2023 IEEE International Conference on Metaverse Computing, Networking and Applications (MetaCom) | 979-8-3503-3333-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/MetaCom57706.2023.00106

Generative AI-empowered Effective Physical-Virtual Synchronization in the Vehicular Metaverse

Minrui Xu[1], Dusit Niyato[1], Hongliang Zhang[2], Jiawen Kang[3], Zehui Xiong[4], Shiwen Mao[5], and Zhu Han[6,7]
[1]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[2]School of Electronics, Peking University, Beijing 100871, China
[3]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[4]Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372, Singapore
[5]Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
[6]Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004, USA
[7]Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea
Email: minrui001@e.ntu.edu.sg, dniyato@ntu.edu.sg, hongliang.zhang@pku.edu.cn, kavinkang@gdut.edu.cn,
zehui_xiong@sutd.edu.sg, smao@ieee.org, hanzhu22@gmail.com

***Abstract*—Metaverse seamlessly blends the physical world and virtual spaces through ubiquitous communication and computing equipment and infrastructure. In intelligent transportation systems, the vehicular Metaverse can provide a fully immersive and hyperreal travel experience (e.g., via augmented reality head-up displays, AR-HUDs) to drivers and passengers in autonomous vehicles (AVs) through roadside units (RSUs). However, providing real-time and immersive services requires effective physical-virtual synchronization between AVs and virtual simulators. This paper proposes a generative AI-empowered physical-virtual synchronization framework for the vehicular Metaverse. In physical-to-virtual synchronization, digital twin (DT) tasks generated by AVs are offloaded for execution in RSUs with future route generation. In virtual-to-physical synchronization, virtual simulators customize diverse and personalized AR content via generative AI models based on user preferences. Furthermore, we propose a multi-task enhanced auction-based mechanism to match and price AVs and virtual simulators for RSUs to provide real-time and effective services. Finally, property analysis and experimental results demonstrate that the proposed mechanism is strategy-proof and adverse-selection free while increasing social surplus.**

***Index Terms*—Vehicular Metaverse, generative artificial intelligence, digital twin, augmented reality, auction theory.**

I. Introduction

As a long-term vision, the Metaverse is an evolution of the mobile Internet towards the advanced three-dimensional visualization stage of digital transformation [1]. By blending physical transportation systems with 3D virtual spaces via multi-dimensional and multi-sensory communications, the vehicular Metaverse can extend the physical space of vehicles via real-time physical-virtual synchronization [2]. For instance, autonomous vehicles (AVs) with large windshields and side windows provide the most convenient and promising interface for users to synchronize and interact with avatars and other objects in virtual space. In physical-to-virtual (P2V) synchronization, vehicles can connect with the digital twin (DT) in virtual space by continuously executing DT tasks [3]. In virtual-to-physical (V2P) synchronization, vehicles can install the windshield and side windows with augmented reality (AR) head-up displays (HUDs), which can blend and display 3D virtual content on AR-HUDs with realistic street views. However, achieving high synchronization accuracy while providing real-time and immersive services with effective physical-virtual synchronization in the vehicular Metaverse is challenging.

On the one hand, to effectively synchronize digital twins and avatars in virtual space, AVs continuously generate computation-intensive DT tasks to synchronize with the virtual space, i.e., the P2V synchronization. However, the local computation resources of AVs might be insufficient to execute these tasks and update the results to RSUs [4]. Therefore, AVs prefer to offload these tasks to RSUs with large-scale computing and communication infrastructure for real-time execution. In addition, RSUs can utilize the information in AVs' DTs to assist with service provisioning. For example, AI models that analyze past routes and current locations in AVs' DTs can predict the future routes of vehicles. This way, the accuracy of location-based services can be improved during the physical-virtual synchronization in the Metaverse.

