arXiv:2602.16174v1 [cs.NI] 18 Feb 2026

通过联邦拆分决策 Transformer 实现元宇宙资源分配的边缘学习

Fatih Temiz、Shavbo Salehi 和 Melike Erol-Kantarci，*IEEE* 会士渥太华大学电气工程与计算机科学学院，加拿大渥太华电子邮箱：{ftemi033、ssale038、melike.erolkantarci}@uottawa.ca

摘要——**基于移动边缘计算（MEC）的无线元宇宙服务为用户提供不受束缚的沉浸式体验，而要实现卓越的体验质量（QoE），必须满足严格的时延约束和视觉质量要求。为此，面向虚拟现实用户的 MEC 智能资源分配需要通过 MEC 服务器之间的协同来利用分布式数据。联邦学习（FL）是一种很有前景的解决方案，可与强化学习（RL）相结合，在各 MEC 服务器上形成通用策略。然而，传统联邦学习需要在 MEC 服务器与云端之间传输完整的模型参数，并且会因简单粗暴的全局聚合而导致性能下降，在异构多无线接入技术环境中尤为明显。为应对这些挑战，本文提出联邦分割决策Transformer（FSDT），这是一种离线强化学习框架，其中 Transformer 模型被划分部署于 MEC 服务器和云端。智能体特定组件（例如基于 MEC 的嵌入层和预测层）支持本地适应，而云端的共享全局层则促进各 MEC 服务器之间的协同训练。实验结果表明，与基线方法相比，FSDT 在异构环境中可将体验质量提升至多 10%，同时将近 98% 的 Transformer 模型参数卸载至云端，从而减轻 MEC 服务器的计算负担。**

索引词——**决策 Transformer（DT）、边缘人工智能、联邦式分割学习（FSL）、元宇宙。**

I. 引言

元宇宙的迅速发展开启了沉浸式体验的新时代，其应用范围远不止娱乐，还延伸至教育、医疗保健、文化和全球社会互动。它通过无线虚拟现实（VR）提供沉浸式服务，针对用户需求进行定制，以提升用户的体验质量（QoE），并克服有线体验所受物理限制带来的障碍。然而，当仅由硬件能力有限的头戴式显示器（HMD）执行高分辨率视场（FoV）渲染任务时，时延便成为一个重大问题 [1]。因此，移动边缘计算（MEC）受到广泛关注，它可将计算密集型任务卸载至边缘网络 [1]，或至少与头戴式显示器协同处理 [2],，从而降低时延、提高视觉质量，避免晕动症。此外，人类视觉具有内在的层级结构，由中央、旁中央和周边区域组成，这些区域感知视觉信息的细节程度和敏感度各不相同 [3]。注视感知流媒体系统可以利用这一层级特性，更合理地分配 MEC 资源，利用头部与眼动追踪，在用户最关注的视场区域提供高保真渲染。

强化学习（RL）在复杂、动态且对时间敏感的环境中展现出巨大潜力，可为 VR 流媒体提供自适应资源分配与调度 [1]。然而，当探索不足时，深度确定性策略梯度（DDPG）等在线强化学习算法可能难以收敛至最优策略 [4]。为增强在线强化学习框架，研究人员提出了数字孪生（DTW）技术：该技术可在数字空间中实现更安全的仿真与训练，并通过持续学习和实时更新强化学习策略来缓解不同步问题 [5]。尽管 DTW 能够支持强化学习，但其开发与维护成本高昂，因为需要大量收集数据并在缓冲区中存储历史数据。构建能够实时复现静态和动态环境的高保真 DTW，势必会产生显著开销。为解决在线强化学习的不足，离线强化学习方法（如决策 Transformer（DT））应运而生，成为颇具前景的替代方案；这类方法利用历史轨迹实现更稳定且样本效率更高的学习 [6],，并逐渐成为无线网络优化的新兴范式，尤其适用于意图驱动网络管理 [7] 和资源分配问题[8]。除了在线与离线之分，联邦学习（FL）还可与强化学习相结合，形成联邦强化学习（FRL）：在分布式节点或 MEC 服务器上分布式训练泛化策略，同时解决单个 MEC 服务器的数据稀缺问题，且无需共享原始数据 [9]。

近期研究探讨了在支持 MEC 的元宇宙场景中应用联邦学习和联邦强化学习，并结合基于视场角渲染的 QoE 模型 [4], [10]。在 [10],中，作者在开放式无线接入网环境中设计了一种网络切片框架，并在多个智能体上训练联邦强化学习模型，以优化功率控制和物理资源块分配。与基准算法相比，他们基于视场角渲染、由体验质量驱动的方法显著降低了延迟，并提升了用户体验。此外， [4], 中的作者提出了 FedPromptDT，这是一种基于提示引导的决策变换器和定制 QoE 模型的联邦式离线强化学习算法，并利用眼动追踪数据集证明，该算法的性能优于在线算法和集中式决策变换器训练。

