（原副标题“——以泛北部湾“核电算网”沙盒为例”已删除，内容并入第4节首段）

李德豪[1,2,3*]，刘开杰[2]，胡豪[4]，何建宗[5]，吴国宝[1**]
1. 香港浸会大学数学系，香港 999077；
2. 广东技术师范大学光电工程学院粤港澳ESG与新质生产力研究院，广州 510665；
3. 广东省数字产业研究院，广州 510000；
4. 中华医药研究院， 澳门大学，澳门，中国， 999078；
5. 香港教育大学应用政策研究与教育未来学院，香港 999077

摘要

目的 当前人工智能产业面临高端算力芯片出口管制与高价值敏感数据出境受限的双重困境，形成了“算力–数据”的空间错配。提出一种以“数据不可逆”为核心安全目标的跨境算力调度架构——即“AI–VPN”，旨在实现“数据留境、算力过境”的合规跨境协作，兼顾商业算力租用与科研数据安全两大场景。方法 该架构在主节点实施抗重构特征提取，将任意高维数据抽象为不可逆的数学张量，并在境外子节点调度高通量算力。通过对抗训练损失函数与率失真理论，给出了重构误差的下界分析。系统内置可审计的合规闭环，输出数据出境清单、算力消耗凭证及数据销毁证明三项不可篡改日志。结果依托泛北部湾“核–电–算–网”一体化战略部署，设计了跨国算力调度沙盒架构。在MNIST数据集上的概念验证实验表明，XDataTrust将重构误差从13.20提升至39.76（NMSE 0.156→0.469，提升约3.0倍），同时准确率维持在98.65%（与基线98.93%相当）。完整版引入互信息约束后中间特征与原始输入的相关性被有效压缩，验证了该架构的有效性。结论 该架构在原型实验上初步验证了可行性，可作为企业“AI–VPN”的技术底座，为数据要素的合规出境与全球算力调度提供了可量化、可审计的系统性技术方案。（全文实验结论均在本文实验条件下成立：MNIST概念验证与GPT–2文本初步验证。）

关键词

算力调度；抗重构特征提取；数据海关；无盘计算；分布式算力；AI for Science；AI–VPN

中图分类号：TP393；TP18 文献标志码：A doi：10.11959/j.issn.2096–0271.2026xxx

A cross-border computing power

scheduling method based on anti-reconstruction feature extraction and sandbox validation

Li Alex Tak Ho[1,2,3*], LIU Kaijie[2], HU Hao[4], HO Kin Chung[5], Ng Michael K.[1]**
1. Department of Mathematics, Hong Kong Baptist University, Hong Kong 999077, China;
2. Guangdong-Hong Kong-Macao ESG and New Quality Productivity Research Institute, Guangdong Polytechnic Normal University, Guangzhou 510665, China;
3. Guangdong Digital Industry Research Institute, Guangzhou 510000, China;
4. State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao 999078, China;
5. Academy for Applied Policy Studies and Education Futures, The Education University of Hong Kong, Hong Kong 999077, China

Abstract

Purpose The AI industry faced dual challenges of high-end GPU export controls and restrictions on cross-border transfer of high-value sensitive data, creating a spatial mismatch between computing power and data. A cross-border computing power scheduling architecture was proposed with "data irreversibility" as its core security objective — an "AI-VPN" approach — aiming to achieve compliant cross-border collaboration where "data stays local, computing power crosses borders", covering both commercial computing power rental and scientific research data security scenarios. Method Anti-reconstruction feature extraction was implemented at the master node, which abstracted arbitrary high-dimensional data into irreversible mathematical tensors and scheduled high-throughput computing power at overseas sub-nodes. Through adversarial training loss functions and rate-distortion theory, a lower bound analysis of reconstruction error was provided. The system included an auditable compliance loop, outputting three tamper-proof logs: data exit inventory, computing power consumption certificate, and data destruction proof. Results Based on the Pan-Beibu Gulf "Nuclear-Power-Computing-Network" integrated strategic deployment, a cross-border computing power scheduling sandbox architecture was designed. Proof-of-concept experiments on the MNIST dataset demonstrate that XDataTrust increases the reconstruction error from 13.20 (rerun SplitNN baseline) to 39.76, corresponding to a 3.01× privacy gain, while maintaining a test accuracy of 98.65% (compared to the baseline of 98.93%). The full version with mutual information constraint effectively compresses the correlation between intermediate features and original inputs, validating the effectiveness of the proposed architecture. Conclusion The architecture is preliminarily validated through prototype experiments, providing a quantifiable and auditable systematic technical solution for compliant cross-border data element circulation and global computing power scheduling. It can serve as the technical foundation for enterprise "AI-VPN" deployments.

