DOI 10.15622/ia.24.1.11

N.V. HUNG, P.T. DAT, N. TAN, N.A. QUAN, L.T.N. TRANG, L.M. NAM

HEVERL – VIEWPORT ESTIMATION USING REINFORCEMENT LEARNING FOR 360-DEGREE VIDEO STREAMING

Nguyen Viet Hung, Pham Tien Dat, Nguyen Tan, Nguyen Anh Quan, Le Thi Huyen Trang, Le Mai Nam. **HEVERL – Viewport Estimation Using Reinforcement Learning for 360-degree Video Streaming.**

Abstract. 360-degree video content has become a pivotal component in virtual reality environments, offering viewers an immersive and engaging experience. However, streaming such comprehensive video content presents significant challenges due to the substantial file sizes and varying network conditions. To address these challenges, view adaptive streaming has emerged as a promising solution, aimed at reducing the burden on network capacity. This technique involves streaming lower-quality video for peripheral views while delivering high-quality content for the specific viewport that the user is actively watching. Essentially, it necessitates accurately predicting the user's viewing direction and enhancing the quality of that particular segment, underscoring the significance of Viewport Adaptive Streaming (VAS). Our research delves into the application of incremental learning techniques to predict the scores required by the VAS system. By doing so, we aim to optimize the streaming process by ensuring that the most relevant portions of the video are rendered in high quality. Furthermore, our approach is augmented by a thorough analysis of human head and facial movement behaviors. By leveraging these insights, we have developed a reinforcement learning model specifically designed to anticipate user view directions and improve the experience quality in targeted regions. The effectiveness of our proposed method is evidenced by our experimental results, which show significant improvements over existing reference methods. Specifically, our approach enhances the Precision metric by values ranging from 0.011 to 0.022. Additionally, it reduces the Root Mean Square Error (RMSE) by 0.008 to 0.013, the Mean Absolute Error (MAE) by 0.012 to 0.018 and the F1-score by 0.017 to 0.028. Furthermore, we observe an increase in overall accuracy of 2.79 to 16.98. These improvements highlight the potential of our model to significantly enhance the viewing experience in virtual reality environments, making 360-degree video streaming more efficient and user-friendly.

Keywords: head-eye movement, reinforcement learning, deep learning, machine learning, video streaming, 360-degree video.

1. Introduction. In recent years, prediction models have gained significant attention in the research community, particularly in the field of 360-degree video streaming. Accurate prediction in this context is crucial as it enhances the viewer's immersion and understanding of the video content. However, achieving high prediction accuracy remains a challenging task, especially under varying network conditions.

Existing research has explored various methods to improve the performance of prediction models for 360-degree videos. For example, reinforcement learning has been used to control model predictions based on data-driven designs, significantly improving performance, as demonstrated in the work of the authors in paper [1].

In the context of 360-degree videos, accurate prediction is excellent since it increases the viewer's understanding and immersion of the video. Significantly, when network conditions change, adapting to meet viewers' needs is difficult. From these research issues [2-4], the prediction models are built based on head movements to adapt to different types of videos on the Viewport Adaptive Streaming (VAS) system. However,the adaptability and self-learning ability are not only low but also dependable on the initial data, so it is still difficult when the data changes continuously.

In the context of 360-degree videos, accurate viewport prediction is essential for adapting to viewers' needs, especially when network conditions fluctuate. Studies such as those by the authors in [2-4] have developed prediction models based on head movements to adapt to different types of videos on the Viewport Adaptive Streaming (VAS) system. However, these models often suffer from limited adaptability and self-learning capabilities, particularly when data changes continuously, making it difficult to maintain accuracy.

Virtual reality (VR) presents additional challenges in this domain. As VR technology becomes more widespread, ensuring users feel fully immersed and interactively engaged is critical. However, video streaming in VR is constrained by factors such as network bandwidth, video resolution, and content complexity. High-speed transmission in the viewer's viewport area, coupled with lower quality in other areas, is a fundamental requirement for VR video streaming. Many studies have attempted to address these challenges by analyzing network conditions and employing optimization methods, but achieving a balance between network optimization and user immersion remains a significant hurdle. Therefore, predicting the viewer's viewport area is valid and applicable, considering the user's perspective without downloading the entire content. This means that the video content will be offloaded, and the network will have more space, which will help to improve the user's viewing area. In fact, [5] research has also shown critical retinal areas of the viewer; these are considered core areas. Based on these areas, we can quickly fix minor problems, such as limiting the quality of areas that are not considered and thereby improving the quality of areas that are considered.

Predicting the audience's view is a real challenge. Because each person will have a completely different view angle when turning their head and moving eyes. One more challenge for this problem is that the heads may not be moved when the viewers move their eyes. Only the movements of the eyes do not provide enough basis for a prediction model since we need both head-eye movements to be analyzed. Figure 1 shows information and forecast areas

in recent times. However, in this paper, we build on the principles of head movements to determine a better viewport position.

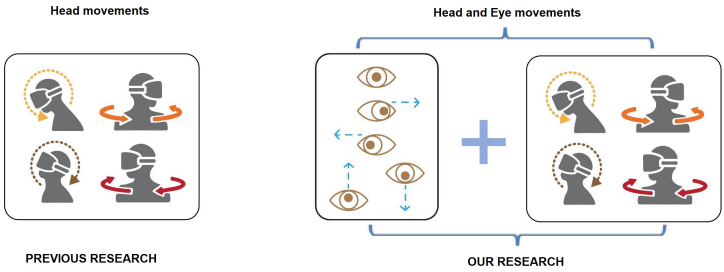

Fig. 1. Head-eye movements

Besides, psychology and perspective on movement are essential issues to analyze and make the right prediction [6-8]. First, the video contents partly affect viewers' psychology. For example, in emotional videos, viewers tend to change their head and eye movements when they have excess feelings. Second, many authors have been also researching perspective effects to evaluate the standard user's field of view. The factor of heads and eyes moving without following the rules also contributes significantly to incorrect prediction orders.

Furthermore, streaming 360-degree videos requires much more bandwidth compared to regular videos. The prediction method is necessary to achieve the user's perceived quality QoE because the user only sees a part of it. Thus, watching adaptive video streaming is an effective method to satisfy video quality [9-12]. However, this performance relies on the view adaptation scheme, view prediction and bandwidth. To overcome these problems, we propose a server-to-client streaming framework based on reinforcement learning, which optimizes 360-degree video streaming in viewport prediction to adapt to changing network conditions. We call this method HEVERL:

– To address these issues, we propose the HEVERL (Head-eye Movement Oriented Viewport Estimation Based on Reinforcement Learning) approach, which represents an advancement in viewport prediction for VR applications. Unlike traditional methods that rely solely on head orientation data, HEVERL incorporates both head-eye movement information to more accurately forecast the user's future viewport. This multi-modal sensing strategy provides a comprehensive understanding of the user's visual attention and behavior within the VR environment, leading to improved prediction accuracy.

– The HEVERL algorithm also introduces a novel content preparation and delivery mechanism that adaptively prefetches and updates the bitrate of previously viewed perspectives based on predicted viewport distribution. This proactive, viewport-aware content optimization enhances the user's perceived quality of experience by addressing network fluctuations and view prediction errors. By integrating prefetching and adaptive bitrate selection for previously visited viewports, HEVERL sets itself apart from traditional VR video streaming solutions, which generally rely on reactive strategies.

In summary, the HEVERL algorithm's dynamic adaptation to fluctuating network conditions and its ability to overcome potential view prediction errors represent a significant advancement in VR video streaming. The algorithm HEVERL may enhance the robustness and reliability of the VR experience, especially in latency-sensitive applications, and marks a step forward in achieving consistently high-quality VR viewing experiences. To provide a better understanding of our research, this report includes the following content: Section 2 discusses the related work. Section 3 describes the suggested viewport estimation technique. Section 4 evaluates the proposed method's performance compared to other methods. Section 5 concludes.

2. Related work

2.1. Streaming Video 360 Degrees. Recent research has focused on 360-degree video streaming, aiming to optimize bandwidth usage without compromising video quality. Studies [13-15] suggest that 360-degree video should be used as standard content to transmit the entire video, ensuring high viewing quality for users in all directions. However, streaming the full video consumes substantial bandwidth, allowing only a portion of the 360-degree video to be viewed at a time.