除了利用分布式MEC资源将联邦学习融入移动边缘之外，分割学习在

6G时代的边缘端人工智能（AI）中发挥着关键作用。通过对模型进行分层划分，分割学习能够更高效地利用分散在网络边缘的计算资源，从而最终提升整体资源利用率 [11]。在 [12], 中，作者针对异构边缘内容缓存网络提出了一种个性化联邦强化学习框架。他们的方法将多头深度Q网络与逐层个性化联邦训练相结合，通过分割模型来平衡全局知识共享与本地个性化。该设计缓解了联邦学习中朴素全局聚合的局限性，而这种聚合往往导致缓存策略同质化，并降低不同MEC服务器上的性能。然而，上述系统仍未受益于依赖先进Transformer架构的离线强化学习范式。

近期，研究者在 [13], 中探索了基于 Transformer 架构的联邦分割训练，尤其是用于视觉 Transformer，随后又拓展至决策Transformer [14]。这些架构主要用于与任务无关的训练，在异构环境中表现出色，使不同智能体能够各自专精，同时受益于共享的全局表征。例如，在视觉应用中，一个智能体可以通过其本地嵌入和预测层执行分类，另一个智能体则专注于分割，但二者都能持续受益于共享的全局Transformer解码器。这种设置有助于更充分地利用分布式数据，同时让每个智能体都能灵活适应自身的任务或环境，而不必牺牲协同学习带来的优势。

在此研究基础上，本文提出一种基于联邦分割决策Transformer（FSDT）的元宇宙应用资源分配框架，以应对MEC数据稀缺、在线强化学习不稳定、简单的全局聚合以及联邦学习中全模型交换效率低下等挑战。FSDT 方法结合联邦学习与分割学习来训练决策Transformer算法，训练期间模型在MEC服务器与云端之间进行分割。该设计通过特定智能体类型的嵌入和预测层保持本地适应性，并借助云端共享的Transformer实现全局协调。本文的主要贡献如下：

1) 我们提出了一种基于模型划分的协同边缘人工智能框架，将计算密集型的全局Transformer解码器部署在云端，同时由分布式 MEC 服务器承载嵌入层和预测层，从而实现面向特定领域的定制。2) 我们针对异构多 RAT网络中的基于 FSDT 的元宇宙资源分配问题进行建模，并采用公开VR眼动追踪数据集。3) 我们开展了大量实验，以评估多种权衡因素，包括在线与离线强化学习算法，以及两种训练模式下的联邦强化学习（FRL），从而在多种设置下为所提出的 FSDT 框架提供全面的基准评估。

II. 系统模型与问题构建

A. 系统模型

在本文中，我们考虑一种异构多RAT MEC系统，该系统根据配备头戴式显示器的无线VR用户的请求，渲染并提供沉浸式 360° 视频流，如图1所示。该系统由共址基站（BS）组成，每个基站均连接至一个 MEC 服务器，用于分配包括 GPU渲染周期、中央处理器编码周期、带宽（BW）和分辨率比例在内的资源，以确保向VR用户及时传输高质量视频。根据所采用的接入技术，基站被划分为 K 种不同的智能体类型：中上频段（*UMB*）、*6 GHz*以下和*Wi-Fi*。在每种智能体类型（即域）内，基站在不同的杂波环境下运行，例如城市微蜂窝（UMi）和室内热点（InH），以反映多样化的传播条件。此外，为支持自适应流媒体，我们采用 [5] 中的方法，将每个 360°视频投影到二维平面上，并划分为图块（例如4×4），如图1所示。对于时隙 t的用户 k，只有视场（FoV） $\mathcal{F}_k$（t）内的图块以高质量传输，而外围图块则采用较低的分辨率。分层视觉使用对应于标清（SD）、高清（HD）和超高清（UHD）的三个质量等级$q \in \{1, 2, 3\}$。图块按照包含 F 帧的图像组（GoP）进行编码，并在每个时隙中传输一个图像组。然后，对于每个用户 k，质量等级 q 下的数据大小为

$$d_{k,q}(t) = M_{k,q}(t)\, \beta_{k,q}(t)\, F, \tag{1}$$

其中 $M_{k,q}(t)$ 表示用户k 在质量 q 下的图块数量（其中 $\sum_q M_{k,q}(t)$ $= M_k(t)$；例如图 1 中的 $3+9+4=16$），$\beta_{k,q}(t)$ 是质量等级 q 下每个图块的比特率，总数据量为 $D_k(t) = \sum_q d_{k,q}(t)$ （比特）。此外，我们将 $\beta_{k,q}(t) = B_{q,\max}\, r_{k,q}(t)$ 建模，其中 $r_{k,q}(t)$ 是分配给用户 k 在质量等级 q 下的分辨率比。该 GoP 瓦片划分设计优先保证视场分辨率，同时降低总体传输与计算成本。

