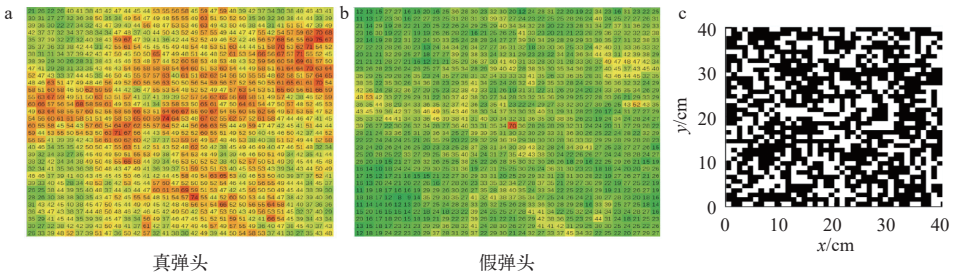


图 11 第 2 种 HDPE 板干扰下真伪核弹头诱发裂变的探测器阵列读数

Fig. 11 Reading of detector array induced fission by real and fake nuclear warheads under the second HDPE board interference

96.3%的^{235}U。

真弹头和伪核弹头在 HDPE 板干扰下探测到的中子通量空间分布有明显差异。通过图 10、11 可明显看出，易裂变物质的丰度和质量发生改变的核弹头的诱发裂变中子通量明显减少，机器学习可以通过其中的各种特征来分辨核弹头是否存在核材料转移盗窃的问题，并且还能判断是否被假的核弹头替代。通过训练，深度学习模型能够识别和提取干扰信号中的关键特征，即便存在干扰的情况，也能有效判断核弹头的真伪。这种方法不仅提高了核军控核查的有效性，同时也保护了核弹头设计的敏感信息，避免核查过程中敏感信息的泄露。

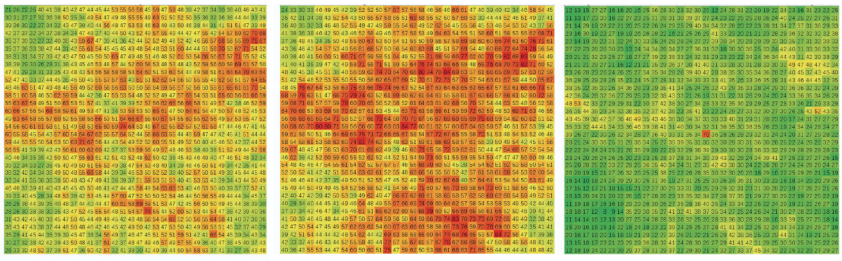
2.7　不同核弹头设计对算法有效性的影响

为全面评估深度学习模型在核军控核查中的性能，特别关注了模型的鲁棒性及其对敏感信息保护的能力。为此，在深度学习模型的验证集中特意引入了散料样本和其他核弹头设计样本，以此来模拟更复杂和具有挑战性的情况。由于核弹头和散料型块设计样本无法公开，本文没有给出具体设计数据。而散料样本由于物理形态和核弹头设计截然不同，但可以令散料的易裂变材料丰度和质量与核弹头样本相同来进行算法检验。本算法的目标是为了保护设计和结构的敏感信息，检验丰度与质量是否满足可以发生链式反应这一核武器基本条件进行核查的。如果散料样本也能准确判别，则可以认为算法可以保护敏感信息，同时可有效进行核军控核查。

散料样本，尽管在物理形态上与核弹头设计相去甚远，但它们所含的易裂变材料在丰度和质量上与核弹头中的材料相同。在核军控核查的背景下，这些散料样本应识别为真核弹头样本。通过在验证集中加入这些散料样本，能够测试模型在面对与核弹头设计差异较大的情况下，是否仍能保持有效的判断能力。

图 12 为真散料样本与真核弹头样本的丰度和质量相同情况下，诱发裂变信号经过相同 HDPE 板干扰后仿真模拟的中子通量空间位置分布。图 12 同时示出裂变层^{235}U 丰度降低为 49% 的伪核弹头的仿真结果。可以看出，真散料样本与核弹头中裂变层丰度和质量相同情况下，其仿真结果稍有不同，但仍具有共性，与伪核弹头数据截然不同。