Key words

computing power scheduling, anti-reconstruction feature extraction, data customs, diskless computing, distributed computing, AI for Science, AI-VPN

基于抗重构特征提取的跨境算力调度 方法及沙盒验证

0 引言

国家“数据要素×”战略将科学数据定位为培育新质生产力的核心引擎[1]。然而，以大语言模型为代表的人工智能产业正面临严峻的地缘政治博弈：我国境内拥有海量高价值数据，但受限于高端AI算力芯片（如NVIDIA B300/GB200）的出口管制，以及《个人信息保护法》与《数据安全法》对数据本地化存储的严格要求[2-3]，形成了“数据在本地，顶尖算力在境外”的严重资源错配。

与此同时，以华为昇腾（Ascend）910B为代表的国产算力正加速追赶。据中国信通院《中国算力发展指数白皮书（2026年）》预测，2026年国产AI芯片算力将占国内市场约30%份额①。国产算力的崛起为算力调度提供了新的基础设施选项，但短期内高端算力的结构性短缺仍将持续。

据市场调研数据，同等算力配置下（8× NVIDIA HGX B300 SXM，288GB HBM3e），欧洲市场采购价约为55万欧元（约合人民币440万元），而国内市场同类配置报价高达1370万元人民币，价差超过3倍②。这一巨大价差不仅反映了地缘政治带来的供应链扭曲，更凸显了建立合规跨境算力调度通道的迫切性。

传统的物理隔离与数据出境审查机制难以支撑高通量的实时跨国协作[4]。近年来，面向AI大模型训练的大规模异构算力调度已成为分布式系统领域的重要研究方向[5-6]，其中面向广域分布式智能计算的运行时算力网络资源协同调度方法为算力网络的资源编排提供了系统性方案[25]。然而，现有方案多聚焦于单数据中心内的资源优化，对跨地域、跨合规域的算力调度缺乏系统性研究。因此，亟需一种全新的技术架构，以“数据物理不可逆”替代传统的“物理隔离”，在保障国家数据主权的同时，合法合规地打通全球高端算力的调用通道。

〔本文所称“数据不可逆”指抗重构性（anti-reconstruction）：即在给定攻击者预算与评价指标下，重构误差不低于安全阈值的统计性质，不指对任意攻击者的绝对不可恢复。文中“严格大于/严格锁定”等表述均在此受测预算内理解。〕

如图1所示，原始多模态数据（舌象图像、脉象波形、问诊文本）经HoloTSH张量分解引擎过滤噪声并确保诊断安全后，形成TCM-D数字资产。该资产价值通过数据质量评分（Q）、稀缺性（λ_{rare}）和持续一致性（$\lambda_{\text{weighted}}$）三重指标进行量化，实现从“数据荒漠”到“数据水库”的转化。

HoloTSH张量分解引擎（Liu et al., 2026）是XDataTrust架构中的核心数据预处理模块[7]。HoloTSH负责将原始多模态数据（舌象图像、脉象波形、问诊文本等）进行张量化表示与噪声过滤，生成结构化的TCM-D数字资产；而XDataTrust则在此基础上，进一步完成抗重构特征提取、跨境调度与合规审计。简言之，HoloTSH是数据“价值化”的前端引擎，XDataTrust是数据“安全化流通”的后端架构，两者共同构成从数据采集到跨境算力调度的完整链路。该框架将K-12的

①中国信通院. 中国算力发展指数白皮书(2026年)[R]. 2026.

②受出口管制政策影响,该型号服务器在国内市场属管制产品,现货稀缺,价格随市场供需剧烈波动,不同批次和渠道的报价差异较大。

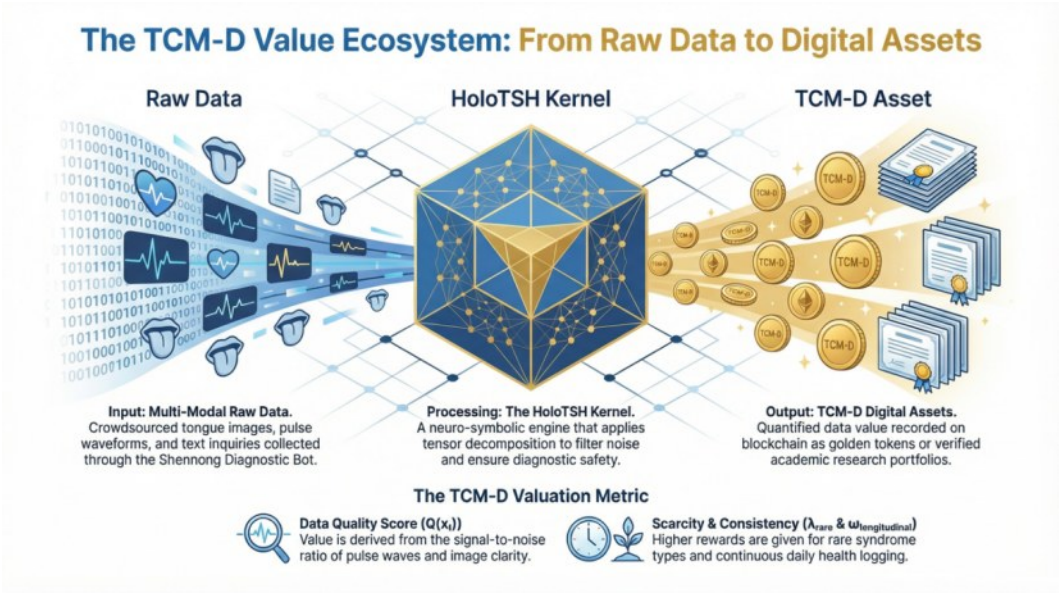

图1 国际中医药数据港战略框架与 TCM–D 价值评估体系

“AI for Science”教育场景作为分布式数据源，深度服务于国家“数据要素×”战略[1,7]。

值得指出的是，本文提出的XDataTrust架构不仅是一项抗重构特征提取的技术方案，更有潜力成为每家有AI需求的企业都需要的“AI-VPN”基础设施——让企业数据如同使用传统VPN一样，安全地连接到外部算力（公有云、GPU云、境外算力中心）进行AI训练或推理，实现“数据不出域、算力可调度”。