1) 通信模型：大尺度信道增益采用带有对数正态阴影衰落的近距离（CI）路径损耗模型进行建模 [15]。距离 d （以米为单位）和载波频率 f_c （以 Hz 为单位）处的路径损耗表示为：

$$PL(d, f)\ [\text{dB}] = 20 \log_{10}\left(\frac{4\pi f d_0}{c}\right) + 10\alpha \log_{10}\left(\frac{d}{d_0}\right) + X_\sigma, \tag{2}$$

其中， $d_0 = 1$ 米为参考距离， α 为路径损耗指数（PLE），$X_\sigma \sim \mathcal{N}(0,\ \sigma^2)$用于建模对数正态阴影衰落， c为光速。

与基站 i 关联的用户 k 的可实现下行链路（DL）数据速率建模如下：

$$R_k = B_k \log_2\left(1 + \frac{P_k g_k^2 h_k}{N_0 B_k \cdot \text{NF}}\right), \tag{3}$$

其中， B_k 为分配的带宽， P_k 为发射功率， g_k为小尺度瑞利衰落系数， h_k 为线性尺度下的路径损耗， N_0 为热噪声功率谱密度，NF

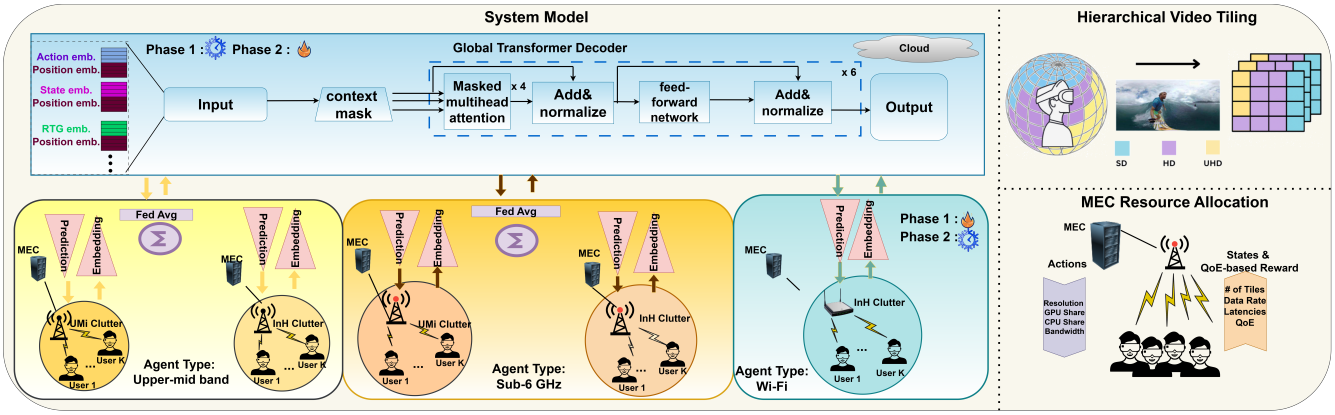

图1：系统模型

为接收机噪声系数（线性尺度）。因此，传输时延为：

$$T_k^{\text{comm}}(t) = \frac{D_k(t)}{C_r R_k}, \tag{4}$$

其中 R_k 为可实现的下行数据速率， C_r 为压缩因子， $D_k(t)$表示用户k 在时隙 t的任务总数据量。

2) 计算模型：在移动边缘计算服务器上，对用户 k 在时隙 t 的每个图像组的处理包括基于GPU的渲染和基于CPU的编码计算任务 [2]。

- **基于GPU的渲染：** 令 γ_{gpu} 表示质量等级 q下以像素/瓦片计的GPU开销， $f_{\text{gpu}\ k}$ 为分配给用户 k的GPU容量（像素/秒）。每个像素包含3个通道（即红绿蓝），共计每像素24比特。则总像素数为

$$P_k(t) = F \sum_q M_{k,q}(t)\, p_{k,q}, \tag{5}$$

其中$p_{k,q}$ 为质量 q下的每瓦片像素数，渲染时延为

$$T_k^{\text{rend}}(t) = \frac{\gamma^{\text{gpu}} P_k(t)}{f_k^{\text{gpu}}}. \tag{6}$$

- **基于CPU的编码：** 令 γ_{cpu} 表示以周期/比特计的CPU开销， $f_{\text{cpu}\ k}$ 为CPU容量（周期/秒）。给定以比特计的总编码数据量 $D_k(t)$，编码时延为

$$T_k^{\text{enc}}(t) = \frac{\gamma^{\text{cpu}} D_k(t)}{f_k^{\text{cpu}}}. \tag{7}$$

因此，用户 k 的端到端时延为

$$T_k^{\text{tot}}(t) = T_k^{\text{rend}}(t) + T_k^{\text{enc}}(t) + T_k^{\text{comm}}(t). \tag{8}$$