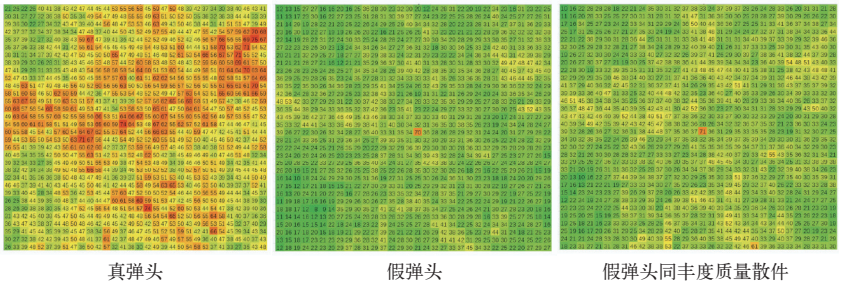
图 13 为假散料样本与将裂变层^{235}U 丰度降低为 49% 的伪核弹头中裂变层丰度和质量相同情况下，诱发裂变信号经过相同 HDPE 板干扰后的蒙特卡罗仿真模拟探测器中子读数分布的对



真弹头　同丰度质量散件　假弹头

图 12　HDPE 板干扰下等同真核弹头散料的探测器阵列读数对比

Fig. 12　Comparison of detector array readings for bulk materials equivalent to real nuclear warheads under HDPE board interference



真弹头　假弹头　假弹头同丰度质量散件

图 13　HDPE 板干扰下等同伪核弹头散料的探测器阵列读数对比

Fig. 13　Comparison of detector array readings for bulk materials equivalent to fake nuclear warheads under HDPE board interference

比。图 13 同时示出裂变层^{235}U 丰度为 96% 的核弹头仿真结果。可以看出，假散料样本与伪核弹头中裂变层丰度和质量相同情况下，其仿真结果稍有不同，但仍具有共性，与真弹头数据截然不同。

3 基于深度学习算法的交互式零知识协议

在完成整个蒙特卡罗模拟仿真后，将每次仿真出的中子通量空间位置分布与是否存在足量的易裂变材料一一对应，即可得到深度学习的训练样本库，如图 14 所示，将左图所示的仿真装置气泡探测器阵列探测的中子通量空间位置分布作为右图神经网络的输入，是否存在足量的易裂变材料作为神经网络的输出（真弹头输出为 1，伪核弹头输出为 0）。这里的深度学习算法是 BP 算法，也就是从输出的误差向前传导使神经网络的阈值和权重向误差更小的方向迭代的算法。通过将数据库中的探测器阵列读数与核弹头的真假作为输入数据和输出数据，每次输入数据都会正向通过神经网络，然后得到 0 或者 1 的结果，把结果和输出作差得到误差，再由误差梯度下降方法更新神经网络的权重和阈值，经过多轮多次训练得到正确率高的神经网络。将上述原理转化为整个深度学习算法的流程：1）初始化，随机初始化网络参数，如权重和偏置；2）数据预处理，对蒙特卡罗仿真得到的数据进行归一化处理；3）前向传播，将输入数据通过神经网络进行前向传播，得到预测结果；4）计算损失，使用损失函数计算预测结果和真实值之间的差异；5）反向传播，根据损失函数计算的梯度，使用反向传播算法对网络参数进行更新；6）迭代优化，重复步骤 3～5，直到网络在验证集上的性能不再提升或达到预设的迭代次数；7）模型评估，使用独立的测试集评估模型的性能；8）超参数调优，根据模型评估结果调整学习率、网络结构等超参数。

将通过蒙特卡罗仿真得到中子气泡探测器的读数作为神经网络模型的输入，核弹头真假作为输出，使用 BP 算法对样本库的训练库进行训练。根据梯度下降原理，随着循环和参数的迭代，深度学习网络各神经元的阈值和权重将会迭代，直到得到本研究所需的可以判别经过 HDPE 板干扰诱发裂变信号是满足链式反应条件的足量丰度和质量的真样本还是不满足链式反应条件的丰度和质量的假样本。

在蒙特卡罗方法的基础上，首先对外置中子源照射下的 HDPE 板进行建模，模拟中子与物质相互作用的过程。裂变信号在通过 HDPE 板时会受到干扰，随后被中子探测器阵列捕获并记录中子数。得到的中子数矩阵结果作为训练样例的输入（x），而待测物体（正反例）是否为真待测物体则标记为 1 和 0，作为训练样例的输出（y）。使用这些数据来训练神经网络，通过代入网络并进行参数调整，循环直至达到指定次数或满足预定的错误率标准。训练完成后，使用测试库对模型进行检验，如果错误率符合要求，则训练成功并输出神经网络模型；若训练未能达到预期效果，则需要重新调整神经网络的起始参数，并重新开始训练过程。算法流程图如图 15 所示。