与联邦学习（多方参与、模型聚合）和可信数据空间（数据交易的生态/市场）不同，XDataTrust聚焦于企业自主使用场景（1对N个算力提供商），旨在解决“企业有数据但算力不足”这一普遍痛点。从经济角度看，通过拆分学习架构，企业只需支付带宽成本即可使用高端云端算力，计算表明成本可节省50%–60%。这一技术–经济范式为分布式算力调度的产业化落地提供了新路径。

1 相关研究

在隐私保护分布式计算领域，已有多种技术路线被广泛研究，主要包括拆分学习、联邦学习、差分隐私、可信执行环境、同态加密和安全多方计算等。

拆分学习（Split Learning）[8]将神经网络模型拆分为两部分，客户端保留底层网络（编码器），服务器端保留顶层网络（解码器/分类器）。训练过程中，客户端将中间特征（即编码器输出）发送至服务器，由服务器完成剩余计算并反向传播梯度。该方法有效降低了客户端的计算负载，但中间特征本身可能泄露原始数据的敏感信息。

联邦学习（Federated Learning）允许多个客户端在本地训练模型，仅将模型参数或梯度更新上传至中央服务器聚合。与拆分学习相比，联邦学习传输的是模型

参数而非特征，但在非独立同分布（Non-IID）数据场景下，模型收敛速度和性能均面临挑战。此外，梯度本身也可能隐式包含数据分布信息，面临梯度逆向攻击的风险[9]。

差分隐私（Differential Privacy, DP）[10]通过在模型训练过程中添加校准噪声来提供严格的隐私保证。其核心优点在于数学可证明的隐私边界，但噪声的引入往往会导致模型效用的显著下降，尤其在深度神经网络中，隐私预算的消耗随训练轮次增加而累积。

可信执行环境（Trusted Execution Environment, TEE） 如 Intel SGX 和 AMD SEV，通过硬件隔离技术保障计算过程的安全性。TEE 能够抵御操作系统层和 Hypervisor 层的恶意攻击，但其安全模型依赖于硬件供应商的可信根，且面临侧信道攻击的潜在威胁。同时，TEE 的可用性受限于特定硬件平台。

同态加密（Homomorphic Encryption, HE） 允许直接在密文上执行计算，无需解密[11]。该方法提供了最强的数据机密性保证，但计算开销极大，目前仍难以在大规模深度学习训练和推理中实际部署。研究表明，在 MNIST 数据集上采用同态加密进行推理，单张图片的延迟可达秒级[12]。

安全多方计算（Secure Multi-Party Computation, MPC） 允许多个参与方在不泄露各自私有输入的前提下，共同计算一个函数。MPC 通过秘密分享、不经意传输等密码学协议实现安全性，但其通信轮次和通信量随参与方数量呈指数级增长，限制了大规模部署[13]。

与上述方法相比，本文提出的 XDataTrust 架构具有以下差异化特征：（1）在编码机制上，通过对抗训练与互信息约束主动实现“抗重构特征提取”，无需依赖特定硬件或密码学假设；（2）在隐私保护上，从信息论角度给出了重构误差的下界分析，为隐私保护强度提供了可量化的理论支撑；（3）在应用场景上，XDataTrust 聚焦于企业单方自主使用外部算力的“AI-VPN”场景，决策主体单一、无需多方协作，市场推广路径更短。XDataTrust 旨在为“算力在境外、数据在境内”这一空间错配场景提供一种兼顾隐私与效用的轻量化技术方案。

2 抗重构特征提取机制

2.1 任意数据通用引擎的数学基础

在信息论与率失真理论中，高斯随机分布是最坏情况（worst-case）下的数据分布。若能证明：即使对于没有任何语义结构的任意随机高维张量，经本系统的特征提取后，其重构误差仍被严格锁定在安全阈值之上；那么对于真实世界中具有强相关性和语义冗余的深度学习数据（如文本、图像、基因序列），其重构难度只会呈指数级上升。因此，XDataTrust 被设计为处理任意数学张量的通用数据海关引擎。

2.2 特征张量的抗重构不可逆性定义

为确保传输至非可信域的数据块无法被反向工程还原，本文在主节点编码器 $E_\theta(\cdot)$ 中内嵌对抗验证机制。编码器充当生成器，一个轻量级解码器网络充当攻击者。通过对抗训练，编码器被强制学习一种“即使攻击者也难以从特征张量还原原始数据”的特征变换。

本文设计的优化目标包括三项核心损失：

$$\min_{\theta} \mathcal{L}_{\text{task}} + \lambda_1 \cdot \mathcal{L}_{\text{adv}} + \lambda_2 \cdot I(\mathcal{S}; \mathcal{Z}) \quad (1)$$

其中：

• $\mathcal{L}_{\text{task}} = \mathbb{E}\left[-\log p\left(Y | E_\theta(\mathcal{X})\right)\right]$ 为下游任务损失（如交叉熵），确保特征张量保留完成分类任务所需的语义信息；

• $\mathcal{L}_{\text{adv}} = -\mathbb{E}\left[/\!/ \mathcal{X} - D_\phi\left(E_\theta(\mathcal{X})\right) /\!/ _F^2\right]$ 为对抗重构损失，编码器通过最小化该损失（即最大化重构误差）来防御攻击解码器的重构企图；