3) 体验质量：在 VR 流媒体中，体验质量会受到分块、卡顿、时延和分辨率切换等多种因素的影响。为准确衡量体验质量，我们纳入了三个主要分量，即时延、基于分块的视频质量和分辨率切换。综合考虑这些因素，用户k 在时隙 t的体验质量可表示如下：

$$\text{QoE}_k = \left(1 - \frac{T_k^{\text{tot}}}{T_{\text{th}}}\right) S_k - \lambda R_{\text{amp},k}, \tag{9}$$

$$S_k = \sum_q \frac{w_q M_q}{M_{\text{total}}} \cdot \log\left(1 + \frac{\beta_q}{\beta_{\text{th},q}}\right), \tag{9a}$$

$$R_{\text{amp},k} = \sum_q \left| r_q^{(t)} - r_q^{(t-1)} \right|, \tag{9b}$$

其中，式（9）中的时延反映了网络状况；式（9a）中的视频质量分量通过权重因子$w_q \in \{$（低：1，中：2，高：3$\}$）[5] 体现了传统服务质量指标之外的感知方面；式（9b）中的$R_{\text{amp}、\ k}$ 是分辨率切换幅度，这是无线虚拟现实中衡量用户体验的关键指标，其有效性已通过与平均意见分的相关性在[16]中得到验证。

B. 问题构建与强化学习框架

1) 问题陈述： 本小节说明了所有用户体验质量最大化问题的构建方式。移动边缘计算资源分配问题可表述如下：

$$\max_{\{f_k^s, b_k^s, g_k^s, r_{k,q}\}} \sum_k \text{QoE}_k, \tag{10}$$

$$\text{s.t.} \quad \sum_k f_k^{sh} \le 1, \quad \sum_k b_k^{sh} \le 1, \quad \sum_k g_k^{sh} \le 1, \quad \forall t \tag{10a}$$

$$0.125 < r_{\text{low}} < 0.25 < r_{\text{med}} < 0.5, \quad 0.75 < r_{\text{high}} < 1, \tag{10b}$$

$$\text{QoE}_{k,t} \ge \text{QoE}_{k,\min}, \quad \forall k, t \tag{10c}$$

$$0 \le f_k^{sh},\ b_k^{sh},\ g_k^{sh},\ r_{k,q} \le 1, \tag{10d}$$

其中，在式 (10a) 中， f_k^{sh}、 b_k^{sh} 和 g_k^{sh}分别表示中央处理器、带宽和图形处理器的分配比例；式 (10d) 通过将所有分配变量和速率变量限制在单位区间内，确保中央处理器、图形处理器和带宽分配有效。此外，式 (10b) 设定了质量限制 [5]，式 (10c) 则对用户级体验质量约束违反进行惩罚。

2) MDP 表述：强化学习智能体的马尔可夫决策过程（MDP）定义如下：

a) 状态空间：在时刻 t，针对 K 个用户，系统状态捕获每个用户的下行链路数据速率、时延组成部分、图块分布和体验质量：

$$\mathbf{s}_t = \left[R_k, T_k^{\mathrm{c}}, T_k^{\mathrm{e}}, T_k^{\mathrm{r}}, T_k^{\mathrm{tot}}, \{M_{k,q}\}_{q\in\{\mathrm{high,med,low}\}}, \mathrm{QoE}_k\right]_{k=1}^{K}, \tag{11}$$

其中，R_k 是根据式 (3) 得出的下行链路传输速率，$T_k^{(\cdot)}$ 分别表示下行链路通信、编码、渲染和总延迟，$M_{k,\ q}$ 表示分配给质量等级 q 的图块数量。

b) 动作空间：在时刻 t，系统选择一个动作，该动作共同决定计算资源（f_k^{sh}，g_k^{sh}）和通信资源（b_k^{sh}）的分配，以及所有 K 用户的分辨率比例（$r_{k,\ q}$）：

$$\mathbf{a}_t = \left[f_k^{\mathrm{sh}}, b_k^{\mathrm{sh}}, g_k^{\mathrm{sh}}, \{r_{k,q}\}_{q\in\{\mathrm{high,med,low}\}}\right]_{k=1}^{K}, \tag{12}$$

例如，对于 4 个用户，动作空间和状态空间的维度分别为 24 和 36。

c) 奖励函数：奖励函数在用户体验质量、公平性和约束违规之间进行权衡：

$$r_t = \sum_{k=1}^{K} \mathrm{QoE}_{k,t} - \omega_{\mathrm{fair}}\mathcal{F}_t - \omega_{\mathrm{viol}}\mathcal{V}_t, \tag{13}$$

其中，$\mathcal{F}_t = \sigma\left(\{f_{k,t}\}_{k=1}^{K}\right) + \sigma\left(\{b_{k,t}\}_{k=1}^{K}\right) + \sigma\left(\{g_{k,t}\}_{k=1}^{K}\right)$ 表示基于用户间资源分配方差的公平性惩罚，$\mathcal{V}_t$ 表示时刻 t 的体验质量阈值违规，w_{fair} 和 w_{viol} 表示系统对公平性和违规的敏感程度。