同时对整体步骤进行设计：在核军控核查的深度学习应用流程中，首先，由被核查方监督并申报其核材料的数据；随后，核查方与被核查方共同开展工作，利用蒙特卡罗方法对多项建模任务进行仿真，产生的大量数据用于训练深度学习程

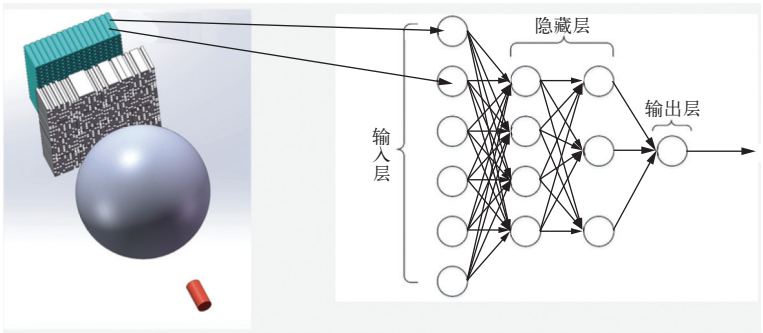


图 14 深度学习神经网络的输入原理图

Fig. 14 Schematic of deep learning neural network input

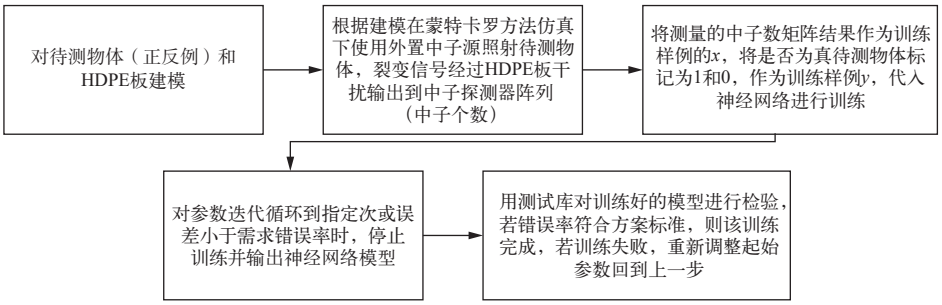


图 15　深度学习算法流程图

Fig. 15　Flow chart of deep learning algorithm

序。在实验准备阶段，双方联合准备 HDPE 棒和气泡探测器，并确保这些器材与深度学习程序中使用的编号一一对应。核查方随机选择一个编号进行抽查，确认 HDPE 棒和气泡探测器正确封装并准备就绪。实验过程中，被核查方根据自己设计的构造，搭建编码墙并按顺序排列探测器，核查方则启动中子源，让中子穿过 HDPE 板并诱发裂变信号。这些信号经过 HDPE 板的干扰后，被探测器捕获并转换为中子计数信息。最终，核查方使用深度学习程序，将这些中子计数信息作为输入，进行真假判断，从而完成对核材料的核查。核查技术方案流程图如图 16 所示。

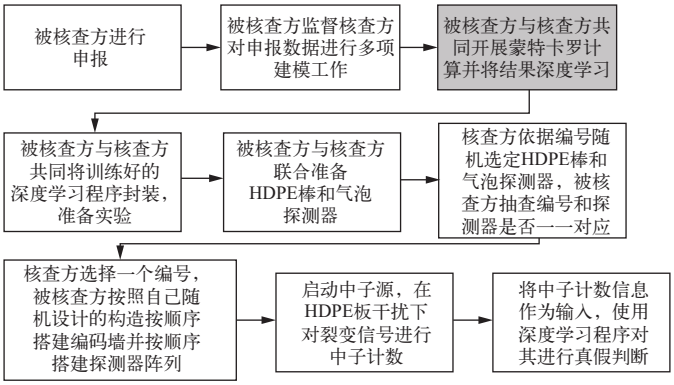


图 16　基于交互式零知识协议的深度学习算法核查技术方案流程图

Fig. 16　Flow chart of deep learning algorithm verification technology based on interactive zero knowledge protocol