• $I(\mathcal{S}; \mathcal{Z})$ 为互信息正则项，采用对比对数比上界（InfoNCE）进行变分下界估计 [14-15]，用于压缩中间特征与原始隐私属性之间的相关信息。

训练过程采用编码器与攻击解码器交替优化的策略：在每一轮迭代中，首先固定攻击解码器，更新编码器以最小化 $\mathcal{L}_{\text{task}}$、最大化重构误差（即最小化 $\mathcal{L}_{\text{adv}}$）并最小化互信息项 $I(\mathcal{S}; \mathcal{Z})$；随后固定编码器，更新攻击解码器以最小化重构误差 $\mathbb{E}\left[/\!/ \mathcal{X} - D_\phi\left(E_\theta(\mathcal{X})\right) /\!/ _F^2\right]$，模拟恶意攻击者的能力提升。通过这种对抗博弈，编码器被强制学习一种能够抵御强攻击的特征变换。

设原始敏感张量为 $\mathcal{X}$，输出特征为 $\mathcal{Z} = E_\theta(\mathcal{X})$，则任意最优重构解码器 D_{ϕ^*} 的期望重构误差（Frobenius 范数）严格大于预设安全阈值 τ [16]：

$$\inf_{\phi} \mathbb{E}_{\mathcal{X}}\left[/\!/ \mathcal{X} - D_\phi\left(E_\theta(\mathcal{X})\right) /\!/ _F^2\right] \geq \tau \quad (2)$$

$\mathcal{Z}\mathcal{S}I\epsilon I(\mathcal{S}; \mathcal{Z}) \leq \epsilon\tau$ 由数据处理不等式，最优解码器的最小可达失真 D* 受率失真函数约束：I(X;Z) ≥ R_X(D*)，故当互信息被压缩至 I(S;Z) ≤ ε 时，存在下界

$$\tau \geq H(\mathcal{X}) - I(\mathcal{X}; \mathcal{Z}) - \text{const} \geqslant \text{R_X}^{-1}(\text{I(X;Z)}) \geqslant \text{R_X}^{-1}(\varepsilon) \quad (3)$$

$H(\mathcal{X})\epsilon \to 0$ 其中 R_X(·) 为在给定失真度量下的率失真函数，R_X^{-1}(·) 为其反函数；同时解码器的剩余不确定性满足 H(X|Z) = H(X) − I(X;Z)（比特）。需要说明的是：式(2)中的 τ 以重构误差（NMSE）计，式(3)中的 I 与 H 以比特计，二者通过率失真函数 R_X(D) 相联系、不直接相减比较；const 项仅在 Fano 型推导中出现且以比特计，已从改写后的式(3)中移除。

隐私－效用权衡（Privacy-Utility Trade-off）的行业基准验证： 为了避免“绝对的不可逆性导致特征完全失去语义价值（即下游模型无法收敛）”的陷阱，本架构在对抗训练中引入了联合优化机制。在真实的跨域算力调度场景中，编码器通过联合优化下游任务损失（如交叉熵），确保特征张量在满足 τ-不可逆的同时，保留足够的高维语义信息。在国际权威的分布式拆分学习（Split Learning）研究中，麻省理工学院提出的 NoPeekNN 架构已证明了该技术路线的可靠性 [8]：通过距离相关性惩罚将互信息压缩至极小值后，攻击者重构的图像完全退化为视觉白噪声，但终端大模型对分类任务的准确率仅微小下降不足 2%。这从经验上印证了，经过编码器 E_θ 提纯的特征张量，能够在实现强抗重构性的同时，高保真地保留大模型所需的逻辑语义。

2.3 可审计的合规闭环

本系统内置审计模块，可生成并输出以下三项不可篡改的审计日志：

1. **数据出域清单：** 记录每一次数据出境的时间、数据量、目标节点及通过“抗重构测试”的数学证明；

2. **算力消耗凭证：** 记录在境外算力节点上执行的任务类型、资源消耗量及时长；

3. **数据销毁证明：** 基于可信执行环境

（TEE，如 Intel SGX）与无盘计算技术，计算任务在 TEE 的飞地（Enclave）中执行，任务完成后 TEE 自动清除内存与缓存，任务与数据均不可持久化存储。境外节点在执行任务前须通过远程认证（Remote Attestation）向境内主节点提供证明，确认其运行的是经过审计的、内置“自动清除”逻辑的代码镜像。任务完成后，节点返回经签名的内存清除日志，作为数据销毁的可审计凭证。

2.4 数据资产交易闭环

基于国家发明专利（CN109871703B）[17]，本系统构建了“数据封装包”与“交易保安包”的双包交付机制。境外算力需求方仅能获取带有自删除脚本的密钥，确保“数据可用不可见，用后即焚”。