III. 基于 FSDT 的元宇宙资源分配

A. FSDT

FSDT 的架构建立在决策变换器范式之上，并经过重新设计以适用于联邦环境 [14]。决策变换器将强化学习建模为序列建模问题，通过自回归建模来学习决策，而非将其视为显式的价值函数或策略优化任务，可表示如下。对于用户 k，上下文马尔可夫决策过程定义为 $(\mathcal{S}_k,\ \mathcal{A}_k,\ \mathcal{P}_k,\ \mathcal{R}_k)$，其中状态为 $s_t^k \in \mathcal{S}_k$，动作是 $a_t^k \in \mathcal{A}_k$，奖励为 $r_t^k = \mathcal{R}_k(s_t^k,\ a_t^k)$，而 $\mathcal{P}(s'|s,\ a)$ 表示状态转移动态。该模型使用无折扣回报目标（RTG），即未来奖励之和，$\hat{R}_t^k = \sum_{t'=t}^{T} r_{t'}^k$，以实现自回归轨迹建模。由此得到如下轨迹表示，可有效用于自回归训练和序列生成：

$$\tau^k = \bigcup_{t=1}^{T}\left(\hat{R}_t^k,\ s_t^k,\ a_t^k\right), \tag{14}$$

其中 T 表示轨迹长度。请注意，离线强化学习的目标是在马尔可夫决策过程中学习一个能够最大化期望回报 $\mathbb{E}\left[\sum_{t=1}^{T} r_t\right]$ 的策略。然后，嵌入层将这些回报目标、状态和动作映射到隐空间，并通过时间嵌入（即位置嵌入）$\pi(t)$加以增强：

$$u_{x,t}^{k_n} := \pi(t) \oplus \varphi_x(x_t^{k_n}), \quad x \in \{r, s, a\}. \tag{15}$$

这些嵌入经过堆叠的 掩码因果自注意力模块处理，如图 1 所示，其中解码器架构与 GPT 架构 [6]高度相似。云端的解码器生成上下文化的词元输出：

$$(v_{r,t}^{k_n},\ v_{s,t}^{k_n},\ v_{a,t}^{k_n}) = G\left(\{u_{r,\tau}^{k_n},\ u_{s,\tau}^{k_n},\ u_{a,\tau}^{k_n}\}_{\tau=t-L}^{t}\right). \tag{16}$$

其中 L 是上下文长度（即用于预测下一个动作的 RTG-状态-动作元组数量）。预测模块P_{k_n} 使用 Sigmoid 激活函数将上下文表示映射为下一个动作。然后，将输出$\hat{a}_t^{k_n}$缩放至有效动作范围。在此 DT 公式的基础上，FSDT 将该方法扩展到了联邦环境。

B. 训练过程

联邦学习和分割学习范式是通信系统中最常见的分布式学习范式之一。在这些方法中，每个客户端（例如 MEC服务器）使用自己的数据集构建本地模型。客户端仅向中央聚合器（例如云端）发送模型更新（例如梯度或嵌入），由中央聚合器将其合并为全局模型，再将其共享给客户端以进一步优化。

算法 1 VR-FSDT 训练

输入： 域全局模型 $\{g_0^k\}_{k=1}^K$，服务器模型 v_0，客户端本地模型 $w^k{}_0{}^n$，通信轮次 R。

输出： 最终 域全局模型 $\{G_R^k\}_{k=1}^K$，服务器 模型 v_R

对于 $r = 1$ 至 R **执行**
  /* 阶段 1：仅更新嵌入-预测 */
  对于 $k = 1$ 至 K **执行**
    对于 $n = 1$ 至 N_k **执行**
      初始化 $w^{k,\ r}{}^n \leftarrow g_{r-1}^k$
      $w^{k,r}{}^n \leftarrow$ 本地更新$(w^{k,\ r}{}^n, v_{r-1})$
    结束循环
    $\bar{w}_r^k \leftarrow \mathrm{FedAvg}(w^{k,\ r}{}^1,\ \ldots\ldots,\ w^{k,\ r}{}^{N_k})$
    $G_r^k \leftarrow$ 域更新$(\bar{w}_r^k, v_{r-1})$
  结束循环
  $v_r \leftarrow v_{r-1}$
  /* 阶段 2：仅更新服务器 Transformer */
  对于 $k = 1$ 到 K **执行**
    对于 $n = 1$ 到 N_k **执行**
      $v_r \leftarrow$ 训练服务器$(G_r^k,\ v_r)$
    结束循环
  结束循环
结束循环

FSDT 的训练在 R 个通信轮次中分两个交替阶段进行（见算法 1），并遵循联邦学习和分割学习的原则。在每一轮中，客户端首先调整其本地模块，以适应自身数据，同时保持服务器端 Transformer 解码器固定（第 2–10 行）。随后，服务器使用所有本地数据训练共享的 Transformer 解码器（第 11–18 行）。目标函数是最小化 DDPG 轨迹数据集中的预测动作与真实动作之间的均方误差（MSE）。