4　结论

本研究旨在开发一种基于深度学习的交互式零知识协议核军控核查方法，以提高核查的准确性和效率，同时确保敏感信息的安全。通过深入分析和一系列仿真实验，可得到如下结论。

1）利用蒙特卡罗仿真技术，成功构建了一个大规模、带有标签的样本库。由该样本库模拟生成了核弹头及其散料在不同条件下诱发裂变中子的通量空间分布，并且与是否有足量易裂变材料这个标签一一对应，为深度学习模型的训练提供了丰富的数据资源。

2）结合密码学零知识协议，由蒙特卡罗仿真构建样本库训练的深度学习算法取代了原模板“黄金弹头”，保护了核武器设计的细节信息不被模板弹头泄露。

3）通过在样本库中加入散料样本和多种核武器设计构形的蒙特卡罗仿真结果，以模拟更具挑战性的核弹头设计未知的场景，提升了训练出的神经网络的泛化能力。

参考文献：

[1] LYU Yanfeng, LYU Xuesheng. Discussion on nuclear arms control based on the international situation in 2022[C]//Proceedings of the 23rd Pacific Basin Nuclear Conference. [S. l.]: [s. n.], 2023.

[2] YAN J, GLASER A. Nuclear warhead verification: A review of attribute and template systems[J]. Science & Global Security, 2015, 23(3): 157-170.

[3] National Academy of Sciences. Monitoring nuclear weapons and nuclear-explosive materials: An assessment of methods and capabilities[M]. Washington D. C.: National Academies Press, 2005.

[4] SPEARS D. Technology R&D for arms control[R]. Washington D. C.: US Department of Energy, Office of Nonproliferation Research and Engineering, 2001.

[5] ANDERSON B. Verification of nuclear weapon dismantlement: Peer review of the UK MoD programme[M]. London: British Pugwash Group, 2012.

[6] GLASER A, BARAK B, GOLDSTON R J. A zero-knowledge protocol for nuclear warhead verification[J]. Nature, 2014, 510: 497-502.

[7] 言杰. 一种随机和交互式的核弹头认证核查协议[C]//中国核学会 2017 年学术年会论文集. 北京: 中国原子能出版社, 2017.

[8] 胡越, 罗东阳, 花奎, 等. 关于深度学习的综述与讨论[J]. 智能系统学报, 2019, 14(1): 1-19.
HU Yue, LUO Dongyang, HUA Kui, et al. A review and discussion on deep learning[J]. Journal of Intelligent Systems, 2019, 14(1): 1-19(in Chinese).

[9] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

[10] GOLDWASSER S, MICALI S, RACKOFF C. The knowledge complexity of interactive proof systems[J]. SIAM Journal on Computing, 1989, 18(1): 186-208.

[11] FETTER S, FROLOV V A, MILLER M, et al. Detecting nuclear warheads[J]. Science & Global Security, 1990, 1(3/4): 225-253.

[12] 张贵英, 吕鹏, 倪邦发, 等. 中子气泡探测器性能改进及初步检验[J]. 原子能科学技术, 2010, 44(10): 1266-1269.
ZHANG Guiying, LV Peng, NI Bangfa, et al. Performance improvement and preliminary testing of neutron bubble detectors[J]. Atomic Energy Science and Technology, 2010, 44(10): 1266-1269(in Chinese).

[13] ING H, MORTIMER A. Space radiation dosimetry using bubble detectors[J]. Advances in Space Research, 1994, 14(10): 73-76.

[14] ING H, NOULTY R A, MCLEAN T D. Bubble detectors—A maturing technology[J]. Radiation Measurements, 1997, 27(1): 1-11.

[15] 张贵英, 倪邦发, 李丽, 等. 利用核径迹技术研制个人中子剂量气泡探测器[J]. 核技术, 2005, 28(9): 663-666.
ZHANG Guiying, NI Bangfa, LI Li, et al. Development of a personal neutron dose bubble detector using nuclear track technology[J]. Journal of Nuclear Technology, 2005, 28(9): 663-666(in Chinese).

[16] MUGHABGHAB S F, National Nuclear Data Center, Brookhaven National Laboratory. Neutron cross sections, Volume 1: Neutron resonance parameters and thermal cross sections, Part B: Z=61-100[M/OL]. (2025-01-18). https://doi.org/10.1016/C2009-0-21857-5.