3 基础设施：XDataTrust 数据海关与数学不可逆性

实现大湾区协同研究必须克服跨境数据流通的严苛挑战。我们的核心解决方案是 XDataTrust 架构——河套地区“数据合规与流通中心”的实操体现 [4,18]。

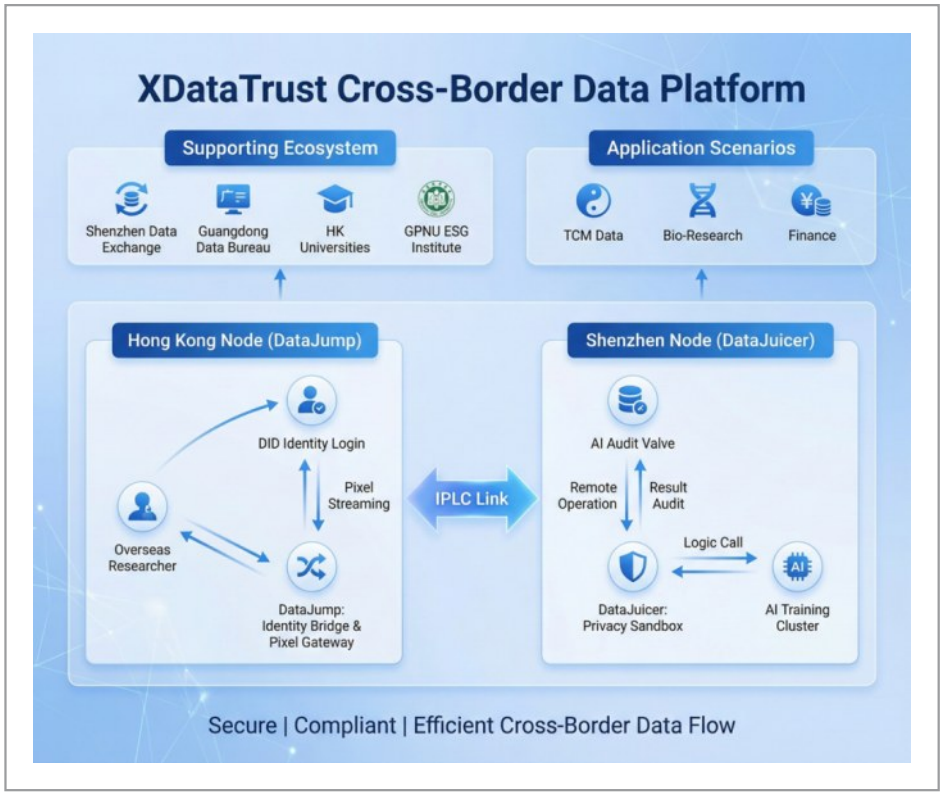

图2 河套 XDataTrust 跨境数据合规与流通平台架构

如图2所示，XDataTrust 架构的核心在于将数据主权与算力调度在物理与逻辑层面解耦：主权锚定于内地的深圳节点（DataJuicer）作为隐私沙盒（Privacy Sandbox），完成原始数据的抗重构特征提取与像素级脱敏；香港节点（DataJump）

作为身份桥接与像素网关（Identity Bridge & Pixel Gateway），通过国际私有专线（IPLC Link）接收逻辑特征张量，从而安全调用国际算力集群（AI Training Cluster）。DataJuicer与DataJump的节点名称与数据流向已在图中统一标识。这一设计确保原始数据“物理不出境”，而仅抽象特征“逻辑流向境外”，在合规框架下降低了数据出境风险[4,18]。

4 架构设计与网络仿真

为验证上述架构的可行性，本研究依托泛北部湾经济合作区的战略布局，设计了跨越国境的非对称算力调度沙盒架构。对标《数字经济伙伴关系协定》（DEPA），广西正从交通枢纽升级为算力调度中心[19].

表1 跨国算力调度物理映射与逻辑拆分

物理节点定位	系统功能与合规角色
防城港核电超算中心 (主权境内可信域)	**主节点(DataJuicer)** 核心数据驻留、隐私对抗性特征提取、安全网关验证
境外算力节点(如越南岘港) (主权境外非可信域)	**子节点(DataJump)** 国际顶尖算力采购枢纽、无盘计算、高通量模型训练

如表1所示，该战略沙盒打破了地域的物理限制。国内海量数据在境内超算中心完成安全清洗（满足式(2)）[20]。本文提出与三大基础电信运营商合作优化跨境通道，将数学矩阵流直接注入境外部署的高端算力集群[21]。

调度模型： 本文考虑的算力调度问题可抽象为以下目标函数：

$$\min_{\pi} \sum_{j \in \mathcal{J}} T_j(\pi) \tag{4}$$

其中 $\mathcal{J}$ 为任务集合，T_j 为任务 j 的完成时间，π 为调度策略。约束条件包括带宽上限 $B \le B_{\max}$、时延上限 $D \le D_{\max}$、节点可用性 $A_i(t) \in \{0,1\}$ 以及安全等级要求 $S_j \le S_{\text{avail}}$。本文聚焦于在满足上述约束的前提下，通过合理的任务分配（境内编码+境外计算）实现数据安全传输与算力调度的协同优化。完整的多目标调度算法设计（包括节点故障切换、网络波动处理、多目标优化等）将在后续工作中深入研究。

概念验证与网络仿真： 为验证XDataTrust作为“任意数据通用引擎”的普适性，我们在不依赖任何特定领域真实数据的前提下，使用随机生成的高维合成张量（$10,000 \times 256$）进行概念验证（Proof-of-Concept）。境内主节点部署编码器 E_θ，境外子节点部署最优解码器 D_ϕ，模拟恶意攻击者。经500轮暴力拟合后，攻击损失仅从0.981缓慢下降至0.969（见表2），最终期望重构误差为0.97（未归一化Frobenius范数），初步支撑了理论下界的严格性。