这种设计平衡了本地适应能力（阶段 1）与全局协调能力（阶段 2）。

阶段 1：在第 t 轮次，服务器将最新的域全局参数发送给类型为 k 的所有客户端。每个客户端 $n \in\{1,\ldots,N_k\}$ 更新其本地嵌入和预测模块（$e^{k}{}_{,t}{}^{n}, p^{k}{}_{,t}{}^{n}$），同时保持服务器 Transformer G_t 固定：

$$\min_{\{e_t^{k,n},\, p_t^{k,n}\}} \sum_{k=1}^{K}\sum_{n=1}^{N_k} \mathrm{MSE}\Big(y_t^{k,n},\, p_t^{k,n}\big(G_t(e_t^{k,n}(x_t^{k,n}))\big)\Big). \tag{17}$$

随后，对客户端更新进行聚合，即联邦平均（FedAvg），并在每种智能体类型内进行（例如，在 UMB 域的各杂波之间）：

$$e_t^k = \frac{1}{N_k}\sum_{n=1}^{N_k} e_t^{k,n}, \quad p_t^k = \frac{1}{N_k}\sum_{n=1}^{N_k} p_t^{k,n}. \tag{18}$$

阶段 2：在（e_t^k, p_t^k）冻结的情况下，服务器利用来自所有客户端类型的嵌入更新其 Transformer G_t，并最小化与式 17 相同的目标函数，但仅针对 G_t。单个云端共享解码器在多接入边缘计算服务器之间联合训练，而嵌入层和预测层保持域特定。训练后复制该解码器，并与域特定层组合，从而得到 UMB、6 GHz 以下频段和 WiFi 模型。

IV. 仿真场景与结果

A. 实验设置与基线：

在我们的仿真中，我们考虑了五种采用多无线接入技术设置的异构 MEC 环境，每种环境服务于四个用户。用户位置随机分布在一个方形区域内，其边长对应于表 I 中规定的最大距离。该表还报告了关键的杂波参数，包括带宽、载波频率f_c、路径损耗指数（α）、阴影衰落方差 σ，以及 P 在非视距条件下的 [17]。此外，数据集取自[18]，其中包含 200+ 个 YouTube 视频（4K、25fps、20–60 秒）。数据集按时间划分为 80% 的训练集和 20% 的测试集。每帧的二维眼动坐标通过切比雪夫距离确定瓦片质量等级（图 1）。以下各小节中的所有结果均为 3 个随机种子下的平均值

表 I：多无线接入技术参数

无线接入技术/杂波	**BW [MHz]**	f_c **[GHz]**	**PLE** α	σ **[dB]**	**距离 [m]**	P **[dBm]**
UMB/UMi	200	6.75	2.56	6.53	230	33
UMB/InH	200	6.75	2.72	9.21	65	33
Sub6GHz/UMi	100	2.90	2.90	2.90	230	30
Sub6GHz/InH	100	2.90	3.10	6.50	65	30
Wi-Fi/InH	160	2.40	2.52	5.75	65	24

1）基线方法：我们评估了五种算法，包括表 II 所示的四种基线方法和 FSDT，其中前缀 C- 和 F- 分别表示集中式和联邦式。离线方法（即 C-DT、F-DT 和 FSDT）利用使用 C-DDPG 生成的预收集轨迹。FSDT 训练 20 轮，每轮包括 100 个客户端梯度时间步（阶段 1），随后是 100 个服务器梯度时间步（阶段 2）。测试时，所有方法均在同一测试数据集上测试 100 个回合。

表 II：基线方法与所提出的 FSDT 的比较。

算法	**训练。模式 聚合 交换信息。**		
C-DDPG [5]	集中式	N/A	N/A
F-DDPG	联邦式	所有层	评论家参数
C-DT [6]	集中式	N/A	N/A
F-DT	联邦式	所有层	完整参数
FSDT	联邦拆分	本地层	嵌入

表 III：FSDT 和 MEC 参数。

FSDT 参数	**值**	**MEC 参数**	**值**
n 注意力头_	4	γ_q^{cpu} (q=1,2,3)	200, 190, 180
n_模块	6	F, M	16, 16
上下文_长度	50	T_{th} （毫秒）， $QoE_{\min}$	200, 4
优化器	AdamW	C_r, NF (分贝)	300, 7
学习_率	10^{-4}	f^{CPU} (周期/秒),f^{GPU} (像素/秒)	15×10^9
权重_衰减	10^{-4}	w_{fair}, w_{viol}	2, 2
隐藏层维度_	256	$\beta_{max,\ q}$ (M比特)	12.4, 1.4, 0.5