On the other hand, based on the preferences of drivers and passengers in AVs, Metaverse virtual simulators provide personalized services with AR content on HUDs, for effective V2P synchronization. However, high-quality AR content is scarce due to the intensive computing and time cost of the content creation process, which leads to low match qualities between AVs and virtual simulators [3]. Fortunately, generative AI, with effective and efficient inference and information creation capabilities, allows for diverse AR content customization. Specifically, based on user preferences in AVs' DTs, RSUs can provide AI-generated content (AIGC) related to the original subjects of virtual simulators with generative AI models [5]–[7]. This way, virtual simulators can provide diverse and scalable AR content to AVs through generative rendering and streaming at RSUs.

As shown in Fig. 1, in this paper, we propose a novel generative AI-empowered physical-virtual synchronization frame-

979-8-3503-3333-6/23/$31.00 ©2023 IEEE
DOI 10.1109/MetaCom57706.2023.00106

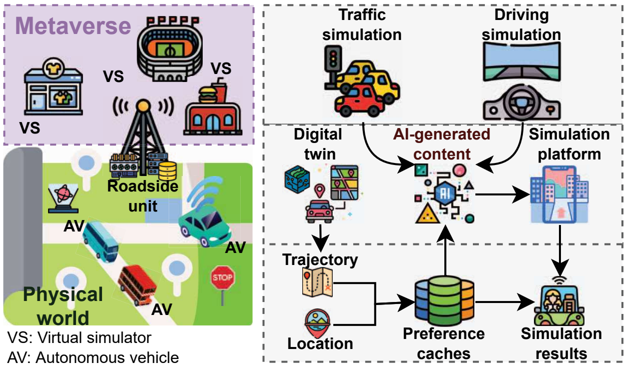


Fig. 1: The generative AI-empowered vehicular Metaverse.

work where generative AI is leveraged to create personalized AR content. In this framework, we design the P2V synchronization that AVs maintain and continuously update the DTs by offloading DT tasks to RSUs for execution. To improve synchronization accuracy, RSUs can predict the future route of AVs for effective location-based synchronization services based on the current location and historical routes of AVs. Moreover, based on the future route and user preferences, RSUs can customize the AR content of virtual simulators by creating diverse content via the generative AI model named TSDreamBooth, which is fine-tuned using the Belgium traffic sign (BelgiumTS) dataset [8]. Finally, we propose a multi-task enhanced auction-based mechanism to satisfy multi-dimensional requirements (e.g., deadlines and quality) of multiple DT tasks simultaneously.

Our main contributions are summarized as follows:

- In the vehicular Metaverse, we propose a novel AI-native physical-virtual synchronization framework, which consists of DT-assisted route generation and preference-aware AR content generation.
- We propose the TSDreamBooth to empower Metaverse virtual simulators to customize diverse simulated AR content based on the future routes and user preferences of AVs.
- To incentivize RSUs for provisioning communication and computing resources, an enhanced auction-based mechanism is proposed to maximize social surplus during synchronization while guaranteeing fully strategy-proof and adverse-selection free of participants.

II. Generative AI-empowered System Model

In the system model, we consider three main roles in the vehicular Metaverse, i.e., AVs, RSUs, and virtual simulators. The set of AVs is represented by the set $\mathcal{I} = 1, \ldots, i, \ldots, I$, the set of RSUs is represented as $\mathcal{J} = \{1, \ldots, j, \ldots, J\}$, and the set of virtual simulators is represented as $\mathcal{K} = \{0, 1, \ldots, k, \ldots, K\}$. We consider the RSUs to possess the communication and computing resources in the system. To facilitate physical-virtual synchronization, uplink and downlink channels are allocated to uploading DT tasks and streaming AR content. Therefore, communication resources at RSUs consist of uplink bandwidth B_j^u and downlink bandwidth B_j^d. Moreover, to provide services such as executing DT tasks and rendering augmented reality layers, each RSU j is equipped with computing resources, including the CPU frequency f_j^C and the GPU frequency f_j^G.

During the synchronization, N DT tasks generated by each vehicle can be represented as $DT_i = (<s_{i,1}^{DT}, e_{i,1}^{DT}, d_{i,1}>, \ldots, <s_{i,N}^{DT}, e_{i,N}^{DT}, d_{i,N}>)$, where $s_{i,n}^{DT}$ is the size of DT data, $e_{i,n}^{DT}$ represents the number of CPU cycles required per unit data, and $d_{i,n}$ denotes the deadline for completing the task. The size of preference caches of AV i within the DT is C_i. Each vehicle $i \in \mathcal{I}$ has its private value v_i for executing its DT task DT_i, drawn from the probability distributions. The values of DT tasks can be interpreted as the characteristics of the autonomous vehicles, such as the level of urgency to align with DT models [9], which may vary for each vehicle during its travel.

Similar to the Internet display advertising [10], we consider two types of virtual simulators in the system, i.e., virtual driving simulators and virtual traffic simulators. Virtual traffic simulators $1, \ldots, K$ provide AR content designed to elicit real-time feedback from drivers and passengers, such as providing information about sales or promotions at nearby shops. The virtual driving simulator 0 delivers driving simulations for improving the performance of autonomous driving. The value of personalized AR content for each synchronizing pair of AV i and virtual traffic simulator k is $U_{i,k}$, which is the product of the common value v_i of AV i and the match quality $m_{i,k}$, i.e., $U_{i,k} = v_i m_{i,k}$. The common values for virtual traffic simulator k are gained from the provisioning of real-time traffic simulations for the synchronizing AV i, which can be represented by the AV i's private value v_i. Additionally, the amount of personalized information determines the match quality $m_{i,k}$ of virtual traffic simulator k. The values of AVs and virtual simulators in synchronizing pairs are positively correlated. Finally, let $U_{\iota,(l)}$ and $m_{\iota,(l)}$ represent the l highest value and match quality for the synchronizing AV ι, respectively.

A. *Multi-task Digital Twin Model*

To synchronize with the vehicular Metaverse, physical entities, i.e., AVs, generate and offload DT synchronizing demands, i.e., DT model updates, to RSUs for real-time execution. Therefore, we consider the demands as tasks that are required to be accomplished by RSUs. The transmission delay for AV i to upload its DT task DT_i to RSU j can be calculated as [9] $t_{i,n,j}^{DT} = \frac{s_{i,n}^{DT}}{R_{i,j}^u}$, where $R_{i,j}^u$ is the uplink transmission rate. The computation delay in processing the DT task DT_i of AV i for RSU j can be calculated as $l_{i,n,j}^{DT} = \frac{s_{i,n}^{DT} e_{i,n}^{DT}}{f_j^C}$. In the proposed system, without loss of generality, we consider that each RSU can accomplish both computing and transmission requirements of DT tasks, i.e., $t_{i,n,j}^{DT} + l_{i,n,j}^{DT} \leqslant d_{i,n}, \forall i \in \mathcal{I}, j \in \mathcal{J}, n = 1, \ldots, N$. With available communication and computing resources, RSUs can provide AR rendering services for virtual simulators. This way, virtual simulators can send their AR content to AVs, i.e., synchronizing from virtual spaces to the physical world.

B. AR Simulation Model

1) Generative AI-based AR Simulation Customization: The process for customizing AR simulation using generative AI is divided into two stages: fine-tuning and inference. To fine-tune a text-to-image diffusion model using AR simulations for a specific subject from virtual simulators, we follow a two-step process. In the first step, the low-resolution text-to-image model is first fine-tuned with input images paired with a text prompt that includes a unique identifier. At the same time, a class-specific prior preservation loss is incorporated to ensure that diverse instances belonging to the subject's class are generated while preserving the model's prior knowledge about the class. In the second step, the super-resolution components are then fine-tuned using pairs of low-resolution and high-resolution images taken from the set of input images, allowing high accuracy in small details of the subject to be maintained. During the fine-tuning of generative AI, virtual simulators input their original AR simulations as training data to train models. Based on the knowledge of AR simulations, e.g., a class of traffic signs, the fine-tuned generative AI model can extract features of these traffic signs for customization.

Therefore, the provisioning of AR simulations is no longer limited to the hit preference caches $h_{i,k}$ [3]. However, due to the limitation of generative AI models, some customized AR content might not be satisfactory, which can be identified by the trained validation models.

The validation models indicate the quality of generative AI models with generative score $G_{i,j,k} \in [0,1]$. For each AR layer of virtual traffic simulator k, the rendering task can be represented by $AR_k = {<}s_k^{AR}, e_k^{AR}{>}$ [11], where s_k^{AR} is the data size of each AR layer and e_k^{AR} is the required GPU cycles per unit data for rendering. Therefore, given the total number of virtual simulators $K+1$, the match quality $m_{i,k}$ and hit preference caches $h_{i,k}$ are drawn independently from a set of distributions $m_{i,k} = h_{i,k} \sim F_{i,k}$. To explain further, given the synchronizing AV ι, virtual traffic simulators $k = 1, \ldots, K$ can measure the match qualities $m_{\iota,k}$ of their traffic simulations. However, the virtual driving simulator 0 that provides AR simulations to the synchronizing AV ι cannot immediately measure its match quality $m_{\iota,0}$. Therefore, asymmetric information exists among virtual simulators that might result in adverse selection [10].

Empowered by generative AI models, the match quality $m_{i,k}$ is no longer limited by the hit preference caches $h_{i,k}$. As generative AI can generate countless and diverse AR content based on user preferences and location datasets, virtual simulators can utilize more computing resources and downlink transmission resources during offline training. During the remaining time of DT execution, the total amount of simulations $Q_{i,n,j,k}$ can be calculated as $Q_{i,n,j,k} = (d_{i,n} - T^{DT}_{i,n,j})R^{AR}_{i,j}/s^{AR}_k$ for task n in DT_i of AV i and its RSU j. Then, the marginal generative AI-empowered match quality of AV i in simulator k via RSU j can be measured as

$$m_{i,n,j,k} = \frac{\log_2(1 + G_{i,j,k}Q_{i,n,j,k})h_{i,k}}{\theta(h_{i,k})}, \qquad (1)$$

where $\theta(h_{i,k})$ is the relative accuracy among the original model w_i and the fine-tuned model $w_{i,k}$ for strongly convex objectives [12]. In particular, $\theta(\cdot) = 1$ indicates no improvement for training in simulation platforms, and $\theta(\cdot) = 0$ indicates the AI model is trained optimally.

2) AR Simulation Rendering: The effective transmission latency of transmitting the AR content AR_k to AV i for task n from RSU j can be calculated as

$$t^{AR}_{i,j,k} = \frac{Q_{i,n,j,k}s^{AR}_k}{R^d_{i,j}}, \qquad (2)$$

where $R^d_{i,j}$ is the downlink transmission rate between AV i and RSU j. Moreover, the effective computation latency in completing the simulation AR_k can be calculated as

$$l^{AR}_{i,j,k} = \frac{Q_{i,n,j,k}s^{AR}_k e^{AR}_k}{f^G_j}, \qquad (3)$$

which depends on the simulation latency in the GPUs of RSU j. Eqs. (2) and (3) imply that the V2P synchronization in generative AI-empowered vehicular Metaverse can fully utilize communication and computing resources.

In the synchronization system, RSUs can use their available computation and communication resources to provide real-time physical-virtual synchronization services for AVs and virtual simulators. However, the total synchronization delay cannot exceed the required deadline of AV i. Let $g^{DT}_{i,j}$ be the allocation variable that AV i is allocated to RSU j and $g^{AR}_{i,j,k}$ be the allocation variable that virtual traffic simulator k is allocated by RSU j to match AV i. The total synchronization delay $T^{total}_{i,j,k}$ required by RSU j to process both the DT task of AV i and the AR rendering the task of virtual traffic simulator k should be less than the required deadline, which can be expressed as

$$\begin{aligned} T^{total}_{i,n,j,k} = {} & g^{DT}_{i,j} \cdot (t^{DT}_{i,n,j} + l^{DT}_{i,n,j}) \\ & + g^{AR}_{i,j,k} \cdot (t^{AR}_{i,n,j,k} + l^{AR}_{i,n,j,k}) \leqslant d_{i,n}, \end{aligned} \qquad (4)$$

$\forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}, n = 1, \ldots, N$. The AR content of virtual traffic simulator k is displayed on AR-HUD of AV i during the processing of DT tasks at RSU j, and thus the expected utilizing duration of AR content can be represented by $T^{total}_{i,j,k}$.

III. Surplus Maximization

In the proposed system, a synchronization market, consisting of the physical and the virtual submarkets, is established to incentivize RSUs to provide communication and computing resources for synchronization between AVs and virtual simulators. Here, we consider physical and virtual entities in the market to be risk neutral, and their surpluses are correlated positively. Therefore, the synchronization mechanism is expected to map the DT values $\mathbf{v} = (v_1, \ldots, v_I)$ and AR values $\mathbf{U} = (I_{1,0}, \ldots, U_{I,K})$ to the payments of AVs

$\mathbf{p}^{DT} = (p_1^{DT}, \ldots, p_I^{DT})$ and the payments of virtual simulators $\mathbf{p}^{AR} = (p_1^{AR}, \ldots, p_K^{AR})$ with the allocation probabilities $\mathbf{g}^{DT} = (g_1^{DT}, \ldots, g_I^{DT})$ and $\mathbf{g}^{AR} = (g_0^{AR}, \ldots, g_K^{AR})$. By accomplishing DT tasks, the total expected surplus for RSUs from AV $i \in \mathcal{I}$ in the physical submarket can be represented by $S^{DT}(\mathbf{g}^{DT}) = \mathbb{E}\left[\sum_{i=1}^{I} v_i \mathbf{g}_{i,j}^{DT}(\mathbf{v})\right]$. Based on the optimal reaction to the dominant strategies of the virtual traffic simulators, the virtual driving simulator can motivate RSU with the expected surplus of $S_F^{AR} = \mathbb{E}[U_{i,0} g_{i,j,0}^{AR}(Q_i)]$. In addition, the total expected surplus provided by virtual traffic simulators is defined by $S_I^{AR}(\mathbf{g}^{AR}) = \mathbb{E}[\sum_{k=1}^{K} U_{i,k} g_{i,j,k}^{AR}(U_i)]$. In conclusion, the total surplus that RSU j can gained from the virtual submarket can be defined as $S^{AR}(\mathbf{z}^{AR}) = \gamma S_F^{AR}(\mathbf{z}^{AR}) + S_P^{AR}(\mathbf{z}^{AR})$, where γ denotes the relative bargaining power of virtual driving simulator 0.

Then, the objective is to maximize the total surplus in the synchronization market, which can be represented as

$$S^{DT} + \sum_{n=1}^{N} T_{i,n,j,k}^{total} \cdot \left(\gamma S_F^{AR} + S_I^{AR}\right). \tag{5}$$

IV. Multi-task Enhanced Mechanism Design

To tackle the multi-task synchronization system, we propose the multi-task enhanced second-score auction-based mechanism, named MTEPViSA, based on the EPViSA proposed in [3]. Similarly, the MTEPViSA allocates and prices the synchronizing AV in the physical submarket by calculating the scoring rule. Therefore, we first define the AIGC-empowered synchronization scoring rule as follows.

Definition 1 (AIGC-empowered Synchronization Scoring Rule)**.** Let o be any offered bidding price, the AIGC-empowered synchronization scoring rule $\Phi^{syn}(o, \mathbf{d})$ is defined as

$$\Phi^{syn}(o, \mathbf{d}) = q - \phi(\mathbf{d}), \tag{6}$$

where $\mathbf{d}$ contains the submitted deadlines of DT tasks and $\phi(\cdot)$ is a non-decreasing function and $\phi(\mathbf{0}) = 0$.