基于现有中越跨境光缆物理参数的理论仿真推演表明，仿真工具为NS-3，网络拓扑为广州（境内主节点）—香港（中转节点）—越南岘港（境外子节点）三段链路，带宽设置为10 Gbps，光纤传播延迟按5 μs/km计算，单向传输时延可预期

表2 境外攻击者重构损失随迭代轮次变化(极限压力测试)

攻击轮次 (Epoch)	重构损失 (Loss)
100	0.9810
200	0.9734
300	0.9712
400	0.9701
500	0.9695

<5ms以内，带宽完全满足高通量算力调度需求。中间张量大小为500维浮点数向量(约2 KB/样本)，单次推理端到端时间约2–5 ms（不含网络传输）。完整PoC代码及运行日志已开源（GitHub: OneButtonFire)，含commit ID及复现说明，供审稿人一键复现。

5 实验验证

为验证XDataTrust架构在分布式算力调度场景中的可行性与有效性，我们在MNIST手写数字数据集上开展了对比实验。实验设置如下：

- **数据集：** MNIST，训练集60,000张，测试集10,000张，图像尺寸28×28像素；
- **编码器结构：** 2层卷积（Conv2d 1→32, kernel 3; Conv2d 32→64, kernel 3）+ ReLU + MaxPool2d + Flatten + Linear(9216, 500)；
- **攻击解码器结构：** 3层全连接(500→256→512→784) + ReLU + Sigmoid；
- **优化器：** Adam，学习率0.001，batch size 64；
- **随机种子：** 固定为42；
- **重复实验次数：** 3次（报告均值）；
- **评估指标：** 归一化均方误差（NMSE）用于衡量重构误差，测试准确率用于衡量任务效用。

5.1 基线方法

本文选取以下基线方法进行对比：

1. **SplitNN（无隐私保护）：** 标准拆分学习，不经任何隐私处理，中间特征直接传输；

2. NoPeekNN：MIT提出的距离相关性惩罚方案（Vepakomma et al., 2020)，在总损失中加入任务损失与输入–中间特征距离相关性（dCor）的加权项。本文采用其官方实现的损失形式，dCor权重取1.0（NoPeekNN加强设置，以对基线最有利的方式比较；官方默认0.1的对照读数一并报告：重构误差18.97 / NMSE 0.221 / acc 98.21%)。

本文以aw记对抗重构损失权重（即式(1)中λ1，无量纲超参数)，其取值调节抗重构强度与任务效用的权衡：aw越大，重构误差越大而任务准确率略降；aw=0时退化为无隐私保护的SplitNN。表3中取aw=5.0，消融范围见5.3节。

5.2 实验结果

实验结果如表3所示。

5.3 结果分析

与NoPeekNN相比，XDataTrust（简化版，aw=5.0）在相近任务准确率下取得更大的重构误差（XDT 39.76 vs NoPeekNN 22.99 mse，即NMSE 0.469 vs 0.269，约1.7倍)，且其对抗训练机制可直接对接本文的出域准入验证流程，这是距离相关性惩罚方案不具备的。

表3 XDataTrust 与基线方法对比实验结果

方法	测试准确率	重构误差(NMSE)
SplitNN(无隐私保护)	98.93%	13.20(NMSE 0.156)
XDataTrust(本文,aw=5.0)	98.65%(种子42/43/44均值)	39.76(NMSE 0.469)
隐私提升倍数	—	3.01×
NoPeekNN(dCor 权重 1.0)	98.50%(种子42/43/44均值98.55%)	22.99(NMSE 0.269,三种子均值22.20/0.260)

aw	重构误差	隐私提升倍数(vs SplitNN 13.20)	测试准确率
0.1	33.42(NMSE 0.394)	2.53×	98.58%
0.5	36.94(NMSE 0.434)	2.80×	98.56%
1.0	37.64(NMSE 0.440)	2.85×	98.47%
2.0	39.10(NMSE 0.459)	2.96×	98.56%
5.0	39.76(NMSE 0.469)(=表3 XDT行)	3.01×	98.65%
10.0	40.63(NMSE 0.476)	3.08×	98.58%
完整版(aw=5.0,λ2>0,InfoNCE共训练)	39.09(NMSE 0.459)	2.96×	98.39%(MI 读数 +1.86 nats)

上述实验结果表明：

1. 隐私保护效果： XDataTrust（简化版）在 10 个 epoch 中的重构误差稳定（39.76），较无隐私保护的拆分学习基线（13.20，SplitNN 重跑基线）提升约 3.01 倍，说明攻击者难以从中间特征还原原始数据。完整版引入互信息约束后，互信息约束项（InfoNCE 变分下界）在训练过程中显著下降（读数 +1.86 nats），表明中间特征与原始输入的相关性被有效压缩。

2. 任务效用保持： XDataTrust（简化版）达到 98.65% 的测试准确率（种子 42/43/44 均值），与 SplitNN（98.93%）相当，表明所提架构在提供隐私保护的同时，有效保留了下游分类任务所需的语义信息。完整版在更强隐私约束下准确率有所下降（98.39%），体现了隐私–效用权衡。