B. 结果

在奖励方面，图2展示了FSDT与基线方法相比的移动平均奖励，其中给出了在线强化学习算法的训练结果以及所有算法的测试结果。结果表明，通过结合离线强化学习、联邦学习和分割学习，FSDT始终优于基线方法，奖励最高提升了10%。这些范式相辅相成，分别带来了关键优势，尤其是通过策略性采样改善学习效果、缓解本地移动边缘计算数据匮乏的问题，以及通过模型分区克服朴素全局聚合的局限性。F-DDPG和C-DDPG在训练和测试期间的表现相当相似。与F-DT相比，FSDT和C-DT在不同随机种子下的表现波动更小，性能分别提高了近10%和5%。

图 3 和图 4 展示了所有多接入边缘计算服务器和用户的体验质量与时延分布。箱线图显示了中位数、四分位距、总体分布范围和均值（菱形）。在体验质量方面，图 3 表明，大多数方案的体验质量中位数介于 5.2 和 5.5 之间。所提出的 FSDT 方法表现最佳，其中位体验质量约为 5.6，上须线达到 6.07，明显优于所有基线方法。相比之下，F-DT 的体验质量中位数最低（例如 5.1），且波动范围更大，表明其性能稳定性较差。此外，C-DDPG 与 F-DDPG 的比较表明，在异构环境下，传统联邦学习并不能提升在线强化学习，因为更新浅层评论家网络对协作学习的帮助有限。此外，比较 C-DT 和 F-DT 也可见相同趋势：传统联邦学习（例如 FedAvg）并未改善体验质量。相比之下，FSDT 改进了 C-DT，因为每个多接入边缘计算服务器既通过 FedAvg 对嵌入层和预测层进行聚合，从域内多接入边缘计算服务器中学习，也通过共享的 Transformer 解码器从域间多接入边缘计算服务器中学习。

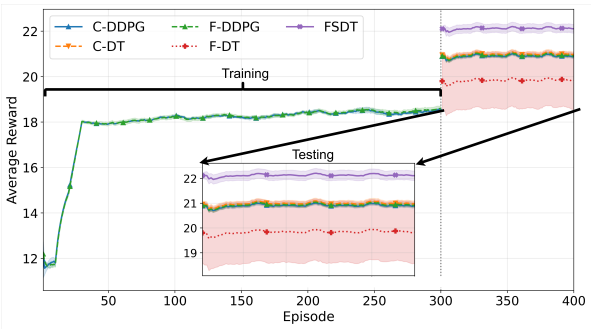

图2：基线方法与FSDT的平均奖励随回合的变化

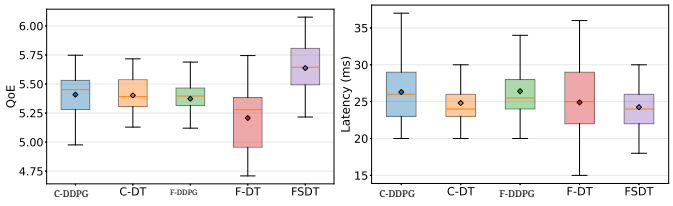

图3：体验质量对比 图4：时延对比

在时延方面，图 4 展示了时延情况，所有方案的平均时延均在 20–30 毫秒范围内。C-DT 和 FSDT 的中位时延最低，约为 22–23 毫秒，其中 FSDT 的分布范围更窄，表明其性能更加稳定。C-DDPG 和 F-DDPG 的中位时延较高，约为 26 毫秒，且须线范围更宽。

C. 复杂度分析

表 IV 总结了 FSDT 模型的子网络。Transformer 解码器占据了 98.79% 的参数，使计算保留在云端，且无需在多接入边缘计算服务器之间频繁交换。该设计通过仅共享动作、状态和回报目标的潜在表示，降低了移动边缘计算服务器负载和训练时间，并增强了隐私保护。传统联邦学习和 FSDT 的训练通信成本可表示为

$$T_{FL} = 2RP_{tot}, \quad T_{FSDT} = 2BR(F+G) + 2R(P_e + P_p), \quad (19)$$

其中，P_{tot}、P_e 和 P_p 分别表示参数总数、嵌入参数数和预测参数数，B 为批次大小（例如 64），F 和 G 分别表示在分割点 [13] 传输的特征和梯度。根据模型超参数的不同，该架构可以减少多接入边缘计算服务器之间的总体通信量。

表 IV：FSDT 分割详情

指标	解码器（云端）	嵌入（边缘端）	预测（边缘端）
参数数量	4,738,560	42,496	15,677
大小 (MB)	18.076	0.162	0.060
内存 (%)	98.79	0.89	0.33

V. 结论

总之，基于 FSDT 的资源分配框架面向异构多无线接入技术环境下的边云协同学习，通过结合离线强化学习、联邦学习和分割学习，展现出卓越的性能。与未拆分模型的基线相比，该框架的结果更为稳定，令本地层和共享全局Transformer解码器均受益。FSDT 不仅实现了更高的体验质量，时延最多降低 10%，还通过将计算负担留在云端提高了训练效率，并增强了隐私保护。未来的工作可探索边缘设备协同或任务无关模型的开发。

致谢

本研究得到了加拿大自然科学与工程研究理事会加拿大研究讲席计划的支持。我们衷心感谢我们研究实验室最近的校友、已故的Han Zhang博士帮助我们获取本研究中使用的数据集。我们将本文献给她，以寄托深切的哀思。

参考文献

[1] X. Liu 与 Y. Deng，“面向 MEC 赋能无线虚拟现实（VR）网络的基于学习的预测、渲染与关联优化”，*IEEE TWC*，第 20 卷，第 10 期，第 6356–6370 页，2021。[2] N. Liu、T. H. Luan、Y. Wang、Y. Liu 与 Z. Su，“元宇宙中面向体验质量的协同 VR 渲染与动态资源租赁”，*IEEETMC*，第 1–16 页，2025。[3] S. Yang、Y. He 与 X. Zheng，“FoVR：带宽受限无线网络中基于注意力的 VR 流媒体”，载于 *2019* 年第 *16* 届 *IEEE SECON* 年会，2019，第 1–9 页。[4] T. Zhou 等，“面向移动边缘计算中定制化 VR 流媒体资源分配的联邦式基于提示的决策 Transformer”，*IEEE TWC*，第 1–1 页，2025。[5] J. Yu、A. Y. Alhilal、T. Zhou、P. Hui 与 D. H. K. Tsang，“面向沉浸式虚拟现实的基于注意力且感知体验质量的数字孪生赋能边缘计算”，*IEEE TWC*，第 23 卷，第 9 期，第 11 276–11 290 页，2024。[6] L. Chen 等，“决策 Transformer：经由序列建模的强化学习”，载于 *NeurIPS*，2021。[7] M. A. Habib、P. E. Iturria-Rivera、Y. Ozcan、M. Elsayed、M. Bavand、R. Gaigalas 与 M. Erol-Kantarci，“利用大语言模型、Informer 与决策 Transformer 的能力实现 6G 中意图驱动的 RAN 管理”，*IEEE TNSE*，第 1–20 页，2025。[8] J. Zhang、J. Li、Z. Wang、L. Shi、S. Jin、W. Chen 与 H. V. Poor，“面向无线通信的决策 Transformer：资源管理的新范式”，*IEEE Wirel. Commun.*，第 32 卷，第 2 期，第 180–186 页，2025。[9] H. Zhang、H. Zhou、M. Elsayed、M. Bavand、R. Gaigalas、Y. Ozcan 与 M. Erol-Kantarci，“针对联邦强化学习赋能小区休眠控制的分布式攻击”，载于*2023 IEEE Globecom* 研讨会，2023，第 485–490 页。[10] R. Mudi 与 H. Elbiaze，“面向元宇宙的 O-RAN 切片中基于联邦强化学习的资源分配”，载于 *IEEE ICC 2025*，2025，01 07页 – 。[11] Z. Lin、G. Qu、X. Chen 与 K. Huang，“6G 边缘网络中的分割学习”，*IEEE Wirel. Comm.*，第 31 卷，第 4 期，第 170–176 页，2024。[12] Z. Li、T. Li、H. Liu 与 T.-T. Chan，“面向异构边缘内容缓存网络的个性化联邦深度强化学习”，载于 *2024* 年第 *22* 届 *WiOpt*，2024，第 313–320 页。[13] S. Park、G. Kim、J. Kim、B. Kim 与 J. C. Ye，“用于 COVID-19 胸部 X 光诊断的联邦式分割任务无关视觉 Transformer”，载于 *NeurIPS*，第 34 卷，2021，第 24 617–24 630 页。[14] Z. Wang、B. Chen、X. Qu、Z. Hong、J. Xiao 与 J. Wang，“结合联邦分割训练的面向多类型智能体控制的任务无关决策 Transformer”，载于 *2024 IJCNN*，2024，第 1–7 页。[15] S. Sun 等，“5G 无线通信大规模传播路径损耗模型的预测精度、灵敏度与参数稳定性研究”，*IEEE TVT*，第 65 卷，2860，2016 5，2843 期 页 – 。[16] J. Li、R. Feng、Z. Liu、W. Sun 与 Q. Li，“无线网络上虚拟现实视频传输的体验质量建模”，载于 *2018 IEEE GLOBECOM*，2018，第 1–7 页。[17] A. Bazzi、R. Bomfin、M. Mezzavilla、S. Rangan、T. S. Rappaport 与 M. Chafii，“面向 6G 的上中频段频谱：愿景、机遇与挑战”，*IEEE Comm. Mag.*，2025。[18] Y. Xu 等，“动态 360° 沉浸式视频中的注视预测”，载于 *2018 IEEE/CVF* 计算机视觉与模式识别会议，2018，第 5333–5342 页。