The auctioneer can calculate the scoring rule based on previous transaction results and current submitted bids and deadlines. In the physical submarket, AVs submit their multi-dimensional bids $\mathbf{b}^{DT} = ((b_1^{DT}, \ldots, b_I^{DT}), \mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_I))$ to the auctioneer. The auctioneer computes the scores $\Phi^{syn} = \Phi^{syn}(b^{DT}, \eta) = (\phi_1^{syn}(b_1^{DT}, \eta_1), \ldots, \phi_I^{syn}(b_I^{DT}, \eta_I))$ to the auctioneer. Then, the auctioneer determines the winning AV in the physical submarket for synchronization according to the calculated scores. The auctioneer allocates the trader with the highest score as the winning physical entity, as $g_i^{DT}(\Phi^{syn}) = 1_{\{\Phi_i^{syn} > \max\{\Phi_{-i}^{syn}\}\}}$. In addition, the payment that the winning AV needs to pay is the bidding price of the second highest score, i.e., $p_i^{DT}(\Phi^{syn}) = g_i^{DT}(\Phi^{syn}) \cdot b_{\arg\max\{\Phi_{-i}^{syn}\}}^{DT}$.

In the virtual submarket, virtual simulators submit their bids $b^{AR} = (b_0^{AR}, b_1^{AR}, \ldots, b_K^{AR})$ to the auctioneer. In the MTEPViSA mechanism, the price scaling factor $\alpha \geqslant 1$ is utilized. First, the auctioneer determines the allocation probabilities for virtual traffic simulators as $g_k^{AR}(b^{AR}) = 1_{b_k^{AR} > \alpha b_{-k}^{AR}}$. Then, the allocation probability of the virtual driving simulator is calculated as $g_0^{AR}(b^{AR}) \leqslant 1 - \sum_{k=1}^{K} g_k^{AR}(b^{AR})$. Based on the price scaling factor, the winning MAR is required to pay $p_k^{AR}(b^{AR}) = g_k^{AR}(b^{AR}) \cdot \rho_k^{AR}$, where

$$\rho_k^{AR} = \begin{cases} T_{i,j,0}^{total} b_0^{AR}, & k = 0, \\ T_{i,j,k}^{total} \alpha \max\{b_{-k}^{AR}\}, & k = 1, \ldots, K. \end{cases} \tag{7}$$

Then, the efficient AIGC-empowered Scoring Rule can be defined as follows.

Definition 2 (Efficient AIGC-empowered Scoring Rule)**.** An efficient synchronization scoring rule is in the form of

$$\Phi^{syn}(o^{DT}, \mathbf{d}^*) = o^{DT} + \mathbf{d}^*[\gamma S_F^{AR}(\mathcal{M}) + S_I^{AR}(\mathcal{M})], \tag{8}$$

where $\mathbf{d}^*[\gamma S_F^{AR}(\mathcal{M}) + S_I^{AR}(\mathcal{M})]$ is the total surplus of virtual simulators by providing Metaverse billboards.

The allocation and pricing rules are effective and efficient when the efficient scoring rule exists [13] and the price scaling factor is selected as $\alpha_\iota = \max(1, \gamma[Q_{\iota,0}]/\mathbb{E}[Q_{\iota,(2)}])$ [10], where ι is the synchronizing AV in the physical submarket. Finally, under the cost-per-time model of rendering AR content and the efficient scoring rule, the MTEPViSA is fully strategy-proof and adverse-selection-free.

V. Experimental Results

In the simulation of the vehicular Metaverse, we consider a physical-virtual synchronization with 30 AVs, 30 virtual simulators, and 1 RSU by default. For each RSU, 20 MHz uplink and 20 MHz downlink channels are allocated for DT task uploading and AR content streaming, respectively. In addition, the CPU frequency of RSU is set to 3.6 GHz, and the GPU frequency is set to 19 GHz. The channel gain between RSUs and AVs is ranged from $[0, 1]$, where U denotes the uniform distribution. The transmission power of AVs is sampled from $U[0, 1]$ mW and the transmission power of RSUs is sampled from $U[0, 5]$ mW. The additive white Gaussian noise at AVs and RSUs is sampled from $\mathcal{N}(0, 1)$, where $\mathcal{N}$ denotes the normal distribution. For each DT task generated by AV, the data size is sampled from $U[0, 1]$ MB, the required CPU cycles per unit data are sampled from $[0, 1]$ Gcycles/MB, and the required deadline is sampled from $U[0.9, 1.1]$ seconds. For each AR simulation, the data size is sampled from $U[0, 0.25]$ MB and the required GPU cycles per unit data are sampled from $U[0, 1]$ Gcycles/MB. The valuation of AVs for accomplishing the DT tasks is sampled from $U[0.1, 1]$, and the number of preferences of AVs is sampled from $Zipf(2)$, where $Zipf$ denotes the Zipf distribution. The relative bargaining power of the virtual driving simulator is set to 1 while the default synchronization accuracy is 0.5. Generative AI based on large text-to-image models, such as stable diffusion [5] and Dreambooth [6], will have a game-changing impact on content creation in the Metaverse. To demonstrate the ability to generate diverse and high-quality images for vehicular Metaverse. As illustrated in Fig. 2, we experiment with modifying background color

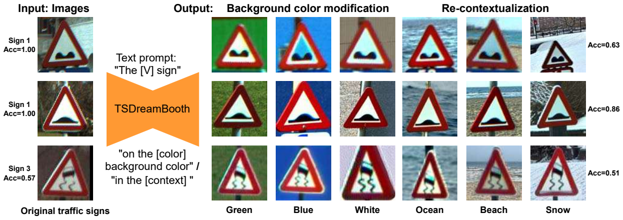


Fig. 2: Synthesized images of traffic signs by TSDreambooth for background modification and re-contextualization tasks.

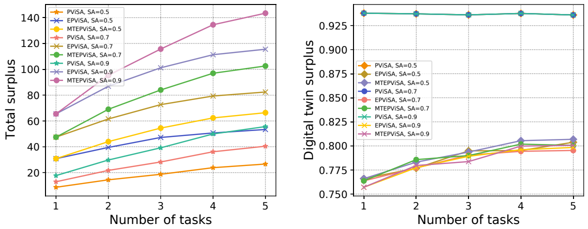
and re-contextualization for traffic signs, which is the iconic task for transportation systems. We first use the training set in BelgiumTS dataset [8] to fine-tune the Dreambooth to the TSDreambooth. Then, we train a validation model based on the pre-trained GoogLeNet to fit the BelgiumTS dataset. Finally, we generate new images based on the testing set in BelgiumTS and evaluate the generative score using the validation model. In Fig. 3, we evaluate the performance of the proposed mechanism under different system settings compared with the PViSA and the EPViSA proposed in [3]. From Fig. 3(b), we can understand the reason for the inefficiency of the PViSA. The PViSA mechanism always selects the AV with the highest valuation in the physical submarket to synchronize while ignoring the potential surplus in the virtual submarket.

VI. Conclusions

In this paper, we have proposed a generative AI-empowered physical-virtual synchronization framework for the vehicular Metaverse. In this framework, we have designed the DT-assisted future route prediction for AVs in the P2V synchronization. In addition, we have considered virtual simulators with generative AI models to customize diverse and scalable simulations in the V2P synchronization. Finally, we have devised the multi-task enhanced auction-based synchronization mechanism to incentivize RSUs to support effective synchronization. The property analysis has illustrated that the proposed mechanism is strategy-proof and adverse-selection free. The experimental results have illustrated that the proposed mechanism can increase the surplus by around 50%.

Acknowledgment

This research is supported in part by the NSFC under grant No. 62102099 and U22A2054, the National Research Foundation (NRF), Singapore and Infocomm Media Development Authority under the Future Communications Research Development Programme (FCP), and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019), under Energy Research Test-Bed and Industry Partnership Funding Initiative, part of the Energy Grid (EG) 2.0 programme, under DesCartes and the Campus for Research Excellence and Technological Enterprise (CREATE) programme, the SUTD SRG-ISTD-2021-165, the SUTD-ZJU IDEA Grant (SUTD-ZJU (VP) 202102), and the Ministry of Education, Singapore, under its SUTD Kickstarter Initiative (SKI 20210204), the NSF under grant CNS-2148382, CNS-2107216, CNS-2128368, CMMI-2222810, US Department of Transportation, Toyota and Amazon.



(a) Surplus v.s. number of tasks.

(b) DT surplus v.s. number of tasks.

Fig. 3: Performance evaluation under different numbers of tasks and generative scores.

References

[1] H. Duan, J. Li, S. Fan, Z. Lin, X. Wu, and W. Cai, "Metaverse for social good: A university campus prototype," in *Proc. of the 29th ACM International Conference on Multimedia*, Virtual Event, China, Oct. 2021, pp. 153–161.

[2] M. Xu, W. C. Ng, W. Y. B. Lim, J. Kang, Z. Xiong, D. Niyato, Q. Yang, X. S. Shen, and C. Miao, "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Communications Surveys & Tutorials*, pp. 1–1, Nov. 2022.

[3] M. Xu, D. Niyato, B. Wright, H. Zhang, J. Kang, Z. Xiong, S. Mao, and Z. Han, "Epvisa: Efficient auction design for real-time physical-virtual synchronization in the metaverse," *arXiv preprint arXiv:2211.06838*, 2022.

[4] Y. Ren, R. Xie, F. R. Yu, T. Huang, and Y. Liu, "Quantum collective learning and many-to-many matching game in the metaverse for connected and autonomous vehicles," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12 128–12 139, Jul. 2022.

[5] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, Jun. 2022, pp. 10 684–10 695.

[6] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," *arXiv preprint arXiv:2208.12242*, 2022.

[7] M. Xu, H. Du, D. Niyato, J. Kang, Z. Xiong, S. Mao, Z. Han, A. Jamalipour, D. I. Kim, V. Leung *et al.*, "Unleashing the power of edge-cloud generative ai in mobile networks: A survey of aigc services," *arXiv preprint arXiv:2303.16129*, 2023.

[8] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition—how far are we from the solution?" in *Proc. of International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, Aug. 2013.

[9] Y. Hui, X. Ma, Z. Su, N. Cheng, Z. Yin, T. H. Luan, and Y. Chen, "Collaboration as a service: Digital twins enabled collaborative and distributed autonomous driving," *IEEE Internet of Things Journal*, vol. 9, no. 19, pp. 18 607–18 619, Mar. 2022.

[10] N. Arnosti, M. Beck, and P. Milgrom, "Adverse selection and auction design for internet display advertising," *American Economic Review*, vol. 106, no. 10, pp. 2852–66, Oct. 2016.

[11] P. Ren, X. Qiao, Y. Huang, L. Liu, C. Pu, S. Dustdar, and J.-L. Chen, "Edge ar x5: An edge-assisted multi-user collaborative framework for mobile web augmented reality in 5g and beyond," *IEEE Transactions on Cloud Computing*, vol. 10, no. 4, pp. 2521–2537, Oct. 2022.

[12] N. H. Tran, W. Bao, A. Zomaya, M. N. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. of IEEE conference on computer communications*, Paris, France, May 2019, pp. 1387–1395.

[13] M. Tang, H. Pang, S. Wang, L. Gao, J. Huang, and L. Sun, "Multi-dimensional auction mechanisms for crowdsourced mobile video streaming," *IEEE/ACM Transactions on Networking*, vol. 26, no. 5, pp. 2062–2075, Aug. 2018.