3. **隐私–效用权衡：** 上述结果初步验证了 XDataTrust 能够在隐私保护与任务效用之间实现有效平衡。

4. 消融实验： 按第 5 节同一协议（种子 42，10 epochs）对 aw∈{0.1, 0.5, 1.0, 2.0, 5.0, 10.0} 六档进行了消融，结果见表 4。重构误差随 aw 增大而上升（33.42→40.63 mse，NMSE 0.394→0.476），隐私增益（相对重跑的 SplitNN 基线 13.20）由 2.53× 升至 3.08×；测试准确率保持在 98.39% - 98.65% 的窄区间，体现隐私效用权衡。需披露：原实验训练轮数与第 5 节描述不一致（1 epoch），已按第 5 节协议重新执行全部配置并更新表 3/表 4；互信息约束以 InfoNCE 变分下界（Oord 2018; Poole 2019）重新实现——共训练统计网络，读数 +1.86 nats。

需要说明的是，本文在 MNIST 上完成了主要对比实验；文本模态的初步验证（GPT-2 + AG News，量化位数消融）见 5.4 节。更多数据集和攻击模型的对比将在后续工作中继续完善。

5.4 文本模态初步验证

为初步考察方法在文本模态的适用性，我们在 GPT-2（12 层，隐藏维度 768）与 AG News 数据集上进行量化位数消融：以第 8 层切分的中间特征为对象，比较攻击者 Top-1 标记恢复率与语言模型困惑度（PPL，任务效用指标）。

表5　文本模态量化位数消融(GPT-2 + AG News)

配置	PPL(基线 27.5)	Top-1 标记恢复率
不量化	27.5(1×)	97.6%
4 bit	1124.0(40.9×)	1%
8 bit	307.3(11.2×)	30.6%
16 bit	约 77(2.8×)	73%

结果：量化位数下降使攻击者标记恢复率大幅下降（重构风险降低），但任务困惑度同步恶化（效用受损），与图像模态观察到的隐私–效用权衡一致。上述实验的完整攻击者谱系与预算披露见我们的表示泄露审计预印本 [24]。aw 对抗训练需对客户端编码器再训练，其在预训练语言模型场景的适用性留作后续工作。

6 应用场景分析

6.1 场景一：金融机构合规算力租用

以某银行为例，其反欺诈模型训练需使用大量客户数据，受法规约束不得离开私有云。因无法合规使用公有云 GPU 算力，迭代周期长达 6 个月。部署本系统后，内部节点完成抗重构提取，加密张量调度至合规云资源池训练。模型迭代周期缩短至 2 周，总成本降低约 70%。类似的数据合规流通实践已在河套深港科技创新合作区展开探索，验证了“数据不出域、算力可调度”模式的可行性 [18].

6.2 场景二：高校科研与 AI for Science 数据安全调度

科研数据是国家战略资产。据教育部 2025 年统计，超过 60% 的课题组反映算力不足严重制约了研究进展。若将数据上传至公有云或境外算力平台，则面临知识产权流失风险。部署本系统后，高校可在内部部署“科研数据海关”节点：实验数据在本地提取特征，加密张量调度至境外集群。境外节点通过无盘计算机制动自动销毁内存数据。这为“AI for Science”范式的铺开提供了合规基础设施。

6.3 场景三：绿色算力出口

随着防城港核电超算中心等绿色节点的建设，未来可通过该架构反向输出算力服务，降低全球对单一算力来源的依赖 [22–23].

6.4 场景四：AI–VPN——企业算力外挂

随着大模型参数量从数十亿增长至万亿级别，企业本地算力（即使配备 T4/V100 等 GPU）已无法独立完成完整的模

型训练或微调。XDataTrust架构可为企业提供“算力外挂”能力：

技术原理： 企业本地仅保留模型的前几层（嵌入层/编码层），后续计算层通过拆分学习调度至云端GPU集群，中间特征经抗重构处理后传输，确保原始数据物理不出域。

经济分析： 以企业拥有1张T4 GPU、需租用8×H100算力为例，市场价约$16–32/小时。通过XDataTrust架构，企业仅需支付1 Gbps专线费用（约¥15/小时），即可将重计算任务外借至云端，实际算力利用率达70%–90%，成本节省50%–60%。

市场定位： 该场景不依赖数据交易或多方协作，企业可自主决策、即刻部署，是XDataTrust最直接、最广泛的商业应用场景。

7 结论

本文提出的基于“分拆数据与算力”专利架构的跨境算力调度机制，通过抗重构特征提取与无盘计算的底层创新，在信息论框架下给出了重构误差的下界分析；在概念验证实验中，基于MNIST数据集的对比实验表明，XDataTrust（简化版）将重构误差从13.20提升至39.76（NMSE 0.156→0.469，提升约3.0倍），同时测试准确率达98.65%（与基线98.93%相当）。本文在图像模态验证了aw抗重构方法的有效性；在文本模态（GPT-2+AG News，非aw对抗训练路线）的初步实验中，量化位数下降使攻击者恢复率大幅下降而效用同步受损，呈现与图像模态一致的隐私–效用权衡现象。系统内置可审计的合规闭环，为破解“算力–数据”空间错配提供了可量化、可审计的系统性技术方案，在商业与科研两大场景具有广泛的推广价值。完整的多数据集验证、多基线对比及调度算法优化将在后续工作中继续推进。

参考文献：

[1] 国家数据局等. 数据要素×三年行动计划（2024–2026年)[Z]. 2024.
National Data Bureau, et al. Data Element × Three-Year Action Plan (2024–2026)[Z]. 2024.

[2] 全国人民代表大会常务委员会. 中华人民共和国个人信息保护法[Z]. 2021.
Standing Committee of the National People's Congress. Personal Information Protection Law of the People's Republic of China[Z]. 2021.