According to the research [16], there are two types of view-adaptive streaming: proposed tile-based streaming and asymmetric panorama image-based streaming. Panorama-based streaming generates multiple versions of a 360-degree video from different perspectives, necessitating video playback based on the user's orientation. While this approach reduces the apparent quality of the viewport and significantly lowers bandwidth usage, it also requires greater flexibility because limited versions result in poor display quality in viewer mode.

In tile-based streaming, the video is divided into multiple encrypted tiles, and different devices request tiles based on the user's perspective. Many algorithms for 360-degree video streaming [17, 18] transmit the Field of View (FoV) in this manner, effectively reducing bandwidth. However, this method is less flexible due to the dynamic changes in the user's perspective. As a result, recent viewport adaptation methods have relied on FoV [19, 20]. These

FoV-based prediction methods have improved significantly by reducing the performance impact caused by network distance to the predicted FoV and uneven bitrate assignment [21-24]. They dramatically reduce tile quality variation within the FoV. However, they still depend heavily on accurate bandwidth calculations, which can be influenced by network conditions, leading to estimation errors and performance degradation.

To overcome these limitations, we propose a reinforcement learning method combined with user behavior analysis to automatically adapt to network conditions and select tiles that optimize the predicted viewport area.

2.2. Synthetic prediction models. In this section, we will present some models built for prediction in recent years.

2.2.1. Head movements. In studies [3,4,25-27], the authors developed segment prediction models based on head movements. While many of these models are similar to our proposed model and aim to enhance the accuracy of predicting future user views in recommender systems, we identified some limitations. Notably, these methods primarily consider head movements while neglecting eye movements. The head can remain stationary while the eyes move. Therefore, experimental methods should account for head-eye movements to improve prediction accuracy.

Regarding head movements, most studies focus on changes in head position, acknowledging that head movements are generally slower compared to eye movements. However, addressing both types of movements presents a significant challenge. Many studies exclusively target head movements, overlooking the crucial role of eye movements. In reality, while the head may turn left or right, the eyes can independently look in different directions. This disparity underscores the importance of algorithmic adaptation to accommodate more complex movements for improved accuracy in prediction models.

2.2.2. Head-eye movements. In the study [28], the author developed cloud streaming for head-mounted displays, allowing viewers to experience the illusion of being in a virtual room by rotating their viewpoint. Additionally, in the study [29], the authors implemented a caching strategy that predicts user views based on cell resolution, aiming to forecast the viewing frequency of 360-degree video tiles. This method is particularly impactful under limited buffering conditions.

In another approach [30], the authors focused on predicting how different segments of a 360-degree video would be viewed on a head-mounted display. This method incorporated overlapping views and utilized techniques such as saliency detection, face detection, and object detection. However, the

algorithm primarily fine-tuned a fixed prediction network, leaving questions unanswered regarding the adaptability to changing movement dynamics.

While studies [28-30] have made significant strides in considering both head-eye movements, there is a pressing need for further research on the Viewport Adaptive Streaming (VAS) system's role in predicting user views. This gap in our understanding presents an exciting opportunity for future exploration and innovation in the field.

2.3. Reinforcement learning-based prediction. Viewport adaptation schemes for 360-degree video rely on estimated frequency width accuracy and are categorized based on throughput and buffering [31, 32]. However, this approach needs more flexibility and performs optimally only under specific network conditions. Therefore, adaptive algorithms are designed based on bitrate and user behavior to address these challenges and enhance adaptability and performance.

Approaches that rely on explicitly storing states and actions rather than using approximate functions are not scalable for real-world cyber environments. In response, D-DASH [33, 34] computes the action value Q using a neural network model (such as RNN or LSTM). D-DASH has shown superior performance and faster convergence compared to traditional Q-learning methods. However, its performance is still contingent on specific states and actions. To tackle this limitation, we propose an RL-based algorithm for decision-making that autonomously adapts to environmental changes.

Furthermore, the correlation between video perspective quality and video bitrate is non-linear. A neural network predicts video quality, while an RL algorithm selects the bitrate. This approach outperforms existing methods by delivering higher video quality and reducing latency.

While the authors have demonstrated that reinforcement learning optimizes adaptive bitrate for videos [35], this approach utilizes deep reinforcement learning (DRL) to train the curriculum autonomously. This enables bitrate decisions for 360-degree videos based on chunk selection and planning. This method has shown superior experimental results compared to state-of-the-art techniques. However, it primarily focuses on bitrate selection and chunk planning decisions, contrasting with our proposed method, which leverages user behavior to automatically adjust the bitrate and determine quality levels specifically for the viewport area.

On the contrary, in a recent study [36], researchers developed a system tailored for adaptation on Facebook's video platform using reinforcement learning (RL) in a live environment. They simulated the RL technique to train the agent effectively. Similarly, [37] introduced an advanced sequential reinforcement learning model to streamline decision-making and enhance the

Quality of Experience (QoE). These studies highlight the effectiveness of RL techniques in optimizing video streaming environments. However, these studies primarily concentrate on improving QoE by addressing factors like buffering, video quality, and timing without incorporating behavioral considerations.

Generally, reinforcement learning methods involve an agent making adaptive decisions in an interactive environment through trial and error [34]. Reinforcement learning empowers the agent to optimize its actions based on feedback, which is crucial for navigating dynamic and uncertain network conditions. However, these methods can be time-consuming, and their effectiveness hinges on the exploration strategy employed. Therefore, our proposed method aims to swiftly predict and make decisions that do not compromise the viewer's perception amidst fluctuating network conditions.

3. Proposed viewport estimation method – HEVERL. HEVERL is an acronym that stands for **H**ead **E**ye Movement Oriented **V**iewport **E**stimation Based on **R**einforcement **L**earning in Figure 2. Before discussing the HEVERL design, we will formulate the video streaming problem using the assumptions and constraints described in Section 3.

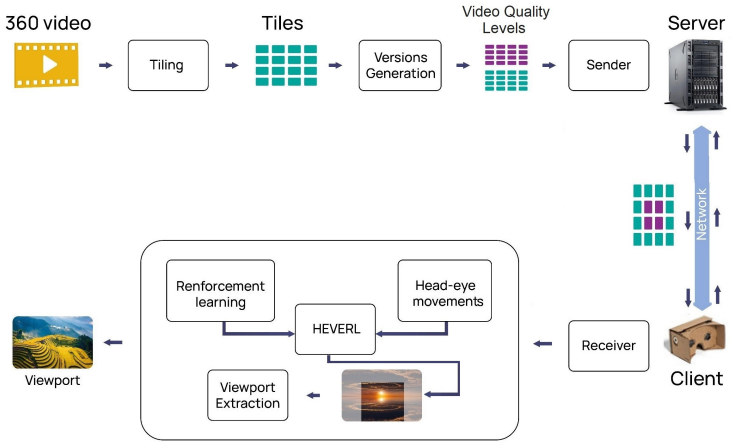

Fig. 2. HEVERL architecture

In this part, we present a problem that needs to be solved by predicting the viewport area that the human movement direction is using. Prediction is done when the direction of the user's movement does not change because it

is easier to predict and increase the quality of that area. However, in reality, prediction is very complicated because the more flexible the user is, the more significant the changes in prediction. Therefore, for each period t, it is necessary to predict the viewport area, and the next point will change, and so on, until the end of period t_n.

Furthermore, these changes will affect the user's perceived quality because when the user's movement direction is in any position, that area will increase in quality and reduce the near-quality area when not noticed. Therefore, this prediction increases the quality of user perception and limits bandwidth consumption in limited network conditions.

The core principle of tiling-based viewport adaptive streaming lies in the spatial partitioning of video content into distinct, granular sections known as tiles. This innovative architectural design deviates from the conventional view of the entire video frame as a single entity. By breaking down the video in this manner, the streaming system can handle the delivery of each tile independently, leading to more advanced adaptation strategies.

Expanding on the tiled structure, the tiling-based approach generates numerous encoded versions for each tile. This extensive range of tile variants empowers the system to enhance video quality based on the user's current viewport or field of view. Tiles that intersect with the user's viewport, called 'visible tiles,' are encoded at a higher quality to deliver an immersive viewing experience. In contrast, tiles outside the user's viewport, known as 'invisible tiles,' are encoded at a lower quality to conserve bandwidth and system resources in Figure 3.

The tiling-based viewport adaptive streaming approach is built on selectively assigning quality to visible and invisible tiles. By delivering the highest quality version of the tiles currently in the user's viewport, the system can provide an optimal visual experience without requiring high-quality data to be transmitted for the entire frame. This targeted quality allocation allows for efficient bandwidth utilization while reducing the risk of stalls or quality degradation during playback, as the system can dynamically adjust tile quality in response to user navigation and viewport changes.

The tiling-based viewport adaptive streaming model represents a significant step forward in video delivery, addressing the challenges of providing high-quality content while maximizing resource utilization. By spatially partitioning the video into tiles and selectively encoding multiple quality versions for each tile, the system can adaptively deliver the most appropriate content to the user based on their current viewport, resulting in a more immersive and bandwidth-efficient streaming experience.

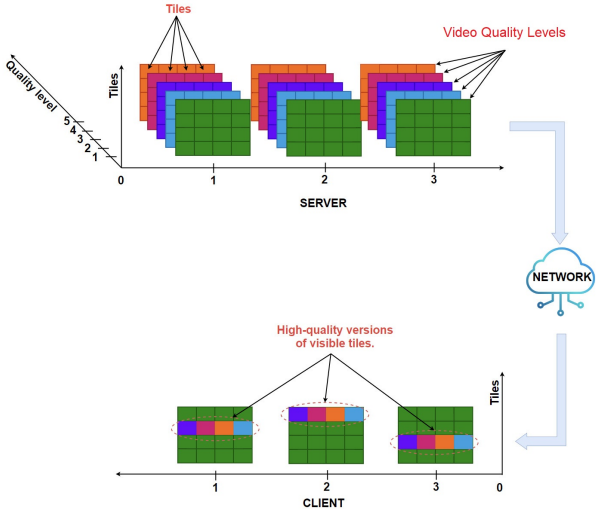

Fig. 3. Tiling-based Viewport Adaptive Streaming

3.1. Design of viewport prediction and selection. In this section, our focus is on designing a predictive model and devising methodologies for computing and categorizing viewport regions using reinforcement learning, illustrated in Figure 4.

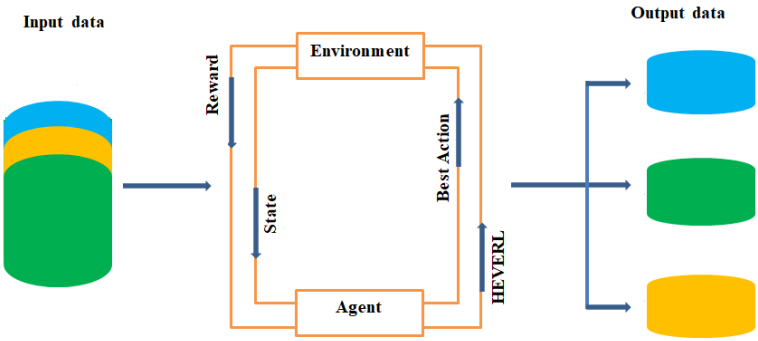

Fig. 4. HEVERL System

The system is structured as follows:

First, the data undergoes preprocessing before input. The data is represented through two states: t and t'. These states are stored as arrays and evolve spatially and temporally.

Second, we configure the environment settings and perform analysis based on these states. Subsequently, the algorithm calculates weights and dynamically predicts the user's viewing area throughout the video. The parameters are determined as follows:

– **Agent**. The Agent's objective is to locate the flag image, depicted in Figure 5. The Agent's path includes obstacles that influence the determination of the necessary route, impacting subsequent decision-making. Figure 5 illustrates how the Agent interacts with the Environment through actions such as left, right, up, and down.

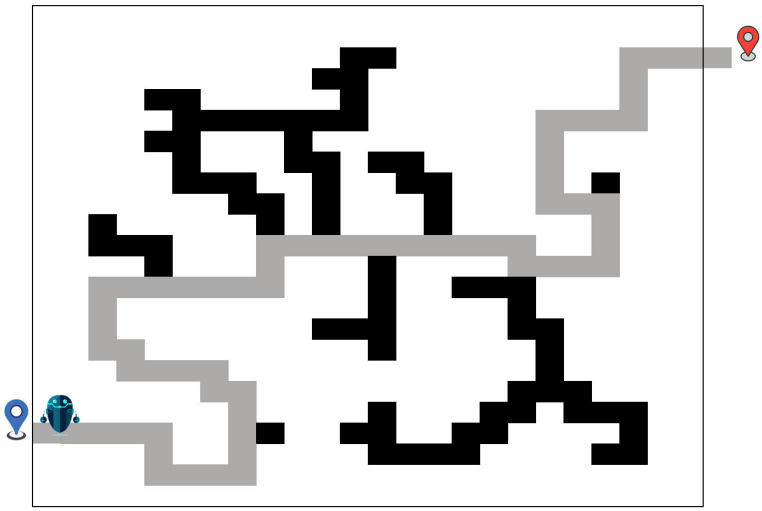

Fig. 5. Agent

– **State**. The state indicates the current position within the environment. Following each action, the environment provides the agent with a corresponding state.

– **Best Action**. The optimal action represents the transition process from the Agent to the environment. When the Agent reaches a forbidden box, the process terminates. The sequence of interactions between the Agent and the environment from start to finish is termed an Episode. Throughout the episode,

the Agent aims to select actions that maximize the Reward. The method by which the Agent selects these actions is known as the Policy.

– **HEVERL**. HEVERL will determine the final value to be saved and prepare for the next step based on the best action selections. Once identified and classified, the results will arrange the viewport sections sequentially and decide where to display information on the user's screen.

On the one hand, our approach utilizes the Markov Decision Process (MDP), a framework that aids agents in making decisions based on specific states. In applying this framework, we assume states possess the Markov property: the transition probability between two states is influenced solely by the preceding state.

Firstly, the concept of "probability of switching between two states" arises because, in reality, actions do not always yield deterministic outcomes. In an ideal scenario, repeating an action would consistently produce identical results. However, real-world processes are often stochastic. For instance, as depicted in Figure 6, if an agent decides to move upward and the environment's response is not deterministic, the outcomes can vary probabilistically. In this example, the agent might experience an 80% chance of returning to the "upper cell" state, with a 10% probability of transitioning to the "left cell" state and the "right cell" state each.

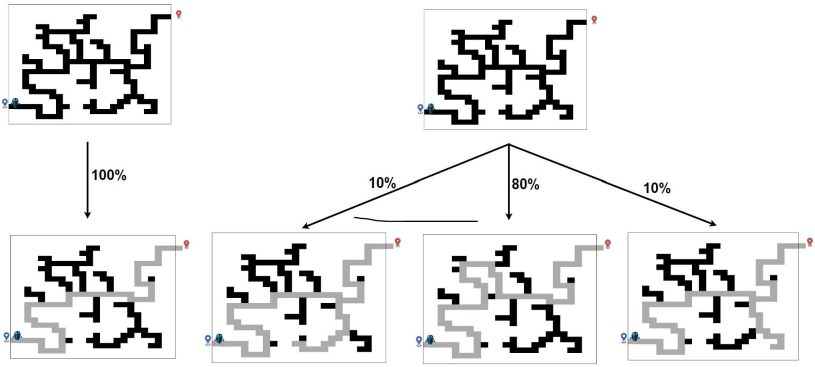

Fig. 6. Example process

3.2. Viewport Estimation Using Reinforcement Learning for 360-degree Video Streaming – HEVERL. HEVERL is the method we propose. It is based on the MDP model and is represented as follows. First, we calculate the Q_{Val} value when performing action h at state t by:

$$Q_{Val} = Q(t,h) = X(t,h) + \alpha \max_{h} Q(t',h). \quad (1)$$

Let X be the reward received when transferring state and $X(t,h)$ is the reward received υ'ith t' is the next state. Let α be the discount coefficient, ensuring that the farther "far away" from the Q_{Val} target, the smaller it is. Besides, let t be the state, and h be the action. This formula demonstrates that the Q_{Val} of action h at state t equals reward $X(t,h)$ plus the largest Q_{Val} of the following states t' when performing action h. As a result, we can create a state-action matrix as a lookup table using only that simple formula. As a result, for each state agent, the action with the highest Q_{Val} should be chosen. However, the Q_{Val} before and after acting will differ because RL is a stochastic process. This distinction is known as Temporal Difference:

$$f(h,t) = X(t,h) + \alpha \max_{h'} Q(t',h) - Q_{a-1}(t,h). \quad (2)$$

Therefore, the matrix Q_{Val} needs to be weighted based on TD by:

$$Q_a(t,h) = Q_{a-1}(t,h) + \sigma f_a(t,h), \quad (3)$$

where σ is the learning rate, through the times the agent performs actions, Q_{Val} will gradually converge. Thus, we aim to choose the appropriate action for a particular state. In other words, we use state as input and output as an action. During this stage, we realized that there is no constant solution using Neural Network (NN). All we need to do is remove the lookup table Q_{Val} and replace it with a simple NN in Figure 7. Besides, we employ a neural network structured with 4 layers. The configuration specifies the number of neurons per layer: 64, 128, 64, and 128 for layers 1, 2, 3, and 4, respectively. On the other hand, we use 3 neurons with x_1 as longitude, x_2 as latitude, and x_3 as the user's head-eye movement speed in Figure 7. In this part, we use x_3 represents the user's head-eye movements speed. It quantifies how quickly the user shifts their gaze. This variable can offer insights into user attention and focus, potentially indicating areas of interest or distraction. It could be measured in degrees per second if tracking angular movement per second for screen-based interactions. Understanding this speed, we can use adaptive content based on user engagement levels. All layers utilize ReLU activation functions, and regularization techniques include a Dropout set to 0.5 and an L2 regularization set to 0.01.

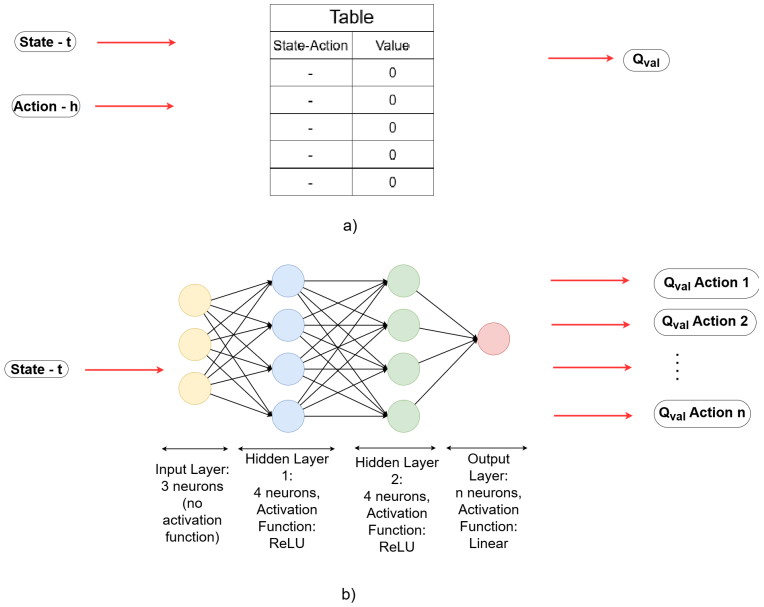

Fig. 7. State - Action

However, the most crucial part of NN is still missing. That is the Loss function. We aim to force the NN to learn how to accurately estimate the Q_{Val} for actions. Therefore, to determine the error between the actual and predicted Q_{Val}. The formula is determined and calculated as follows:

$$Loss\_function = (X + \alpha max_{h'} Q(t', h'; \varphi') - Q(t, h; \varphi))^2. \quad (4)$$

On the other hand, our HEVERL algorithm is proposed to perform as follows:

– **Step 01**: the setup environment injects a state into the network is t; The output is the Q_{Val} of the corresponding actions;

– **Step 02**: the agent chooses an action with a Policy and executes that action;

– **Step 03**: the environment returns state t' and reward x as the result of action h and saves the experience tuple $[t, h, x, t']$into memory;

– **Step 04**: sample the experiences into several batches and train the NN;

– **Step 05**: repeat until the end of M (M = 1000) episodes.

After performing the aforementioned steps, we calculate the predicted positions, which may fluctuate between different states. Experiments also indicate that our algorithm has shown improvement compared to conventional methods.

4. Performance Evaluation

4.1. Experimental Settings. To experiment, we use five videos: the Video Turtle describes People releasing baby turtles into the sea on the beach during the day. The Bar video describes the Bar as Light, with users moving and the bartender at work. The Video Ocean is described as follows: Under the ocean, people go underwater to see whales. Besides, there are two videos, Sofa and Po. Riverside is described as People sitting on sofas in the living room to talk, and Riverside videos outdoors during the day, with human activities. Each video contains traces of corresponding head-eye movements, and the information is also confirmed to change even when there is no head movement in Figure 8.

On the one hand, our dataset originates from the CSV file referenced [38]. We use two columns to indicate the viewer's position in latitude and longitude, normalized to a range of 0 to 1. Longitude values are scaled by multiplying by 2π, and latitude values by π to determine their on-screen positions. To display these positions accurately on an image, multiply longitude by the desired width and latitude by the desired height. Using these longitude and latitude coordinates, we can pinpoint the exact position of the observer. Besides, according to the authors in the article [38], head-eye movement data were collected from panoramic (360-degree) videos using head-eye tracking technologies. Head motion sensing technologies utilize accelerometers, gyroscope sensors, and kinematic trackers. Eye movement sensing technologies employ infrared eye trackers and eye-tracking glasses. 360-degree videos are recorded for users to view in virtual reality environments. Data from head-eye tracking sensors are recorded simultaneously with the video to provide information about how users interact with the content.

In this study, we utilized a dataset from [38] comprising head-eye movements data collected from 57 participants, including 25 women, whose ages ranged from 19 to 44 years (mean age: 25.7 years). Each participant viewed five distinct 360-degree videos for 20 seconds. The gaze data, sampled at 250Hz, yielded approximately 285,000 samples per video, totaling 1,425,000 samples across all videos. For model development, 80% of the data was

allocated for training and 20% for testing, ensuring comprehensive exposure during training and robust evaluation of model performance.

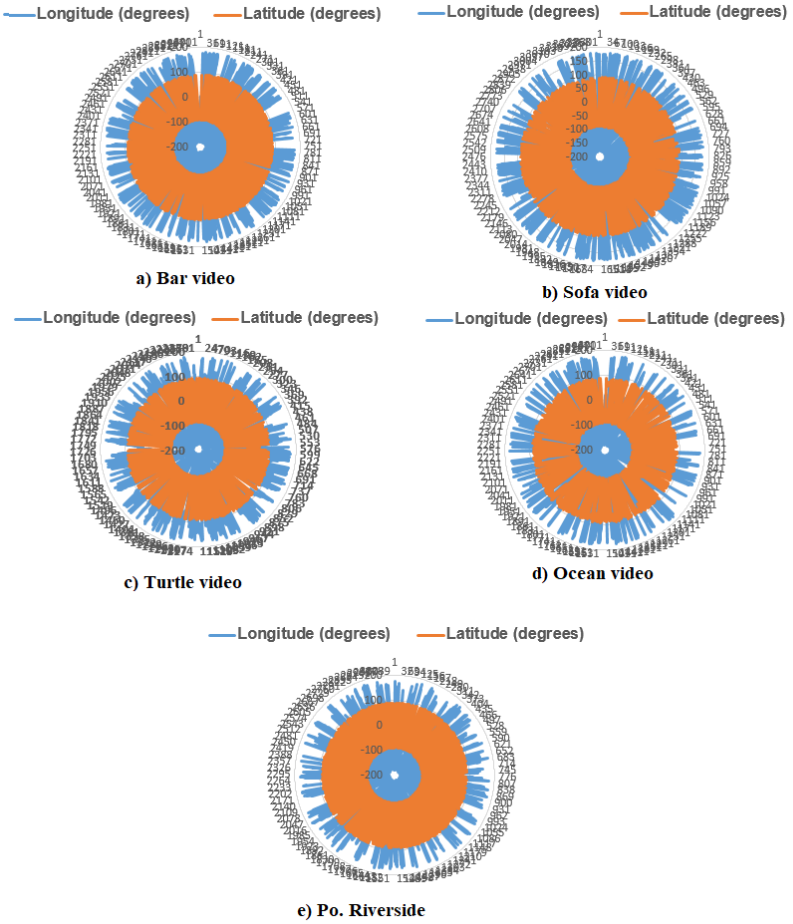

Fig. 8. Head-eye movements Dataset [38]

On the other hand, we experimented with the Windows 10 computer operating system, a Python-written experiment on a PC running 64-bit Windows 11, with 8192 MB RAM and an Intel® CoreTM i5-10400F Processor (6 Core, 12 Thread) CPU to measure the training time of different solutions. The proposed method HEVERL will be evaluated alongside other methods

by calculating the Root Mean Square Error (RMSE) based on precision calculation in the context of VAS.

The values defined in Table 1 should be replaced by the following abbreviations: TP for true positives, TN for true negatives, FP for false positives, and FN for false negatives.

– **Accuracy**. Accuracy is useful when the dataset's classes are well-balanced, with a similar number of instances in each class. However, accuracy can be misleading in imbalanced datasets, where one class significantly outnumbers the others.

$$Act = \frac{TN + TP}{TN + TP + FN + FP}. \tag{5}$$

– **Precision**. Indicates the precision with which Positive issues are detected.

$$Prec = \frac{TP}{TP + FP}. \tag{6}$$

– **Recall**. Recall measures the ability to find all the positive samples.

$$Recall = \frac{TP}{TP + FN}. \tag{7}$$

– **F1-Score**. F1-Score is the harmonic mean of precision and recall, providing a balance between the two.

$$F1 - Score = 2 * \frac{Prec * Recall}{Prec + Recall}. \tag{8}$$

Table 1. Definition of parameters

Values	Description
True Positives (TP)	True Positives are received True Positive;
False Positives (FP)	True Negatives are obtained False as Positive;
True negatives (TN)	True Negatives are received True Negatives;
False negatives (FN)	True Positives are received False as Negative.

Root Mean Square Error – RMSE. RMSE is one of the two leading performance indicators for a regression model. It computes the average difference between values predicted by a model and actual values. It estimates how well the model can predict the target value (accuracy):

$$RMSE = \sqrt{\frac{\sum_{a=1}^{H}(Prec_a - Act_a)^2}{H}}. \quad (9)$$

Mean Absolute Error – MAE. MAE is the average absolute magnitude of the errors between predicted and observed (true) viewport positions.

$$MAE = \frac{1}{H}\sum_{a=1}^{H}|Prec_a - Act_a|, \quad (10)$$

where:

– Let $Prec_a$ be the prediction rating,

– Let Act_a be the actual rating in the testing data set,

– H represents the number of rating prediction pairs between the testing data and the prediction result.

4.2. Viewport prediction performance. The viewport prediction performance of the HEVERL model is compared to the current reference models, including GLVP [3], AEVE [4], and GRU [39], in terms of Precision, RMSE, and MAE. This comparison aims to evaluate the viewport prediction capabilities of HEVERL against the benchmark models, intending to identify the advantages and effectiveness of the HEVERL model in applications that rely on accurate viewport prediction. Assessing these key performance metrics provides insights into the relative strengths and improvements offered by the HEVERL approach compared to the existing reference techniques.

The viewport prediction performance of HEVERL is compared with the current reference models such as GLVP [4], GRU [39], and AEVE [4] in terms of Precision, RMSE (Root Mean Square Error), MAE (Mean Absolute Error) in Table 2.

Table 2. HEVERL compared to the reference methods

Methods	Accuracy	Precision	Recall	F1-score	RMSE	MAE
GRU	71.23	0.865	0.860	0.861	0.248	0.147
GLVP	69.26	0.876	0.871	0.872	0.244	0.140
AEVE	83.45	0.869	0.862	0.864	0.249	0.144
HEVERL	**86.24**	**0.887**	**0.893**	**0.889**	**0.236**	**0.128**

The study compares the viewport prediction performance of HEVERL, a new proposed model, to existing reference models. Viewport prediction is essential in many applications, including adaptive streaming and

virtual/augmented reality, because it allows for efficient resource utilization and a better user experience.

In terms of precision, the study assesses each model's ability to predict the user's viewport. A higher Precision value indicates improved predictive performance. The results show that HEVERL outperforms the reference models in accurately predicting the user's viewport.

The study also examines the models' root mean square error (RMSE) and mean absolute error (MAE). These metrics are crucial in assessing the disparity between the predicted and actual viewport coordinates. Lower RMSE and MAE values indicate a higher level of predictive performance. The findings reveal that HEVERL exhibits lower RMSE and MAE than the reference models, suggesting that it delivers more accurate viewport predictions with fewer errors.

The study's results demonstrate that the HEVERL model is highly effective in viewport prediction. This model holds significant promise as a tool for optimizing resource allocation and enhancing the overall user experience in various applications that rely on accurate viewport prediction. Significantly, it surpasses the current reference models in terms of Precision, RMSE, and MAE.

4.3. Training time evaluation. Table 3 illustrates the performance of four algorithms (AEVE, GRU, GLVP, and HEVERL) across five datasets (Bar, Ocean, Po Riversides, Sofa, and Turtle). The performance metrics indicate that these algorithms yield favorable results, with average processing times below 100ms for the entire video. This demonstrates the algorithms' effectiveness in aiding decision-making processes.

Table 3. Training time overview

Methods	Bar	Ocean	Po. Riversides	Sofa	Turtle
AEVE	0.0953	0.0644	0.0722	0.0904	0.0933
GRU	0.1020	0.0766	0.0708	0.0921	0.0983
GLVP	0.1030	0.0951	0.0649	0.0954	0.0913
HEVERL	0.0885	0.0971	0.0863	0.1100	0.0782

However, it is critical to consider not only raw performance metrics but also the algorithm's consistency and stability. An algorithm that performs well on average but has a high degree of variability in results may be less desirable than one with slightly lower peak performance but is more stable and reliable.

Choosing the best algorithm is a nuanced decision based on the problem's requirements and constraints. If the goal is to maximize performance across all data sets, the HEVERL algorithm is the top choice. However, the algorithm's performance in specific data sets or use cases may be more relevant. A thorough understanding of the problem context and desired outcomes is

required before making a definitive recommendation on the best algorithm for training time evaluation.

5. Conclusions. In this paper, we tackle the difficult task of viewport prediction in the context of VR video streaming. The proposed solution outperformed four reference methods in several critical evaluation metrics, such as Precision, Root Mean Square Error (RMSE), and Mean Absolute Error. By accurately predicting the user's current and future viewport, the authors' approach has the potential to significantly improve VR content delivery, lowering latency and improving overall viewing quality. Accurate viewport prediction is a critical enabler for optimizing bandwidth utilization and selectively streaming high-quality content only for the regions of interest, ultimately increasing user satisfaction and engagement with more immersive and enjoyable VR services across various domains, such as gaming, education, and training.

Our strategy focuses on exploring and optimizing techniques to enhance the performance of Reinforcement Learning models within the VAS system, aiming to predict and improve user experience quality. Moving forward, we plan to conduct additional experiments to validate the effectiveness of this approach, while also investigating solutions for integrating and deploying these optimized models across real-world virtual reality platforms.

References

1. Pan X., Chen X., Zhang Q., Li N. Model predictive control: A rein-forcement learning-based approach. Journal of Physics: Conference Series. IOP Publishing. 2022. vol. 2203. no. 1. DOI: 10.1088/1742-6596/2203/1/012058.
2. Feng X., Swaminathan V., Wei S. Viewport prediction for live 360-degree mobile video streaming using user-content hybrid motion tracking. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2019. vol. 3. no. 2. pp. 1–22. DOI: 10.1145/3328914.
3. Nguyen H., Dao T.N., Pham N.S., Dang T.L., Nguyen T.D., Truong T.H. An accurate viewport estimation method for 360 video streaming using deep learning. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9. no. 4. DOI: 10.4108/eetinis.v9i4.2218.
4. Nguyen D. An evaluation of viewport estimation methods in 360-degree video streaming. 7th International Conference on Business and Industrial Research (ICBIR). IEEE, 2022. pp. 161–166. DOI: 10.1109/ICBIR54589.2022.9786513.
5. Nguyen V.H., Pham N.N., Truong C.T., Bui D.T., Nguyen H.T., Truong T.H. Retina-based quality assessment of tile-coded 360-degree videos. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9. no. 32. DOI: 10.4108/eetinis.v9i32.1058.
6. Lee E.-J., Jang Y.J., Chung M. When and how user comments affect news readers' personal opinion: perceived public opinion and perceived news position as mediators. Digital Journalism. 2020. vol. 9. no. 1. pp. 42–63. DOI: 10.1080/21670811.2020.1837638.
7. Nguyen H.V., Tan N., Quan N.H., Huong T.T., Phat N.H. Building a chatbot system to analyze opinions of english comments. Informatics and Automation. 2023. vol. 22. no. 2.

pp. 289–315.
8. Raja U.S., Carrico A.R. A qualitative exploration of individual experiences of environmental virtual reality through the lens of psychological distance. Environmental Communication. 2021. vol. 15. no. 5. pp. 594–609. DOI: 10.1080/17524032.2020.1871052.
9. Jiang Z., Zhang X., Xu Y., Ma Z., Sun J., Zhang Y. Reinforcement learning based rate adaptation for 360-degree video streaming. IEEE Transactions on Broadcasting. 2021. vol. 67. no. 2. pp. 409–423. DOI: 10.1109/TBC.2020.3028286.
10. Nguyen V.H., Bui D.T., Tran T.L., Truong C.T., Truong T.H. Scalable and resilient 360-degree-video adaptive streaming over http/2against sudden network drops. Computer Communications. 2024. vol. 216. pp. 1–15. DOI: 10.1016/j.comcom.2024.01.001.
11. Kan N., Zou J., Li C., Dai W., Xiong H. Rapt360: Reinforcement learning-based rate adaptation for 360-degree video streaming with adaptive prediction and tiling. IEEE Transactions on Circuits and Systems for Video Technology. 2022. vol. 32. no. 3. pp. 1607–1623. DOI: 10.1109/TCSVT.2021.3076585.
12. Hung N.V., Chien T.D., Ngoc N.P., Truong T.H. Flexible http-based video adaptive streaming for good QoE during sudden bandwidth drops. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2023. vol. 10. no. 2. DOI: 10.4108/eetinis.v10i2.2994.
13. Wong E.S., Wahab N.H.A., Saeed F., Alharbi N. 360-degree video bandwidth reduction: Technique and approaches comprehensive review. Applied Sciences. 2022. vol. 12. no. 15. DOI: 10.3390/app12157581.
14. Lampropoulos G., Barkoukis V., Burden K., Anastasiadis T. 360-degree video in education: An overview and a comparative social media data analysis of the last decade. Smart Learning Environments. 2021. vol. 8. DOI: 10.1186/s40561-021-00165-8.
15. Ng K.-T., Chan S.-C., Shum H.-Y. Data compression and transmission aspects of panoramic videos. IEEE Transactions on Circuits and Systems for Video Technology. 2005. vol. 15. no. 1. pp. 82–95. DOI: 10.1109/TCSVT.2004.839989.
16. Xie L., Xu Z., Ban Y., Zhang X., Guo Z. 360ProbDASH: Improving QoE of 360 video streaming using tile-based http adaptive streaming. Proceedings of the 25th ACM international conference on Multimedia. 2017. pp. 315–323. DOI: 10.1145/3123266.3123291.
17. Hosseini M., Swaminathan V. Adaptive 360 VR video streaming: Divide and conquer. IEEE International Symposium on Multimedia (ISM). IEEE, 2016. pp. 107–110.
18. El-Ganainy T., Hefeeda M. Streaming virtual reality content. arXiv preprint arXiv:1612.08350. 2016. DOI: 10.48550/arXiv.1612.08350.
19. Xu M., Song Y., Wang J., Qiao M., Huo L., Wang Z. Predicting head movement in panoramic video: A deep reinforcement learning approach. IEEE transactions on pattern analysis and machine intelligence. 2019. vol. 41. no. 11. pp. 2693–2708. DOI: 10.1109/TPAMI.2018.2858783.
20. Hu H.-N., Lin Y.-C., Liu M.-Y., Cheng H.-T., Chang Y.-J., Sun M. Deep 360 pilot: Learning a deep agent for piloting through 360deg sports videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1396–1405.
21. Bao Y., Wu H., Zhang T., Ramli A.A., Liu X. Shooting a moving target: Motion-prediction-based transmission for 360-degree videos. IEEE International Conference on Big Data. IEEE. 2016. pp. 1161–1170. DOI: 10.1109/BigData.2016.7840720.
22. Petrangeli S., Swaminathan V., Hosseini M., De Turck F. An http/2-based adaptive streaming framework for 360 virtual reality videos. Proceedings of the 25th ACM international conference on Multimedia. 2017. pp. 306–314. DOI: 10.1145/3123266.3123453.

23. Hung N.V., Tien B.D., Anh T.T.T., Nam P.N., Huong T.T. An efficient approach to terminate 360-video stream on http/3. AIP Conference Proceedings. AIP Publishing. 2023. vol. 2909. no. 1.
24. Yu J., Liu Y. Field-of-view prediction in 360-degree videos with attention-based neural encoder-decoder networks. Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems. 2019. pp. 37–42. DOI: 10.1145/3304113.3326118.
25. Park S., Bhattacharya A., Yang Z., Das S.R., Samaras D. Mosaic: Advancing user quality of experience in 360-degree video streaming with machine learning. IEEE Transactions on Network and Service Management. 2021. vol. 18. no. 1. pp. 1000–1015. DOI: 10.1109/TNSM.2021.3053183.
26. Lee D., Choi M., Lee J. Prediction of head movement in 360-degree videos using attention model. Sensors. 2021. vol. 21. no. 11. DOI: 10.3390/s21113678.
27. Chen X., Kasgari A.T.Z., Saad W. Deep learning for content-based personalized viewport prediction of 360-degree VR videos. IEEE Networking Letters. 2020. vol. 2. no. 2. pp. 81–84. DOI: 10.1109/LNET.2020.2977124.
28. Vielhaben J., Camalan H., Samek W., Wenzel M. Viewport forecasting in 360 virtual reality videos with machine learning. IEEE international conference on artificial intelligence and virtual reality (AIVR). IEEE. 2019. pp. 74–747. DOI: 10.1109/AIVR46125.2019.00020.
29. Uddin M.M., Park J. Machine learning model evaluation for 360°video caching. IEEE World AI IoT Congress (AIIoT). IEEE. 2022. pp. 238–244. DOI: 10.1109/AIIoT54504.2022.9817292.
30. Fan C.-L., Yen S.-C., Huang C.-Y., Hsu C.-H. Optimizing fixation prediction using recurrent neural networks for 360° video streaming in head-mounted virtual reality. IEEE Transactions on Multimedia. 2020. vol. 22. no. 3. pp. 744–759. DOI: 10.1109/TMM.2019.2931807.
31. Yaqoob A., Bi T., Muntean G.-M. A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. IEEE Communications Surveys and Tutorials. 2020. vol. 22. no. 4. pp. 2801–2838. DOI: 10.1109/COMST.2020.3006999.
32. Liu X. Deng Y. Learning-based prediction, rendering and association optimization for mec-enabled wireless virtual reality (VR) networks. IEEE Transactions on Wireless Communications. 2021. vol. 20. no. 10. pp. 6356–6370. DOI: 10.1109/TWC.2021.3073623.
33. Gadaleta M., Chiariotti F., Rossi M., Zanella A. D-DASH: A deep q-learning framework for dash video streaming. IEEE Transactions on Cognitive Communications and Networking. 2017. vol. 3. no. 4. pp. 703–718. DOI: 10.1109/TCCN.2017.2755007.
34. Souane N., Bourenane M., Douga Y. Deep reinforcement learning-based approach for video streaming: Dynamic adaptive video streaming over HTTP. Applied Sciences. 2023. vol. 13. no. 21. DOI: 10.3390/app132111697.
35. Xie Y., Zhang Y., Lin T. Deep curriculum reinforcement learning for adaptive 360 ○ video streaming with two-stage training. IEEE Transactions on Broadcasting. 2023. vol. 70. no. 2. pp. 441–452. DOI: 10.1109/tbc.2023.3334137.
36. Du L., Zhuo L., Li J., Zhang J., Li X., Zhang H. Video quality of experience metric for dynamic adaptive streaming services using dash standard and deep spatial-temporal representation of video. Applied Sciences. 2020. vol. 10. no. 5. DOI: 10.3390/app10051793.
37. Mao H., Chen S., Dimmery D., Singh S., Blaisdell D., Tian Y., Alizadeh M., Bakshy E. Real-world video adaptation with reinforcement learning. arXiv preprint arXiv:2008.12858. 2020. DOI: 10.48550/arXiv.2008.12858.

38. David E.J., Gutiérrez J., Coutrot A., Da Silva M.P., Callet P.L. A dataset of head and eye movements for 360 videos. Proceedings of the 9th ACM Multimedia Systems Conference. 2018. pp. 432–437.

39. Wu C., Zhang R., Wang Z., Sun L. A spherical convolution approach for learning long term viewport prediction in 360 immersive video. Proceedings of the AAAI Conference on Artificial Intelligence. 2020. vol. 34. no. 01. pp. 14003–14040. DOI: 10.1609/aaai.v34i01.7377.

Nguyen Viet Hung — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, network security, artificial intelligence, traffic engineering in next-generation networks, QoE/QoS guarantee for network services, green networking, applications. The number of publications — 23. hungnv@eaut.edu.vn; Ky Phu - Ky Anh, Ha Tinh, Viet Nam; office phone: +84(098)911-2079.

Pham Tien Dat — Research assistant, East Asia University of Technology. Research interests: applications, networks. The number of publications — 1. 20212452@eaut.edu.vn; Vu Ninh – Kien Xuong, Thai Binh, Viet Nam; office phone: +84(036)239-6558.

Nguyen Tan — Research assistant, East Asia University of Technology. Research interests: applications, data analysis. The number of publications — 3. tan25102000@gmail.com; Trung Dung - Tien Lu, Hung Yen, Viet Nam; office phone: +84(035)919-0216.

Nguyen Anh Quan — Research assistant, East Asia University of Technology. Research interests: applications, networks. The number of publications — 1. anhq46724@gmail.com; Gia Lam, Hanoi, Viet Nam; office phone: +84(096)278-4293.

Le Thi Huyen Trang — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 2. tranglth@eaut.edu.vn; Phuong Tri - Thi Tran Phung - Dan Phuong, Hanoi, Viet Nam; office phone: +84(032)889-9334.

Le Mai Nam — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: software engineering, optimization mathematics, applications. The number of publications — 1. namlm@eaut.edu.vn; Phuong Trung - Thanh Oai, Hanoi, Viet Nam; office phone: +84(098)208-2117.

УДК 006.72 **DOI** 10.15622/ia.24.1.11

Н. Хунг, Ф.Т. Дат, Н. Тан, Н.А. Куан, Л. Транг, Л.М. Нам,

ОЦЕНКА ОБЛАСТИ ПРОСМОТРА С ИСПОЛЬЗОВАНИЕМ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ ДЛЯ ПОТОКОВОЙ ПЕРЕДАЧИ ВИДЕО В ФОРМАТЕ 360 ГРАДУСОВ

Хунг Н., Дат Ф.Т., Тан Н., Куан Н.А., Транг Л., Нам Л.М. **Оценка области просмотра с использованием обучения с подкреплением для потоковой передачи видео в формате 360 градусов.**

Аннотация. Видеоконтент в формате 360 градусов стал ключевым компонентом в средах виртуальной реальности, предлагая зрителям захватывающий и увлекательный опыт. Однако потоковая передача такого комплексного видеоконтента сопряжена со значительными трудностями, обусловленными существенными размерами файлов и переменчивыми сетевыми условиями. Для решения этих проблем в качестве перспективного решения, направленного на снижение нагрузки на пропускную способность сети, появилась адаптивная потоковая передача просмотра. Эта технология предполагает передачу видео более низкого качества для периферийных зон просмотра, а высококачественный контент – для конкретной зоны просмотра, на которую активно смотрит пользователь. По сути, это требует точного прогнозирования направления просмотра пользователя и повышения качества этого конкретного сегмента, что подчеркивает значимость адаптивной потоковой передачи просмотра (VAS). Наше исследование углубляется в применение методов пошагового обучения для прогнозирования оценок, требуемых системой VAS. Таким образом, мы стремимся оптимизировать процесс потоковой передачи, обеспечивая высокое качество отображения наиболее важных фрагментов видео. Кроме того, наш подход дополняется тщательным анализом поведения движений головы и лица человека. Используя эти данные, мы разработали модель обучения с подкреплением, специально предназначенную для прогнозирования направлений взгляда пользователя и повышения качества изображения в целевых областях. Эффективность предлагаемого нами метода подтверждается нашими экспериментальными результатами, которые показывают значительные улучшения по сравнению с существующими эталонными методами. В частности, наш подход повышает метрику прецизионности на значения в диапазоне от 0,011 до 0,022. Кроме того, он снижает среднеквадратичную ошибку (RMSE) в диапазоне от 0,008 до 0,013, среднюю абсолютную ошибку (MAE) – от 0,012 до 0,018 и оценку F1 – от 0,017 до 0,028. Кроме того, мы наблюдаем увеличение общей точности с 2,79 до 16,98. Эти улучшения подчеркивают потенциал нашей модели для значительного улучшения качества просмотра в средах виртуальной реальности, делая потоковую передачу видео на 360 градусов более эффективной и удобной для пользователя.

Ключевые слова: движение головы и глаз, обучение с подкреплением, глубокое обучение, машинное обучение, потоковая передача видео, видео на 360 градусов.

Литература

1. Pan X., Chen X., Zhang Q., Li N. Model predictive control: A rein-forcement learning-based approach. Journal of Physics: Conference Series. IOP Publishing. 2022. vol. 2203. no. 1. DOI: 10.1088/1742-6596/2203/1/012058.
2. Feng X., Swaminathan V., Wei S. Viewport prediction for live 360-degree mobile video streaming using user-content hybrid motion tracking. Proceedings of the ACM on

Interactive, Mobile, Wearable and Ubiquitous Technologies. 2019. vol. 3. no. 2. pp. 1–22. DOI: 10.1145/3328914.

3. Nguyen H., Dao T.N., Pham N.S., Dang T.L., Nguyen T.D., Truong T.H. An accurate viewport estimation method for 360 video streaming using deep learning. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9. no. 4. DOI: 10.4108/eetinis.v9i4.2218.
4. Nguyen D. An evaluation of viewport estimation methods in 360-degree video streaming. 7th International Conference on Business and Industrial Research (ICBIR). IEEE, 2022. pp. 161–166. DOI: 10.1109/ICBIR54589.2022.9786513.
5. Nguyen V.H., Pham N.N., Truong C.T., Bui D.T., Nguyen H.T., Truong T.H. Retina-based quality assessment of tile-coded 360-degree videos. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9. no. 32. DOI: 10.4108/eetinis.v9i32.1058.
6. Lee E.-J., Jang Y.J., Chung M. When and how user comments affect news readers' personal opinion: perceived public opinion and perceived news position as mediators. Digital Journalism. 2020. vol. 9. no. 1. pp. 42–63. DOI: 10.1080/21670811.2020.1837638.
7. Nguyen H.V., Tan N., Quan N.H., Huong T.T., Phat N.H. Building a chatbot system to analyze opinions of english comments. Informatics and Automation. 2023. vol. 22. no. 2. pp. 289–315.
8. Raja U.S., Carrico A.R. A qualitative exploration of individual experiences of environmental virtual reality through the lens of psychological distance. Environmental Communication. 2021. vol. 15. no. 5. pp. 594–609. DOI: 10.1080/17524032.2020.1871052.
9. Jiang Z., Zhang X., Xu Y., Ma Z., Sun J., Zhang Y. Reinforcement learning based rate adaptation for 360-degree video streaming. IEEE Transactions on Broadcasting. 2021. vol. 67. no. 2. pp. 409–423. DOI: 10.1109/TBC.2020.3028286.
10. Nguyen V.H., Bui D.T., Tran T.L., Truong C.T., Truong T.H. Scalable and resilient 360-degree-video adaptive streaming over http/2against sudden network drops. Computer Communications. 2024. vol. 216. pp. 1–15. DOI: 10.1016/j.comcom.2024.01.001.
11. Kan N., Zou J., Li C., Dai W., Xiong H. Rapt360: Reinforcement learning-based rate adaptation for 360-degree video streaming with adaptive prediction and tiling. IEEE Transactions on Circuits and Systems for Video Technology. 2022. vol. 32. no. 3. pp. 1607–1623. DOI: 10.1109/TCSVT.2021.3076585.
12. Hung N.V., Chien T.D., Ngoc N.P., Truong T.H. Flexible http-based video adaptive streaming for good QoE during sudden bandwidth drops. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2023. vol. 10. no. 2. DOI: 10.4108/eetinis.v10i2.2994.
13. Wong E.S., Wahab N.H.A., Saeed F., Alharbi N. 360-degree video bandwidth reduction: Technique and approaches comprehensive review. Applied Sciences. 2022. vol. 12. no. 15. DOI: 10.3390/app12157581.
14. Lampropoulos G., Barkoukis V., Burden K., Anastasiadis T. 360-degree video in education: An overview and a comparative social media data analysis of the last decade. Smart Learning Environments. 2021. vol. 8. DOI: 10.1186/s40561-021-00165-8.
15. Ng K.-T., Chan S.-C., Shum H.-Y. Data compression and transmission aspects of panoramic videos. IEEE Transactions on Circuits and Systems for Video Technology. 2005. vol. 15. no. 1. pp. 82–95. DOI: 10.1109/TCSVT.2004.839989.
16. Xie L., Xu Z., Ban Y., Zhang X., Guo Z. 360ProbDASH: Improving QoE of 360 video streaming using tile-based http adaptive streaming. Proceedings

of the 25th ACM international conference on Multimedia. 2017. pp. 315–323. DOI: 10.1145/3123266.3123291.

17. Hosseini M., Swaminathan V. Adaptive 360 VR video streaming: Divide and conquer. IEEE International Symposium on Multimedia (ISM). IEEE, 2016. pp. 107–110.
18. El-Ganainy T., Hefeeda M. Streaming virtual reality content. arXiv preprint arXiv:1612.08350. 2016. DOI: 10.48550/arXiv.1612.08350.
19. Xu M., Song Y., Wang J., Qiao M., Huo L., Wang Z. Predicting head movement in panoramic video: A deep reinforcement learning approach. IEEE transactions on pattern analysis and machine intelligence. 2019. vol. 41. no. 11. pp. 2693–2708. DOI: 10.1109/TPAMI.2018.2858783.
20. Hu H.-N., Lin Y.-C., Liu M.-Y., Cheng H.-T., Chang Y.-J., Sun M. Deep 360 pilot: Learning a deep agent for piloting through 360deg sports videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1396–1405.
21. Bao Y., Wu H., Zhang T., Ramli A.A., Liu X. Shooting a moving target: Motion-prediction-based transmission for 360-degree videos. IEEE International Conference on Big Data. IEEE. 2016. pp. 1161–1170. DOI: 10.1109/BigData.2016.7840720.
22. Petrangeli S., Swaminathan V., Hosseini M., De Turck F. An http/2-based adaptive streaming framework for 360 virtual reality videos. Proceedings of the 25th ACM international conference on Multimedia. 2017. pp. 306–314. DOI: 10.1145/3123266.3123453.
23. Hung N.V., Tien B.D., Anh T.T.T., Nam P.N., Huong T.T. An efficient approach to terminate 360-video stream on http/3. AIP Conference Proceedings. AIP Publishing. 2023. vol. 2909. no. 1.
24. Yu J., Liu Y. Field-of-view prediction in 360-degree videos with attention-based neural encoder-decoder networks. Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems. 2019. pp. 37–42. DOI: 10.1145/3304113.3326118.
25. Park S., Bhattacharya A., Yang Z., Das S.R., Samaras D. Mosaic: Advancing user quality of experience in 360-degree video streaming with machine learning. IEEE Transactions on Network and Service Management. 2021. vol. 18. no. 1. pp. 1000–1015. DOI: 10.1109/TNSM.2021.3053183.
26. Lee D., Choi M., Lee J. Prediction of head movement in 360-degree videos using attention model. Sensors. 2021. vol. 21. no. 11. DOI: 10.3390/s21113678.
27. Chen X., Kasgari A.T.Z., Saad W. Deep learning for content-based personalized viewport prediction of 360-degree VR videos. IEEE Networking Letters. 2020. vol. 2. no. 2. pp. 81–84. DOI: 10.1109/LNET.2020.2977124.
28. Vielhaben J., Camalan H., Samek W., Wenzel M. Viewport forecasting in 360 virtual reality videos with machine learning. IEEE international conference on artificial intelligence and virtual reality (AIVR). IEEE. 2019. pp. 74–747. DOI: 10.1109/AIVR46125.2019.00020.
29. Uddin M.M., Park J. Machine learning model evaluation for 360°video caching. IEEE World AI IoT Congress (AIIoT). IEEE. 2022. pp. 238–244. DOI: 10.1109/AIIoT54504.2022.9817292.
30. Fan C.-L., Yen S.-C., Huang C.-Y., Hsu C.-H. Optimizing fixation prediction using recurrent neural networks for 360° video streaming in head-mounted virtual reality. IEEE Transactions on Multimedia. 2020. vol. 22. no. 3. pp. 744–759. DOI: 10.1109/TMM.2019.2931807.
31. Yaqoob A., Bi T., Muntean G.-M. A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. IEEE Communications Surveys and Tutorials. 2020. vol. 22. no. 4. pp. 2801–2838. DOI: 10.1109/COMST.2020.3006999.

32. Liu X. Deng Y. Learning-based prediction, rendering and association optimization for mec-enabled wireless virtual reality (VR) networks. IEEE Transactions on Wireless Communications. 2021. vol. 20. no. 10. pp. 6356–6370. DOI: 10.1109/TWC.2021.3073623.
33. Gadaleta M., Chiariotti F., Rossi M., Zanella A. D-DASH: A deep q-learning framework for dash video streaming. IEEE Transactions on Cognitive Communications and Networking. 2017. vol. 3. no. 4. pp. 703–718. DOI: 10.1109/TCCN.2017.2755007.
34. Souane N., Bourenane M., Douga Y. Deep reinforcement learning-based approach for video streaming: Dynamic adaptive video streaming over HTTP. Applied Sciences. 2023. vol. 13. no. 21. DOI: 10.3390/app132111697.
35. Xie Y., Zhang Y., Lin T. Deep curriculum reinforcement learning for adaptive 360 ○ video streaming with two-stage training. IEEE Transactions on Broadcasting. 2023. vol. 70. no. 2. pp. 441–452. DOI: 10.1109/tbc.2023.3334137.
36. Du L., Zhuo L., Li J., Zhang J., Li X., Zhang H. Video quality of experience metric for dynamic adaptive streaming services using dash standard and deep spatial-temporal representation of video. Applied Sciences. 2020. vol. 10. no. 5. DOI: 10.3390/app10051793.
37. Mao H., Chen S., Dimmery D., Singh S., Blaisdell D., Tian Y., Alizadeh M., Bakshy E. Real-world video adaptation with reinforcement learning. arXiv preprint arXiv:2008.12858. 2020. DOI: 10.48550/arXiv.2008.12858.
38. David E.J., Gutiérrez J., Coutrot A., Da Silva M.P., Callet P.L. A dataset of head and eye movements for 360 videos. Proceedings of the 9th ACM Multimedia Systems Conference. 2018. pp. 432–437.
39. Wu C., Zhang R., Wang Z., Sun L. A spherical convolution approach for learning long term viewport prediction in 360 immersive video. Proceedings of the AAAI Conference on Artificial Intelligence. 2020. vol. 34. no. 01. pp. 14003–14040. DOI: 10.1609/aaai.v34i01.7377.

Хунг Нгуен Вьет — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, сетевая безопасность, искусственный интеллект, организация трафика в сетях нового поколения, гарантия качества сетевых услуг, экологичные сети, приложения. Число научных публикаций — 23. hungnv@eaut.edu.vn; Ки Фу - Ки Ань, Хатинь, Вьетнам; р.т.: +84(098)911-2079.

Дат Фам Тянь — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: приложения, сети. Число научных публикаций — 1. 20212452@eaut.edu.vn; Ву Нинь – Кьен Сюонг, Тхайбинь, Вьетнам; р.т.: +84(036)239-6558.

Тан Нгуен — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: приложения, анализ данных. Число научных публикаций — 3. tan25102000@gmail.com; Чынг Зунг - Тиен Лу, Хынгйен, Вьетнам; р.т.: +84(035)919-0216.

Куан Нгуен Ань — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: приложения, сети. Число научных публикаций — 1. anhq46724@gmail.com; Зялай, Ханой, Вьетнам; р.т.: +84(096)278-4293.

Транг Ле Тхи Хуэйен — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 2. tranglth@eaut.edu.vn; Фуонг Три - Тхи Тран Пхунг - Дан Фуонг, Ханой, Вьетнам; р.т.: +84(032)889-9334.

Нам Ле Май — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: разработка программного обеспечения, математика оптимизации, прикладные программы. Число научных публикаций — 1. namlm@eaut.edu.vn; Фуонг Чунг - Тхань Оай, Ханой, Вьетнам; р.т.: +84(098)208-2117.