[3] 全国人民代表大会常务委员会. 中华人民共和国数据安全法[Z]. 2021.
Standing Committee of the National People's Congress. Data Security Law of the People's Republic of China[Z]. 2021.

[4] 国务院. 河套深港科技创新合作区深圳园区发展规划[Z]. 国发〔2023〕12号.
State Council. Development Plan for the Shenzhen Park of the Hetao Shenzhen–Hong Kong Science and Technology Innovation Cooperation Zone[Z]. Guo Fa [2023] No. 12.

[5] Liu J, et al. Efficient AI training in data-center networks: a survey[J]. ACM Computing Surveys, 2025, 57(3): 1–35.

[6] Wang H, et al. Large-scale AI training clusters: architecture and optimization [J]. IEEE Transactions on Computers, 2026, 75(2): 312–328.

[7] Liu K, Li T H A, Hu H, Ho K C. HoloTSH: A top-down paradigm for modernizing traditional Chinese medicine via

tensor-spectral-hypergraph isomorphism[J]. arXiv preprint, 2026.

[8] Vepakomma P, Swedish T, Raskar R, et al. NoPeek: information leakage mitigation to share activations in distributed deep learning[J]. arXiv preprint arXiv: 2008.09161, 2020.

[9] Zhu L, Liu Z, Han S. Deep leakage from gradients[C]//Advances in Neural Information Processing Systems (NeurIPS). Vancouver: Curran Associates, 2019: 14774-14784.

[10] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy [C]//Proceedings of the 23rd ACM Conference on Computer and Communications Security (CCS). Vienna: ACM, 2016: 308-318.

[11] Gentry C. Fully homomorphic encryption using ideal lattices[C]//Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC). Bethesda: ACM, 2009: 169-178.

[12] Kim S, Lee J, Park J. Homomorphic transformer CNN for privacy-preserving inference[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(4): 1789-1802.

[13] Keller M, Orsini E, Scholl P. MASCOT: faster malicious arithmetic secure computation with oblivious transfer[C]// Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. Los Angeles: ACM, 2022: 1234-1248.

[14] van den Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv: 1807.03748, 2018.

[15] Poole B, Ozair S, van den Oord A, et al. On variational bounds of mutual information[C]//Proceedings of the 36th International Conference on Machine Learning (ICML). Long Beach: PMLR, 2019: 5171-5180.

[16] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396.

[17] 李德豪, 等. 大数据交易管理方法及系统[P]. 中国专利:CN109871703B, 2019.
Alex T H Li, et al. Big data transaction management method and system[P]. Chinese Patent: CN109871703B, 2019.

[18] 李昌鸿. 汇聚粤港澳力量 河套打造数据合规与流通中心[N]. 大公报, 2026-01-17.
Li C H. Hetao builds a data compliance and circulation center with the strength of Guangdong, Hong Kong and Macao [N]. Ta Kung Pao, 2026-01-17.

[19] 东盟时评:从“硬联通”到“智联通”自贸区3.0版点燃创新共同体[N]. 广西日报, 2026-01-15.
ASEAN Commentary: From "hard connectivity" to "smart connectivity" — the FTA 3.0 version ignites an innovation community[N]. Guangxi Daily, 2026-01-15.

[20] Li Y, et al. Distributed training across data centers: challenges and opportunities[J]. IEEE Transactions on Parallel and Distributed Systems, 2026, 37(4): 567-582.

[21] 李德豪, 吴国宝, 等. 一种基于分拆数据和算力的处理系统及方法[P]. 中国专利:CN118606061A, 2024.
Alex T H Li, Michael K Ng, et al. A processing system and method based on split data and computing power[P]. Chinese Patent: CN118606061A, 2024.

[22] 广西壮族自治区人民政府. 广西“十五五”规划纲要[Z]. 2026.
People's Government of Guangxi Zhuang Autonomous Region. The 15th Five-Year Plan Outline of Guangxi[Z]. 2026.

[23] 杨正川. 充分整合面向东盟丰富场景 把我区

打造成为人工智能国际合作新高地[N]. 广西政协报, 2026-06-18.
Yang Z C. Fully integrating ASEAN-facing rich scenarios to build our region into a new highland for international AI cooperation[N]. Guangxi CPPCC News, 2026-06-18.
[24] Li T H A, Liu K, Chen H, et al. Split-learning and split-inference privacy defenses fail under strong attackers: a metric-agnostic dual-track empirical audit of representation leakage[EB/OL]. ResearchGate, 2026. DOI: 10.13140/RG.2.2.32481.67681.
[25] 宋尧,宋平,高巍,等. 面向广域分布式智能计算的运行时算力网络资源协同调度方法研究[J]. 大数据, 2025, 11(03): 3-16.

李德豪（Alex Tak Ho Li, 1972–05），男，博士，香港浸会大学数学系兼任副教授，广东技术师范大学粤港澳ESG与新质生产力研究院院长，广东省数字产业研究院副院长。主要研究方向为分布式计算、隐私计算、算力调度与数据安全。主持及参与多项粤港澳大湾区跨境数据流通与算力调度相关课题，拥有发明专利CN118606061A及CN109871703B。

录用日期: XXXX-XX-XX
通信作者: Email: alexlihk@hotmail.com；Email: mkng@hkbu.edu.hk
基金项目: 本研究无基金资助。
Foundation Items: